# In-cell discontinuous reconstruction path-conservative methods for non conservative hyperbolic systems - Second-order extension

Ernesto Pimentel-García, Manuel Jesús Castro, Christophe Chalons, Tomás Morales de Luna, Carlos Parés

Journal of Computational Physics 459 (2022) 111152

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

# In-cell discontinuous reconstruction path-conservative methods for non conservative hyperbolic systems - Second-order extension

Ernesto Pimentel-García [a,*], Manuel J. Castro [a], Christophe Chalons [b], Tomás Morales de Luna [c], Carlos Parés [a]

[a] Departamento de Análisis Matemático, Estadística e Investigación Operativa, y Matemática aplicada, Universidad de Málaga, Bulevar Louis Pasteur, 31, 29010, Málaga, Spain
[b] Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035 Versailles, France
[c] Departamento de Matemáticas, Universidad de Córdoba, Campus de Rabanales, 14071, Córdoba, Spain

ARTICLE INFO

*Article history:*
Received 12 April 2021
Received in revised form 2 January 2022
Accepted 14 March 2022
Available online 18 March 2022

*Keywords:*
In-cell reconstruction
Path-conservative methods
Shock-capturing methods
Finite volume methods
MUSCL-Hancock
Nonconservative hyperbolic systems

ABSTRACT

We are interested in the numerical approximation of discontinuous solutions in non conservative hyperbolic systems. An extension to second-order of a new strategy based on in-cell discontinuous reconstructions to deal with this challenging topic is presented. This extension is based on the combination of the first-order in-cell reconstruction with the standard MUSCL-Hancock reconstruction. The first-order strategy allowed in particular to capture exactly the isolated shocks and this new second-order extension keep this property. Moreover, the well-balanced property of the method is also studied. Several numerical tests are proposed to validate the methods for the Coupled-Burgers system, Gas dynamics equations in Lagrangian coordinates and the modified shallow water system.



## 1. Introduction

We consider first order quasi-linear PDE systems

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

in which the unknown $\mathbf{u}(x,t)$ takes values in an open convex set $\Omega$ of $\mathbb{R}^N$, and $\mathcal{A}(\mathbf{u})$ is a smooth locally bounded map from $\Omega$ to $\mathcal{M}_{N\times N}(\mathbb{R})$. The system is supposed to be strictly hyperbolic and the characteristic fields $R_i(\mathbf{u})$, $i = 1, \ldots, N$, are supposed to be either genuinely nonlinear:

$$\nabla\lambda_i(\mathbf{u}) \cdot R_i(\mathbf{u}) \neq 0, \quad \forall \mathbf{u} \in \Omega,$$

or linearly degenerate:

$$\nabla\lambda_i(\mathbf{u}) \cdot R_i(\mathbf{u}) = 0, \quad \forall \mathbf{u} \in \Omega.$$

* Corresponding author.
*E-mail addresses:* erpigar@uma.es (E. Pimentel-García), mjcastro@uma.es (M.J. Castro), christophe.chalons@uvsq.fr (C. Chalons), tomas.morales@uco.es (T. Morales de Luna), pares@uma.es (C. Parés).

https://doi.org/10.1016/j.jcp.2022.111152

Here, $\lambda_1(\mathbf{u}), \dots, \lambda_N(\mathbf{u})$ represent the eigenvalues of $\mathcal{A}(\mathbf{u})$ (in increasing order) and $R_1(\mathbf{u}), \dots, R_N(\mathbf{u})$ a set of associated eigenvectors.

Under some hypotheses of regularity for $\mathcal{A}(\mathbf{u})$, the theory introduced by Dal Maso, LeFloch, and Murat [14] allows one to define the nonconservative product $\mathcal{A}(\mathbf{u})\,\partial_x\mathbf{u}$ as a bounded measure for functions $\mathbf{u}$ with bounded variation. To do this, a family of Lipschitz continuous paths $\Phi : [0,1] \times \Omega \times \Omega \to \Omega$ has to be prescribed, which must satisfy certain regularity and compatibility conditions, in particular

$$\Phi(0; \mathbf{u}_l, \mathbf{u}_r) = \mathbf{u}_l, \qquad \Phi(1; \mathbf{u}_l, \mathbf{u}_r) = \mathbf{u}_r, \quad \forall \mathbf{u}_l, \mathbf{u}_r \in \Omega, \tag{1.2}$$

and

$$\Phi(s; \mathbf{u}, \mathbf{u}) = \mathbf{u}, \quad \forall \mathbf{u} \in \Omega. \tag{1.3}$$

The interested reader is addressed to [14] for a rigorous and complete presentation of this theory. The family of paths can be understood as a tool to give a sense to integrals of the form

$$\int_a^b \mathcal{A}(\mathbf{u}(x))\,\partial_x\mathbf{u}(x)\,dx,$$

for functions $\mathbf{u}$ with jump discontinuities. More precisely, given a bounded variation function $\mathbf{u} : [a,b] \mapsto \Omega$, we define:

$$\int_a^b \mathcal{A}(\mathbf{u}(x))\,\partial_x\mathbf{u}(x)\,dx = \int_a^b \mathcal{A}(\mathbf{u}(x))\,\partial_x\mathbf{u}(x)\,dx + \sum_m \int_0^1 \mathcal{A}(\Phi(s; \mathbf{u}_m^-, \mathbf{u}_m^+))\frac{\partial\Phi}{\partial s}(s; \mathbf{u}_m^-, \mathbf{u}_m^+)\,ds. \tag{1.4}$$

In this definition, $\mathbf{u}_m^-$ and $\mathbf{u}_m^+$ represent, respectively, the limits of $\mathbf{u}$ to the left and right of its $m$th discontinuity. Observe that, in (1.4), the family of paths has been used to determine the Dirac measures placed at the discontinuities of $\mathbf{u}$.

If such a mathematical definition of the nonconservative products is assumed to define the concept of weak solution, the generalized Rankine-Hugoniot condition:

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

has to be satisfied across an admissible discontinuity. Here, $\sigma$ is the speed of propagation of the discontinuity, and $\mathbf{u}^-$ and $\mathbf{u}^+$ are the left and right limits of the solution at the discontinuity.

Once the family of paths has been prescribed, a concept of entropy is required, as it happens for systems of conservation laws, that may be given by an entropy pair or by Lax's entropy criterion.

Since the concept of weak solution depends on the family of paths, which is a priori arbitrary, the crucial question is how to choose the 'good' family of paths. In fact, when the hyperbolic system is the vanishing-viscosity limit of the parabolic problems

$$\partial_t\mathbf{u}^\varepsilon + \mathcal{A}(\mathbf{u}^\varepsilon)\,\partial_x\mathbf{u}^\varepsilon = \varepsilon(\mathcal{R}(\mathbf{u}^\varepsilon)\partial_x\mathbf{u}^\varepsilon)_x, \tag{1.6}$$

where $\mathcal{R}(\mathbf{u})$ is any positive-definite matrix, the adequate family of paths should be related to the *viscous profiles*: a function $\mathbf{v}$ is said to be a viscous profile for (1.6) linking the states $\mathbf{u}^-$ and $\mathbf{u}^+$ if it satisfies

$$\lim_{\chi\to-\infty}\mathbf{v}(\chi) = \mathbf{u}^-, \quad \lim_{\chi\to+\infty}\mathbf{v}(\chi) = \mathbf{u}^+, \quad \lim_{\chi\to\pm\infty}\mathbf{v}'(\chi) = 0 \tag{1.7}$$

and there exists $\sigma \in \mathbb{R}$ such that the traveling wave

$$\mathbf{u}^\varepsilon(x,t) = \mathbf{v}\left(\frac{x-\sigma t}{\varepsilon}\right), \tag{1.8}$$

is a solution of (1.6) for every $\epsilon$. It can be easily verified that, in order to be a viscous profile, $\mathbf{v}$ has to solve the equation

$$-\xi\mathbf{v}' + \mathcal{A}(\mathbf{v})\,\mathbf{v}' = (\mathcal{R}(\mathbf{v})\,\mathbf{v}')', \tag{1.9}$$

with boundary conditions (1.7). If there exists a viscous profile linking the states $\mathbf{u}^-$ and $\mathbf{u}^+$, the good choice for the path connecting the states would be, after a reparameterization, the viscous profile $\mathbf{v}$.

