



# A new comment on the computation of non conservative products using Roe-type path conservative schemes

Christophe Chalons, Frédéric Coquel

## ► To cite this version:

Christophe Chalons, Frédéric Coquel. A new comment on the computation of non conservative products using Roe-type path conservative schemes. 2016. <hal-01424306>

**HAL Id: hal-01424306**

**<https://hal.archives-ouvertes.fr/hal-01424306>**

Submitted on 2 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A new comment on the computation of non conservative products using Roe-type path conservative schemes

Christophe Chalons\*and Frédéric Coquel†

December 21, 2016

## Abstract

We are interested in the numerical approximation of the discontinuous solutions of non conservative hyperbolic systems. We more precisely consider a non conservative formulation of the usual gas dynamics equations and show how to slightly modify the so-called Roe-type path-conservative schemes to properly capture the underlying shock discontinuities. Numerical evidences are proposed. The present note follows a first comment on the computation of non conservative products in [1].

## 1 Introduction

We are interested in hyperbolic equations of the form

$$\begin{cases} \partial_t \mathbf{u} + \mathcal{A}(\mathbf{u})\partial_x \mathbf{u} = 0, & x \in \mathbb{R}, \quad t \in \mathbb{R}^{+,*}, \\ \mathbf{u}(x, 0) = \mathbf{u}_0(x), \end{cases} \quad (1)$$

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

$$\partial_t \mathcal{U}(\mathbf{u}) + \partial_x \mathcal{F}(\mathbf{u}) \leq 0, \quad (2)$$

where  $(\mathcal{U}, \mathcal{F})$  is an entropy-entropy flux pair (meaning  ${}^T \nabla \mathcal{U} \mathcal{A} = {}^T \nabla \mathcal{F}$  with  $\mathcal{U}$  strictly convex). From a theoretical point of view, it is well known that the solutions can develop discontinuities in finite time even if  $\mathbf{u}_0$  is smooth. It is therefore necessary to understand (1)-(2) in a weak sense in order to define a notion of discontinuous weak solution which makes sense mathematically and physically. If (1) is a system of conservation laws, that is if there exists a flux

---

\*Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035 Versailles, France. E-mail: christophe.chalons@uvsq.fr

†CNRS et Centre de Mathématiques Appliquées, U.M.R. 7641, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France. E-mail: frederic.coquel@cmap.polytechnique.fr

function  $\mathbf{f} : \mathbb{R}^p \rightarrow \mathbb{R}^p$  such that  $\mathcal{A} = \nabla \mathbf{f}$ , the notion of weak solution in the sense of distributions is used to define the solutions of (1)-(2). Assuming in addition that the characteristic fields are linearly degenerate or genuinely non linear, existence and uniqueness results have been obtained for initial data with bounded variation and close enough to a constant state (see for instance [27]). From a numerical point of view, there exists in this situation a large number of finite volume methods to approximate the smooth and discontinuous solutions of (1)-(2). The most famous of them are certainly the Godunov and Godunov-type methods whose fundamental ingredient is the (exact or approximate) Riemann solution, namely the solution with an initial data made of two constant states separated by a discontinuity (see for instance [21], [28]). The situation is much more intricate when the matrix  $\mathcal{A}$  is not the Jacobian matrix of a flux function since the theory of distributions does not apply. The non conservative product  $\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}$  has indeed a loose sense at a point of discontinuity of the solution  $\mathbf{u}$ . To overcome this difficulty, Dal Maso, LeFloch and Murat [19] introduced for functions with bounded total variation a notion of weak solution that allows to define the non conservative product  $\mathcal{A}(\mathbf{u})\partial_x \mathbf{u}$  and that generalizes the notion of weak solution in the sense of distributions for conservative systems. It is based on the theoretical notion of "path". Importantly, a path acts as a new degree of freedom of the model and a first difficulty consists in choosing the family of paths. Similarly to the conservative setting, this definition leads to existence and uniqueness results of weak solutions to (1)-(2) under certain assumptions (see for instance [23]). At the numerical level, imposing the path in the numerical method is still an open and very ambitious problem because of the sensitiveness of the traditional methods with respect to the choice of the path and the usual discretisation parameters, see for instance [22] and the more recent review [24]. Several attempts to impose the so-called "path" within the numerical scheme have been tried. Without any attempt to be exhaustive, they proved to be very successful in [3, 4, 5], [12, 13] (see also the references therein) but the strategy unfortunately highly relies on the structure of the models studied in the papers and generalizing the approach is not clear at all. More recently, C. Parés and collaborators developed in [26] and later works the so-called "path-conservative" schemes. Since then, it has been proved that this approach generally fails in converging to the right solutions, see [1], [10]. Actually, such a failure happens for any numerical method in which the small scale effects are not controlled, either if it is path-conservative or not. At last, a different strategy from the one proposed in the present contribution and that significantly reduces these convergence errors for the same problem has been introduced in [20] and generalized to general nonconservative systems within the framework of path-conservative methods in [11].

In [1], the authors more precisely consider the usual gas dynamics equations in Lagrangian coordinates and show that path-conservative schemes are not convergent when applied to a non conservative version of these equations. In the present paper, which structure follows closely Abgrall and Karni's [1], we consider the same equations and show how to slightly modify the usual "path-conservative" schemes to compute correctly the solutions of the non conservative

formulation of the gas dynamics equations. It is based on a slight modification of the underlying averaging procedure by introducing modified cells as already introduced in several quite recent contributions, see for instance [30], [17], [14], [15], [8], for a non exhaustive list. We also refer the reader to [2] for a hybrid Roe-Glimm scheme which spirit is similar in some sense to the strategy we derive here.

The content of the first four sections is well-known but necessary to set up the problem and the notations for the sake of clarity. The main contribution is thus contained in Sections 5 and 6 where the method is described and numerical experiments are proposed. At last, Section 7 contains some concluding remarks.

## 2 The model under consideration

We consider the following gas dynamics equations in Lagrangian coordinates :

$$\begin{cases} \partial_t \tau - \partial_x u = 0, \\ \partial_t u + \partial_x p = 0, \\ \partial_t E + \partial_x pu = 0, \end{cases} \quad (3)$$

where  $t$  is the time,  $x$  the spatial coordinate,  $\tau > 0$  represents the inverse of a density,  $u$  is the velocity and  $p = p(\tau, e) > 0$  is the pressure of the fluid. Here  $e > 0$  denotes the internal specific energy and satisfies  $E = e + u^2/2$ . In order to illustrate our purpose, it will be sufficient to focus on the case of a perfect gas equation of state. It will much simplify the forthcoming calculations but more general equations of state could be considered as well. Therefore we set  $p(\tau, e) = (\gamma - 1)e/\tau$  where  $\gamma > 1$  is the so-called adiabatic coefficient.

From a mathematical point of view, it is well known that (3) is strictly hyperbolic with eigenvalues  $\lambda_0 = 0$  and  $\lambda_{\pm} = \pm c$ ,  $c = \sqrt{\gamma p/\tau}$ . Moreover, the characteristic field associated with  $\lambda_0$  is linearly degenerate and the characteristic fields associated with  $\lambda_{\pm}$  are genuinely nonlinear [21]. For the sake of conciseness, we introduce the following short form

$$\partial_t \mathbf{u}(x, t) + \partial_x \mathbf{f}(\mathbf{u}(x, t)) = 0, \quad (4)$$

where the flux function  $\mathbf{f}$  finds a natural definition with respect to the conservative unknowns  $\mathbf{u} = (\tau, u, E)$ . At last, we supplement (4) with an initial data  $\mathbf{u}_0$  at time  $t = 0$ , that is to say  $\mathbf{u}(x, t = 0) = \mathbf{u}_0(x)$ .

