COMMUN. MATH. SCI.
Vol. 18, No. 1, pp. 1–30

PATH-CONSERVATIVE IN-CELL DISCONTINUOUS RECONSTRUCTION SCHEMES FOR NON CONSERVATIVE HYPERBOLIC SYSTEMS[*]

CHRISTOPHE CHALONS[†]

Abstract. We are interested in the numerical approximation of discontinuous solutions in non-conservative hyperbolic systems. We introduce the basics of a new strategy based on in-cell discontinuous reconstructions to deal with this challenging topic, and apply it to a 2x2 non-conservative toy model, and a 3x3 gas dynamics system in Lagrangian coordinates. The strategy allows in particular to compute *exactly* isolated shocks. Numerical evidences are proposed.

Keywords. Non-conservative hyperbolic systems; Path conservative schemes; In-cell discontinuous reconstructions; Discontinuous solutions; Shock waves.

AMS subject classifications. 35L40; 65M99; 76M12; 76N15.

1. Introduction

In this paper, we are interested in the numerical approximation of non-conservative hyperbolic systems of the form

$$\begin{cases} \partial_t \mathbf{u} + \mathcal{A}(\mathbf{u}) \partial_x \mathbf{u} = 0, \quad x \in \mathbb{R}, \quad t \in \mathbb{R}^{+,\star}, \\ \mathbf{u}(x,0) = \mathbf{u}_0(x), \end{cases} \tag{1.1}$$

where $\mathbf{u}(x,t) \in \mathbb{R}^p$ is the unknown and $\mathbf{u}_0$ the initial data, supplemented with an initial condition

$$\mathbf{u}(x,0) = \mathbf{u}_0(x), \quad x \in \mathbb{R}, \tag{1.2}$$

and the validity of an entropy inequality

$$\partial_t \mathcal{U}(\mathbf{u}) + \partial_x \mathcal{F}(\mathbf{u}) \leq 0. \tag{1.3}$$

Here $(\mathcal{U},\mathcal{F})$ is an entropy-entropy flux pair, that is to say ${}^T\nabla\mathcal{U}\mathcal{A} = {}^T\nabla\mathcal{F}$ with $\mathcal{U}$ strictly convex. By non-conservative, we mean that $\mathcal{A}$ is not a Jacobian matrix. This does not prevent some of the equations of (1.1) from being in conservation form, but we assume that they are not all conservation laws. It turns out that the theoretical and numerical study of such systems is a very difficult task as briefly recalled now.

Theoretical aspects. Let us first review the main theoretical aspects of non-conservative hyperbolic systems. Generally speaking, hyperbolic systems develop discontinuous solutions for large times (see [28]), so that solutions in a weak sense are considered. When the model is made of conservation laws, solutions are usually defined in the sense of distributions and, under the validity of an entropy inequality, existence and uniqueness results are proved for initial data close to a constant state (see for instance Liu [32,33], Glimm [22], Lax [27,28], or LeFloch [31] for a review and extensions). In the case of a non-conservative system made of one or several non-conservation laws, the distribution theory does not apply anymore. Dal Maso, LeFloch and Murat proposed in [19] a definition of the non-conservative product $\mathcal{A}(\mathbf{u})\partial_x\mathbf{u}$ which extends the

[*] Received: November 24, 2018; Accepted (in revised form): August 2, 2019. Communicated by Giovanni Russo.

[†] Laboratoire de Mathématiques de Versailles, Université de Versailles Saint-Quentin (UVSQ), CNRS, Université Paris-Saclay, 78035 Versailles, France (christophe.chalons@uvsq.fr). https://chalons.perso.math.cnrs.fr/

notion of weak solution of conservation laws. More precisely, they introduce the *paths theory* to define $\mathcal{A}(\mathbf{u})\partial_x\mathbf{u}$ thanks to a family of paths $\phi:[0,1]\times\Omega\times\Omega\to\Omega$ satisfying the consistency property

$$\phi(0,\mathbf{u},\mathbf{u})=\mathbf{u},\quad \phi(1,\mathbf{u},\mathbf{u})=\mathbf{u},\quad \text{for all}\,(\mathbf{u},\mathbf{u})\in\Omega\times\Omega.$$

Under specific assumptions given in [19], the non-conservative product $\mathcal{A}(\mathbf{u})\partial_x\mathbf{u}$ at a given point x_0 separating two constant states $\mathbf{u}_0=(u_0,v_0)$ and $\mathbf{u}_1=(u_1,v_1)$ is defined by

$$[\mathcal{A}(\mathbf{u})\partial_x\mathbf{u}]_\phi=\int_0^1\mathcal{A}(\phi(s,\mathbf{u}_0,\mathbf{u}_1))\frac{\partial\phi}{\partial s}(s,\mathbf{u}_0,\mathbf{u}_1)ds\,\delta_{x_0}. \tag{1.4}$$

Said differently, the shock is admissible provided that the generalized Rankine-Hugoniot relation

$$-\sigma(\mathbf{u}_1-\mathbf{u}_0)+\int_0^1\mathcal{A}(\phi(s,\mathbf{u}_0,\mathbf{u}_1))\frac{\partial\phi}{\partial s}(s,\mathbf{u}_0,\mathbf{u}_1)ds=0 \tag{1.5}$$

holds true, where σ denotes the speed of propagation of the shock. Similarly to the conservative setting, this definition leads to existence and uniqueness results of weak solutions to (1)-(2) but a first difficulty is clearly to define the relevant path according to the physics of the model under consideration.

Numerical aspects (in brief). The numerical approximation of discontinuous solutions in non-conservative systems is a very difficult task. The main reasons are the deep sensitiveness of the standard methods with respect to the choice of the path and the usual discretisation parameters, see for instance [9,12,24,29] and the references therein, as well as the lack of a Lax-Wendroff-type convergence result. In particular, it is not guaranteed that the converged solution satisfies the path theoretical requirement (1.5). The literature is large on the topic but the proposed schemes are often not satisfying in the sense that either they work only for some very particular systems or small amplitude shocks, or they involve some random sampling techniques which are difficult to extend in several space dimensions. Without any attempt to be exhaustive, we refer, for instance, the reader to [4,5,7,11,13,15,21,35] and the references therein where different models and numerical approaches have been considered. Among these methods, the most recent and complete theory is probably the so-called path-conservative schemes theory, developed by C. Pares [35] and collaborators. However, it was proved in [1,12] that the consistency definition provided by the path-conservative formalism is not always enough to ensure the convergence to the expected solution. This is especially true in the case of small-scale dependent solutions of interest in the present paper, again because of a lack of control of the numerical diffusion. Nevertheless, we will see that when combined with a suitable in-cell discontinuous reconstruction strategy, the path-conservative formalism allows to control the numerical diffusion in numerical shocks.

General context. The present contribution follows a series of recent works on this topic, and more precisely the two comments on the computation of non-conservative products recently given in [1] and [15]. In a few words, the authors consider in [1] the gas dynamics equations in Lagrangian coordinates and show numerically that path-conservative schemes are not convergent to the correct solution when applied to a non-conservative version of these equations. This fact was explained theoretically in [12]. In [15], the authors consider the same set of equations and show how to slightly modify the usual path-conservative schemes to compute correctly the solutions of this non-conservative formulation. The proposed modification is based on a new averaging procedure of the path-conservative schemes and relies on both the introduction of modified averaging cells and a random sampling at each time step. The numerical results are really convincing and a convergence result is proved for isolated shocks. This shows that if the averaging procedure is dealt with care, then the path-conservative approximate Riemann solvers can be a powerful tool for the purpose of computing non-conservative shocks. This was actually the main message of [15]. However and as already stated above, the averaging procedure proposed in [15] relies on a random sampling and it is well-known from the work by Collela [18] that the computation of shocks with Glimm's random choice type methods is difficult to extend in several space dimensions. Therefore it could be a strong limitation for future works.

Objective of the paper. The aim of the present contribution is to propose a new averaging strategy based on in-cell discontinuous reconstruction in order to get rid of random sampling and modified cells. As we will see, it allows to follow isolated shocks *exactly*, and provides convergent results to the correct solution for more general initial data. In-cell discontinuous reconstruction techniques used in the present paper were developed by F. Lagoutière and B. Després to reduce the numerical diffusion in the transport of discontinuous solutions of linear and nonlinear equations, see for instance [20, 25, 26], before being extended to different settings. In particular, in [10] and [3], the authors define a conservative scheme which is based on in-cell discontinuous reconstructions of nonclassical shocks for approximating the solutions of *nonconvex* scalar conservation laws and *non-genuinely nonlinear* systems of conservation laws. Again, the striking feature of the strategy is to allow for a perfect control of the numerical diffusion associated with the nonclassical discontinuities. More precisely, it allows for the exact computation of such isolated simple waves. In [16, 17] and [39], the authors succeeded in extending this approach based on in-cell reconstructions to constrained (scalar or systems of) conservation laws in traffic modeling. In the present contribution, we aim at considering a first step towards the extension of in-cell discontinuous reconstructions towards non-conservative systems. Despite the present contribution being only the very beginning in the development of this strategy applied to non-conservative systems, we believe that it might be considered as a relevant alternative to numerical methods involving random sampling, which are so far the only ones for which convergence results can be proved.

Outline of the paper. The outline of the paper is as follows. In Section 2, we consider a non-conservative toy model and show how the in-cell discontinuous reconstruction strategy can be used to define a relevant projection onto the set of piecewise constant solutions at each time step and therefore to properly compute the shock discontinuities. Note that for this toy model, the exact Riemann solver will be used to define the in-cell reconstructions. At last, Section 3 considers the non-conservative gas dynamics equations in Lagrangian coordinates and shows how the discontinuous reconstruction strategy can be combined with the use of an approximate Riemann solver while keeping the same accuracy in the shock computations. The last section gives the main conclusions and perspectives of this work.

2. Application to a non-conservative toy model

In this section, we are interested in the numerical approximation of the weak solutions of the following non-conservative system of two partial differential equations:

$$\begin{cases} \partial_t u + \partial_x \dfrac{u^2}{2} + u\partial_x v = 0, \\ \partial_t v + \partial_x \dfrac{v^2}{2} + v\partial_x u = 0, \end{cases} \quad (x,t) \in \mathbb{R}\times\mathbb{R}^+, \tag{2.1}$$

where $\mathbf{u} = (u,v)^t$ belongs to the state space $\Omega = \{\mathbf{u} \in \mathbb{R}^2, u+v>0\}$. This system can be given the condensed form (1.1) where the non-Jacobian matrix $\mathcal{A}$ is defined by

$$\mathcal{A}(\mathbf{u}) = \begin{pmatrix} u & u \\ v & v \end{pmatrix}. \tag{2.2}$$

Such a model, which consists of two coupled Burgers equations, is probably the simplest example of a non-conservative model. It has already been studied for instance in [5] and can be understood as a simplified two-fluid model where u and v denote the velocity of each fluid.
Let us state useful properties of the model (see again [5], or [23] for the basic definitions), the proof of which is left to the reader.

Lemma 2.1. *System (1.1) is strictly hyperbolic over Ω with eigenvalues*

$$\lambda_1(\mathbf{u}) = 0 < \lambda_2(\mathbf{u}) = u+v,$$

and eigenvectors

$$r_1(\mathbf{u}) = (1,-1)^t, \quad r_2(\mathbf{u}) = (u,v)^t.$$

The first characteristic field is linearly degenerate and the second characteristic field is genuinely nonlinear. Moreover, the Riemann invariants are respectively given by

$$I_1(\mathbf{u}) = u+v, \quad I_2(\mathbf{u}) = u/v.$$

Remark 2.1. We have implicitly assumed $v \neq 0$ in the definition of I_2. In the case $v=0$ and for $\mathbf{u} \in \Omega$, the Riemann invariant is given by $I_2(\mathbf{u}) = v/u$.

Lemma 2.2. *Smooth solutions of (1.1) obey the following additional conservation laws*

$$\partial_t(u+v) + \partial_x \frac{(u+v)^2}{2} = 0, \quad \partial_t\Big(\frac{v}{u+v}\Big) = 0. \tag{2.3}$$

More generally, for any convex function f from $\mathbb{R}$ to $\mathbb{R}$, smooth solutions of (1.1) satisfy

$$\partial_t f(u+v) + \partial_x\Big(\int^{u+v} s f'(s)ds\Big) = 0. \tag{2.4}$$

In other words, the mapping $(u,v) \to f(u+v)$ is an entropy of (1.1).

In the forthcoming developments, the initial-value problem (1.1)-(1.2) is supplemented with the validity of the entropy inequality

$$\partial_t f(u+v) + \partial_x\Big(\int^{u+v} s f'(s)ds\Big) \leq 0 \tag{2.5}$$

in the usual distributional sense and for any convex function f from $\mathbb{R}$ to $\mathbb{R}$. As discussed in the introduction, such an entropy inequality is sufficient to prove existence and uniqueness of solutions close to a constant state when the system is conservative. Here, we are clearly in a non-conservative setting and according to the path theory of [19], an additional information encompassed in the so-called paths is needed for the problem to be well-posed.