The main difference with the conservative case is that now every choice of viscous term $\mathcal{R}$ leads to different jump conditions, while for standard conservative systems the usual Rankine-Hugoniot conditions are always recovered independently of the choice of the viscous term.

The definition of numerical methods for system (1.1) that converge to the correct weak solutions is not a simple task. It is well known that, although Lax's equivalence Theorem ensures that consistency and stability implies convergence for linear systems and methods, this is not the case in general for nonlinear problems. For instance, in the case of systems of conservation laws, stable conservative methods may converge to solutions that are not admissible weak solutions: this is the case for Roe method that may converge to weak solutions that are not entropy solutions. In order to ensure the convergence to the right weak solutions, besides consistency and stability, entropy has to be well controlled: for instance, entropy-fix techniques have to be added to Roe method (see for example [17]). In the case of nonconservative systems, consistency, stability, and control of the entropy are not enough: the numerical viscosity and, in general, the numerical dissipation effects, have to be well-controlled as well (see [19] for a review on this topic).

The design of finite-difference or finite-volume methods satisfying these four properties is difficult in general. Nevertheless, different techniques have been introduced to overcome, at least partially, this convergence issue: [4], [3], [2], [5], [8], [12], [13], [15], [22], [11]. In particular the path-conservative entropy stable methods introduced in [8] and extended to DG high-order methods in [18] significantly reduce the convergence error: to do this, entropy-conservative numerical methods are first introduced that are stabilized by means of a discretization of the viscous term of the regularized equation (1.6).

More recently, in [11], an in-cell discontinuous reconstruction technique has been added to first-order path-conservative methods that allows one to capture correctly weak solutions with isolated shock waves.

The goal of this article is to extend the in-cell discontinuous reconstruction methods introduced in [11] to second-order accuracy. To do this, these numerical methods will be first written as high-order path-conservative schemes (see for example [6,10]) and then, depending of the smoothness of the numerical solution, a standard MUSCL-Hancock reconstruction (see [25] and [26]) or a discontinuous one is used in the cell to update the numerical solution. Moreover, the well-balanced property of the method is also analyzed.

The paper is organized as follows: In Section 2 a brief introduction to path-conservative methods is given, then in Section 3 the new family of second-order in-cell discontinuous reconstruction methods is presented. First the semi-discrete method is introduced including the description of the reconstructions in the cells; then a temporal discretization based on a second order Taylor development is introduced. The shock-capturing property of the method is then enunciated and proved. Next, the well-balanced property of the method is studied. Section 4 is devoted to show numerically the efficiency of the proposed numerical scheme. More precisely, first the Coupled-Burgers nonconservative system introduced as a toy problem in [7] is considered: the application of the method to this system is described and several numerical tests are shown to validate the methods. Next, we focus on the Gas dynamics equations in Lagrangian coordinates and the modified shallow water system introduced in [9]. These systems were used in [1] and [9] respectively to illustrate the convergence issue of path-conservative methods when small-scale effects are not controlled: the method proposed in this paper is applied to these system to put on evidence that the convergence issue is corrected. The paper finishes with some conclusions and an Appendix where we describe the reconstruction procedure for non-isolated shocks for the modified Shallow Water system.

## 2. Path-conservative methods

According to [22], a first order numerical method for solving (1.1) is said to be path-conservative if it can be written in the form

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \frac{\Delta t}{\Delta x}\left(\mathcal{D}^+_{j-1/2} + \mathcal{D}^-_{j+1/2}\right), \tag{2.1}$$

where the following notation is used:

- $\Delta x$ and $\Delta t$ are the space and time steps respectively. They are supposed to be constant for simplicity.
- $I_j = \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$ are the computational cells, whose length is $\Delta x$.
- $t_n = n\Delta t$, $n = 0, 1\ldots$.
- $\mathbf{u}_j^n$ is the approximation of the average of the exact solution at the $j$th cell at time $t_n$, that is,

$$\mathbf{u}_j^n \approx \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{u}(x, t_n)\, dr. \tag{2.2}$$

- Finally,

$$\mathcal{D}^{\pm}_{j+1/2} = \mathcal{D}^{\pm}\left(\mathbf{u}_j^n, \mathbf{u}_{j+1}^n\right),$$

where $\mathcal{D}^-$ and $\mathcal{D}^+$ two Lipschitz continuous functions from $\Omega \times \Omega$ to $\Omega$ that satisfy

$$\mathcal{D}^{\pm}(\mathbf{u}, \mathbf{u}) = 0, \quad \forall \mathbf{u} \in \Omega, \tag{2.3}$$

and

$$\mathcal{D}^-(\mathbf{u}_l,\mathbf{u}_r)+\mathcal{D}^+(\mathbf{u}_l,\mathbf{u}_r)=\int_0^1 \mathcal{A}\big(\Phi(s;\mathbf{u}_l,\mathbf{u}_r)\big)\frac{\partial\Phi}{\partial s}(s;\mathbf{u}_l,\mathbf{u}_r)\,ds, \tag{2.4}$$

for every set $\mathbf{u}_l,\mathbf{u}_r\in\Omega$.

The definition of path-conservative methods is a *formal concept of consistency* for weak solutions defined on the basis of the family of paths $\Phi$. In fact, this is a natural extension of the definition of conservative methods for systems of conservation laws: it can be easily shown that, if (1.1) is a system of conservation laws, i.e. if $\mathcal{A}(\mathbf{u})$ is the Jacobian of a flux function $F(\mathbf{u})$, then every method that is path-conservative for any family of paths can be rewritten as a conservative method (see [10] for a recent review).

This framework makes it easy to extend many well-known conservative schemes to nonconservative systems. Let us show two examples:

- Godunov method:

$$\mathcal{D}_G^-(\mathbf{u}_l,\mathbf{u}_r)=\int_0^1 \mathcal{A}(\Phi(s;\mathbf{u}_l,\mathbf{u}_0))\frac{\partial\Phi}{\partial s}(s;\mathbf{u}_l,\mathbf{u}_0)\,ds, \tag{2.5}$$

$$\mathcal{D}_G^+(\mathbf{u}_l,\mathbf{u}_r)=\int_0^1 \mathcal{A}(\Phi(s;\mathbf{u}_0,\mathbf{u}_r))\frac{\partial\Phi}{\partial s}(s;\mathbf{u}_0,\mathbf{u}_r)\,ds, \tag{2.6}$$

where $\mathbf{u}_0$ is the value at $x=0$ of the self-similar solution of the Riemann problem

$$\begin{cases}\partial_t\mathbf{u}+\mathcal{A}(\mathbf{u})\partial_x\mathbf{u}=0,\\ \mathbf{u}(x,0)=\begin{cases}\mathbf{u}_l & \text{if } x<0,\\ \mathbf{u}_r & \text{otherwise.}\end{cases}\end{cases} \tag{2.7}$$

If the family of paths satisfies some conditions of compatibility with the solutions of the Riemann problems, the method can be interpreted in terms of the averages of the exact solutions of local Riemann problems in the cells, as it happens for system of conservation laws: see [20].