A particular initial data is given by a Riemann initial data connecting two constant states  $\mathbf{u}_L$  and  $\mathbf{u}_R$  in the phase space ( $\tau$  and  $p$  are positive) as follows

$$\mathbf{u}_0(x) = \begin{cases} \mathbf{u}_L & \text{if } x < 0, \\ \mathbf{u}_R & \text{if } x > 0. \end{cases} \quad (5)$$

*Non conservative formulation.* Easy calculations show that for smooth solutions system (3) can be equivalently recast under the following non conservation form

$$\begin{cases} \partial_t \tau - \partial_x u = 0, \\ \partial_t u + \partial_x p = 0, \\ \partial_t e + p \partial_x u = 0, \end{cases} \quad (6)$$

where only the last equation has been changed. Using a chain rule argument, we also introduce a short form for (6), namely

$$\partial_t \mathbf{v}(x, t) + \mathcal{A}(\mathbf{v})(x, t) \partial_x \mathbf{v}(x, t) = 0, \quad (7)$$

where  $\mathbf{v} = (\tau, u, e)$  and

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

Note that the mapping  $\mathbf{u} \rightarrow \mathbf{v}$  is one-to-one so that the proposed change of variable is admissible. In particular, one can naturally define  $\mathbf{v}_0$  at time  $t = 0$  from  $\mathbf{u}_0$ , and thus  $\mathbf{v}(x, t = 0) = \mathbf{v}_0(x)$ .

### 3 Jump relations and admissible shock solutions

It is well-known that the solutions of hyperbolic systems generally develop discontinuities in finite time. Considering system (4), a discontinuity

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

between two states  $\mathbf{u}_-$  and  $\mathbf{u}_+$  and propagating with velocity  $\sigma$  is a weak solution if and only if it satisfies the Rankine-Hugoniot jump relations

$$-\sigma(\mathbf{u}_+ - \mathbf{u}_-) + \mathbf{f}(\mathbf{u}_+) - \mathbf{f}(\mathbf{u}_-) = 0.$$

After some calculations, these relations write equivalently

$$\begin{cases} \sigma(\tau_+ - \tau_-) + (u_+ - u_-) = 0, \\ -\sigma(u_+ - u_-) + (p_+ - p_-) = 0, \\ -\sigma(e_+ - e_-) + \frac{1}{2}(p_+ + p_-)(u_+ - u_-) = 0, \end{cases} \quad (8)$$

while usual admissibility criteria based on entropy consideration or related Lax criterion impose the following compatibility condition  $\sigma(\tau_+ - \tau_-) > 0$ , for shock discontinuities ( $\sigma \neq 0$ ). We refer for instance the reader to [21] for these more than classical considerations. Note however that the former criterion will be a key ingredient in the definition of our numerical strategy.

*Non conservative formulation.* If one now considers the non conservative formulation (7), the theory of distributions to define weak solutions does not make sense. To overcome this difficulty, Dal Maso, LeFloch and Murat [19] introduced a generalization of the notion of weak solution in the sense of distributions that allows to define the non conservative products. More precisely, it is based on the theoretical notion of path. A path  $\phi : [0, 1] \times \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$  associated with (7) satisfies the properties  $\phi(0, \mathbf{v}_-, \mathbf{v}_+) = \mathbf{v}_-$  and  $\phi(1, \mathbf{v}_-, \mathbf{v}_+) = \mathbf{v}_+$  and

serves to define the generalized Rankine-Hugoniot relations associated with a discontinuity

$$\mathbf{v}(x, t) = \begin{cases} \mathbf{v}_- & \text{if } x < \sigma t \\ \mathbf{v}_+ & \text{if } x > \sigma t \end{cases}$$

between two states  $\mathbf{v}_-$  and  $\mathbf{v}_+$  and propagating with velocity  $\sigma$  by setting

$$-\sigma(\mathbf{v}_+ - \mathbf{v}_-) + \int_0^1 \mathcal{A}(\phi(s, \mathbf{v}_-, \mathbf{v}_+)) \partial_s \phi(s, \mathbf{v}_-, \mathbf{v}_+) ds = 0. \quad (9)$$

As expected, this definition does not depend on the path when the matrix  $\mathcal{A}$  is the Jacobian matrix of a flux function like in formulation (4) and leads to the previous and usual Rankine-Hugoniot relations in such a situation.

A very simple choice of path which will be considered in the following is defined for all  $\mathbf{v}_-$  and  $\mathbf{v}_+$  such that  $\sigma(\tau_+ - \tau_-) > 0$  and is linear with respect to  $\tau$ ,  $u$  and  $p$ , namely

$$\begin{cases} \tau(s) = \tau_- + s(\tau_+ - \tau_-), \\ u(s) = u_- + s(u_+ - u_-), \\ p(s) = p_- + s(p_+ - p_-), \end{cases}$$

for all  $s \in [0, 1]$ . Easy calculations then show that the generalized jump relations (9) boil down to (8) so that both conservative and non conservative formulas are equivalent for such a choice of path, see [1].

## 4 Original Roe-type path conservative schemes

We briefly recall in this section the Roe-type path conservative schemes proposed in [26] to approximate the solutions of non conservative systems. It was shown in [1] that such a strategy fails in producing good approximations of admissible shock solutions when applied to system (7). The next section shows how to slightly modify this approach to make the numerical solutions perfectly match the exact ones. Let us first introduce some notations.

We denote by  $\Delta x$  and  $\Delta t$  the space and time steps, respectively. We define the mesh interfaces  $x_{j+1/2} = j\Delta x$  for  $j \in \mathbb{Z}$  and the intermediate times  $t^n = n\Delta t$  for  $n \in \mathbb{N}$ . We also define the mid-points  $x_j = (x_{j-1/2} + x_{j+1/2})/2$  for  $j \in \mathbb{Z}$ . At each time  $t^n$ ,  $\mathbf{v}_j^n$  represents an approximation of the exact solution on the interval  $[x_{j-1/2}, x_{j+1/2}]$ ,  $j \in \mathbb{Z}$ , and we look for an approximation of the solution at time  $t^{n+1}$ . At time  $t = 0$  we set

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

According to [29] and [26], Roe-type path conservative schemes associated with (7) and a given path  $\phi$  are based on a Roe linearization  $\mathcal{A}_\phi$  satisfying the properties

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

3. for all  $\mathbf{v}_L$  and  $\mathbf{v}_R$ ,

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

In the sequel, we will follow [25] and set

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

with

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

which is easily seen to satisfy the three properties. The approximate Riemann solution constructed from the Roe linearization, namely the exact solution to the linear Riemann problem

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

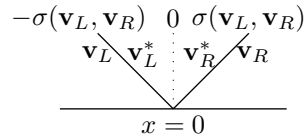
is then given by

$$\mathbf{v}(x/t; \mathbf{v}_L, \mathbf{v}_R) = \begin{cases} \mathbf{v}_L & \text{if } x/t < -\sigma(\mathbf{v}_L, \mathbf{v}_R), \\ \mathbf{v}_L^* & \text{if } -\sigma(\mathbf{v}_L, \mathbf{v}_R) < x/t < 0, \\ \mathbf{v}_R^* & \text{if } 0 < x/t < \sigma(\mathbf{v}_L, \mathbf{v}_R), \\ \mathbf{v}_R & \text{if } x/t > \sigma(\mathbf{v}_L, \mathbf{v}_R), \end{cases} \quad (10)$$

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  $\mathcal{A}_\phi(\mathbf{v}_L, \mathbf{v}_R)$ , respectively, namely

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

and  $\sigma(\mathbf{v}_L, \mathbf{v}_R) = c(\bar{\mathbf{v}}(\mathbf{v}_L, \mathbf{v}_R)) = \sqrt{\gamma \bar{p} / \bar{\tau}}(\mathbf{v}_L, \mathbf{v}_R)$ . 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.