A first example of family of paths. Following Volpert [40], one can choose for ϕ the straight lines family given by

$$\phi(s,\mathbf{u},\mathbf{u}) = \mathbf{u} + s(\mathbf{u}-\mathbf{u}), \quad \forall\, \mathbf{u},\mathbf{v} \in \Omega, \quad \forall\, s \in [0,1].$$

In this case, (1.5) writes

$$\begin{cases} -\sigma(u_1-u_0) + \dfrac{(u_1+u_0)}{2}\big((u_1+v_1)-(u_0+v_0)\big) = 0, \\ -\sigma(v_1-v_0) + \dfrac{(v_1+v_0)}{2}\big((u_1+v_1)-(u_0+v_0)\big) = 0. \end{cases} \tag{2.6}$$

Let us assume that $\mathbf{u}_0 \neq \mathbf{u}_1$. Then, if $u_0+v_0 \neq u_1+v_1$, it is easy to check that (2.6) can be equivalently written

$$\begin{cases} v_1 u_0 = v_0 u_1, \\ \sigma = \frac{1}{2}\big((u_0+v_0)+(u_1+v_1)\big), \end{cases} \tag{2.7}$$

while in the case $u_0+v_0 = u_1+v_1$, meaning that the value of the first Riemann invariant I_1 is the same, we get $\sigma = 0$ and the discontinuity is a contact discontinuity associated with the first characteristic field.

Note that the condition $u_0+v_0>0$ implies existence and uniqueness of $\mathbf{u}_1=(u_1,v_1)$ satisfying (2.7) for any given $\mathbf{u}_0=(u_0,v_0)$ and σ. Conversely, condition $u_1+v_1>0$ implies existence and uniqueness of $\mathbf{u}_0=(u_0,v_0)$ satisfying (2.7) for any given $\mathbf{u}_1=(u_1,v_1)$ and σ. In the following, we will use the notation

$$\begin{cases} \mathbf{u}_0 = \varphi(\mathbf{u}_1,\sigma), \\ \sigma = \frac{1}{2}\big((u_0+v_0)+(u_1+v_1)\big), \end{cases}$$

where φ will be called a kinetic function.

A second example of family of paths. Following LeFloch [30] and Sainsaulieu [37], ϕ can also be implicitly defined by adding a second-order diffusion tensor to (2.1):

$$\begin{cases} \partial_t u + \partial_x \dfrac{u^2}{2} + u\partial_x v = \varepsilon_1 \partial_{xx}(u+v), \quad \varepsilon_1 > 0, \\ \partial_t v + \partial_x \dfrac{v^2}{2} + v\partial_x u = \varepsilon_2 \partial_{xx}(u+v), \quad \varepsilon_2 > 0. \end{cases} \tag{2.8}$$

In this case, a shock discontinuity $(\sigma,\mathbf{u}_0,\mathbf{u}_1)$ is said to be admissible if there exists a travelling wave solution of (2.8) such that:

$$\begin{gathered} \mathbf{u}(x,t) = \overline{\mathbf{u}}(\xi), \quad \xi = x-\sigma t, \\ \lim_{\xi\to-\infty} \overline{\mathbf{u}}(\xi) = \mathbf{u}_0, \quad \lim_{\xi\to+\infty} \overline{\mathbf{u}}(\xi) = \mathbf{u}_1. \end{gathered} \tag{2.9}$$

It is shown in LeFloch [30] how to derive a family of paths consistent with this definition. Berthon [5] used this definition and showed for system (2.1) that for any $\mathbf{u}_0$ in Ω and σ in $](u_0+v_0)/2,(u_0+v_0)[$, there exists a unique state $\mathbf{u}_1 \neq \mathbf{u}_0$ in Ω and a unique travelling wave solution (up to a translation) satisfying (2.9) and such that the generalized Rankine-Hugoniot conditions (1.5) write

$$\begin{cases} v_1 = \dfrac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}\big(2\sigma-(u_0+v_0)\big)+\dfrac{\varepsilon_1 v_0-\varepsilon_2 u_0}{\varepsilon_1+\varepsilon_2}e^{\left(2-2(u_0+v_0)/\sigma\right)}, \\ \sigma = \dfrac{1}{2}\big((u_0+v_0)+(u_1+v_1)\big), \end{cases} \tag{2.10}$$

or equivalently

$$\begin{cases} v_0 = \dfrac{\varepsilon_2}{\varepsilon_1+\varepsilon_2}\big(2\sigma-(u_1+v_1)\big)+\dfrac{\varepsilon_1 v_1-\varepsilon_2 u_1}{\varepsilon_1+\varepsilon_2}e^{\left(2-2(u_1+v_1)/\sigma\right)}, \\ \sigma = \dfrac{1}{2}\big((u_0+v_0)+(u_1+v_1)\big). \end{cases} \tag{2.11}$$

We note in particular that the exit state $\mathbf{u}_1$ actually depends on the shape of the diffusion tensor, and more precisely on the ratio $\varepsilon_2/\varepsilon_1$. This is a characteristic of non-conservative systems as illustrated in various contributions on this subject, see for instance Raviart and Sainsaulieu [36], Sainsaulieu [37], Berthon and Coquel [6, 8], Chalons and Coquel [13, 14]. We also refer to Berthon, Coquel and LeFloch [9]. Here again, we will use the notation

$$\begin{cases} \mathbf{u}_0 = \varphi(\mathbf{u}_1,\sigma), \\ \sigma = \frac{1}{2}\big((u_0+v_0)+(u_1+v_1)\big). \end{cases}$$

2.1. A path-conservative in-cell discontinuous numerical scheme.

Let us now turn to the numerical approximation of the solutions of our toy model. We first introduce some notations and briefly recall the usual Godunov scheme. As we will see, this scheme fails in approximating the correct shock solutions defined by a family of paths ϕ, but it will be useful for approximating the smooth parts of the solutions, in particular the rarefaction waves. We thus motivate and describe the proposed in-cell discontinuous reconstruction strategy which allows in particular to compute exactly any isolated admissible shock. This property is the key property to explain the success of the approach for general initial data.

We introduce a constant space step Δx and constant time step Δt and we set $\nu = \Delta t/\Delta x$. The mesh interfaces are defined by $x_{j+1/2} = j\Delta x$ for $j\in\mathbb{Z}$ and the intermediate times by $t^n = n\Delta t$ for $n\in\mathbb{N}$. As usual in the finite volume framework, we seek at each time t^n for an approximation $\mathbf{u}_j^n$ of the solution in the interval $[x_{j-1/2}, x_{j+1/2})$, $j\in\mathbb{Z}$. Therefore, a piecewise constant approximate solution $x\to\mathbf{u}_\nu(x,t^n)$ of the solution $\mathbf{u}$ is given by

$$\mathbf{u}_\nu(x,t^n) = \mathbf{u}_j^n \text{ for all } x\in C_j = [x_{j-1/2}; x_{j+1/2}),\ j\in\mathbb{Z},\ n\in\mathbb{N}.$$

When $n=0$, we set

$$\mathbf{u}_j^0 = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{u}_0(x)dx, \text{ for all } j\in\mathbb{Z}.$$

2.1.1. Failure of the classical Godunov scheme.

The classical Godunov scheme is composed of two steps: a first step in which the solution evolves in time according to the PDE model under consideration, and a second step of projection onto piecewise constant functions.

Step 1: Evolution in time. In this first step, one solves the following Cauchy problem

$$\begin{cases} \partial_t \mathbf{u}(x,t) + \mathcal{A}(\mathbf{u}(x,t))\partial_x \mathbf{u}(x,t) = 0, \ x \in \mathbb{R}, \\ \mathbf{u}(x,0) = \mathbf{u}_\nu(x,t^n), \end{cases} \tag{2.12}$$

with the given family of paths for times $t \in [0,\Delta t]$. Recall that $x \to \mathbf{u}_\nu(x,t^n)$ is piecewise constant. Then, under the usual CFL restriction

$$\frac{\Delta t}{\Delta x} \max\{|\lambda_i(\mathbf{u})|, \ i = 1,2\} \leq \frac{1}{2}, \tag{2.13}$$

for all the $\mathbf{u}$ under consideration, the solution of (2.12) is known by gluing together the solutions of the Riemann problems set at each interface :

$$\mathbf{u}(x,t) = \mathbf{u_r}(\frac{x - x_{j+1/2}}{t}; \mathbf{u}_j^n, \mathbf{u}_{j+1}^n) \text{ for all } (x,t) \in [x_j, x_{j+1}] \times [0,\Delta t], \tag{2.14}$$

where $(x,t) \to \mathbf{u_r}(\frac{x}{t}; \mathbf{u}_L, \mathbf{u}_R)$ denotes the self-similar solution of the Riemann problem

$$\begin{cases} \partial_t \mathbf{u}(x,t) + \mathcal{A}(\mathbf{u}(x,t))\partial_x \mathbf{u}(x,t) = 0, \ x \in \mathbb{R}, \ t \in \mathbb{R}^{+,\star} \\ \mathbf{u}(x,0) = \begin{cases} \mathbf{u}_L \text{ if } x < 0, \\ \mathbf{u}_R \text{ if } x > 0, \end{cases} \end{cases}$$

given in the Appendix, whatever $\mathbf{u}_L$ and $\mathbf{u}_R$ are in the phase space Ω. Recall that this solution actually depends on the family of paths under consideration.

Step 2: Projection. In order to get a piecewise constant approximate solution on each cell $\mathcal{C}_j$ at time t^{n+1}, the solution $x \to \mathbf{u}(x,\Delta t)$ given by (2.14) is simply averaged on $\mathcal{C}_j$, as expressed by the following update formula:

$$\mathbf{u}_j^{n+1} = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{u}(x,\Delta t) dt, \ j \in \mathbb{Z}. \tag{2.15}$$

In the following, it will be useful to write (2.15) equivalently as

$$\mathbf{u}_j^{n+1} = \frac{1}{2}\Big(\mathbf{u}_{j,L}^{n+1} + \mathbf{u}_{j,R}^{n+1}\Big), \ j \in \mathbb{Z}, \tag{2.16}$$

with

$$\mathbf{u}_{j,L}^{n+1} = \frac{2}{\Delta x} \int_{x_{j-1/2}}^{x_j} \mathbf{u_r}(\frac{x - x_{j-1/2}}{\Delta t}; \mathbf{u}_{j-1}^n, \mathbf{u}_j^n) dx \tag{2.17}$$

and

$$\mathbf{u}_{j,R}^{n+1} = \frac{2}{\Delta x} \int_{x_j}^{x_{j+1/2}} \mathbf{u_r}(\frac{x - x_{j+1/2}}{\Delta t}; \mathbf{u}_j^n, \mathbf{u}_{j+1}^n) dx. \tag{2.18}$$

As illustrated on Figure 2.1 obtained with initial data

$$\mathbf{u}_0(x) = (u,v)_0(x) = \begin{cases} (6,5) \text{ if } x < 0.5, \\ (0.7,0.3) \text{ if } x > 0.5, \end{cases} \tag{2.19}$$

and the second family of path with $\epsilon_2 = \epsilon_1$, the numerical results provided by this scheme are not satisfactory when a shock is present in the solution in the sense that the intermediate state is different from the exact one. On the contrary, if we consider for instance $\epsilon_2 = 10\epsilon_1$ and

$$\mathbf{u}_0(x) = (u,v)_0(x) = \begin{cases} (1,2) \text{ if } x < 0.5, \\ (5,1) \text{ if } x > 0.5, \end{cases} \tag{2.20}$$

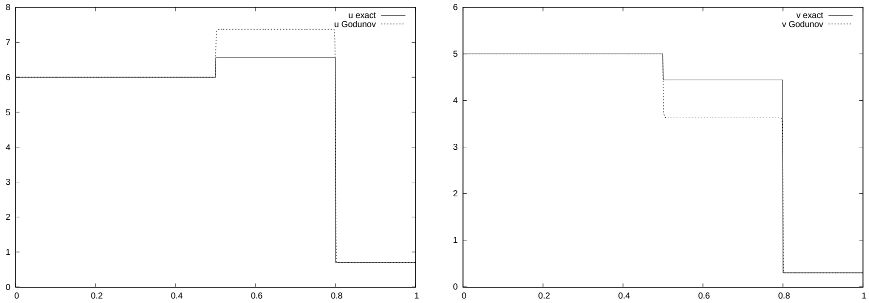

FIG. 2.1. *u (left) and v (right) - Contact discontinuity followed by a shock wave - Final time* $t=0.05$ *- 1000-point mesh*

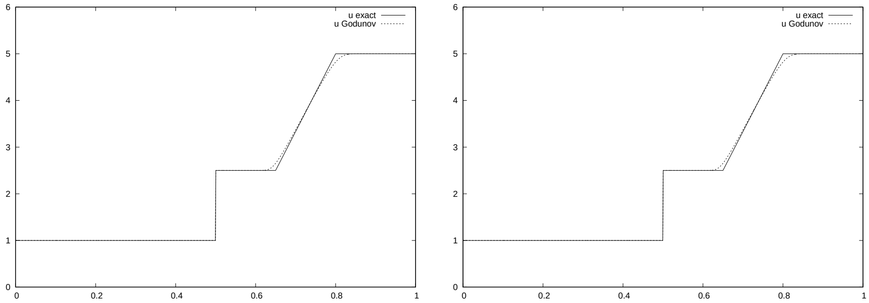

FIG. 2.2. *u (left) and v (right) - Contact discontinuity followed by a rarefaction wave - Final time* $t=0.05$ *- 1000-point mesh*

leading to a rarefaction wave, it works correctly in the sense that the intermediate state is now correct, see Figure 2.2.