- Roe method:

$$\mathcal{D}_R^\pm(\mathbf{u}_l,\mathbf{u}_r)=\mathcal{A}_\Phi^\pm(\mathbf{u}_l,\mathbf{u}_r)\,(\mathbf{u}_r-\mathbf{u}_l), \tag{2.8}$$

where $\mathcal{A}_\Phi(\mathbf{u}_l,\mathbf{u}_r)$ is a Roe linearization of $\mathcal{A}(\mathbf{u})$ in the sense defined by Toumi in [24], i.e. a function $\mathcal{A}_\Phi:\Omega\times\Omega\mapsto\mathcal{M}_{N\times N}(\mathbb{R})$ satisfying the following properties:
  - for each $\mathbf{u}_l,\mathbf{u}_r\in\Omega$, $\mathcal{A}_\Phi(\mathbf{u}_l,\mathbf{u}_r)$ has $N$ distinct real eigenvalues $\lambda_1(\mathbf{u}_l,\mathbf{u}_r),\ldots,\lambda_N(\mathbf{u}_l,\mathbf{u}_r)$;
  - $\mathcal{A}_\Phi(\mathbf{u},\mathbf{u})=\mathcal{A}(\mathbf{u})$, for every $\mathbf{u}\in\Omega$;
  - for any $\mathbf{u}_l,\mathbf{u}_r\in\Omega$,

$$\mathcal{A}_\Phi(\mathbf{u}_l,\mathbf{u}_r)\,(\mathbf{u}_r-\mathbf{u}_l)=\int_0^1 \mathcal{A}(\Phi(s;\mathbf{u}_l,\mathbf{u}_r))\frac{\partial\Phi}{\partial s}(s;\mathbf{u}_l,\mathbf{u}_r)\,ds. \tag{2.9}$$

As usual $\mathcal{A}_\Phi^\pm(\mathbf{u}_l,\mathbf{u}_r)$ represent the matrices whose eigenvalues are the positive/negative parts of $\lambda_1(\mathbf{u}_l,\mathbf{u}_r),\ldots,\lambda_N(\mathbf{u}_l,\mathbf{u}_r)$ with same eigenvectors.

First order path-conservative numerical schemes can be extended to high-order by using reconstruction operators:

$$\mathbf{u}_j'(t)=-\frac{1}{\Delta x}\left(\mathcal{D}^-_{j+\frac12}(t)+\mathcal{D}^+_{j-\frac12}(t)+\int_{x_{j-\frac12}}^{x_{j+\frac12}}\mathcal{A}(P_j^t(x))\frac{\partial}{\partial x}P_j^t(x)\,dx\right), \tag{2.10}$$

where $P_j^t(x)$ is the smooth approximation of the solution at the $j$th-cell provided by a high-order reconstruction operator from the sequence of cell values $\{\mathbf{u}_j(t)\}$ and

$$\mathcal{D}^\pm_{j+1/2}(t)=\mathcal{D}^\pm_{j+1/2}(\mathbf{u}^-_{j+1/2}(t),\mathbf{u}^+_{j+1/2}(t)),$$

where $\mathbf{u}^-_{j+1/2}(t)=P_j^t(x_{j+\frac12})$ and $\mathbf{u}^-_{j+1/2}(t)=P_{j+1}^t(x_{j+\frac12})$ (see [10] for details).

In [9] it was shown that, if the numerical solutions provided by a path-conservative method converge uniformly in the sense of graphs as $\Delta x \to 0$, the limit is a weak solution according to the chosen family of paths. Nevertheless, this notion of convergence is too strong and the numerical solutions provided by finite-difference or finite-volume methods do not converge usually in this sense. This is not to say that path-conservative methods do not converge: in practice, it can be observed that numerical methods like the extensions of Godunov or Roe schemes described in the previous section converge in $L^1$-norm under the usual CFL condition. What happens is that the limit may be a weak solution according to a different family of paths, i.e. it is a classical solution in the smoothness regions but its discontinuities satisfy a jump condition (1.5) different of the expected one: see [9], [1]. In fact, the family of paths that controls the jump conditions satisfied by the limits of the numerical solutions is related to the viscous profiles of the equivalent equation of the method: see [9]. If, for instance, the family of paths is based on the viscous profiles related to a regularization (1.6), the leading terms in the equivalent equation that represent the numerical viscosity of the scheme may not match the viscous term in (1.6).

Observe that, as it has been mentioned before, the definition of path-conservative method is a formal notion of consistency. Nevertheless, as pointed out before, this consistency, together with stability and control of the entropy, is not enough to ensure the convergence towards the correct weak solution: the numerical viscosity and, in general, the numerical dissipation effects, have to be well-controlled. Let us stress, before finishing this section, that:

- The convergence to wrong weak solutions is not due to the consistency property, but to the lack of control of the small-scale effects in the numerical solutions.
- This convergence issue affects to every methods in which the small-scale effects are not controlled, whether its consistency is based on the notion of path-conservative method or not.
- It is possible to design path-conservative methods that overcome, at least partially, this difficulty, as shown in [11] or [8].

## 3. Second-order in-cell discontinuous reconstruction path-conservative methods

In this section, a new numerical method of the form (2.10) is described. The scheme is based on a first-order path-conservative numerical method with fluctuation functions $\mathcal{D}^{\pm}$, which is combined with a particular novel reconstruction operator. A standard second-order reconstruction operator in smoothness regions is used, while a discontinuous reconstruction operator close to discontinuities is performed, so that numerical viscosity is removed in the non-smooth regions.

### 3.1. Semi-discrete method

Once the numerical approximations $\mathbf{u}_j^n$ of the averages of the solutions have been computed at time $t_n = n\Delta t$, the first step is to mark the cells $I_j$ where a discontinuity is present. More explicitly, the cells such that the solution of the Riemann problem consisting of (1.1) with initial conditions

$$\mathbf{u}(x,0) = \begin{cases} \mathbf{u}_{j-1}^n & \text{if } x < 0, \\ \mathbf{u}_{j+1}^n & \text{if } x > 0, \end{cases} \tag{3.1}$$

involves a shock wave. Let us denote by $\mathcal{M}_n$ the set of indices of the marked cells, i.e.

$$\mathcal{M}_n = \{j \text{ s.t. the solution of the Riemann problem (1.1), (3.1) involves a shock wave}\}. \tag{3.2}$$

To advance in time the following semi-discrete numerical method is considered:

$$\mathbf{u}_j'(t) = -\frac{1}{\Delta x}\left( \mathcal{D}^-_{j+\frac{1}{2}}(t) + \mathcal{D}^+_{j-\frac{1}{2}}(t) + \int\!\!\!\!\!\!-_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{A}(P_j^n(x,t))\frac{\partial}{\partial x} P_j^n(x,t)dx \right), \quad t \geq t_n, \tag{3.3}$$

where

$$\mathbf{u}_j(t) \approx \frac{1}{\Delta x}\int\!\!\!\!\!\!-_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{u}(x,t)dx,$$

$$\mathcal{D}^{\pm}_{j+1/2}(t) = \mathcal{D}^{\pm}_{j+1/2}(\mathbf{u}^-_{j+1/2}(t), \mathbf{u}^+_{j+1/2}(t)),$$

with

$$\mathbf{u}^-_{j+1/2}(t) = P_j^n(x_{j+\frac{1}{2}}, t), \quad \mathbf{u}^+_{j+1/2}(t) = P_{j+1}^n(x_{j+\frac{1}{2}}, t),$$

and $P_j^n(x,t)$ is defined as follows:

- If $j-1, j, j+1 \notin \mathcal{M}_n$ then $P_j^n$ is the MUSCL-Hancock reconstruction (see [25] and [26]), i.e. the approximation of the first degree Taylor polynomial of the solution given by:

$$P_j^n(x,t) = \mathbf{u}_j^n + \widetilde{\partial_x \mathbf{u}}_j^n (x - x_j) - \mathcal{A}(\mathbf{u}_j^n)\widetilde{\partial_x \mathbf{u}}_j^n (t - t^n).$$