Once the solution is defined, we denote by  $x \rightarrow \tilde{\mathbf{v}}(x, t)$  the piecewise constant approximate solution obtained by glueing together the Roe-type approximate solutions at each interface, that is to say

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

for all  $(x, t) \in [x_j, x_{j+1}) \times [0, \Delta t)$ ,  $j \in \mathbb{Z}$ ,  $n \in \mathbb{N}$ . The Roe-type path conservative scheme is then defined according to [26] as any Godunov-type scheme by averaging the solution on each cell  $[x_{j-1/2}, x_{j+1/2})$ , namely

$$\mathbf{v}_j^{n+1} = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \tilde{\mathbf{v}}(x, \Delta t) dx, \quad (11)$$

under the CFL restriction

$$\Delta t \max_{j \in \mathbb{Z}} |\sigma(\mathbf{v}_j^n, \mathbf{v}_{j+1}^n)| \leq \frac{\Delta x}{2}. \quad (12)$$

Easy calculations give the following form for the update formula,

$$\mathbf{v}_j^{n+1} = \mathbf{v}_j^n - \frac{\Delta t}{\Delta x} (D_{j-1/2}^+ + D_{j+1/2}^-), \quad D_{j+1/2}^\pm = \mathcal{A}_\phi^\pm(\mathbf{v}_j^n, \mathbf{v}_{j+1}^n)(\mathbf{v}_{j+1}^n - \mathbf{v}_j^n),$$

where  $\pm$  refers to the positive and negative part of the matrices. In the sequel, we will also use the notation  $\sigma_{j+1/2}^n = \sigma(\mathbf{v}_j^n, \mathbf{v}_{j+1}^n)$ .

## 5 Modified Roe-type path conservative schemes

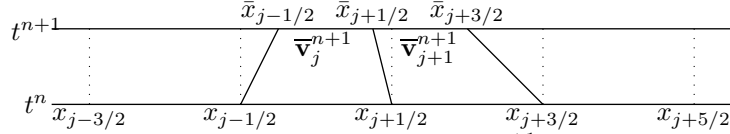
In this section, we present a *slight* modification of the original Roe-type path conservative scheme in order to compute the admissible shocks properly. By *slight*, it is meant that only the average step of the method will be changed. The proposed strategy was already introduced in several quite recent contributions, see for instance [30], [17], [14], [15], [8], for a non exhaustive list. It involves modified cells and random sampling, which objective is to get sharp discontinuities, *i.e.* without numerical diffusion. Controlling the numerical diffusion associated with the average step of Roe-type path conservative schemes, and more generally of any Godunov-type scheme, is a key point to approximate shocks in non conservative systems. We refer for instance the reader to [3], [4], [5], [7], [12], [13] for more details.

As already underlined in the previous section, the Roe-type path conservative scheme defines the update value at the new time step by simply averaging *on each fixed cells*  $(x_{j-1/2}, x_{j+1/2})$  and at time  $t = \Delta t$  the solution made of the juxtaposition of the Roe-type approximate Riemann solutions defined by (9), according to (11). Instead, we propose to average the solution at time  $t = \Delta t$  *on modified cells*  $(\bar{x}_{j-1/2}, \bar{x}_{j+1/2})$  defined by  $\bar{x}_{j+1/2} = x_{j+1/2} + s_{j+1/2} \Delta t$ , where  $s_{j+1/2}$  has to be defined. Note that  $s_{j+1/2}$  is expected to represent an approximation of the speed of propagation of an admissible shock in the Riemann problem set at interface  $x_{j+1/2}$ , so that we will have  $s_{j+1/2}^n = -\sigma_{j+1/2}^n$  or  $s_{j+1/2}^n = +\sigma_{j+1/2}^n$ . More precisely, we propose to set

$$\bar{\mathbf{v}}_j^{n+1} = \frac{1}{\Delta x_j} \int_{\bar{x}_{j-1/2}}^{\bar{x}_{j+1/2}} \tilde{\mathbf{v}}(x, \Delta t) dx, \quad (13)$$



with  $\overline{\Delta x}_j = \bar{x}_{j+1/2} - \bar{x}_{j-1/2}$ , as illustrated on the figure below.



In order to recover the initial mesh and define  $\mathbf{v}_j^{n+1}$  without introducing numerical diffusion, we then proceed with a random sampling procedure similar to the one proposed in the random choice method [18]. More precisely, let be given  $(a_n)_n$  a well-distributed random sequence within  $(0, 1)$  (e.g. the van der Corput sequence), we simply set

$$\mathbf{v}_j^{n+1} = \begin{cases} \bar{\mathbf{v}}_{j-1}^{n+1} & \text{if } a_{n+1} \in (0, \frac{\Delta t}{\Delta x} s_{j-1/2}^{n,+}) \\ \bar{\mathbf{v}}_j^{n+1} & \text{if } a_{n+1} \in [\frac{\Delta t}{\Delta x} s_{j-1/2}^{n,+}, 1 + \frac{\Delta t}{\Delta x} s_{j+1/2}^{n,-}) \\ \bar{\mathbf{v}}_{j+1}^{n+1} & \text{if } a_{n+1} \in [1 + \frac{\Delta t}{\Delta x} s_{j+1/2}^{n,-}, 1) \end{cases} \quad (14)$$

where we have set  $s_{j+1/2}^{n,+} = \max(s_{j+1/2}^n, 0)$  and  $s_{j+1/2}^{n,-} = \min(s_{j+1/2}^n, 0)$ .

*Definition of  $s_{j+1/2}^n$ .* In agreement with the selection principle  $\sigma(\tau_+ - \tau_-) > 0$  introduced in Section 3, we simply set

$$s_{j+1/2}^n = \begin{cases} -\sigma_{j+1/2}^n & \text{if } (\tau_{j+1}^n - \tau_j^n) \leq 0, \\ \sigma_{j+1/2}^n & \text{if } (\tau_{j+1}^n - \tau_j^n) > 0. \end{cases}$$

On the other hand, if the Riemann solution at interface  $x_{j+1/2}$  consists in a stationary contact discontinuity ( $u_j = u_{j+1}$  and  $p_j = p_{j+1}$ ), which we write here

$$|u_j - u_{j+1}| \leq \epsilon \quad \text{and} \quad |p_j - p_{j+1}| \leq \epsilon \quad (15)$$

for a given threshold  $\epsilon$ , we naturally set  $s_{j+1/2}^n = 0$ . In practice, we set  $\epsilon = 1e-8$ .

We thus have the following result.