As clearly explained, for instance, in [5, 6, 8, 13, 14], the main reason of this failure is the excessive numerical diffusion of the Godunov scheme across the shocks, which disagrees with the underlying regularization operator at the discrete level. In other words, the numerical diffusion plays a crucial role and must be controlled to make the approximate and exact solutions coincide. If the numerical diffusion does not exactly mimic the action of the regularization operator, the numerical solutions disagree with the exact solutions. This is observed with the usual Godunov scheme but also with any standard finite volume scheme.

The sensitiveness with respect to the numerical diffusion is typical of non-conservative systems, but also appear in conservative systems (when the matrix $\mathcal{A}$ is the Jacobian matrix of a flux function), when the system is hyperbolic but has at least one characteristic field that is neither genuinely nonlinear, nor linearly degenerate, or when it is not hyperbolic but mixed hyperbolic-elliptic. Such systems need also to be closed by a *kinetic relation*, which is similar to the previous notion of *path*, and can give rise to the so-called nonclassical shock waves, see for instance [31]. From a numerical point of view, similar issues to those already discussed come out and approximating nonclassical shocks is challenging because of the dependence on the underlying diffusion mechanisms. Again, standard techniques are useless and a deeper analysis shows that the failure can be related to the (un)control of the numerical diffusion.

In order to overcome this difficulty, a new numerical approach was first proposed in [10] (see also [3]) to compute nonclassical solutions to scalar conservation laws. The proposed scheme is fully conservative on fixed meshes and has the property of exactly capturing isolated nonclassical shocks. For such isolated discontinuities, the underlying numerical diffusion thus reduces to the minimum, namely at one point, unlike standard finite difference schemes. The method is based on an in-cell discontinuous reconstruction technique performed in each computational cell that may contain a nonclassical shock. The next section proposes to extend this approach to the present setting of a non-conservative system in order to properly compute the underlying (in some sense nonclassical) shocks on a fixed mesh and satisfy at the same time the family of paths.

2.1.2. In-cell discontinuous reconstruction. *Overview of the strategy.* In the previous sections, it was shown that the Godunov scheme is not a good candidate when shocks are present in the solution, but it works correctly when the solution is smooth. Therefore, we will first propose to keep on using the Godunov scheme "far away" from shock discontinuities. On the contrary, in the vicinity of shock discontinuities, we will follow the same approach as in [10] which consists in adding details in the piecewise constant representation of the approximate solution on each cell C_j. More precisely, we will reconstruct discontinuities in the relevant cells C_j and use them to define $\mathbf{u}_j^{n+1}$ instead of simply using the constant values $\mathbf{u}_{j-1}^n$, $\mathbf{u}_j^n$ and $\mathbf{u}_{j+1}^n$ like in the Godunov scheme. As we will see hereafter, such an approach will allow to *exactly* compute isolated shock discontinuities in the sense that for such solutions $\mathbf{u}_j^n$ will equal the average of the *exact* solution on the cell C_j. The corresponding numerical discontinuity will then be diffused on one cell at most. Such a sharp control of the numerical diffusion is at the core of the success of the strategy.

The reconstruction procedure. It is now a matter of defining which cells are to be concerned with the reconstruction procedure as well as the reconstructed discontinuities themselves, but also the strategy to evaluate $\mathbf{u}_j^{n+1}$ using the new details provided by the discontinuous reconstructions. Let us consider the cell C_j and proceed as follows. Assume that at time t^n,

$$(u+v)_{j-1}^n > (u+v)_{j+1}^n. \tag{2.21}$$

According to the Riemann solver, we consider that a shock discontinuity is expected to appear locally around the cell C_j and to develop at the next times $t > t^n$. Indeed, such a shock is present in the Riemann solution associated with the inital states $\mathbf{u}_{j-1}^n$ and $\mathbf{u}_{j+1}^n$. Hence and with clear notations, we are tempted to introduce in the cell C_j the left and right states $\mathbf{u}_{j,l} = \mathbf{u}_*(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)$ and $\mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}^n$ of the shock which is expected to be present in the Riemann solution associated with $\mathbf{u}_{j-1}^n$ and $\mathbf{u}_{j+1}^n$. Since we are considering the cell C_j, we require that the reconstructed discontinuity between $\mathbf{u}_{j,l}$ and $\mathbf{u}_{j,r}$ is located inside C_j at a position

$$\bar{x}_j^u = x_{j-1/2} + d_j^{n,u} \Delta x, \tag{2.22}$$

for the u component, and

$$\bar{x}_j^v = x_{j-1/2} + d_j^{n,v} \Delta x, \tag{2.23}$$

for the v component, for some $d_j^{n,u}$ and $d_j^{n,v}$ in $[0,1]$. Note indeed that in general, we will consider that the positions of the discontinuities may be different for both components u and v, see Figure 2.3. Regarding the position of the discontinuities in the cell, it

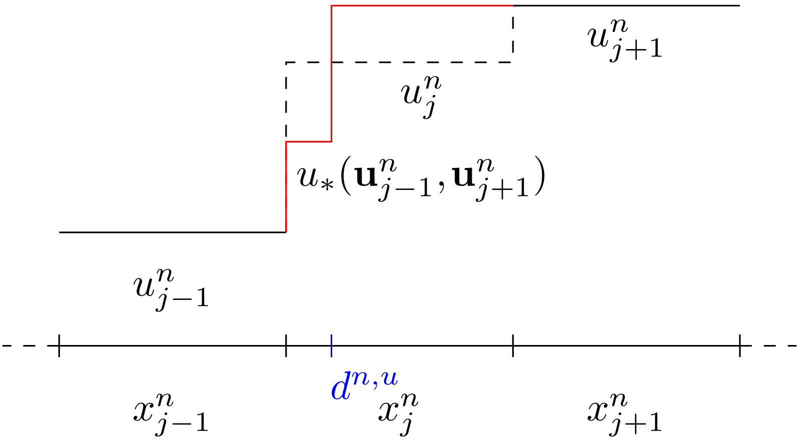

FIG. 2.3. *Reconstruction of a shock in cell C_j. Example of the u component, assuming that the cell j starts at 0 and has length 1; otherwise, just replace $d^{n,u}$ by $x_{j-1/2}+d^{n,u}\Delta x$. A similar drawing could be done for the v component.*

is natural to impose that the reconstruction procedure has to be conservative, which writes

$$d_j^{n,u}u^n_{j,l}+(1-d_j^{n,u})u^n_{j,r}=u^n_j, \tag{2.24}$$

or equivalently,

$$d_j^{n,u}=\frac{u^n_{j,r}-u^n_j}{u^n_{j,r}-u^n_{j,l}}, \tag{2.25}$$

for the u component, and

$$d_j^{n,v}v^n_{j,l}+(1-d_j^{n,v})v^n_{j,r}=v^n_j, \tag{2.26}$$

or equivalently,

$$d_j^{n,v}=\frac{v^n_{j,r}-v^n_j}{v^n_{j,r}-v^n_{j,l}}, \tag{2.27}$$

for the v component. Clearly, it is possible to reconstruct such discontinuities inside the cell C_j provided that

$$0\leq d_j^{n,u}=\frac{u^n_{j,r}-u^n_j}{u^n_{j,r}-u^n_{j,l}}\leq 1, \tag{2.28}$$

and

$$0\leq d_j^{n,v}=\frac{v^n_{j,r}-v^n_j}{v^n_{j,r}-v^n_{j,l}}\leq 1, \tag{2.29}$$

which gives two additional conditions for the in-cell reconstruction procedure to make sense.

To conclude the definition of the reconstruction strategy, let us mention that still according to the Riemann solver, it is natural to consider that the speed of propagation $\sigma_{j,l,r}$ of the reconstructed discontinuity equals $\sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)$ for both components u and v, where of course $\sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)$ denotes the exact value of the speed of propagation of the shock in the Riemann solution associated with $\mathbf{u}_{j-1}^n$ and $\mathbf{u}_{j+1}^n$.

Update formulas. At this stage, the reconstructed discontinuity is completely defined, as well as the reconstruction criteria (2.21), (2.28) and (2.29) for this reconstruction to take place. It thus remains to define the update formula for $\mathbf{u}_j^{n+1}$, as well as the influence of the reconstruction on the update formulas of $\mathbf{u}_{j-1}^{n+1}$ and $\mathbf{u}_{j+1}^{n+1}$. Since $\sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n) > 0$, note from now on that for the sake of simplicity and in order to avoid dealing with the interaction of two reconstructed discontinuities in adjacent cells, no reconstruction will be considered in the cell C_{j+1}.

The cell C_j. In this cell, we consider that the system under consideration is completely solved by the reconstructed discontinuity, and thus writes

$$\partial_t u = -[\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}]_\phi^u \delta_{x-\bar{x}_j^u = \sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)t}$$

for the u component, and

$$\partial_t v = -[\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}]_\phi^v \delta_{x-\bar{x}_j^v = \sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)t}$$

for the v component, where, with clear notations, $[\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}]_\phi^{u,v}$ is given by

$$-[\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}]_\phi^u = -\sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)(u_{j,r}^n - u_{j,l}^n)$$

for the u component, and

$$-[\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}]_\phi^v = -\sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)(v_{j,r}^n - v_{j,l}^n)$$

for the v component. Integrating in space and time, we get for the u component

$$u_j^{n+1} = u_j^n - \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \int_{t^n}^{t^n+\Delta t} [\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}]_\phi^u \delta_{x-\bar{x}_j^u = \sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)t},$$

namely

$$u_j^{n+1} = u_j^n - \frac{\sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)(u_{j,r}^n - u_{j,l}^n)}{\Delta x} \times \min(\Delta t, \Delta t^u) \tag{2.30}$$

where Δt^u is the time needed by the reconstructed discontinuity in u to reach the interface $x_{j+1/2}$, that is to say

$$\Delta t^u = \frac{1 - d_j^{n,u}}{\sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)} \Delta x.$$

For the v component

$$v_j^{n+1} = v_j^n - \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \int_{t^n}^{t^n+\Delta t} [\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}]_\phi^v \delta_{x-\bar{x}_j^v = \sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)t},$$

namely

$$v_j^{n+1} = v_j^n - \frac{\sigma(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)(v_{j,r}^n - v_{j,l}^n)}{\Delta x} \times \min(\Delta t, \Delta t^v) \tag{2.31}$$

where Δt^v is the time needed by the reconstructed discontinuity in v to reach the interface $x_{j+1/2}$,

$$\Delta t^v = \frac{1-d_j^{n,v}}{\sigma(\mathbf{u}_{j-1}^n,\mathbf{u}_{j+1}^n)}\Delta x.$$

Formulas (2.30) and (2.31) are equivalent to setting $\mathbf{u}_j^{n+1}=\mathbf{u}_{j,l}^n$ if Δt is greater than the times needed by the reconstructed discontinuities on u and v to reach the interface $x_{j+1/2}$. If not, they are equivalent to averaging the reconstructed discontinuities at their new position in the cell C_j after moving at velocity $\sigma_{j,l,r}$ for a time of length Δt.

The cell C_{j+1}. If Δt is greater than the times needed by the reconstructed discontinuities on u and v to reach the interface $x_{j+1/2}$, it is clear that the reconstructed discontinuities are expected to influence the update formulas on the cell C_{j+1}. However, under the CFL condition (2.13), the reconstructed discontinuities in u and v in the cell C_j cannot reach the middle point x_{j+1} of the cell C_{j+1} and may thus influence the half interval $[x_{j+1/2},x_{j+1})$ only. Since no reconstruction is considered in the cell C_{j+1}, we consider the usual update formula

$$\mathbf{u}_{j+1}^{n+1}=\frac{1}{2}\left(\mathbf{u}_{j+1,L}^{n+1}+\mathbf{u}_{j+1,R}^{n+1}\right)$$

but with (component by component)

$$\mathbf{u}_{j+1,L}^{n+1}=\mathbf{u}_{j+1}^n-\frac{2\sigma(\mathbf{u}_{j-1}^n,\mathbf{u}_{j+1}^n)(\mathbf{u}_{j,r}^n-\mathbf{u}_{j,l}^n)}{\Delta x}\times\left(\Delta t-\min(\Delta t,\Delta t^{\mathbf{u}})\right)$$

in order to take into account the propagation of the reconstructed discontinuities inside the half interval $[x_{j+1/2},x_{j+1})$. Note that compared to the usual Godunov scheme, the value of $\mathbf{u}_{j+1,R}^{n+1}$ is unchanged.

To conclude the proposed numerical scheme, let us underline that when no reconstruction takes place in the cells C_{j-1} and C_j, we use the classical Godunov scheme, namely

$$\mathbf{u}_j^{n+1}=\frac{1}{2}\left(\mathbf{u}_{j,L}^{n+1}+\mathbf{u}_{j,R}^{n+1}\right)$$

with the definitions (2.17) and (2.18).