Here, $\widetilde{\partial_x \mathbf{u}}_j^n$ is the *minmod* approximation of the first order spacial derivative of $\mathbf{u}$ at $x_j$ at time $t_n$, whose $k$th component is given by

$$\left(\widetilde{\partial_x \mathbf{u}}_j^n\right)_k = \text{minmod}\left(\alpha \frac{u_{j+1,k}^n - u_{j,k}^n}{\Delta x}, \frac{u_{j+1,k}^n - u_{j-1,k}^n}{2\Delta x}, \alpha \frac{u_{j,k}^n - u_{j-1,k}^n}{\Delta x}\right),$$

where $u_{j,k}^n$ represents the $k$th component of $\mathbf{u}_j^n$, $\alpha$ is a parameter with $1 \le \alpha < 2$ and

$$\text{minmod}(a,b,c) = \begin{cases} \min\{a,b,c\} & \text{if } a,b,c > 0, \\ \max\{a,b,c\} & \text{if } a,b,c < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that for the Taylor polynomial we have used the approximation:

$$\partial_t \mathbf{u}(x_i,t_n) = -\mathcal{A}(\mathbf{u}(x_i,t_n))\partial_x \mathbf{u}(x_i,t_n) \approx -\mathcal{A}(\mathbf{u}_j^n)\widetilde{\partial_x \mathbf{u}}_j^n.$$

- If $j \in \mathcal{M}_n$ then

$$P_j^n(x,t) = \begin{cases} \mathbf{u}_{j,l}^n & \text{if } x \le x_{j-1/2} + d_j^n \Delta x + \sigma_j^n (t - t_n), \\ \mathbf{u}_{j,r}^n & \text{otherwise,} \end{cases}$$

where $d_j^n$ is chosen so that

$$d_j^n u_{j,l,k}^n + (1 - d_j^n) u_{j,r,k}^n = u_{j,k}^n, \tag{3.4}$$

for some index $k \in \{1, \dots, N\}$; and $\sigma_j^n$, $\mathbf{u}_{j,l}^n$, and $\mathbf{u}_{j,r}^n$ are chosen so that if $\mathbf{u}_{j-1}^n$ and $\mathbf{u}_{j+1}^n$ may be linked by an admissible discontinuity with speed $\sigma$, then

$$\mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n, \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}^n, \quad \sigma_j^n = \sigma. \tag{3.5}$$

Observe that this in-cell discontinuous reconstruction can only be done if $0 \le d_j^n \le 1$, i.e. if

$$0 \le \frac{u_{j,r,k}^n - u_{j,k}^n}{u_{j,r,k}^n - u_{j,l,k}^n} \le 1,$$

otherwise the index $j$ is removed from the set $\mathcal{M}_n$ and the MUSCL-Hancock reconstruction is applied in the cell. Moreover, if $d_j^n = 1$ and $\sigma_j^n > 0$ (resp. $d_j^n = 0$ and $\sigma_j^n < 0$) the cell is unmarked and the cell $I_{j+1}$ (resp. $I_{j-1}$) is marked if necessary: note that in these cases, the discontinuity leaves the cell $I_j$ for any $t > t_n$.

- Otherwise (i.e. if $j \notin \mathcal{M}_n$ but $j-1 \in \mathcal{M}_n$ or $j+1 \in \mathcal{M}_n$) then

$$P_j^n(x,t) = \mathbf{u}_j^n.$$

**Remark 3.1.** In the case $j \in \mathcal{M}_n$, if one of the equations of system (1.1), say the $k$th one, is a conservation law, the index $k$ is selected in (3.4), so that the corresponding variable is conserved. Moreover, if there is a linear combination of the unknowns $\sum_{k=1}^N \alpha_k u_k$ that is conserved, (3.4) may be replaced by:

$$d_j^n \sum_{k=1}^N \alpha_k u_{j,l,k}^n + (1 - d_j^n) \sum_{k=1}^N \alpha_k u_{j,r,k}^n = \sum_{k=1}^N \alpha_k u_{j,k}^n. \tag{3.6}$$

If there are more than one conservation laws, the index $k$ corresponding to one of them is selected in (3.4).

### 3.2. Choice of $\sigma_j^n$, $\mathbf{u}_{j,l}^n$, $\mathbf{u}_{j,r}^n$

Two different strategies are considered here: the first one is based on the exact solutions of the Riemann problems and the second one on a Roe linearization:

- **First strategy:** Assume that the solutions for Riemann problems are explicitly known. Then, for any given marked cell $j$, we may choose $\sigma_j^n$, $\mathbf{u}_{j,l}^n$, $\mathbf{u}_{j,r}^n$ as the speed, the left, and the right states of (one of the) discontinuous waves appearing in the solution of the Riemann problem with initial data $\mathbf{u}_{j-1}^n$, $\mathbf{u}_{j+1}^n$. Observe that, if the solution of the Riemann problem consists of only one discontinuous wave of speed $\sigma$ linking $\mathbf{u}_{j-1}^n$ and $\mathbf{u}_{j+1}^n$, then necessarily $\sigma_j^n = \sigma$, $\mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n$, $\mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}^n$ and (3.5) is satisfied.
- **Second strategy:** If a Roe matrix is available, for any given marked cell $j$, we may choose $\sigma_j^n$, $\mathbf{u}_{j,l}^n$, $\mathbf{u}_{j,r}^n$ as the speed, the left, and the right states of a non-trivial wave appearing in the solution of the linearized Riemann problem with initial data $\mathbf{u}_{j-1}^n$, $\mathbf{u}_{j+1}^n$. To do this, first the coordinates $\{\alpha_k\}$ of $\mathbf{u}_{j+1}^n - \mathbf{u}_{j-1}^n$ in the basis of eigenvectors of $\mathcal{A}_\Phi(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)$, are computed, i.e.

$$\mathbf{u}_{j+1}^n - \mathbf{u}_{j-1}^n = \sum_{k=1}^{N} \alpha_k R_k(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n).$$

  Next, an index $k^*$ such that $\alpha_{k^*} \neq 0$ is selected. Then, $\sigma_j^n$, $\mathbf{u}_{j,l}^n$, and $\mathbf{u}_{j,r}^n$ are chosen as follows:

$$\sigma_j^n = \lambda_{k^*}(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n), \quad \mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n + \sum_{k=1}^{k^*-1} \alpha_k R_k(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n), \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j,l}^n + \alpha_{k^*} R_{k^*}(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n).$$

  Observe that, if $\mathbf{u}_{j-1}^n$ and $\mathbf{u}_{j+1}^n$ can be linked by an admissible discontinuous wave of speed $\sigma$, then the Roe property implies

$$\mathcal{A}_\Phi(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)\left(\mathbf{u}_{j+1}^n - \mathbf{u}_{j-1}^n\right) = \sigma\left(\mathbf{u}_{j+1}^n - \mathbf{u}_{j-1}^n\right),$$

  so that $\sigma$ is an eigenvalue of the Roe matrix and $\mathbf{u}_{j+1}^n - \mathbf{u}_{j-1}^n$ an associated eigenvector. Therefore, the solution of the linearized Riemann problem consists of only one wave of speed $\sigma$ linking $\mathbf{u}_{j-1}^n$ and $\mathbf{u}_{j+1}^n$. Therefore, $\sigma_j^n = \sigma$, $\mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n$, $\mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}^n$ and (3.5) is again satisfied.

Notice that in both cases, (3.5) is always satisfied regardless of the discontinuous wave that is selected to build the discontinuous reconstruction. As a general strategy, the discontinuous wave whose amplitude is maximal can be selected: this is done for instance in Subsection 4.3. Nevertheless, in some cases, the specific knowledge of the problem may lead to a different choice, as it will be seen in Subsection 4.2.

These two strategies can be easily extended to any approximate Riemann solver.

### 3.3. Time step

The time step $\Delta t_n$ is chosen as follows:

$$\Delta t_n = \min(\Delta t_n^c, \Delta t_n^r). \tag{3.7}$$

Here

$$\Delta t_n^c = CFL \cdot \min_j \left(\frac{\Delta x}{\max_l |\lambda_{j,l}|}\right) \tag{3.8}$$

where $CFL \in (0,1)$ is the stability parameter and $\lambda_{j,l}, \ldots, \lambda_{j,N}$ represent the eigenvalues of $\mathcal{A}(\mathbf{u}_j^n)$; and

$$\Delta t_n^r = \min_{j \in \mathcal{M}_n} \begin{cases} \dfrac{1-d_j^n}{|\sigma_j^n|}\Delta x, & \text{if} \quad \sigma_j^n > 0, \\ \dfrac{d_j^n}{|\sigma_j^n|}\Delta x, & \text{if} \quad \sigma_j^n < 0. \end{cases} \tag{3.9}$$

Observe that, besides the stability requirement, this choice of time step ensures that no discontinuous reconstruction leaves a marked cell.