**Theorem 5.1** *Under the CFL restriction (12), the scheme defined by (13)-(14) is consistent with (7) in the following sense :*

- (i) *Constant state : Assume that  $\mathbf{v} := \mathbf{v}_{j-1}^n = \mathbf{v}_j^n = \mathbf{v}_{j+1}^n$ , then  $\mathbf{v}_j^{n+1} = \mathbf{v}$ .*
- (ii) *Isolated shock discontinuity : Let  $\mathbf{v}_-$  and  $\mathbf{v}_+$  be two constant states that can be connected by an admissible discontinuity in the sense of Section 3. We denote by  $\sigma$  the speed of propagation. Assume that  $\mathbf{v}_j^0 = \mathbf{v}_-$  if  $j \leq 0$  and  $\mathbf{v}_j^0 = \mathbf{v}_+$  if  $j > 0$ . Then the scheme defined by (13)-(14) is equivalent to Glimm's random choice scheme and then converges to the exact solution of (7) given by  $\mathbf{v}(x, t) = \mathbf{v}_-$  if  $x < \sigma t$  and  $\mathbf{v}(x, t) = \mathbf{v}_+$  if  $x > \sigma t$ . In particular, we have  $\mathbf{v}_j^n \in \{\mathbf{v}_-, \mathbf{v}_+\} \forall j \in \mathbb{Z}$  and  $\forall n \in \mathbb{N}$  so that no spurious state is introduced in the numerical solution.*

**Proof.** The proof of the first point is trivial. Regarding the second point, it suffices to recall that the Roe linearization leads to an exact solution when it corresponds to an isolated shock wave. The rest of the proof is immediate and the convergence result of Glimm’s scheme can be found for instance in [27].

The second point of the theorem shows in particular that the proposed scheme is able to compute properly and without numerical diffusion isolated shock waves associated with the non conservative system (7). Actually, it coincides with the random choice method in such a specific case. The next section devoted to numerical experiments shows that the numerical solutions perfectly match the exact ones also for more general initial data still guaranteeing no numerical diffusion.

## 6 Numerical experiments in Lagrangian and Eulerian coordinates

Several test cases are proposed in this section to illustrate the behavior of the proposed numerical scheme. The adiabatic coefficient is set to be equal to  $\gamma = 1.4$ . We compare the solutions given by two numerical schemes, namely the original path-conservative scheme and the proposed numerical scheme. We use a mesh with 300 cells per unit of length and a CFL restriction of 0.45. The results are displayed with the  $x$  coordinate on the horizontal axis, and the unknowns specified in the figure captions on the vertical axis. The results are first given in Lagrangian coordinates, see Figures 1, 2, 3 and 4, but also in Eulerian coordinates, see Figures 7, and 8, following the direct extension of the proposed method described in appendix A. Note that the first two test cases below have been cooked up in such a way that exact solutions are either an isolated discontinuity or two shock discontinuities starting from the same right state, which explains the very specific values. In addition, these values can be understood in dimensionless form. The last test case in Eulerian coordinates is more realistic, it is inspired from the first test case of [2] and involves a very large pressure jump across the shock.

*Test 1.* The first test case consists in an isolated shock discontinuity :

$$(\tau, u, p)_0(x) = \begin{cases} (2.09836065573770281, 2.3046638387921279, 1.0) & \text{if } x < 0, \\ (8.0, 0.0, 0.1) & \text{if } x > 0. \end{cases}$$

The speed of propagation equals 0.3905124837953326544238. The final time of simulation is  $t = 0.5$ . As expected, we observe on Figure 1 that the original path-conservative scheme fails in producing good numerical results and creates spurious additional waves. On the contrary, Figures 1 and 2 show that the proposed numerical scheme captures the discontinuity without numerical diffusion since again, it coincides with the random choice method in this case. It is thus convergent and first-order accurate as proved in [27].

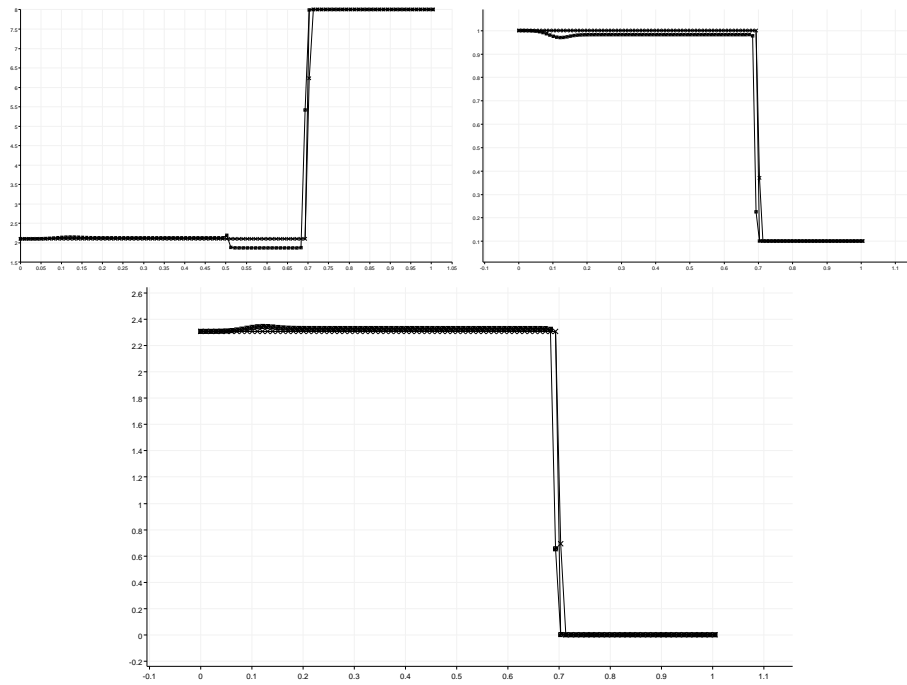


Figure 1: **Test 1 in Lagrangian coordinates** :  $\tau$  (Left),  $p$  (Right) and  $u$  (Bottom). Exact solution (solid line), our scheme (cross), path-conservative scheme (square)

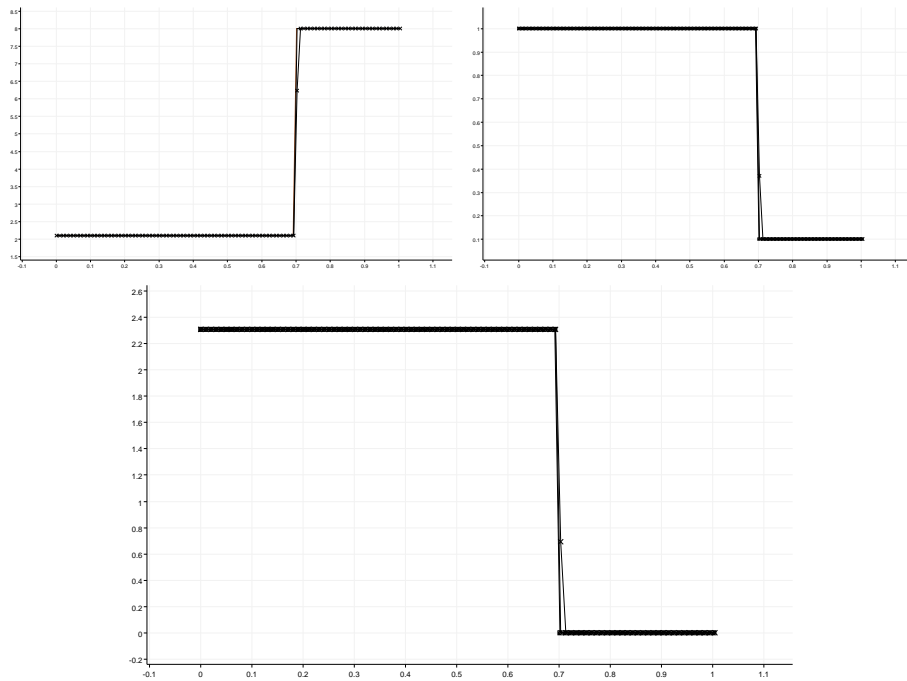


Figure 2: **Test 1 in Lagrangian coordinates** :  $\tau$  (Left),  $p$  (Right) and  $u$  (Bottom). Exact solution (solid line), our scheme with 300 points (cross), our scheme with 800 points (square)