Summary. To sum up, the update value of a given cell C_j is kept unchanged with respect to the Godunov scheme if no reconstruction takes place in the cells C_{j-1} and C_j, the update value of a given cell C_j is completely changed if a reconstruction takes place in the cell C_j, and the update value of a given cell C_j is partially changed if no reconstruction takes place in the cell C_j but a reconstruction takes place in the cell C_{j-1}. In this case, $\mathbf{u}_{j,L}^{n+1}$ is changed but not $\mathbf{u}_{j,R}^{n+1}$.

At last, recall that a reconstruction is considered in the cell C_j if and only if the criteria (2.21), (2.28) and (2.29) are satisfied and the criteria (2.21), (2.28) and (2.29) adapted to the cell C_{j-1} are not satisfied. Following [10], let us prove an important property of the proposed scheme, which explains the very good results obtained in the next section. The result states that isolated shock discontinuities are exactly captured by the scheme and contain no spurious numerical diffusion.

THEOREM 2.1. *Assume that $\mathbf{u}_j^0=\mathbf{u}_L$ if $j\le 0$, $\mathbf{u}_j^0=\mathbf{u}_R$ if $j\ge 1$ and that $\mathbf{u}_L$ and $\mathbf{u}_R$ are two constant states in the phase space Ω such that they can be joined by an admissible* *shock discontinuity. In other words, the Riemann solution associated with these left and right states is given by* $\mathbf{u}(x,t)=\mathbf{u}_L$ *if* $x<\sigma t$ *and* $\mathbf{u}(x,t)=\mathbf{u}_R$ *if* $x>\sigma t$ *where* σ *is the speed of propagation given by*

$$\sigma=\frac{1}{2}\big((u_L+v_L)+(u_R+v_R)\big)$$

according to the exact Riemann solver. Then the proposed scheme provides an exact numerical solution on each cell C_j *in the sense that*

$$\mathbf{u}_j^n=\frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}}\mathbf{u}(x,t^n)dx, \qquad j\in\mathbb{Z},\ n\in\mathbb{N}. \tag{2.32}$$

In particular, the numerical discontinuity is diffused on one cell at most.

Proof. Let us first notice that there is no relevant reconstruction in the first time iteration since the only cells which could be affected are $j=0$ and $j=1$ but by conservation we necessarily have $d_0^{0,u}=d_0^{0,v}=1$ and $d_1^{0,u}=d_1^{0,v}=0$. In other words, considering a reconstructed discontinuity in these cells gives back the original averaged value. The Godunov scheme is then used during the first step and as an immediate consequence, equality (3.13) is proved for the first iterate by definition of the Godunov scheme. Note that we have in particular

$$\mathbf{u}_1^1=\mathbf{u}_R-\sigma\frac{\Delta t}{\Delta x}(\mathbf{u}_R-\mathbf{u}_L)$$

component by component.

Let us now see what happens in the next time iteration. It is first clear from above that only C_1 is to be concerned with a reconstructed discontinuity between $\mathbf{u}_L$ and $\mathbf{u}_R$. Interestingly, by conservativity the reconstructed discontinuities in u and v are necessarily located at the exact position of the solution, namely at the position $x=x_{j-1/2}+\sigma\Delta t$. In other words, we reconstruct the exact solution at time $t=\Delta t$. To get the required identity (2.32) for the second iterate, it is sufficient to focus on the two cells C_1 and C_2 (the other ones are trivial) and for instance on the u variable (the v variable can be dealt with in a similar way). Let us first assume that $\Delta t^u\leq\Delta t$. The numerical scheme gives

$$u_1^2=u_1^1-\frac{\sigma(u_R-u_L)}{\Delta x}\times\Delta t,$$

that is to say

$$u_1^2=u_R-\sigma\frac{\Delta t}{\Delta x}(u_R-u_L)-\frac{\sigma(u_R-u_L)}{\Delta x}\times\Delta t=u_R-\sigma\frac{2\Delta t}{\Delta x}(u_R-u_L)$$

and

$$u_2^2=u_2^1=u_R,$$

which clearly coincides with the average of the exact solution after two time steps Δt on the cells C_1 and C_2. Let us now assume that $\Delta t^u\geq\Delta t$ so that the exact shock will pass through the interface $x_{1+1/2}$ and be located at position

$$x=x_{1+1/2}+\sigma(\Delta t-\Delta t^u)$$

in the cell C_2. On the other hand, the numerical scheme gives

$$u_1^2 = u_1^1 - \frac{\sigma(u_R - u_L)}{\Delta x} \times \Delta t^u,$$

that is to say

$$u_1^2 = u_R - \sigma\frac{\Delta t}{\Delta x}(u_R - u_L) - \frac{\sigma(u_R - u_L)}{\Delta x} \times \frac{\Delta x - \sigma \Delta t}{\sigma} = u_L$$

and

$$u_2^2 = \frac{1}{2}\Big(u_R - \frac{2\sigma(u_R - u_L)}{\Delta x} \times (\Delta t - \Delta t^u) + u_R\Big),$$

or equivalently

$$u_2^2 = u_R - \sigma\frac{(\Delta t - \Delta t^u)}{\Delta x}(u_R - u_L)$$

which again clearly coincides with the average of the exact solution on the cells C_1 and C_2 after two time steps. And the process is going on in a similar way for the next time iterations, which proves the result. □

2.2. Numerical experiments. In this section, we illustrate the behavior of the proposed scheme based on in-cell discontinuous reconstructions.

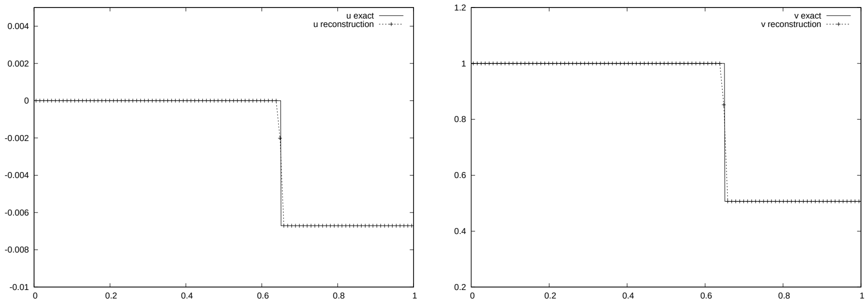

FIG. 2.4. *u (left) and v (right) - Isolated shock - Final time $t = 0.2$ - 100-point mesh*

Test 1. In this first test case, we consider an isolated shock associated with the second family of path with $\epsilon_1 = \epsilon_2$ and associated with the left and right states of the following initial data,

$$\begin{aligned}\mathbf{u}_0(x) = (u,v)_0(x) &= \begin{cases} (u_L, v_L) \text{ if } x < 0.5, \\ (u_R, v_R) \text{ if } x > 0.5, \end{cases} \\ &= \begin{cases} (0,1) \text{ if } x < 0.5, \\ (-0.00670855951629595, 0.50670855951629590) \text{ if } x > 0.5. \end{cases}\end{aligned} \tag{2.33}$$

The speed of propagation is $\sigma = 3/4$. As we can see on Figure 2.4, and in agreement with our theorem, the numerical solution is exact and contains only one point of numerical

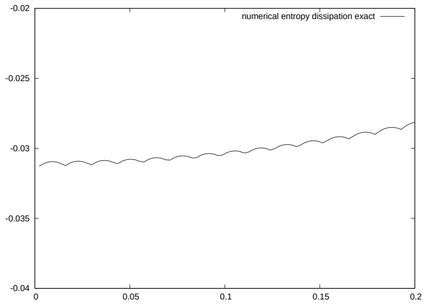

FIG. 2.5. *D^n - Isolated shock - 100-point mesh*

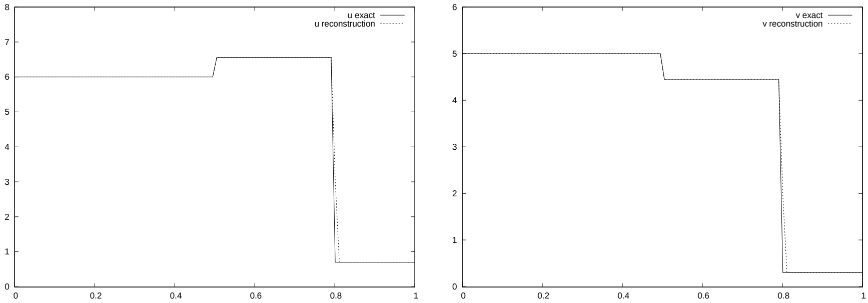

FIG. 2.6. *u (left) and v (right) - Contact discontinuity followed by a shock - Final time $t = 0.05$ - 100-point mesh*

diffusion. On Figure 2.5, we plot the numerical entropy dissipation D^n with respect to the time t^n and defined by

$$D^n = \frac{1}{2}\Big(\sum_j \Delta x(u_j^{n+1} + v_j^{n+1})\Big)^2 - \sum_j \Delta x \frac{(u_j^n + v_j^n)^2}{2} + \Delta t\Big(\frac{(u_R + v_R)^3}{3} - \frac{(u_L + v_L)^3}{3}\Big)$$

where the sum is taken over the mesh cells. We observe that it is nonpositive as expected.

Test 2. The second test case is the same as the one considered on Figure 2.1, and we now clearly see that the proposed strategy allows to properly compute the nonconservative shock and the intermediate state, even with a coarse mesh made of 100 points. The results are given on Figure 2.6.

Test 3. The last test case considers a periodic simulation associated with the initial data given by

$$\mathbf{u}_0(x) = (u,v)_0(x) = \begin{cases} (u_L, v_L) & \text{if } x < 0.4 \text{ or } x > 0.6, \\ (u_R, v_R) & \text{otherwise,} \end{cases} \tag{2.34}$$

with $(u_L, v_L) = (0,1)$ and $(u_R, v_R) = (5,2)$, and again the second family of paths with $\epsilon_1 = \epsilon_2$. On Figure 2.7 we compare the numerical quantities $u+v$ obtained with our scheme and the classical Godunov scheme. Recall that this quantity is conserved so that both methods are expected to give the same solution. Again, we clearly see that the new scheme is less diffusive than the Godunov one at the point of discontinuity of the N-wave profile.

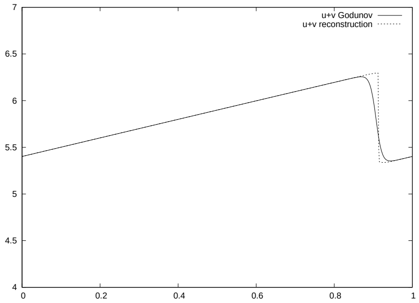

FIG. 2.7. *$u+v$ - Periodic simulation - Final time $t=1$ - 600-point mesh*

3. Application to the gas dynamics equations in Lagrangian coordinates

In this section, we apply the in-cell reconstruction technique to the following gas dynamics equations in Lagrangian coordinates:

$$\begin{cases} \partial_t \tau - \partial_x u = 0, \\ \partial_t u + \partial_x p = 0, \\ \partial_t E + \partial_x pu = 0, \end{cases} \tag{3.1}$$

where $\tau > 0$ represents the inverse of a density, u is the velocity and $p = p(\tau, e) > 0$ is the pressure. Here $e > 0$ denotes the internal energy and satisfies $E = e + u^2/2$. For the sake of simplicity, we consider a perfect gas equation of state $p(\tau, e) = (\gamma - 1)e/\tau$ where $\gamma > 1$. Recall that (3.1) is strictly hyperbolic with eigenvalues $\lambda_0 = 0$ and $\lambda_\pm = \pm c$, $c = \sqrt{\gamma p/\tau}$, and that the characteristic field associated with λ_0 is linearly degenerate and the ones associated with $\lambda_\pm$ are genuinely nonlinear [23]. On the other hand, the admissible solutions of (3.1) are selected by the Lax entropy inequalities, which here are equivalent to $\sigma(\tau_+ - \tau_-) > 0$ where τ_+ and τ_- are the left and right states of the underlying discontinuity, and σ its speed of propagation.

At this stage, (3.1) is written in a classical conservative form which does not raise any difficulty from a numerical point of view since usual Godunov-type schemes can be used, see [23] again. However, the following non-conservative formulation of (3.1) can be easily obtained

$$\begin{cases} \partial_t \tau - \partial_x u = 0, \\ \partial_t u + \partial_x p = 0, \\ \partial_t e + p \partial_x u = 0, \end{cases} \tag{3.2}$$

where only the last equation on the total energy has been replaced with an equation on the internal energy. Setting $\mathbf{u} = (\tau, u, e)$, the matrix $\mathcal{A}(\mathbf{u})$ is given by

$$\mathcal{A}(\mathbf{u}) = \begin{pmatrix} 0 & -1 & 0 \\ \partial_\tau p(\tau, e) & 0 & \partial_e p(\tau, e) \\ 0 & p(\tau, e) & 0 \end{pmatrix}.$$

In order to define the admissible solutions of (3.2), we consider again the path theory of Dal Maso, LeFloch and Murat. Here, a very simple choice of path is defined for all

$\mathbf{u}_0$ and $\mathbf{u}_1$ such that $\sigma(\tau_1-\tau_0)>0$ in a linear way with respect to τ, u and p, namely

$$\begin{cases} \tau(s)=\tau_0+s(\tau_1-\tau_0), \\ u(s)=u_0+s(u_1-u_0), \\ p(s)=p_0+s(p_1-p_0), \end{cases}$$

for all $s\in[0,1]$. Actually, it turns out that easy calculations show that the generalized jump relations (1.5) of the path theory boil down to the classic Rankine-Hugoniot relations applied to (3.1), namely

$$\begin{cases} \sigma(\tau_1-\tau_0)+(u_1-u_0)=0, \\ -\sigma(u_1-u_0)+(p_1-p_0)=0, \\ -\sigma(E_1-E_0)+(p_1u_1-p_0u_0)=0, \end{cases} \tag{3.3}$$

or equivalently

$$\begin{cases} \sigma(\tau_1-\tau_0)+(u_1-u_0)=0, \\ -\sigma(u_1-u_0)+(p_1-p_0)=0, \\ -\sigma(e_1-e_0)+\dfrac{1}{2}(p_1+p_0)(u_1-u_0)=0. \end{cases} \tag{3.4}$$

In other words and with such a choice of path, both conservative and non-conservative formulations (3.1) and (3.2) select the same solutions.