### 3.4. Fully discrete method

Once the time step is chosen, (3.3) is integrated in the interval $[t^n, t^{n+1}]$, with $t^{n+1} = t^n + \Delta t_n$, to obtain:

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \frac{1}{\Delta x} \int_{t^n}^{t^{n+1}} \left( \mathcal{D}^-_{j+\frac{1}{2}}(t) + \mathcal{D}^+_{j-\frac{1}{2}}(t) + \fint_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{A}(P_j^n(x,t)) \partial_x P_j^n(x,t) dx \right) dt,$$

and the mid-point rule is used to approximate the integrals in time:

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \frac{\Delta t_n}{\Delta x} \left( \mathcal{D}^-_{j+\frac{1}{2}}(t^{n+\frac{1}{2}}) + \mathcal{D}^+_{j-\frac{1}{2}}(t^{n+\frac{1}{2}}) + \fint_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{A}(P_j^n(x,t^{n+1/2})) \partial_x P_j^n(x,t^{n+1/2}) \, dx \right). \tag{3.10}$$

The computation of the dashed integral in this expression depends on the cell:

1. If $j-1, j, j+1 \notin \mathcal{M}_n$ the mid-point rule is used again to approximate the integral:

$$\begin{aligned} \fint_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{A}(P_j^n(x,t^{n+1/2})) \partial_x P_j^n(x,t^{n+1/2}) \, dx &\approx \Delta x \mathcal{A}(\mathbf{u}_j^{n+\frac{1}{2}}) \partial_x P_j^n(x_j, t^{n+1/2}) \\ &= \Delta x \mathcal{A}(\mathbf{u}_j^{n+\frac{1}{2}}) \widetilde{\partial_x \mathbf{u}}_j^n, \end{aligned} \tag{3.11}$$

   where

$$\mathbf{u}_j^{n+\frac{1}{2}} = P_j^n(x_j, t^{n+\frac{1}{2}}) = \mathbf{u}_j^n - \frac{\Delta t}{2} \mathcal{A}(\mathbf{u}_j^n) \widetilde{\partial_x \mathbf{u}}_j^n.$$

2. If $j \in \mathcal{M}_n$, taking into account the definition of the dashed integrals (1.4), one has:

$$\fint_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{A}(P_j^n(x,t^{n+1/2})) \partial_x P_j^n(x,t^{n+1/2}) \, dx = \int_0^1 \mathcal{A}(\Phi(s; \mathbf{u}_{j,l}^n, \mathbf{u}_{j,r}^n)) \partial_s \Phi(s; \mathbf{u}_{j,l}^n, \mathbf{u}_{j,r}^n) \, ds. \tag{3.12}$$

   Observe that, if $\mathbf{u}_{j,l}^n$ and $\mathbf{u}_{j,r}^n$ can be linked by a shock whose speed is $\sigma_j^n$, then the generalized Rankine-Hugoniot condition (1.5) leads to

$$\fint_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{A}(P_j^n(x,t^{n+1/2})) \partial_x P_j^n(x,t^{n+1/2}) \, dx = \sigma_j^n \left( \mathbf{u}_{j,r}^n - \mathbf{u}_{j,l}^n \right). \tag{3.13}$$

3. If $j \notin \mathcal{M}_n$ but $j-1 \in \mathcal{M}_n$ or $j+1 \in \mathcal{M}_n$ then

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathcal{A}(P_j^n(x,t^{n+1/2})) \partial_x P_j^n(x,t^{n+1/2}) \, dx = 0. \tag{3.14}$$

The final expression of the fully discrete numerical method is then as follows:

$$\mathbf{u}_j^{n+1} = \mathbf{u}_j^n - \frac{\Delta t_n}{\Delta x} \left( \mathcal{D}^-_{j+\frac{1}{2}}(t^{n+\frac{1}{2}}) + \mathcal{D}^+_{j-\frac{1}{2}}(t^{n+\frac{1}{2}}) + \mathcal{D}_j \right), \tag{3.15}$$

where

$$\mathcal{D}_j = \begin{cases} \Delta x \mathcal{A}(\mathbf{u}_j^{n+\frac{1}{2}}) \widetilde{\partial_x \mathbf{u}}_j^n & \text{if } j-1, j, j+1 \notin \mathcal{M}_n; \\ \displaystyle\int_0^1 \mathcal{A}(\Phi(s; \mathbf{u}_{j,l}^n, \mathbf{u}_{j,r}^n)) \partial_s \Phi(s; \mathbf{u}_{j,l}^n, \mathbf{u}_{j,r}^n) \, ds & \text{if } j \in \mathcal{M}_n; \\ 0 & \text{otherwise.} \end{cases} \tag{3.16}$$

Observe that the numerical method satisfies the following properties:

- Far from discontinuities it coincides with the standard MUSCL-Hancock.
- Close to discontinuities it corresponds to the numerical method introduced in [11]. Nevertheless, there is a slight variation when compared with [11]: in that reference, the discontinuities are allowed to leave the marked cells and the contribution to the neighbor cells are then taken into account. While this technique allows one to avoid additional restrictions to the time step, it makes more difficult the implementation of the numerical method. Nevertheless, the technique proposed here may as well be implemented as in [11].

### 3.5. Shock-capturing property

Let us prove that isolated shock waves are exactly captured by the scheme and contain no spurious numerical diffusion. Although the proof is essentially the same as in [11], it is included for the sake of completeness.

**Theorem 3.2.** *Assume that $\mathbf{u}_l$ and $\mathbf{u}_r$ can be joined by an entropy shock of speed $\sigma$. Then, the numerical method provides an exact numerical solution of the Riemann problem with initial conditions*

$$\mathbf{u}(x,0) = \begin{cases} \mathbf{u}_l & \text{if } x < 0, \\ \mathbf{u}_r & \text{otherwise,} \end{cases}$$

*in the sense that*

$$\mathbf{u}_j^n = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{u}(x, t^n)\, dx, \quad \forall j, n \tag{3.17}$$

*where $\mathbf{u}(x,t)$ is the exact solution.*

**Proof.** Let us suppose that $0 \in I_{j^*}$ and $0 = x_{j^*-1/2} + d\Delta x$, with $0 \le d \le 1$. Then the initial cell averages are:

$$\mathbf{u}_j^0 = \begin{cases} \mathbf{u}_l & \text{if } j < j^*; \\ d\mathbf{u}_l + (1-d)\mathbf{u}_r & \text{if } j = j^*; \\ \mathbf{u}_r & \text{otherwise.} \end{cases}$$

If $0 < d < 1$ the only marked cell at time $t^0 = 0$ is $I_{j^*}$, i.e. $\mathcal{M}_0 = \{j^*\}$. The only non-constant reconstruction is then $P_0^0$ and the equalities

$$\mathbf{u}_j^1 = \mathbf{u}_j^0 = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{u}(x, t^1)\, dx, \quad \forall j \neq j^*$$

can be easily deduced from the definition of the numerical method.