*Test 2.* The second test case is a given Riemann problem leading to three simple waves :

$$(\tau, u, p)_0(x) = \begin{cases} (5.0, 3.323013993227, 0.481481481481) & \text{if } x < 0, \\ (8.0, 0.0, 0.1) & \text{if } x > 0. \end{cases}$$

The 3-shock is the same as in the previous test case. The density of the 1-shock jumps from 5.0 to 3.0 and the discontinuity propagates with velocity  $-0.509175077217$ . The final time of simulation is  $t = 0.5$ . Again, we observe on Figure 3 that the original path-conservative scheme fails in producing good numerical results (the constant states are not properly restored), while the proposed numerical scheme provides the right states and captures the shock discontinuities without numerical diffusion. We also observe on Figure 3 that the position of the 2-shock is shifted with respect to the real position, which comes from the proposed random sampling. Actually, the position gets closer and closer to the right position when the mesh is refined as observed on Figure 4 for a 800-point mesh. Regarding the intermediate constant states in the solutions, Figure 5 proposes a zoom of the plateau by giving the errors between the exact and approximate values at given space positions. It shows that both values are extremely close for the three components  $\tau$ ,  $u$  and  $p$ . At last, Figure 6 gives the  $L^1$ -errors for different meshes and for the three variables  $\tau$ ,  $u$  and  $p$ . The rate of convergence appears to be less than one and we postulate that it comes from the numerical diffusion introduced across the contact discontinuity.

We follow this section by giving the results of the same two test cases but in Eulerian coordinates, see appendix A for more details. We plot  $\rho$ ,  $u$  and  $p$  and the final times are  $t = 0.1$  for both test cases. Note that the final time is shorter than in Lagrangian coordinates in order for the waves to remain in the computational domain. The meshes are made of 300 points and the solutions given by our scheme are now compared to the ones given by the usual method of Roe applied to the conservative equations. The latter allows for a proper computation of the left and right states of the shocks with the usual numerical diffusion in between. The objective is indeed to compare the numerical diffusion of both strategies. As it can be clearly seen, these numerical results again indicate that the non conservative shocks are properly computed by our scheme and with no numerical diffusion, compared to the usual (conservative) Roe's method. Note also that on Figure 7, the usual method of Roe applied to the conservative equations suffers from startup errors, leading to little wiggles.

At last, we conclude this section with a more realistic test case (Test 3) inspired from the first test case of [2] and involving a very large pressure jump in the initial data and across the shock wave. More precisely, we take

$$(\rho, u, p)_0(x) = \begin{cases} (1185, 0, 2.0e11) & \text{if } x < 0, \\ (1185, 0, 1.0e5) & \text{if } x > 0, \end{cases}$$

where the density, velocity and pressure units are respectively  $kg/m^3$ ,  $m/s$ , and  $Pa$ . The final time of simulation is  $t = 2e-5$  and the mesh is made of 500

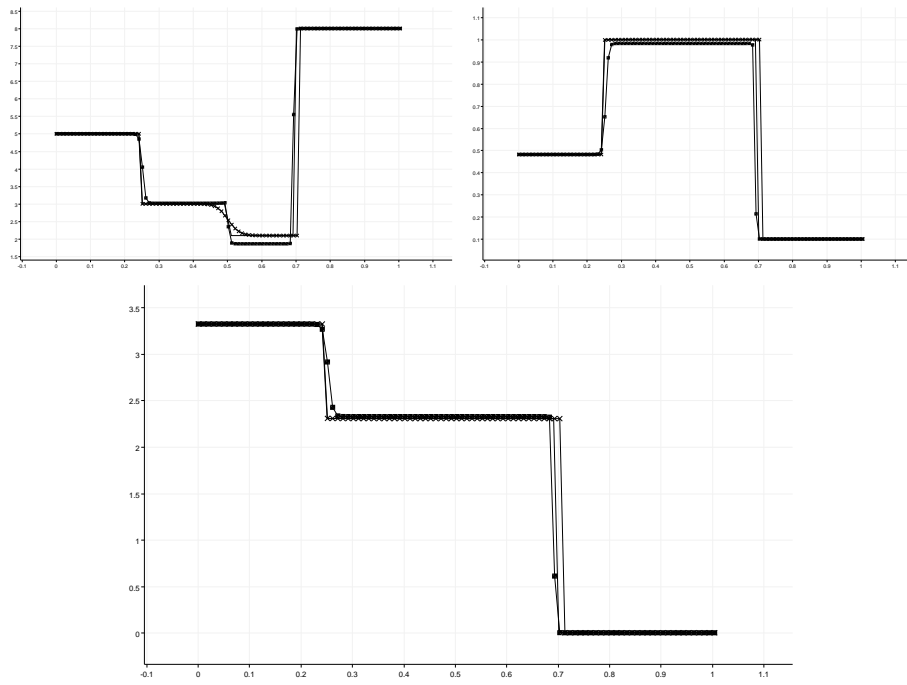


Figure 3: **Test 2 in Lagrangian coordinates** :  $\tau$  (Left),  $p$  (Right) and  $u$  (Bottom). Exact solution (solid line), our scheme (cross), path-conservative scheme (square)

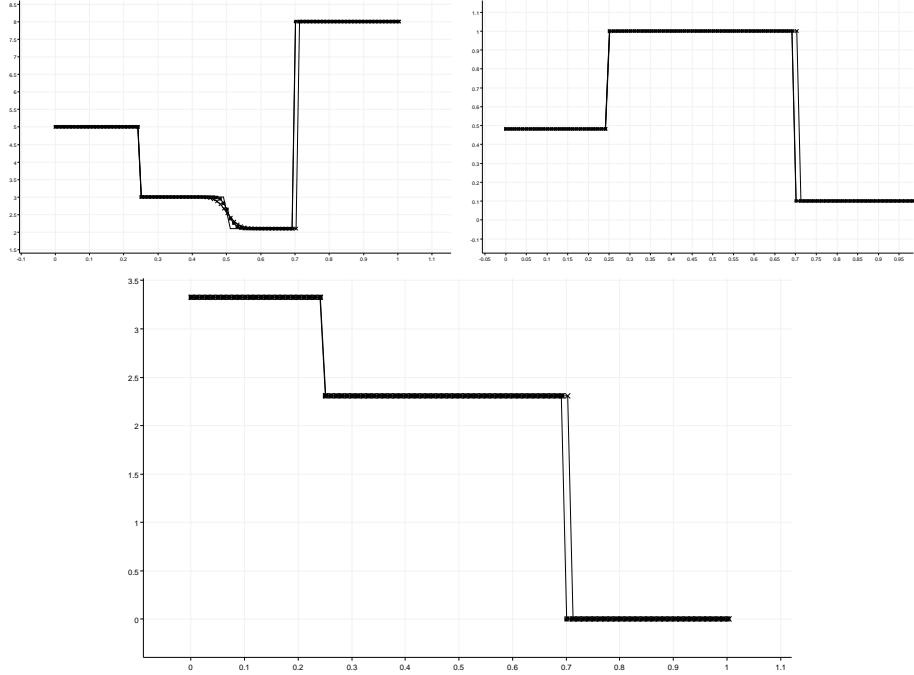


Figure 4: **Test 2 in Lagrangian coordinates** :  $\tau$  (Left),  $p$  (Right) and  $u$  (Bottom). Exact solution (solid line), our scheme with 300 points (cross), our scheme with 800 points (square)