3.1. A Roe-type path-conservative approximate Riemann solver. We begin with the definition of a Roe-type path-conservative approximate Riemann solver associated with (3.2) and a given path ϕ. According to [38] and [35], it is based on a Roe linearization $\mathcal{A}_\phi$ such that

(1) for all $\mathbf{u}_L$ and $\mathbf{u}_R$, $\mathcal{A}_\phi(\mathbf{u}_L,\mathbf{u}_R)$ has 3 distinct eigenvalues,

(2) for all $\mathbf{u}$, $\mathcal{A}_\phi(\mathbf{u},\mathbf{u})=\mathcal{A}(\mathbf{u})$,

(3) for all $\mathbf{u}_L$ and $\mathbf{u}_R$,

$$\mathcal{A}_\phi(\mathbf{u}_L,\mathbf{u}_R)(\mathbf{u}_R-\mathbf{u}_L)=\int_0^1 \mathcal{A}(\phi(s,\mathbf{u}_L,\mathbf{u}_R))\partial_s\phi(s,\mathbf{u}_L,\mathbf{u}_R)ds.$$

The three properties are satisfied if we set

$$\mathcal{A}_\phi(\mathbf{u}_L,\mathbf{u}_R)=\mathcal{A}(\overline{\mathbf{u}}), \quad \overline{\mathbf{u}}=\overline{\mathbf{u}}(\mathbf{u}_L,\mathbf{u}_R)=(\overline{\tau},\overline{u},\overline{e})$$

with

$$\overline{\tau}=\frac{\tau_L+\tau_R}{2}, \quad \overline{u}=\frac{u_L+u_R}{2}, \quad \overline{e}=\frac{\overline{p}\,\overline{\tau}}{\gamma-1} \quad \text{and} \quad \overline{p}=\frac{p_L+p_R}{2},$$

see [34]. The approximate Riemann solution constructed from the Roe linearization is the solution of

$$\begin{cases} \partial_t\mathbf{u}(x,t)+\mathcal{A}_\phi(\mathbf{u}_L,\mathbf{u}_R)\partial_x\mathbf{u}(x,t)=0, \\ \mathbf{u}(x,t=0)=\begin{cases} \mathbf{u}_L \text{ if } x<0, \\ \mathbf{u}_R \text{ if } x>0, \end{cases} \end{cases}$$

given by

$$\mathbf{u}(x/t;\mathbf{u}_L,\mathbf{u}_R)=\begin{cases} \mathbf{u}_L \text{ if } \quad x/t<-\sigma(\mathbf{u}_L,\mathbf{u}_R), \\ \mathbf{u}_L^* \text{ if } -\sigma(\mathbf{u}_L,\mathbf{u}_R)<x/t<0, \\ \mathbf{u}_R^* \text{ if } \ 0<x/t<\sigma(\mathbf{u}_L,\mathbf{u}_R), \\ \mathbf{u}_R \text{ if } \quad x/t>\sigma(\mathbf{u}_L,\mathbf{u}_R), \end{cases} \tag{3.5}$$

where the left and right intermediate states are easily obtained from the left and right eigenvectors l_k and r_k, $k=1,2,3$ of $\mathcal{A}_\phi(\mathbf{u}_L,\mathbf{u}_R)$, respectively, namely

$$\mathbf{u}_L^* = (\mathbf{u}_R, l_1) r_1 + \sum_{k=2}^{3} (\mathbf{u}_L, l_k) r_k, \quad \mathbf{u}_R^* = \sum_{k=1}^{2} (\mathbf{u}_R, l_k) r_k + (\mathbf{u}_L, l_3) r_3,$$

and $\sigma(\mathbf{u}_L,\mathbf{u}_R) = c(\overline{\mathbf{u}}(\mathbf{u}_L,\mathbf{u}_R)) = \sqrt{\gamma\overline{p}/\overline{\tau}}(\mathbf{u}_L,\mathbf{u}_R)$. For the sake of clarity, it will be useful to have in mind the wave pattern of this solution, which is recalled on the next figure.

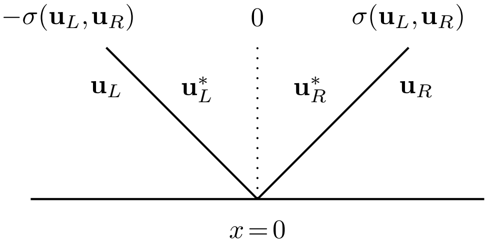
$-\sigma(\mathbf{u}_L,\mathbf{u}_R)$
0
$\sigma(\mathbf{u}_L,\mathbf{u}_R)$
$\mathbf{u}_L$
$\mathbf{u}_L^*$
$\mathbf{u}_R^*$
$\mathbf{u}_R$
$x=0$

FIG. 3.1. *Approximate Riemann solution constructed from the Roe linearization*

Note that once the solution is defined, one can denote by $x \to \tilde{\mathbf{u}}(x,t)$ the piecewise constant approximate solution obtained by glueing together the Roe-type approximate solutions at each interface, that is to say

$$\tilde{\mathbf{u}}(x,t) = \mathbf{u}((x - x_{j+1/2})/t; \mathbf{u}_j^n, \mathbf{u}_{j+1}^n)$$

for all $(x,t) \in [x_j, x_{j+1}) \times [0,\Delta t)$, $j \in \mathbb{Z}$, $n \in \mathbb{N}$. One can also define a Roe-type path-conservative scheme according to [35] as any Godunov-type scheme by averaging the solution on each cell $[x_{j-1/2}, x_{j+1/2})$, namely

$$\mathbf{u}_j^{n+1} = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \tilde{\mathbf{u}}(x,\Delta t) dx = \frac{1}{2}(\mathbf{u}_{j,L}^{n+1} + \mathbf{u}_{j,R}^{n+1}), \tag{3.6}$$

with

$$\begin{aligned}\mathbf{u}_{j,L}^{n+1} &= \frac{2}{\Delta x} \int_{x_{j-1/2}}^{x_j} \tilde{\mathbf{u}}(x,\Delta t) dx \\ &= \frac{2}{\Delta x}\Big(\sigma(\mathbf{u}_{j-1}^n,\mathbf{u}_j^n)\Delta t \mathbf{u}_R^*(\mathbf{u}_{j-1}^n,\mathbf{u}_j^n) + (\frac{\Delta x}{2} - \sigma(\mathbf{u}_{j-1}^n,\mathbf{u}_j^n)\Delta t)\mathbf{u}_j^n\Big)\end{aligned}$$

and

$$\begin{aligned}\mathbf{u}_{j,R}^{n+1} &= \frac{2}{\Delta x} \int_{x_j}^{x_{j+1/2}} \tilde{\mathbf{u}}(x,\Delta t) dx \\ &= \frac{2}{\Delta x}\Big(\sigma(\mathbf{u}_j^n,\mathbf{u}_{j+1}^n)\Delta t \mathbf{u}_L^*(\mathbf{u}_j^n,\mathbf{u}_{j+1}^n) + (\frac{\Delta x}{2} - \sigma(\mathbf{u}_j^n,\mathbf{u}_{j+1}^n)\Delta t)\mathbf{u}_j^n\Big),\end{aligned}$$

under the CFL restriction

$$\Delta t \max_{j\in\mathbb{Z}} |\sigma(\mathbf{u}_j^n,\mathbf{u}_{j+1}^n)| \leq \frac{\Delta x}{2}. \tag{3.7}$$

In the sequel, we will also use the notation $\sigma_{j+1/2}^n = \sigma(\mathbf{u}_j^n,\mathbf{u}_{j+1}^n)$. However, such a Roe-type path conservative scheme fails in computing correctly the discontinuous solutions of our system, see [1], and we now aim at applying the in-cell discontinuous reconstruction method instead. As we will see, such a strategy allows to obtain a perfect agreement between the exact and numerical solutions, and even the exact capturing of isolated discontinuities.

3.2. A path-conservative in-cell discontinuous numerical scheme. The design principle is the same as for the toy model in Section 2.1. The main differences here are the following: First, the exact Riemann solver will be replaced with a Roe-type approximate Riemann solver and second, the local Riemann solutions at each interface may contain two discontinuities propagating with velocities having opposite signs, unlike the toy model for which only one discontinuity propagating with a positive speed could occur. Apart from this, the idea is actually the same, namely to keep on using the classic Godunov-type scheme given above "far away" from shock discontinuities, and to add details in the piecewise constant representation of the approximate solution in the vicinity of shock discontinuities.

Reconstruction procedure. Let us first define which cells j are to be concerned with the reconstruction procedure. We consider the cell C_j and proceed as follows. Assume that at time t^n,

$$u^n_{j-1} > u^n_{j+1}. \tag{3.8}$$

According to the entropy condition $\sigma(\tau_+ - \tau_-) > 0$ and the Rankine-Hugoniot relation $\sigma(\tau_+ - \tau_-) = -(u_+ - u_-)$ across a shock discontinuity, we consider that a shock discontinuity is expected to appear locally around the cell C_j when (3.8) holds true. This is quite natural since such a shock is actually present in the Riemann solution associated with the inital states $\mathbf{u}^n_{j-1}$ and $\mathbf{u}^n_{j+1}$ and which will develop at the next times $t > t^n$. Hence, we are tempted to introduce in the cell C_j a discontinuity given by the Roe-type path-conservative approximate Riemann solver proposed in the previous section. More precisely and with clear notations the left and right states $\mathbf{u}_{j,l}$ and $\mathbf{u}_{j,r}$ of the reconstructed solution are defined by

$$\mathbf{u}_{j,l} = \mathbf{u}^n_{j-1} \quad \text{and} \quad \mathbf{u}^n_{j,r} = \mathbf{u}^*_L(\mathbf{u}^n_{j-1}, \mathbf{u}^n_{j+1}) \quad \text{if} \quad (\tau^n_{j+1} - \tau^n_{j-1}) < 0,$$

and

$$\mathbf{u}_{j,l} = \mathbf{u}^*_R(\mathbf{u}^n_{j-1}, \mathbf{u}^n_{j+1}) \quad \text{and} \quad \mathbf{u}^n_{j,r} = \mathbf{u}^n_{j+1} \quad \text{if} \quad (\tau^n_{j+1} - \tau^n_{j-1}) > 0.$$

The speed of propagation $\sigma_{j,l,r}$ of the reconstructed discontinuity on the cell j is naturally defined by $-\sigma(\mathbf{u}^n_{j-1}, \mathbf{u}^n_{j+1})$ if $(\tau^n_{j+1} - \tau^n_{j-1}) < 0$ and $\sigma(\mathbf{u}^n_{j-1}, \mathbf{u}^n_{j+1})$ if $(\tau^n_{j+1} - \tau^n_{j-1}) > 0$, where of course $\pm\sigma(\mathbf{u}^n_{j-1}, \mathbf{u}^n_{j+1})$ refer to the speeds of propagation of the discontinuities in the Roe-type approximate Riemann solver associated with the initial states $\mathbf{u}^n_{j-1}$ and $\mathbf{u}^n_{j+1}$.

Since we are considering the cell C_j, we also require that the reconstructed discontinuity associated with those left and right states is located inside C_j at a position

$$\bar{x}^\alpha_j = x_{j-1/2} + d^{n,\alpha}_j \Delta x, \tag{3.9}$$

with $\alpha = \tau, u, e$ and for some $d^{n,\alpha}_j$ in $[0,1]$ which may vary with α. In order to define $d^{n,\alpha}_j$, it is natural to impose that the reconstruction procedure is conservative, namely

$$d^{n,\alpha}_j \alpha^n_{j,l} + (1 - d^{n,\alpha}_j)\alpha^n_{j,r} = \alpha^n_j, \tag{3.10}$$

or equivalently,

$$d_j^{n,\alpha} = \frac{\alpha_{j,r}^n - \alpha_j^n}{\alpha_{j,r}^n - \alpha_{j,l}^n}, \tag{3.11}$$

for the $\alpha = \tau, u, e$ component. Clearly, it is possible to reconstruct the discontinuities provided that

$$0 \leq d_j^{n,\alpha} \leq 1, \tag{3.12}$$

which gives three additional conditions for the in-cell reconstruction procedure to make sense.