Let us compute $\mathbf{u}_{j^*}^1$. Observe that, in order to have (3.4), necessarily $d_0^0 = d$. Therefore, since $\mathbf{u}_l$ and $\mathbf{u}_r$ can be linked by an admissible discontinuity of speed $\sigma$, using (3.5) one has:

$$P_0^0(x,t) = \begin{cases} \mathbf{u}_l & \text{if } x \le \sigma t, \\ \mathbf{u}_r & \text{otherwise.} \end{cases}$$

Observe that $P_0^0$ coincides with the exact solution. We have now:

$$\begin{aligned} \mathbf{u}_{j^*}^1 &= \mathbf{u}_{j^*}^0 - \frac{\Delta t_0}{\Delta x}\left(\mathcal{D}^-_{\frac{1}{2}}(t^{\frac{1}{2}}) + \mathcal{D}^+_{-\frac{1}{2}}(t^{\frac{1}{2}}) + \mathcal{D}_0\right) \\ &= \mathbf{u}_{j^*}^0 - \frac{\Delta t_0}{\Delta x}\mathcal{D}_0 \\ &= \mathbf{u}_{j^*}^0 - \frac{\Delta t_0}{\Delta x}\sigma(\mathbf{u}_r - \mathbf{u}_l) \\ &= \left(d + \frac{\sigma\Delta t}{\Delta x}\right)\mathbf{u}_l + \left(1 - d - \frac{\sigma\Delta t}{\Delta x}\right)\mathbf{u}_r, \end{aligned}$$

where it has been used that

$$\mathbf{u}^-_{j-1+2}(t^{1/2}) = \mathbf{u}^+_{j-1+2}(t^{1/2}) = \mathbf{u}_l,$$
$$\mathbf{u}^-_{j+1+2}(t^{1/2}) = \mathbf{u}^+_{j+1+2}(t^{1/2}) = \mathbf{u}_r,$$

so that

$$\mathcal{D}^+_{-\frac{1}{2}}(t^{\frac{1}{2}}) = \mathcal{D}^-_{\frac{1}{2}}(t^{\frac{1}{2}}) = 0.$$

On the other hand, due to the time step restrictions one has

$$x_{j^*-1/2} \le x_{j^*-1/2} + d\Delta x + \sigma\,\Delta t = \sigma\,\Delta t \le x_{j^*+1/2}.$$

Thus, it can be easily checked that:

$$\frac{1}{\Delta x}\int_{x_{j^*-1/2}}^{x_{j^*+1/2}} \mathbf{u}(x, t^1)\,dx = \mathbf{u}^1_{j^*},$$

and (3.17) has been proved for $n = 1$.

If $d = 1$ (resp. $d = 0$) the only marked cell is $I_{j+1}$ (resp. $I_{j-1}$) and the proof is similar.

The proof of the equality (3.17) for $n \ge 2$ is similar to the case $n = 1$.

### 3.6. Well-balanced property

In this section, we analyze under what conditions the proposed numerical method is well-balanced. Note that the stationary solutions of System (1.1) verify

$$\mathcal{A}(\mathbf{u})\partial_x\mathbf{u} = 0, \quad \forall x. \tag{3.18}$$

Observe that, if there exists a smooth stationary solution $\mathbf{u}$ such that $\partial_x\mathbf{u}(x) \neq 0$ for every $x$, then 0 is an eigenvalue of $\mathcal{A}(\mathbf{u}(x))$ and $\partial_x\mathbf{u}(x)$ is an associated eigenvector for every $x$: let us suppose without loss of generality, that $\lambda_N(\mathbf{u}(x)) = 0$. Therefore, $x \rightarrowtail \mathbf{u}(x)$ is an integral curve of the $N$th characteristic field and, since the value of $\lambda_N$ is constant through the integral curve, it is linearly degenerate. Let us denote by $\Gamma$ the set of all the integral curves $\gamma$ of the $N$th characteristic field.

We assume here that the chosen family of paths $\Phi$ satisfies the following property:

(P) Given two states $\mathbf{u}_l$ and $\mathbf{u}_r$ that belong to the same integral curve $\gamma \in \Gamma$, the path linking them is a parameterization of the arc of this curve that connects the two states.

As a consequence of this property, the stationary contact discontinuity

$$\mathbf{u}(x,t) = \begin{cases} \mathbf{u}_l, & x < x^*, \\ \mathbf{u}_r, & x > x^*, \end{cases} \tag{3.19}$$

is an admissible weak solution of the system.

According to [23] we introduce the following

**Definition 3.3.** A numerical scheme (2.1) is said to be well-balanced if, given any pair of states $\mathbf{u}_l$ and $\mathbf{u}_r$ belonging to $\gamma \in \Gamma$ one has

$$\mathcal{D}^{\pm}(\mathbf{u}_l, \mathbf{u}_r) = 0. \tag{3.20}$$

Notice that, if a numerical method satisfying (3.20) is applied to the initial condition

$$\mathbf{u}^0_i = \mathbf{u}^*(x_i), \quad \forall i,$$

where $\mathbf{u}^*$ is a stationary solution, then

$$\mathbf{u}^n_i = \mathbf{u}^0_i, \quad \forall i.$$

Moreover, in [23] and [22] it has been shown that Godunov and Roe methods are well-balanced according to this definition if the property (P) holds. We will assume here that the first-order path-conservative method satisfies (3.20), that is, it is well-balanced.

Now, we must prescribe how the marking process is performed in the in-cell discontinuous reconstruction to preserve the well-balanced property of the method.

*3.6.1. First-order in-cell well-balanced discontinuous reconstruction path-conservative methods*

Once the numerical approximations $\mathbf{u}_j^n$ have been computed at time $t_n = n\Delta t$, let us consider the set of indices:

$$\mathcal{C}_n = \{j \text{ s.t. the solution of the Riemann problem (1.1), (3.1) consists of only a stationary contact discontinuity}\}, \tag{3.21}$$

i.e. $j \in \mathcal{C}_n$ if $\mathbf{u}_{j-1}^n$, $\mathbf{u}_{j+1}^n$ belong to the same curve $\gamma \in \Gamma$.

Observe that, since $\mathcal{M}_n \cap \mathcal{C}_n = \emptyset$, the cells in which a stationary contact discontinuity is detected are not marked so that the scheme reduces to the standard first-order path-conservative method. Therefore, the method is well-balanced.

*3.6.2. Second-order in-cell well-balanced discontinuous reconstruction path-conservative methods*

In order to obtain a second-order in-cell well-balanced discontinuous reconstruction path-conservative method, it is enough to add the following case to the ones in Subsection 3.1:

4. If $j \in \mathcal{C}_n$ then

$$P_j^n(x,t) = \mathbf{u}_j^n.$$

In effect, if the numerical method is applied to the initial condition

$$\mathbf{u}_j^0 = \mathbf{u}^*(x_j), \quad \forall j,$$

where $\mathbf{u}^*$ a stationary solution, then all the indices belong to $\mathcal{C}_n$ and the numerical method reduces to the chosen first-order path-conservative methods. Therefore, the method is well-balanced.

*3.6.3. Systems with source terms*

An important particular case of systems with a linearly degenerate field associated to the null eigenvalues is given by problems with source term:

$$\partial_t \mathbf{u} + \mathcal{A}(\mathbf{u})\partial_x \mathbf{u} = S(\mathbf{u})H_x, \quad x \in \mathbb{R}, \quad t \in \mathbb{R}^+, \tag{3.22}$$

where $S(\mathbf{u})$ is a smooth locally bounded map from $\Omega$ to $\mathbb{R}^N$, and $H$ is a known piecewise continuous function. If $H$ is considered as an artificial unknown that satisfies the equation

$$\partial_t H = 0,$$

System (3.22) can be rewritten as follows:

$$\partial_t \mathbf{U} + \widetilde{\mathcal{A}}(\mathbf{U})\partial_x \mathbf{U} = 0, \quad x \in \mathbb{R}, \quad t \in \mathbb{R}^+, \tag{3.23}$$

where

$$\mathbf{U} = \begin{bmatrix} \mathbf{u} \\ H \end{bmatrix}, \quad \widetilde{\mathcal{A}}(\mathbf{U}) = \left[\begin{array}{c|c} \mathcal{A}(\mathbf{u}) & -S(\mathbf{u}) \\ \hline 0 & 0 \end{array}\right],$$

and $\lambda_{N+1}(\mathbf{U}) = 0$ is an eigenvalue of the matrix for every $\mathbf{U}$: the numerical methods introduced in the two previous subsections can be thus applied. Nevertheless, instead of applying the MUSCL-Hancock reconstruction to $H$, its exact value is used, i.e. the second-order reconstruction will be as follows:

$$\widetilde{P}_j^n(x,t) = \begin{bmatrix} P_j^n(x,t) \\ H(x) \end{bmatrix},$$

where $P_j^n$ is the MUSCL-Hancock reconstruction of the $\mathbf{u}$ variables.