	$ \tau - \tau_e (x = 0.38, t = 0.5)$	$ \tau - \tau_e (x = 0.6, t = 0.5)$
300 points	5.57e-04	9.12e-04
800 points	1.18e-04	3.46e-05
	$ u - u_e (x = 0.5, t = 0.5)$	$ p - p_e (x = 0.5, t = 0.5)$
300 points	8.24e-05	2.45e-05
800 points	1.16e-05	2.10e-06

Figure 5: **Test 2** : Zoom on the plateau errors for Figure 4

	$\ \tau - \tau_e\ _{L^1}$	$\ u - u_e\ _{L^1}$	$\ p - p_e\ _{L^1}$
100 points	0.11547288	0.03454186	0.01494763
200 points	0.08453120	0.02339142	0.00920177
400 points	0.06201163	0.01734427	0.00678331
800 points	0.02879795	0.00704321	0.00290254
1600 points	0.01742338	0.00432291	0.00168837

Figure 6: **Test 2** :  $L^1$ -errors

points. As expected, we observe on Figure 9 that the proposed method gives similar results compared to the usual method of Roe applied to the conservative system, but with less numerical diffusion across the shock.

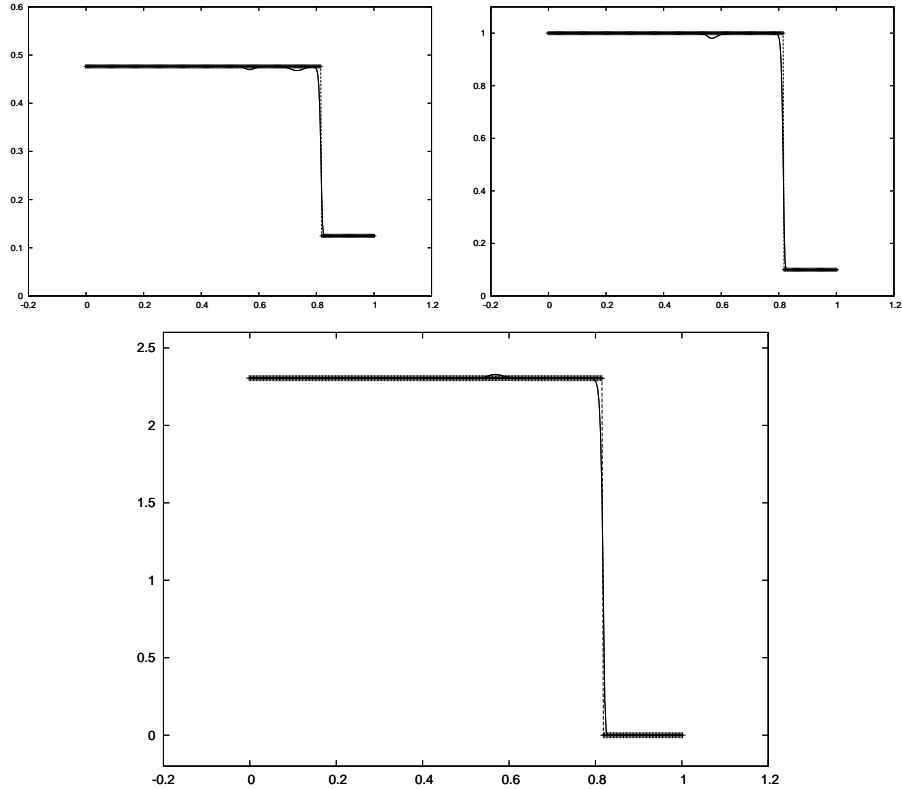


Figure 7: **Test 1 in Eulerian coordinates** :  $\rho$  (Left),  $p$  (Right) and  $u$  (Bottom). Conservative Roe's scheme (solid line) and our scheme (with lines points) with 300 points

## 7 Concluding remarks

A key property in the proof of Theorem 5.1 lies in the fact that the Roe scheme is able to provide an exact solution in the case of an isolated shock discontinuity. Note however that Roe's scheme is not entropic without correction for general systems. This therefore emphasizes the need for the development of such approximate Riemann solvers which are entropic and able to exactly reproduce isolated discontinuities. We refer the reader to [9], [16] for contributions on this topic.



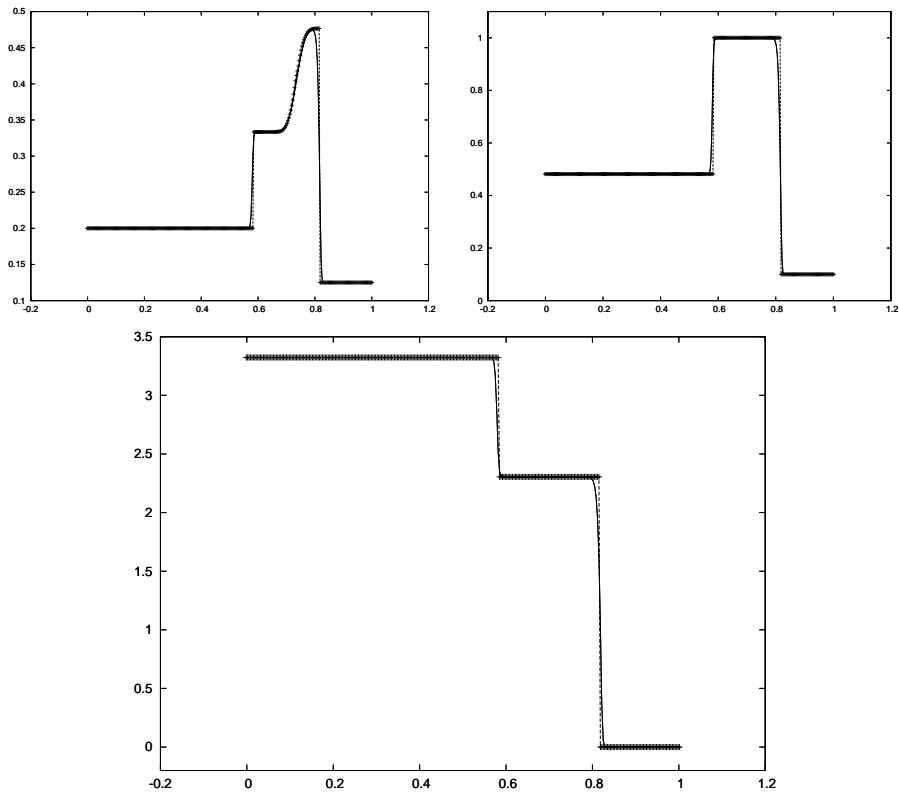


Figure 8: **Test 2 in Eulerian coordinates** :  $\rho$  (Left),  $p$  (Right) and  $u$  (Bottom). Conservative Roe's scheme (solid line) and our scheme (with lines points) with 300 points