At last, for the sake of simplicity and in order to avoid dealing with the interaction of two reconstructed discontinuities in adjacent cells, no reconstruction will be considered in the cell C_j if (3.8) and (3.12) adapted to the cell C_{j+1} hold true and $\sigma_{j+1,l,r} < 0$, while no reconstruction will be considered in the cell C_{j-1} if (3.8) and (3.12) adapted to the cell C_{j-1} hold true and $\sigma_{j-1,l,r} > 0$.

Remark 3.1. In practice, we also impose to the reconstructed states to be admissible in the sense $\tau > 0$ and $e > 0$, which is not guaranteed by the Roe approximate solver.

Update formulas. Let us now give the update formulas for $\mathbf{u}_j^{n+1}$, as well as the influence of the in-cell reconstruction on the update formulas for $\mathbf{u}_{j-1}^{n+1}$ and $\mathbf{u}_{j+1}^{n+1}$ since $\sigma_{j,l,r}$ may be positive or negative. We follow exactly the same approach as for the toy model, which leads to the reconstructed discontinuity propagating with a positive speed in the cell C_j, and that no reconstruction will be considered in the cell C_{j-1} if the reconstructed discontinuity propagates with a negative speed in the cell C_j.

The case $\sigma_{j,l,r} > 0$ and the cell C_j. We set

$$\alpha_j^{n+1} = \alpha_j^n - \frac{\sigma_{j,l,r}(\alpha_{j,r}^n - \alpha_{j,l}^n)}{\Delta x} \times \min(\Delta t, \Delta t^\alpha)$$

where Δt^α is the time needed by the reconstructed discontinuity in $\alpha = \tau, u, e$ to reach the interface $x_{j+1/2}$, namely

$$\Delta t^\alpha = \frac{1 - d_j^{n,\alpha}}{\sigma_{j,l,r}} \Delta x.$$

The case $\sigma_{j,l,r} > 0$ and the cell C_{j+1}. Under the CFL condition (2.13), the reconstructed discontinuities in the cell C_j cannot reach the middle point x_{j+1} of the cell C_{j+1} and may thus influence only the half interval $[x_{j+1/2}, x_{j+1})$. Since no reconstruction is considered in the cell C_{j+1}, we consider the usual update formula

$$\mathbf{u}_{j+1}^{n+1} = \frac{1}{2}\left(\mathbf{u}_{j+1,L}^{n+1} + \mathbf{u}_{j+1,R}^{n+1}\right)$$

but with (component by component)

$$\alpha_{j+1,L}^{n+1} = \alpha_{j+1}^n - \frac{2\sigma_{j,l,r}(\alpha_{j,r}^n - \alpha_{j,l}^n)}{\Delta x} \times \left(\Delta t - \min(\Delta t, \Delta t^\alpha)\right).$$

Note that compared to the usual Godunov scheme, the value of $\alpha_{j+1,R}^{n+1}$ will be changed if and only if an in-cell reconstruction takes place in cell C_{j+2}.

The case $\sigma_{j,l,r}>0$ and the cell C_{j-1} (and no reconstruction in this cell). Under the CFL condition (2.13), the reconstructed discontinuities in the cell C_j cannot influence the cell C_{j-1} farther than x_{j-1}. Since we consider the case where no reconstruction is considered in the cell C_{j-1}, we consider the usual update formula

$$\mathbf{u}_{j-1}^{n+1}=\frac{1}{2}\Big(\mathbf{u}_{j-1,L}^{n+1}+\mathbf{u}_{j-1,R}^{n+1}\Big)$$

but with (component by component)

$$\alpha_{j-1,R}^{n+1}=\alpha_{j-1}^{n}-\frac{2\sigma_{j,l,r}(\alpha_{j-1}^{n}-\alpha_{j,l}^{n})}{\Delta x}\Delta t.$$

The case $\sigma_{j,l,r}<0$ and the cell C_j. We follow along the same lines as above which leads to the same update formulas for τ, u and e, namely

$$\alpha_j^{n+1}=\alpha_j^n-\frac{\sigma_{j,l,r}(\alpha_{j,r}^{n}-\alpha_{j,l}^{n})}{\Delta x}\times\min(\Delta t,\Delta t^{\alpha})$$

component by component, where Δt^{α} is now the time needed by the reconstructed discontinuity in α to reach the interface $x_{j-1/2}$, namely

$$\Delta t^{\alpha}=\frac{d_j^{n,\alpha}}{|\sigma_{j,l,r}|}\Delta x.$$

The case $\sigma_{j,l,r}<0$ and the cell C_{j-1}. Under the CFL condition (2.13) and as before, the reconstructed discontinuities in the cell C_j cannot reach the middle point x_{j-1} of the cell C_{j-1} and may thus influence the half interval $[x_{j-1},x_{j-1/2})$ only. Since no reconstruction is considered in the cell C_{j-1}, we consider the usual update formula

$$\mathbf{u}_{j-1}^{n+1}=\frac{1}{2}\Big(\mathbf{u}_{j-1,L}^{n+1}+\mathbf{u}_{j-1,R}^{n+1}\Big)$$

but with (component by component)

$$\alpha_{j-1,R}^{n+1}=\alpha_{j-1}^{n}-\frac{2\sigma_{j,l,r}(\alpha_{j,r}^{n}-\alpha_{j,l}^{n})}{\Delta x}\times\big(\Delta t-\min(\Delta t,\Delta t^{\alpha})\big).$$

Note that compared to the usual Godunov scheme, the value of $\alpha_{j-1,L}^{n+1}$ will be changed if and only if an in-cell reconstruction with positive speed of propagation takes place in cell C_{j-2}.

The case $\sigma_{j,l,r}<0$ and the cell C_{j+1} (and no reconstruction in this cell). Under the CFL condition (2.13), the reconstructed discontinuities in the cell C_j cannot influence the cell C_{j+1} farther than x_{j+1}. Since we consider the case where no reconstruction is considered in the cell C_{j+1}, we consider the usual update formula

$$\mathbf{u}_{j-1}^{n+1}=\frac{1}{2}\Big(\mathbf{u}_{j-1,L}^{n+1}+\mathbf{u}_{j-1,R}^{n+1}\Big)$$

but with (component by component)

$$\alpha_{j+1,L}^{n+1}=\alpha_{j,r}^{n}-\frac{2\sigma_{j,l,r}(\alpha_{j+1}^{n}-\alpha_{j,r}^{n})}{\Delta x}\Delta t.$$

The case with no reconstruction in the cells C_j and $C_{j\pm1}$.

At last and to conclude the proposed numerical scheme, let us mention that when no reconstruction takes place in the cells C_{j-1}, C_j and C_{j+1}, we use the classical Godunov-type scheme, namely

$$\mathbf{u}_j^{n+1} = \frac{1}{2}\left(\mathbf{u}_{j,L}^{n+1} + \mathbf{u}_{j,R}^{n+1}\right).$$

Similar to the toy model, one can easily prove that the scheme satisfies by construction the following theorem.

Theorem 3.1. *Assume that $\mathbf{u}_j^0 = \mathbf{u}_L$ if $j \leq 0$, $\mathbf{u}_j^0 = \mathbf{u}_R$ if $j \geq 1$ and that $\mathbf{u}_L$ and $\mathbf{u}_R$ are two constant states in the phase space Ω such that they can be joined by an admissible (entropic) shock discontinuity. In other words, the Riemann solution associated with these left and right states is given by $\mathbf{u}(x,t) = \mathbf{u}_L$ if $x < \sigma t$ and $\mathbf{u}(x,t) = \mathbf{u}_R$ if $x > \sigma t$ where σ is the speed of propagation given by*

$$\sigma = \pm\sqrt{-\frac{p_R - p_L}{\tau_R - \tau_L}}.$$

Then the proposed scheme provides an exact numerical solution on each cell C_j in the sense that

$$\mathbf{u}_j^n = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{u}(x,t^n)dx, \qquad j \in \mathbb{Z},\ n \in \mathbb{N}. \tag{3.13}$$

In particular, the numerical discontinuity is diffused on one cell at most.

3.3. Numerical experiments. We now propose several test cases to illustrate the behavior of the scheme. The adiabatic coefficient is set to $\gamma = 1.4$. We compare the solutions with the ones given by the original path-conservative scheme applied to (3.1) or by a classical conservative scheme applied to (3.2). The domain is $[0,1]$ and the CFL restriction is 0.45. The first two cases are such that exact solutions are either an isolated discontinuity or two shock discontinuities starting from the same right state. The last test case is inspired from the first test case of [2] and has a large pressure jump.

Test 1. The first test case is an isolated shock associated with the initial data

$$(\tau,u,p)_0(x) = \begin{array}{l} (2.09836065573770281, 2.3046638387921279, 1.0) \text{ if } x < 0.5, \\ (8.0, 0.0, 0.1) \text{ otherwise.} \end{array}$$

The speed of propagation is 0.39051248379533265 44238 and the final time of the simulation is $t = 0.5$. We clearly see on Figure 3.2 that the original path-conservative scheme fails while Figure 3.3 shows a perfect agreement between our scheme and the exact solution. Recall that our scheme is exact in this case and therefore captures the discontinuity with only one point of numerical diffusion.

Test 2. The second test case is a Riemann problem leading to three waves, namely two shocks and one contact discontinuity, and corresponds to the following initial data,

$$(\tau,u,p)_0(x) = \begin{array}{l} (5.0, 3.323013993227, 0.481481481481) \text{ if } x < 0.5, \\ (8.0, 0.0, 0.1) \text{ otherwise.} \end{array}$$

The 3-shock is the same as in the previous test case. The density of the first shock goes from 5.0 to 3.0 and its speed of propagation is 0.509175077217. The final time is 0.5. Again, we observe on Figures 3.4 and 3.5 that the original path-conservative scheme

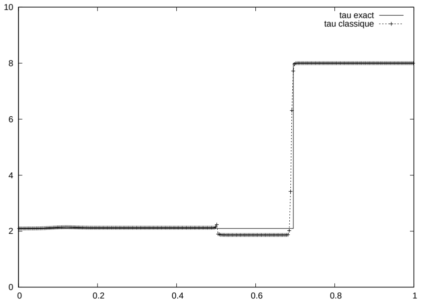

FIG. 3.2. *τ - Test 1 - Classical path-conservative scheme - Final time $t = 0.5$ - 300-point mesh*

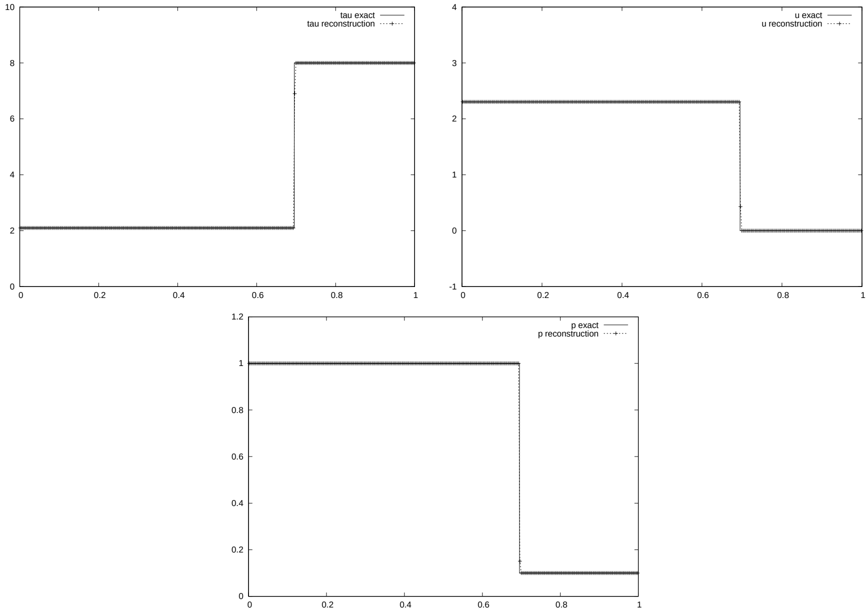

FIG. 3.3. *τ (top left), u (top right) and p (bottom) - Test 1 - Our scheme - Final time $t = 0.5$ - 300-point mesh*

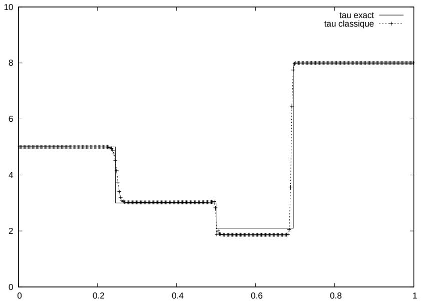

FIG. 3.4. *τ - Test 2 - Classical path-conservative scheme - Final time $t = 0.5$ - 300-point mesh*

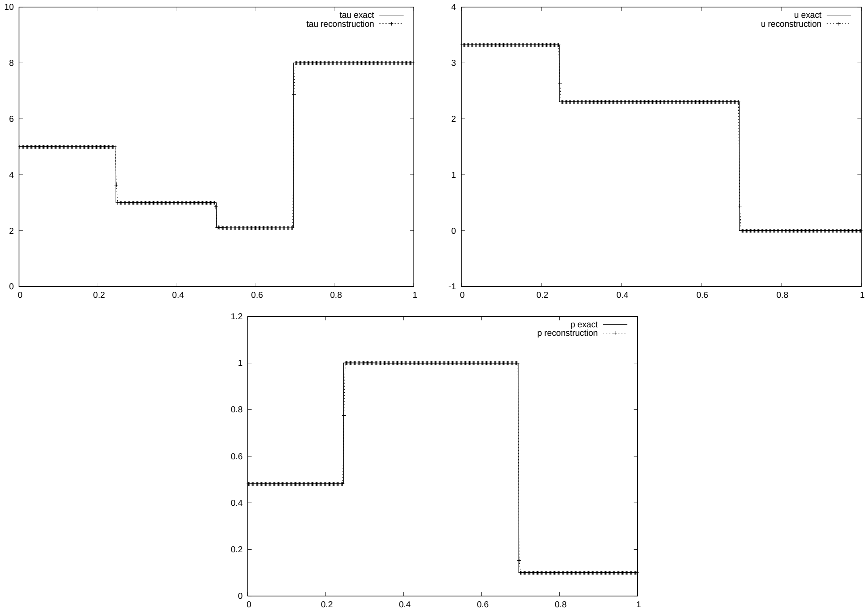

FIG. 3.5. *τ (top left), u (top right) and p (bottom) - Test 2 - Our scheme - Final time $t = 0.5$ - 300-point mesh*

fails and the new one succeeds and properly computes the shocks without numerical diffusion.