## 4. Numerical tests

The following numerical methods will be applied here to three nonconservative systems:

- O1_noDisRec: standard first-order path-conservative Roe or Godunov (it will be indicated between parentheses) methods;
- O1_DisRec: first-order path-conservative method with discontinuous reconstruction;
- O2_noDisRec: Standard second-order extension of the first order path-conservative method based on the MUSCL-Hancock reconstruction;
- O2_DisRec: second-order path-conservative method that combines MUSCL-Hancock and discontinuous reconstruction;

### 4.1. Coupled Burgers system

#### 4.1.1. Equations

Let us first consider the toy system

$$\begin{cases} \partial_t u + \partial_x \left( \dfrac{u^2}{2} \right) + u \partial_x v \ = \ 0, \\ \partial_t v + \partial_x \left( \dfrac{v^2}{2} \right) + v \partial_x u \ = \ 0, \end{cases} \qquad (x,t) \in \mathbb{R} \times \mathbb{R}^+, \tag{4.1}$$

introduced in [7], 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 written in the form (1.1) with

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

The system is strictly hyperbolic in $\Omega$ with eigenvalues

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

whose characteristic fields, given by the eigenvectors

$$R_1(\mathbf{u}) = [1, -1]^T, \quad R_2(\mathbf{u}) = [u, v]^T,$$

are respectively linearly degenerate and genuinely nonlinear.

The sum $u + v$ satisfies the standard Burgers equation

$$\partial_t (u + v) + \partial_x \left( \frac{1}{2} (u + v)^2 \right) = 0,$$

and thus the variable $u + v$ is conserved.

#### 4.1.2. Simple waves

Once the family of paths has been chosen, the simple waves of this system are:

- Stationary contact discontinuities linking states $\mathbf{u}_l$, $\mathbf{u}_r$ such that

$$u_l + v_l = u_r + v_r.$$

- Rarefactions waves joining states $\mathbf{u}_l$, $\mathbf{u}_r$ such that

$$u_l + v_l < u_r + v_r, \quad \frac{u_l}{v_l} = \frac{u_r}{v_r}.$$

- Shock waves joining states $\mathbf{u}_l$ and $\mathbf{u}_r$ such that

$$u_l + v_l > u_r + v_r$$

that satisfy the jump condition:

$$\sigma [u] = \left[ \frac{u^2}{2} \right] + \int_0^1 \phi_u(s; \mathbf{u}_l, \mathbf{u}_r) \partial_s \phi_v(s; \mathbf{u}_l, \mathbf{u}_r) \, ds,$$

$$\sigma [v] = \left[ \frac{v^2}{2} \right] + \int_0^1 \phi_v(s; \mathbf{u}_l, \mathbf{u}_r) \partial_s \phi_u(s; \mathbf{u}_l, \mathbf{u}_r) \, ds.$$

As usual, for any variable $\phi$, $[\phi]$ stands for the jump on the variable $\phi_r - \phi_l$. Remark that this leads, independently of the choice of the family of paths, to

$$\sigma = \frac{u_l + v_l + u_r + v_r}{2},$$

If, for instance, the family of straight segments is chosen

$$\phi_u(s;\mathbf{u}_l,\mathbf{u}_r)=u_l+s(u_r-u_l);\quad \phi_v(s;\mathbf{u}_l,\mathbf{u}_r)=v_l+s(v_r-v_l), \tag{4.2}$$

the jump conditions reduce to:

$$\sigma[u]=\left(\frac{u_l+u_r}{2}\right)(u_r-u_l+v_r-v_l),$$
$$\sigma[v]=\left(\frac{v_l+v_r}{2}\right)(u_r-u_l+v_r-v_l),$$

and two states can be joined by an admissible shock if

$$u_l+v_l>u_r+v_r,\quad \frac{u_l}{v_l}=\frac{u_r}{v_r}.$$

A Roe matrix is given in this case by:

$$\mathcal{A}(\mathbf{u}_l,\mathbf{u}_r)=\begin{bmatrix}0.5(u_l+u_r) & 0.5(u_l+u_r)\\ 0.5(v_l+v_r) & 0.5(v_l+v_r)\end{bmatrix}. \tag{4.3}$$

As it will be seen in Test 1, the corresponding Roe method captures correctly the discontinuities of the weak solutions, what puts on evidence that being path-conservative is not in itself a barrier to the convergence to the right solutions. Nevertheless this is not true for other choices of family of paths. Let us consider, for instance, the family of paths given by the viscous profiles of the regularized system:

$$\begin{cases}\partial_t u+\partial_x\left(\dfrac{u^2}{2}\right)+u\partial_x v &= \epsilon u_{xx},\\ \partial_t v+\partial_x\left(\dfrac{v^2}{2}\right)+v\partial_x u &= \epsilon v_{xx},\end{cases}\qquad (x,t)\in\mathbb{R}\times\mathbb{R}^+, \tag{4.4}$$

introduced in [4]: see this reference for the expression of the corresponding family of paths.

#### 4.1.3. Cell-marking criterion and in-cell discontinuous reconstruction

It will be seen in Test 2 that Godunov's method does not converge to the right weak solutions. In [11] the in-cell discontinuous reconstruction technique has been used to correct this issue with good results. To apply this technique, a cell is marked if

$$u^n_{j-1}+v^n_{j-1}>u^n_{j+1}+v^n_{j+1}.$$

Strategy 1 (based on the exact solutions of the Riemann problems) is followed here to select the discontinuous reconstruction (see Subsection 3.2). More precisely, in a marked cell the left and right states are chosen as follows:

$$\sigma^n_j=\frac{1}{2}(u^n_{j-1}+v^n_{j-1}+u^n_{j+1}+v^n_{j+1}),\quad \mathbf{u}^n_{j,l}=\mathbf{u}^*(\mathbf{u}^n_{j-1},\mathbf{u}^n_{j+1}),\quad \mathbf{u}^n_{j,r}=\mathbf{u}^n_{j+1},$$

where $\mathbf{u}^*(\mathbf{u}^n_{j-1},\mathbf{u}^n_{j+1})$ represents the state at the left of the shock wave appearing in the solution of the Riemann problem. Finally, the conserved variable $u+v$ is chosen to determine $d^n_j$, i.e.

$$d^n_j(u^n_{j,l}+v^n_{j,l})+(1-d^n_j)(u^n_{j,r}+v^n_{j,r})=(u^n_j+v^n_j).$$

This method is extended here to second order following Section 3.

#### 4.1.4. Numerical tests

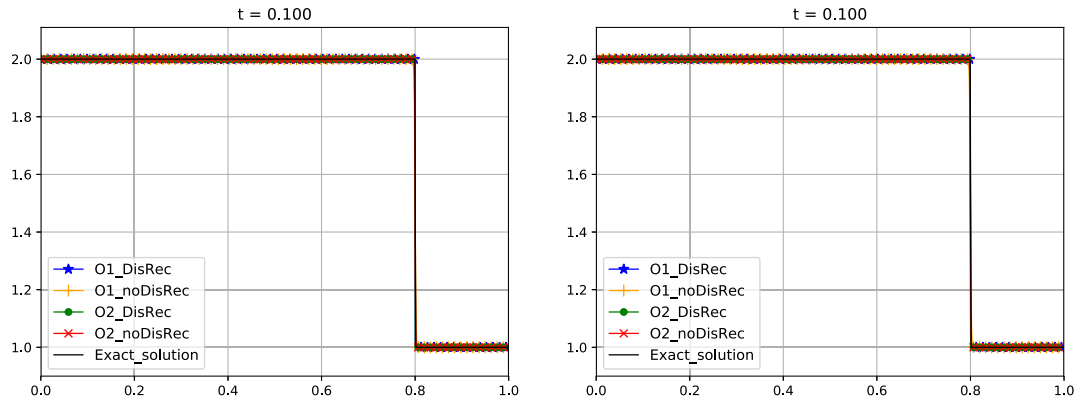
*Test 1: Coupled Burgers' equations with straight segment paths*

In this test case we consider the definition of weak solution related to the family of straight segments (4.2) and the corresponding Roe matrix (4.3). Let us consider the following initial condition

$$\mathbf{u}_0(x)=[u_0(x),v_0(x)]^T=\begin{cases}[2.0,2.0]^T & \text{if } x<0.5,\\ [1.0,1.0]^T & \text{otherwise.}\end{cases}$$

The solution of the Riemann problem in this case consists of a shock wave joining the left and right states.