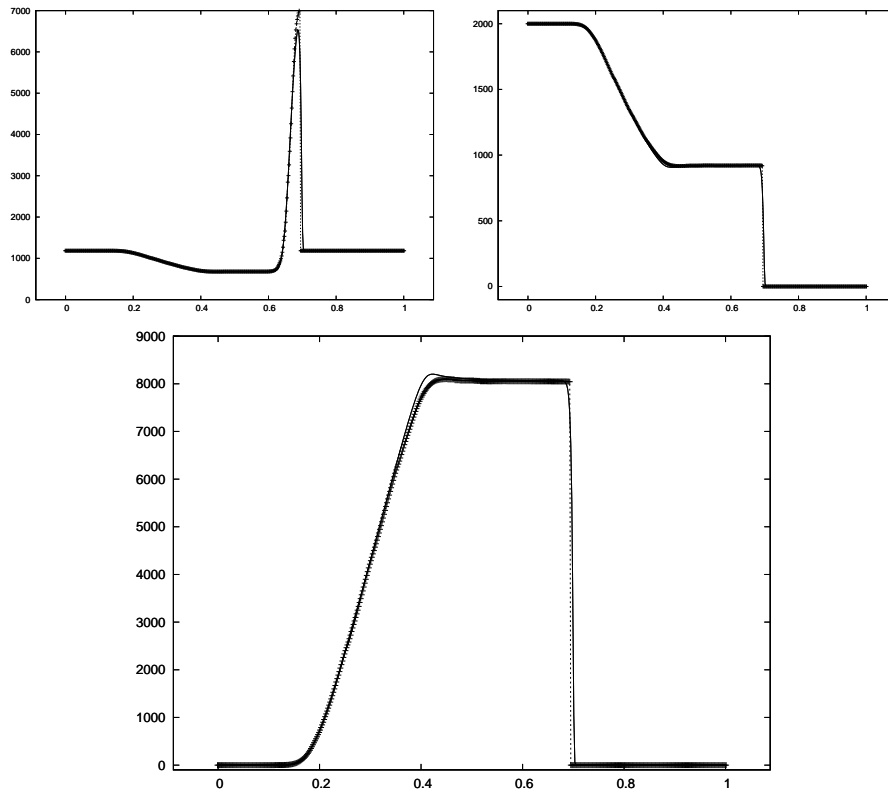


Figure 9: **Test 3 in Eulerian coordinates** :  $\rho$  (Left),  $p \times 10^{-8}$  (Right) and  $u$  (Bottom). Conservative Roe's scheme (solid line) and our scheme (with lines points) with 500 points

The proposed numerical scheme is only statistically conservative on the conservative variables  $\tau$  and  $u$  because of the random sampling, but it does not prevent the method from converging, numerically, to the right solution.

In one space dimension, the random sampling strategy is not mandatory. It could be easily replaced by considering moving meshes with non constant space steps, see for instance [30], which would give a conservative scheme on a moving mesh. In several space dimensions, it is well-known from the work by Colella [18] that the random sampling strategy cannot be used as it stands if one simply considers directional splitting methods. An alternative is currently under preparation.

At last, the proposed method is currently being applied to more general systems, in particular *really* non conservative systems (by *really* it is meant that they do not admit a natural conservative form) of interest for various applications.

## A The non conservative formulation in Eulerian coordinates

We briefly give in this section the basic material used in the previous section to obtain the numerical results in Eulerian coordinates. As we will see, the approach followed is a direct extension of the Lagrangian framework.

First of all, the conservative form of the gas dynamics equations in Eulerian coordinates writes

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x (\rho u^2 + p) = 0, \\ \partial_t \rho E + \partial_x (\rho E u + p u) = 0, \end{cases}$$

where  $\rho$  is the density and the other quantities have already been defined before. The non conservative formulation under consideration in this section is the following natural extension of the Lagrangian framework, namely

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x (\rho u^2 + p) = 0, \\ \partial_t \rho e + \partial_x \rho e u + p \partial_x u = 0, \end{cases}$$

where only the last equation has been changed. It can be given the following short form

$$\partial_t \mathbf{w}(x, t) + \mathcal{B}(\mathbf{w})(x, t) \partial_x \mathbf{w}(x, t) = 0,$$

where  $\mathbf{w} = (\rho, \rho u, \rho e)$  and

$$\mathcal{B}(\mathbf{w}) = \begin{pmatrix} 0 & 1 & 0 \\ -u^2 & 2u & \gamma - 1 \\ -\gamma e u & \gamma e & u \end{pmatrix}.$$

If we denote by  $\mathbf{w}_-$  and  $\mathbf{w}_+$  the left and right states of a shock discontinuity, and  $\sigma$  its speed of propagation, the Rankine-Hugoniot relations associated with the conservative formulation are (see for instance [21])

$$\begin{cases} -\sigma(\rho_+ - \rho_-) + (\rho_+ u_+ - \rho_- u_-) = 0, \\ -\sigma(\rho_+ u_+ - \rho_- u_-) + (\rho_+ u_+^2 - \rho_- u_-^2) + (p_+ - p_-) = 0, \\ -\sigma(\rho_+ e_+ - \rho_- e_-) + (\rho_+ e_+ u_+ - \rho_- e_- u_-) + \frac{1}{2}(p_+ + p_-)(u_+ - u_-) = 0, \end{cases}$$

while the entropy criteria imposes in particular that  $(\rho_+ - \rho_-) > 0$  for a 1-shock discontinuity associated with the first eigenvalue  $u - \sqrt{\gamma(\gamma-1)e}$  and  $(\rho_+ - \rho_-) < 0$  for a 3-shock discontinuity associated with the third eigenvalue  $u + \sqrt{\gamma(\gamma-1)e}$ . Recall that the second eigenvalue  $\lambda_2 = u$  is associated with a linearly degenerate characteristic field giving rise to contact discontinuities only. Like in the Lagrangian framework, the theory of Dal Maso, LeFloch and Murat [19] applies and leads to the generalized Rankine-Hugoniot relations

$$-\sigma(\mathbf{w}_+ - \mathbf{w}_-) + \int_0^1 \mathcal{B}(\phi(s, \mathbf{w}_-, \mathbf{w}_+)) \partial_s \phi(s, \mathbf{w}_-, \mathbf{w}_+) ds = 0$$

and the very simple choice of path given by

$$\begin{cases} \rho(s) = \rho_- + s(\rho_+ - \rho_-), \\ u(s) = u_- + s(u_+ - u_-), \\ p(s) = p_- + s(p_+ - p_-), \end{cases}$$

for all  $s \in [0, 1]$  allows to recover the original conservative Rankine-Hugoniot relations.

As a Roe linearization  $\mathcal{B}_\phi$ , we take

$$\mathcal{B}_\phi(\mathbf{w}_L, \mathbf{w}_R) = \begin{pmatrix} 0 & 1 & 0 \\ -\bar{u}^2 & 2\bar{u} & \gamma - 1 \\ -\gamma\bar{e}\bar{u} & \gamma\bar{e} & \bar{u} \end{pmatrix}.$$

where

$$\bar{\rho} = \frac{\rho_L + \rho_R}{2}, \quad \bar{u} = \frac{u_L + u_R}{2}, \quad \bar{\rho e} = \frac{\rho_L e_L + \rho_R e_R}{2}, \quad \tilde{e} = \frac{\bar{\rho e}}{\bar{\rho}},$$

and

$$\bar{u} = \alpha u_L + (1 - \alpha) u_R, \quad \alpha = \frac{\sqrt{\rho_L}}{\sqrt{\rho_L} + \sqrt{\rho_R}}.$$

It is interesting to note that  $\mathcal{B}_\phi(\mathbf{w}_L, \mathbf{w}_R)$  is not the matrix  $\mathcal{B}$  applied to an average state since the averages involved in its second and third rows are not the same. However, such a choice is natural in order to get the third required property (which can be easily checked) of a Roe linearization, namely

3. for all  $\mathbf{w}_L$  and  $\mathbf{w}_R$ ,

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

Since the proposed averages are clearly consistent, it is also easy to check that the second property holds true, namely that

2. for all  $\mathbf{w}$ ,  $\mathcal{B}_\phi(\mathbf{w}, \mathbf{w}) = \mathcal{B}(\mathbf{w})$ .