On Figure 3.6, we plot the numerical energy dissipation with respect to time and defined by

$$D^n = \sum_j \Delta x E_j^{n+1} - \sum_j \Delta x E_j^n + \Delta t (p_R u_R - p_L u_L)$$

where the sum is taken over the mesh cells. This quantity is clearly zero for a conservative scheme. It is also expected to converge to zero with the mesh size for a convergent non-conservative scheme since our choice of path is equivalent to the classical Rankine-Hugoniot relations applied to the conservative system. It is actually the case for our scheme based on in-cell reconstructions. Interestingly, we observe that the energy dissipation oscillates around zero for a given mesh (the amplitude goes to zero with the mesh size).

Test 3. At last, we conclude this section with a more difficult test case taken from [2] and with a large pressure jump in the initial data given by

$$(\tau, u, p)_0(x) = \begin{array}{l} (1/1185, 0, 2.0e11) \text{ if } x < 0.5, \\ (1/1185, 0, 1.0e5) \text{ otherwise,} \end{array}$$

where the density, velocity and pressure units are respectively kg/m^3, m/s, and Pa. The final time of simulation is $2e-8$ and the mesh is made of 1000 points. We compare on Figure 3.7 the solution given by our scheme with the one given by the classical path-conservative scheme but applied directly to the conservative variable τ, u and E, so that it approximates correctly the solution in this case since both the system and the

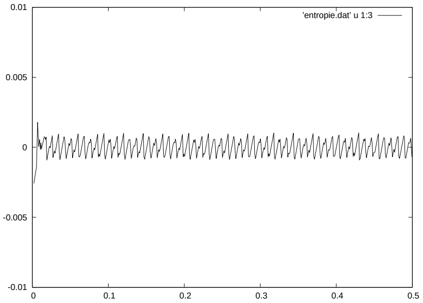

FIG. 3.6. *Energy dissipation - Test 2 - Our scheme - Final time* $t=0.5$ *- 300-point mesh*

scheme are conservative. Here again, we see that our (non-conservative) scheme gives similar results and thus is also able to properly approximate the exact solution.

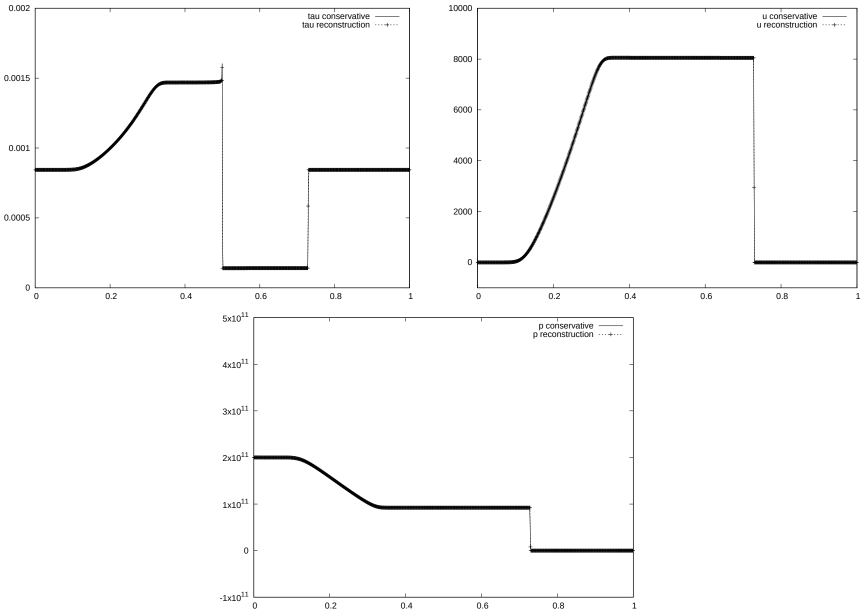

FIG. 3.7. τ *(top left),* u *(top right) and* p *(bottom) - Test 3 - Final time* $t=2e-8$ *- 1000-point mesh*

4. Conclusion and perspectives

We have introduced the basics of the so-called path-conservative in-cell discontinuous reconstruction schemes for the numerical approximation of shock solutions in non-conservative systems. By basics, we mean that it has been applied to quite simple systems, namely a toy model and the non-conservative gas dynamics equations in Lagrangian coordinates. The first (respectively second) system has one (resp. two) characteristic fields leading to shocks, but in both cases the sign of the corresponding characteristic speed is known a priori. The next steps are to consider systems for which the sign of the characteristic speeds depends on the state value, like for instance the gas dynamics equations in Eulerian coordinates and also to consider non-conservative systems which do not admit an equivalent conservative formulation like the one considered in the present paper. One can think for instance of systems arising in turbulence modeling or geophysical flows.

It is also important to note that a key property to make the in-cell discontinuous reconstruction approach successful, and in particular to have the validity of the theorem stating that it gives the exact solution in the case of isolated shocks, lies in the fact that the underlying Riemann solver (exact or of path-conservative Roe type) is able to provide an exact solution in such a case of an isolated shock discontinuity. This therefore emphasizes the need for the development of approximate and entropy-satisfying Riemann solvers which are able to exactly reproduce isolated shocks, which is to be proposed in a forthcoming contribution too.

At last, the scheme is first-order accurate in its present form, although it is ∞-accurate for isolated shocks. The extension to higher order of accuracy is also a current investigation. In particular, using the general high-order path-conservative formalism provides a nice opportunity to extend the present approach and design new high-order finite volume solvers that do not introduce any numerical viscosity on the propagation of isolated shocks, and to explore the extension to multidimensional problems.

Acknowledgments. The author is thankful to Frédéric Coquel and Pascal Jaisson for useful discussions on this work. The author is also very grateful to Manuel J. Castro Díaz and Tomás Morales de Luna for pointing out strong connections between the path-conservative methods and the numerical methods proposed here, and suggesting extensions to high-order accuracy and multidimensional setting in that direction.

Appendix. Exact Riemann solver for the non-conservative toy model. In this appendix, we briefly give the solution to the Riemann problem (2.1)-(2.5) with initial condition given by

$$\mathbf{u}(x,0)=\mathbf{u}_0(x)=\begin{cases}\mathbf{u}_L \text{ if } x<0,\\ \mathbf{u}_R \text{ if } x>0,\end{cases} \tag{5.1}$$

for two constant states $\mathbf{u}_L$ and $\mathbf{u}_R$ in Ω. By Lemma 2.1, this solution is expected to be made of two simple waves, namely a stationary contact discontinuity associated with λ_1 from $\mathbf{u}_L$ to an intermediate state $\mathbf{u}_\star$ and a nonlinear wave associated with λ_2 from $\mathbf{u}_\star$ to $\mathbf{u}_R$. The latter is either a shock discontinuity satisfying the generalized Rankine-Hugoniot relations (1.5) and the entropy inequality (2.5) in the sense of distributions, or a rarefaction wave. Let us go further into details.

Contact discontinuities. As is customary, the set $\mathcal{C}^1(\mathbf{u}_L)$ of admissible states $\mathbf{u}_\star$ that can be joined to $\mathbf{u}_L$ on the right by a contact discontinuity associated with λ_1 is defined thanks to the Riemann invariants. Here we get

$$\mathcal{C}^1(\mathbf{u}_L)=\{\mathbf{u}_\star=(u_\star,v_\star)^t\in\Omega, I_1(\mathbf{u}_\star)=I_1(\mathbf{u}_L)\}$$

or equivalently

$$\mathcal{C}^1(\mathbf{u}_L)=\{\mathbf{u}_\star=(u_\star,v_\star)^t\in\Omega, u_\star+v_\star=u_L+v_L\}.$$

Given $\mathbf{u}_\star$ in $\mathcal{C}^1(\mathbf{u}_L)$, the stationary contact discontinuity solution of (1.1) is then defined by

$$\mathbf{u}(x,t)=\begin{cases}\mathbf{u}_L \text{ if } x<0,\\ \mathbf{u}_\star \text{ if } x>0.\end{cases}$$

Rarefaction waves. The set $\mathcal{R}^1(\mathbf{u}_R)$ of admissible states $\mathbf{u}_\star$ that can be joined to $\mathbf{u}_R$ on the left by a rarefaction wave associated with λ_2 is also defined thanks to the Riemann invariants, together with the compatibility condition $\lambda_2(\mathbf{u}_\star) \leq \lambda_2(\mathbf{u}_R)$. More precisely, we have

$$\mathcal{R}^1(\mathbf{u}_R) = \{\mathbf{u}_\star = (u_\star, v_\star)^t \in \Omega, I_2(\mathbf{u}_\star) = I_2(\mathbf{u}_R), \lambda_2(\mathbf{u}_\star) \leq \lambda_2(\mathbf{u}_R)\}$$

or equivalently

$$\mathcal{R}^1(\mathbf{u}_R) = \{\mathbf{u}_\star = (u_\star, v_\star)^t \in \Omega, v_R u_\star = u_R v_\star, u_\star + v_\star \leq u_R + v_R\}.$$

Given $\mathbf{u}_\star$ in $\mathcal{R}^1(\mathbf{u}_R)$, the rarefaction fan solution of (1.1) is then defined by

$$\mathbf{u}(x,t) = \begin{cases} \mathbf{u}_\star & \text{if } \xi \leq \lambda_2(\mathbf{u}_\star) = u_\star + v_\star, \\ \mathbf{u}_\star(\xi) & \text{if } \lambda_2(\mathbf{u}_\star) \leq \xi \leq \lambda_2(\mathbf{u}_R), \\ \mathbf{u}_R & \text{if } \xi \geq \lambda_2(\mathbf{u}_R) = u_R + v_R, \end{cases}$$

where we have set $\xi = x/t$ for $t > 0$ and where $\mathbf{u}_\star(\xi)$ is defined by

$$\begin{cases} \xi = \lambda_2(\mathbf{u}_\star(\xi)) \\ I_2(\mathbf{u}_\star(\xi)) = I_2(\mathbf{u}_R) \end{cases}$$

or equivalently

$$\begin{cases} \xi = u_\star(\xi) + v_\star(\xi) \\ v_R u_\star(\xi) = u_R v_\star(\xi). \end{cases}$$

We refer for instance to [23] for more details.