In regards to the first property

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

it is not always satisfied since the eigenvalues of  $\mathcal{B}_\phi(\mathbf{w}_L, \mathbf{w}_R)$  are shown to be

$$\mu_1 = \bar{u} - \sqrt{\gamma(\gamma-1)}\bar{e}, \quad \mu_2 = \bar{u}, \quad \mu_3 = \bar{u} + \sqrt{\gamma(\gamma-1)}\bar{e}.$$

As a consequence, while clearly  $\mu_1 < \mu_3$ , it might happen that  $\mu_1 = \mu_2$  or  $\mu_2 = \mu_3$  since  $\bar{u} \neq \bar{u}$  in general. Property 1. is thus satisfied under the condition  $\mu_1 < \mu_2 < \mu_3$ . In practice, this condition turned out to be satisfied for the test cases proposed in the numerical section.

To conclude, let us underline that the objective of this section, and more generally of this paper, was to show that a relevant projection step might give a numerical strategy to be able to properly compute the non conservative shocks and that, it can be considered in both Lagrangian or Eulerian coordinates. However and considering the present section in Eulerian coordinates, the proposed approach is certainly not the best one since Property 1. is only conditionally satisfied. Considering a Lagrange-Projection strategy [21] (based on the Roe linearization introduced in the Lagrangian framework and followed by a transport step) or the development of a new approximate Riemann solver which is entropic and exact for isolated shock waves, might be more relevant. Going further into the details of this strategy will be the purpose of a forthcoming contribution, see the concluding section above.

## References

- [1] R. Abgrall, S. Karni, A comment on the computation of non conservative products, *J. Comput. Phys.*, vol 229, (2010), pp 2759–2763.
- [2] R. Abgrall, H. Kumar, Numerical approximation of a compressible multi-phase system, *Commun. Comput. Phys.*, 15 (2014), pp. 1237-1265.
- [3] C. Berthon, Contribution à l'analyse numérique des équations de Navier-Stokes compressibles à deux entropies spécifiques. Applications à la turbulence compressible, Thèse de doctorat de l'Université Pierre et Marie Curie Paris 6, 1999.
- [4] C. Berthon, Schéma nonlinéaire pour l'approximation numérique d'un système hyperbolique non conservatif, *C. R. Acad. Sci. Paris, Ser. I*, 335, pp. 1069-1072 (2002).
- [5] C. Berthon and F. Coquel, Nonlinear projection methods for multi-entropies Navier-Stokes systems, *Math. Comput.*, 76, pp. 1163-1194 (2007)
- [6] C. Berthon and F. Coquel, Shock layers for turbulence models, *M3AS*, 18, No 8, pp. 1443-1479 (2008).

- [7] C. Berthon, F. Coquel, P. G. LeFloch, Why many theories of shock waves are necessary : Kinetic relations for nonconservative systems, Proc. Royal Soc. Edinburgh, 142 (2012), pp. 1-37.
- [8] M. Bachmann, P. Helluy, J. Jung, H. Mathis and S. Muller, Random Sampling Remap For Compressible Two-Phase Flows, Computers and Fluids, 86, 275–283, 2013.
- [9] F. Bouchut, T. Morales, Semi-discrete entropy satisfying approximate Riemann solvers. The case of the Suliciu relaxation approximation, Journal of Scientific Computing, 41(3):483509, (2009)
- [10] M. J. Castro, P. G. LeFloch, M. L. Munoz, and C. Parés, Why many theories of shock waves are necessary: Convergence error in formally path-consistent schemes, J. Comput. Phys. Vol. 227(17), pp. 8107–8129 (2008)
- [11] M.J. Castro, U.S. Fjordholm, S. Mishra, C. Parés, Entropy conservative and entropy stable schemes for nonconservative hyperbolic systems, SIAM Journal on Numerical Analysis 51 (2013), 1371-1391.
- [12] C. Chalons, F. Coquel, Navier-Stokes equations with several independent pressure laws and explicit predictor-corrector schemes, Numerisch Math., vol 101(3), pp 451-478 (2005)
- [13] C. Chalons, F. Coquel, Euler equations with several independent pressure laws and entropy satisfying explicit projection schemes, Math. Models Methods Appl. Sci. (M3AS), vol 16, No 9, pp 1469-1504 (2006)
- [14] C. Chalons, F. Coquel, Capturing infinitely sharp discrete shock profiles with the Godunov scheme, Proceedings of the Eleventh International Conference on Hyperbolic Problems. Theory, Numerics, Applications, S. Benzoni-Gavage and D. Serre (Eds), Springer, pp 363-370 (2008)
- [15] C. Chalons, F. Coquel, Computing material fronts with a Lagrange-Projection approach Volume 1, 346356, Ser. Contemp. Appl. Math. CAM, 17, World Sci. Publishing, Singapore, (2012), Proceedings of HYP2010, available at <http://hal.archives-ouvertes.fr/hal-00548938/>
- [16] C. Chalons, F. Coquel, Modified Suliciu relaxation system and exact resolution of isolated shock waves, Math. Models Methods Appl. Sci. (M3AS), Vol. 24, No. 5 937971 (2014)
- [17] C. Chalons, P. Goatin, Godunov scheme and sampling technique for computing phase transitions in traffic flow modeling, Interface Free Bound. vol 10, pp 195-219 (2008)
- [18] Collela P., Glimm’s method for gas dynamics, SIAM J. Sci. Stat. Comput., vol 3, pp 76–110 (1982).

- [19] G. Dal Maso, P.-G. LeFloch, F. Murat, Definition and weak stability of a non conservative product, *J. Math. Pures Appl.*, vol 74, (1995), pp 483–548.
- [20] U. S. Fjordholm, S. Mishra, Accurate numerical discretizations of non-conservative hyperbolic systems, *Math. Mod. and Num. Anal.*, 46 (2012), 187-206.
- [21] E. Godlewsky, P.-A. Raviart, Numerical approximation of hyperbolic systems of conservation laws, Springer (1995).
- [22] T.Y. Hou and P.G. LeFloch, Why nonconservative schemes converge to wrong solutions: error analysis, *Math. of Comput.* 62 (1994), 497–530.
- [23] P.-G. LeFloch, T.-P. Liu, Existence theory for nonlinear hyperbolic systems in nonconservative form, *Forum Math.*, pp. 261280 (1993).
- [24] P.G. LeFloch, S. Mishra, Numerical methods with controlled dissipation for small-scale dependent shocks, *Acta Numerica* 23 (2014), 743-816.
- [25] C. D. Munz, On Godunov-Type Schemes for Lagrangian Gas Dynamics, *SIAM Journal on Numerical Analysis*, Vol. 31, No. 1 (1994), pp. 17–42.
- [26] C. Parés, Numerical Methods for Nonconservative Hyperbolic Systems: A Theoretical Framework, *SIAM Journal on Numerical Analysis*, Vol. 44, No. 1 (2006), pp. 300–321.
- [27] D. Serre, *Systèmes de Lois de Conservation I et II*, Diderot, Paris (1996).
- [28] E.-F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. Third Edition. Springer-Verlag Berlin Heidelberg, 2009, 721 pages.
- [29] I. Tóuimi, A weak formulation of Roe’s approximate Riemann solver, *J. Comput. Phys.*, 102 (1992), pp. 360–373.
- [30] X.-G. Zhong, T.Y. Hou, P.-G. Lefloch, Computational methods for propagating phase boundaries. *J. Comput. Phys.* 124 (1996), 192216.