Shock discontinuities. As motivated above, the set $\mathcal{S}^2(\mathbf{u}_R)$ of admissible states $\mathbf{u}_\star$ in Ω that can be joined to $\mathbf{u}_R$ on the left by a shock discontinuity propagating at velocity σ is admissible provided that both the generalized Rankine-Hugoniot relations (1.5) and the entropy inequality (2.5) hold true in the distributional sense. More precisely, $\mathbf{u}_\star$ has to satisfy

$$\begin{cases} -\sigma(\mathbf{u}_R - \mathbf{u}_\star) + \displaystyle\int_0^1 \mathcal{A}(\phi(s, \mathbf{u}_\star, \mathbf{u}_R)) \frac{\partial \phi}{\partial s}(s, \mathbf{u}_\star, \mathbf{u}_R) ds = 0, \\ -\sigma\big(f(u_R + v_R) - f(u_\star + v_\star)\big) + \displaystyle\int_{u_\star + v_\star}^{u_R + v_R} s f'(s) ds \leq 0, \end{cases} \tag{5.2}$$

where ϕ and f respectively denote a family of paths and any convex function. Adding the two components of the generalized Rankine-Hugoniot relations in (5.2) gives

$$-\sigma\big((u_R + v_R) - (u_\star + v_\star)\big) + \frac{1}{2}\big((u_R + v_R)^2 - (u_\star + v_\star)^2\big),$$

which in passing does not depend on the family of paths ϕ anymore, and then

$$\sigma = \frac{(u_\star + v_\star) + (u_R + v_R)}{2}.$$

Note that we have implicitly assumed that $u_\star + v_\star \neq u_R + v_R$ in order to deal with a true shock discontinuity and not a contact discontinuity. Then, since

$$-\sigma\big(f(u_R + v_R) - f(u_\star + v_\star)\big) + \int_{u_\star + v_\star}^{u_R + v_R} s f'(s) ds$$

$$\begin{aligned}
&= \int_{u_\star+v_\star}^{u_R+v_R} (s-\sigma) f'(s) ds \\
&= -\int_{u_\star+v_\star}^{u_R+v_R} \frac{(s-\sigma)^2}{2} f''(s) ds + \frac{(u_R+v_R-\sigma)^2}{2} f'(u_R+v_R) - \frac{(u_\star+v_\star-\sigma)^2}{2} f'(u_\star+v_\star),
\end{aligned}$$

the definition of σ above and the mean value theorem give

$$\begin{aligned}
&-\sigma\big(f(u_R+v_R)-f(u_\star+v_\star)\big) + \int_{u_\star+v_\star}^{u_R+v_R} s f'(s) ds \\
&= -\frac{(\tilde{s}-\sigma)^2}{2}\big(f'(u_R+v_R)-f'(u_\star+v_\star)\big) \\
&\quad + \frac{1}{8}\big((u_R+v_R)-(u_\star+v_\star)\big)^2 \big(f'(u_R+v_R)-f'(u_\star+v_\star)\big)
\end{aligned}$$

for some $\tilde{s}$ in between $(u_\star+v_\star)$ and (u_R+v_R), that is to say

$$\begin{aligned}
&-\sigma\big(f(u_R+v_R)-f(u_\star+v_\star)\big) + \int_{u_\star+v_\star}^{u_R+v_R} s f'(s) ds \\
&= -\frac{1}{2}\big(f'(u_R+v_R)-f'(u_\star+v_\star)\big) \times \big(\tilde{s}-(u_\star+v_\star)\big) \times \big(\tilde{s}-(u_R+v_R)\big).
\end{aligned}$$

By convexity of f, it is thus clear that the entropy inequality in (5.2) is equivalent to

$$u_\star + v_\star \geq u_R + v_R.$$

The set $\mathcal{S}^2(\mathbf{u}_R)$ is then defined by

$$\begin{aligned}
\mathcal{S}^2(\mathbf{u}_R) = \{ & \mathbf{u}_\star = (u_\star, v_\star)^t \in \Omega, u_\star + v_\star \geq u_R + v_R, \\
& -\sigma(\mathbf{u}_R - \mathbf{u}_\star) + \int_0^1 \mathcal{A}(\phi(s,\mathbf{u}_\star,\mathbf{u}_R)) \frac{\partial \phi}{\partial s}(s,\mathbf{u}_\star,\mathbf{u}_R) ds = 0 \}
\end{aligned}$$

for a given family of paths. Given $\mathbf{u}_\star$ in $\mathcal{S}^2(\mathbf{u}_R)$, the shock solution of (1.1) is then defined by

$$\mathbf{u}(x,t) = \begin{cases} \mathbf{u}_\star & \text{if } x < \sigma t, \\ \mathbf{u}_R & \text{if } x > \sigma t. \end{cases}$$

The Riemann solution. Glueing together the simple waves associated with λ_1 and λ_2 and for a given family of paths ϕ, we get that the Riemann solution to (2.1)-(2.5)-(5.1) is given as follows:

- if $(u_L+v_L) \leq (u_R+v_R)$

$$\mathbf{u}(x,t) = \begin{cases} \mathbf{u}_L & \text{if } \xi < 0, \\ \mathbf{u}_\star & \text{if } 0 < \xi < \lambda_2(\mathbf{u}_\star) = u_\star + v_\star, \\ \mathbf{u}_\star(\xi) & \text{if } \lambda_2(\mathbf{u}_\star) \leq \xi \leq \lambda_2(\mathbf{u}_R), \\ \mathbf{u}_R & \text{if } \xi \geq \lambda_2(\mathbf{u}_R) = u_R + v_R, \end{cases}$$

with $\xi = x/t$ and where $\mathbf{u}_\star$ and $\mathbf{u}_\star(\xi)$ are respectively defined by

$$\begin{cases} u_L + v_L = u_\star + v_\star, \\ v_R u_\star = u_R v_\star, \end{cases}$$

which gives in particular

$$u_\star = u_R \frac{u_L + v_L}{u_R + v_R}, \quad v_\star = v_R \frac{u_L + v_L}{u_R + v_R},$$

and

$$\begin{cases} \xi = u_\star(\xi) + v_\star(\xi), \\ v_R u_\star(\xi) = u_R v_\star(\xi). \end{cases}$$

- if $(u_L + v_L) \geq (u_R + v_R)$

$$\mathbf{u}(x,t) = \begin{cases} \mathbf{u}_L \text{ if } \xi < 0, \\ \mathbf{u}_\star \text{ if } 0 < \xi < \sigma, \\ \mathbf{u}_R \text{ if } \xi \geq \sigma, \end{cases}$$

with $\xi = x/t$ and where σ and $\mathbf{u}_\star$ are defined by

$$\begin{cases} u_L + v_L = u_\star + v_\star, \\ -\sigma(\mathbf{u}_R - \mathbf{u}_\star) + \int_0^1 \mathcal{A}(\phi(s,\mathbf{u}_\star,\mathbf{u}_R)) \frac{\partial \phi}{\partial s}(s,\mathbf{u}_\star,\mathbf{u}_R) ds = 0. \end{cases}$$

In particular, we still have $\sigma = \frac{(u_\star + v_\star) + (u_R + v_R)}{2}$.

REFERENCES

[1] R. Abgrall and S. Karni, *A comment on the computation of non-conservative products*, J. Comput. Phys., 229:2759–2763, 2010. 1, 3.1

[2] R. Abgrall and H. Kumar, *Numerical approximation of a compressible multiphase system*, Comm. Comput. Phys., 15(5):1237–1265, 2014. 3.3, 3.3

[3] N. Aguillon, *Capturing nonclassical shocks in nonlinear elastodynamic with a conservative finite volume scheme*, Interface. Free Bound., 18(2):137–159, 2016. 1, 2.1.1

[4] B. Audebert and F. Coquel, *Hybrid Godunov-Glimm method for a nonconservative hyperbolic system with kinetic relations*, in A. de Castro, D. Gómez, P. Quintela, and P. Salgado (eds.), Numer. Math. Adv. Appl., Springer Berlin Heidelberg, 646–653, 2006. 1

[5] C. Berthon, *Schéma nonlinéaire pour l'aproximation numérique d'un système hyperbolique non conservatif*, C.R. Acad. Sci. Paris., Ser. I, 335:1069–1072, 2002. 1, 2, 2, 2.1.1

[6] C. Berthon and F. Coquel, *Multiple solutions for compressible turbulent flow models*, Commun. Math. Sci., 4:497–511, 2006. 2, 2.1.1

[7] C. Berthon and F. Coquel, *Nonlinear projection methods for multi-entropies Navier-Stokes systems*, Math. Comput., 76:1163–1194, 2007. 1

[8] C. Berthon and F. Coquel, *Shock layers for turbulence models*, Math. Mod. Meth. Appl. Sci., 18:1443–1479, 2008. 2, 2.1.1

[9] C. Berthon, F. Coquel, and P.G. LeFloch, *Why many theories of shock waves are necessary: kinetic relations for non-conservative systems*, Proc. Roy. Soc. Edinb. A, 142:1–37, 2012. 1, 2

[10] B. Boutin, C. Chalons, F. Lagoutière, and P.G. LeFloch, *Convergent and conservative schemes for nonclassical solutions based on kinetic relations. I*, Interface. Free Bound., 10(3):399–421, 2008. 1, 2.1.1, 2.1.2, 2.1.2

[11] M.J. Castro, U.S. Fjordholm, S. Mishra, and C. Parés, *Entropy conservative and entropy stable schemes for nonconservative hyperbolic systems*, SIAM J. Numer. Anal., 51(3):1371–1391, 2013. 1

[12] M.J. Castro, P.G. LeFloch, M.L. Muñoz-Ruiz, and C. Parés, *Why many theories of shock waves are necessary: Convergence error in formally path-consistent schemes*, J. Comput. Phys., 227(17):8107–8129, 2008. 1

[13] C. Chalons and F. Coquel, *Navier-Stokes equations with several independent pressure laws and explicit predictor-corrector schemes*, Numerisch Math., 101:451–487, 2005. 1, 2, 2.1.1

[14] C. Chalons and F. Coquel, *The Riemann problem for the multi-pressure Euler system*, J. Hyperbolic Differ. Equ., 2:745–782, 2005. 2, 2.1.1

[15] C. Chalons and F. Coquel, *A new comment on the computation of non conservative products using Roe-type path conservative schemes*, J. Comput. Phys., 335:592–604, 2017. 1

[16] C. Chalons, M.L. Delle Monache, and P. Goatin, *A conservative scheme for non-classical solutions to a strongly coupled PDE-ODE problem*, Interface. Free Bound., 19(4):553–570, 2017. 1

[17] C. Chalons, P. Goatin, and N. Seguin, *General constrained conservation laws. Application to pedestrian flow modeling*, Netw. Heterog. Media, 8(2):433–463, 2013. 1

[18] P. Collela, *Glimm's method for gas dynamics*, SIAM J. Sci. Stat. Comput., 3:76–110, 1982. 1

[19] G. Dal Maso, P.-G. LeFloch, and F. Murat, *Definition and weak stability of a non conservative product*, J. Math. Pures Appl., 74:483–548, 1995. 1, 2

[20] B. Després and F. Lagoutière, *Contact discontinuity capturing schemes for linear advection and compressible gas dynamics*, J. Sci. Comput., 16(4):479–524, 2001. 1

[21] U. Fjordholm and S. Mishra, *Accurate numerical discretizations of non-conservative hyperbolic systems*, M2AN Math. Mod. Numer. Anal., 46(1):187–206, 2012. 1

[22] J. Glimm, *Solutions in the large time for nonlinear hyperbolic systems of equations*, Comm. Pure Appl. Math., 18:697–715, 1965. 1

[23] E. Godlewski and P.-A. Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, Springer, 1995. 2, 3, 4

[24] T.Y. Hou and P.G. Le Floch, *Why nonconservative schemes converge to wrong solutions: Error analysis*, Math. Comput., 62(206):497–530, 1994. 1

[25] F. Lagoutière, *A non-dissipative entropic scheme for convex scalar equations via discontinuous cell-reconstruction*, C.R. Math. Acad. Sci. Paris, 338(7):549–554, 2004. 1

[26] F. Lagoutière, *Stability of reconstruction schemes for scalar hyperbolic conservations laws*, Commun. Math. Sci., 6(1):57–70, 2008. 1

[27] P.D. Lax, *Hyperbolic systems of conservation laws. II*, Comm. Pure Appl. Math., 10:537–566, 1957. 1

[28] P.D. Lax, *Conservation Laws and the Mathematical Theory of Shock Waves*, CBMS Monograph, Soc. Indust. Appl. Math., Philadelphia, 67(2):153, 1973. 1

[29] P. LeFloch and S. Mishra, *Numerical methods with controlled dissipation for small-scale dependent shocks*, Acta Numer., 23:743–816, 2014. 1

[30] P.-G. LeFloch, *Shock waves for nonlinear hyperbolic systems in nonconservative form*, Institute for Mathematics and its Applications, Minneapolis, 593, 1989. 2, 2

[31] P.G. LeFloch, *Hyperbolic Systems of Conservation Laws: The Theory of Classical and Nonclassical Shock Waves*, Lecture Notes in Mathematics, ETH Zurich, Birkhauser, 2002. 1, 2.1.1

[32] T.-P. Liu, *The Riemann problem for general systems of conservation laws*, J. Diff. Eqs., 18:218–234, 1975. 1

[33] T.-P. Liu, *The deterministic version of the Glimm scheme*, Comm. Math. Phys., 57:135–148, 1977. 1

[34] C. Munz, *On Godunov-type schemes for Lagrangian gas dynamics*, SIAM J. Numer. Anal., 31(1):17–42, 1994. 3.1

[35] C. Parès, *Numerical methods for non-conservative hyperbolic systems: a theoretical framework*, SIAM J. Numer. Anal., 44:300–321, 2006. 1, 3.1, 3.1

[36] P.-A. Raviart and L. Sainsaulieu, *A nonconservative hyperbolic system modeling spray dynamics. Part 1: solution of the Riemann problem*, Math. Models Meth. Appl. Sci., 5:297–333, 1995. 2

[37] L. Sainsaulieu, *Travelling waves solutions of convection-diffusion systems whose convection terms are weakly nonconservative*, SIAM J. Appl. Math., 55:1552–1576, 1995. 2, 2

[38] I. Toumi, *A weak formulation of Roe's approximate Riemann solver*, J. Comput. Phys., 102(2):360–373, 1992. 3.1

[39] S. Villa, P. Goatin, and C. Chalons, *Moving bottlenecks for the Aw-Rascle-Zhang traffic flow model*, Discrete Contin. Dyn. Syst. Ser. B, 22(10):3921–3952, 2017. 1

[40] A.-I. Volpert, *The spaces BV and quasilinear equations*, Math. Sbornik, 73:225–267, 1967. 2