Fig. 1 compares the exact solution with the numerical approximations at time $t=0.1$ obtained with O$p$_noDisRec and O$p$_DisRec, $p=1,2$ using a 1000-cell mesh and CFL=0.5: notice that, in this particular case, the standard path-conservative




(a) Variable $u$ (b) Variable $v$

**Fig. 1.** Coupled Burgers system. Test 1: exact solution and numerical solutions obtained at time $t = 0.1$ with 1000 cells. Left: variable $u$. Right: variable $v$.

methods capture correctly the right weak solution. The same comparison has been done for a number of different Riemann problems and, in all cases, the numerical solutions converge to the weak solution.

*Test 2: Isolated shock wave*

From now on, the family of paths given by the viscous profiles of the regularized equation (4.4) is considered. Let us consider the following initial condition taken from [8]

$$\mathbf{u}_0(x) = [u_0(x), v_0(x)]^T = \begin{cases} [7.99, 11.01]^T & \text{if } x < 0.5, \\ [0.25, 0.75]^T & \text{otherwise.} \end{cases}$$

The solution of the Riemann problem consists of a shock wave joining the left and right states.

Fig. 2 compares the exact solution with the numerical approximations at time $t = 0.03$ obtained with O$p$_noDisRec(Godunov) and O$p$_DisRec(Godunov), $p = 1, 2$ using a 100-cell mesh: as it can be seen Godunov's method and its second order extension do not capture the discontinuity properly what is not the case for the methods based on the discontinuous reconstruction. Fig. 3 compares the numerical and exact solutions using a 1000-cell mesh. We remark that the differences between them do not disappear for the standard path-conservative methods as the mesh is refined. This clearly indicates a convergence failure. CFL = 0.5 has been considered.

**Remark 4.1.** If a path-conservative method based on the family of straight segments like the one considered in Test1 is used in this case, the numerical solutions converge to the entropy weak solution corresponding to that family of paths that is different to the one corresponding to the viscous profiles of the regularized problem (4.4).

*Test 3: Contact discontinuity + shock wave*

We consider now the initial condition

$$\mathbf{u}_0(x) = [u_0(x), v_0(x)]^T = \begin{cases} [5, 1]^T & \text{if } x < 0.5, \\ [1, 2]^T & \text{otherwise.} \end{cases}$$

The solution of the corresponding Riemann problems consists of a stationary contact discontinuity followed by a shock. Fig. 4 shows the exact and the numerical solutions at time $t = 0.05$ using a 1000-cell mesh and CFL = 0.5. The conclusions are the same: the in-cell discontinuous reconstruction methods of order 1 and 2 get the exact solution while the standard Godunov methods do not.

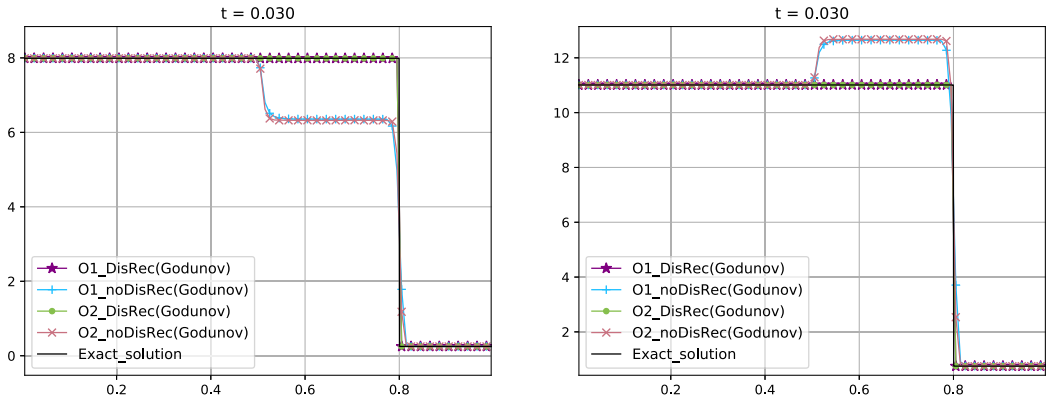
*Test 4: Contact discontinuity + rarefaction*

We consider the initial condition

$$\mathbf{u}_0(x) = [u_0(x), v_0(x)]^T = \begin{cases} [1, 2]^T & \text{if } x < 0.5, \\ [5, 1]^T & \text{otherwise.} \end{cases}$$

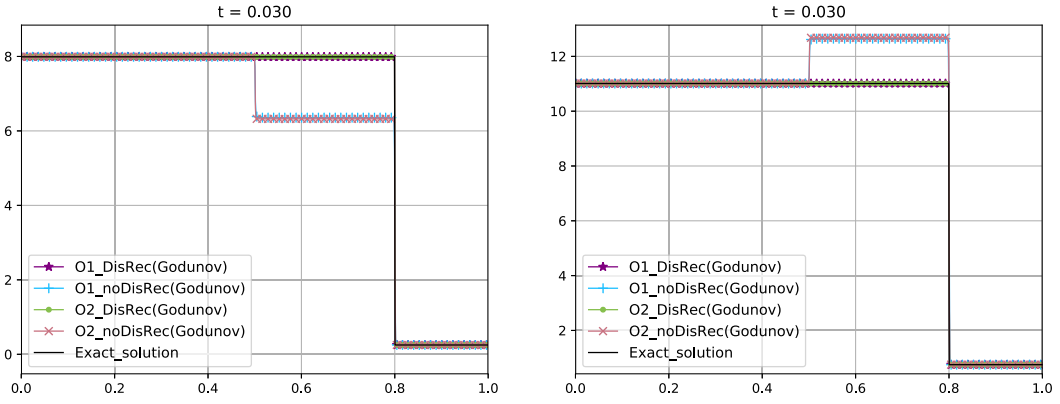
The solution of the corresponding Riemann problem consists of a stationary contact discontinuity followed by a rarefaction.

Fig. 5 shows the exact and the numerical solutions at time $t = 0.05$ using a 1000-cell mesh. In this case all the methods converge to the exact solution but the second order one captures better the solution, as expected.




(a) Variable $u$ (b) Variable $v$

**Fig. 2.** Coupled Burgers system. Test 2: exact solution and numerical solutions obtained at time $t = 0.03$ with 100 cells. Left: variable $u$. Right: variable $v$.




(a) Variable $u$ (b) Variable $v$

**Fig. 3.** Coupled Burgers system. Test 2: exact solution and numerical solutions obtained at time $t = 0.03$ with 1000 cells. Left: variable $u$. Right: variable $v$.

**Table 1**
$L^1$ errors $\|\Delta \cdot\|_1$ at time $t = 1$ for the Coupled Burgers model with initial conditions (4.5).

| $\|\Delta u\|_1$ (1st) | $\|\Delta v\|_1$ (1st) | $\|\Delta u\|_1$ (2nd) | $\|\Delta v\|_1$ (2nd) |
|---|---|---|---|
| 3.40e-19 | 8.88e-19 | 1.17e-18 | 1.78e-18 |

*Test 5: Stationary solution*

We consider the initial condition

$$\mathbf{u}_0(x) = [u_0(x), v_0(x)]^T = [\sin(x), 1 - \sin(x)]^T, \tag{4.5}$$

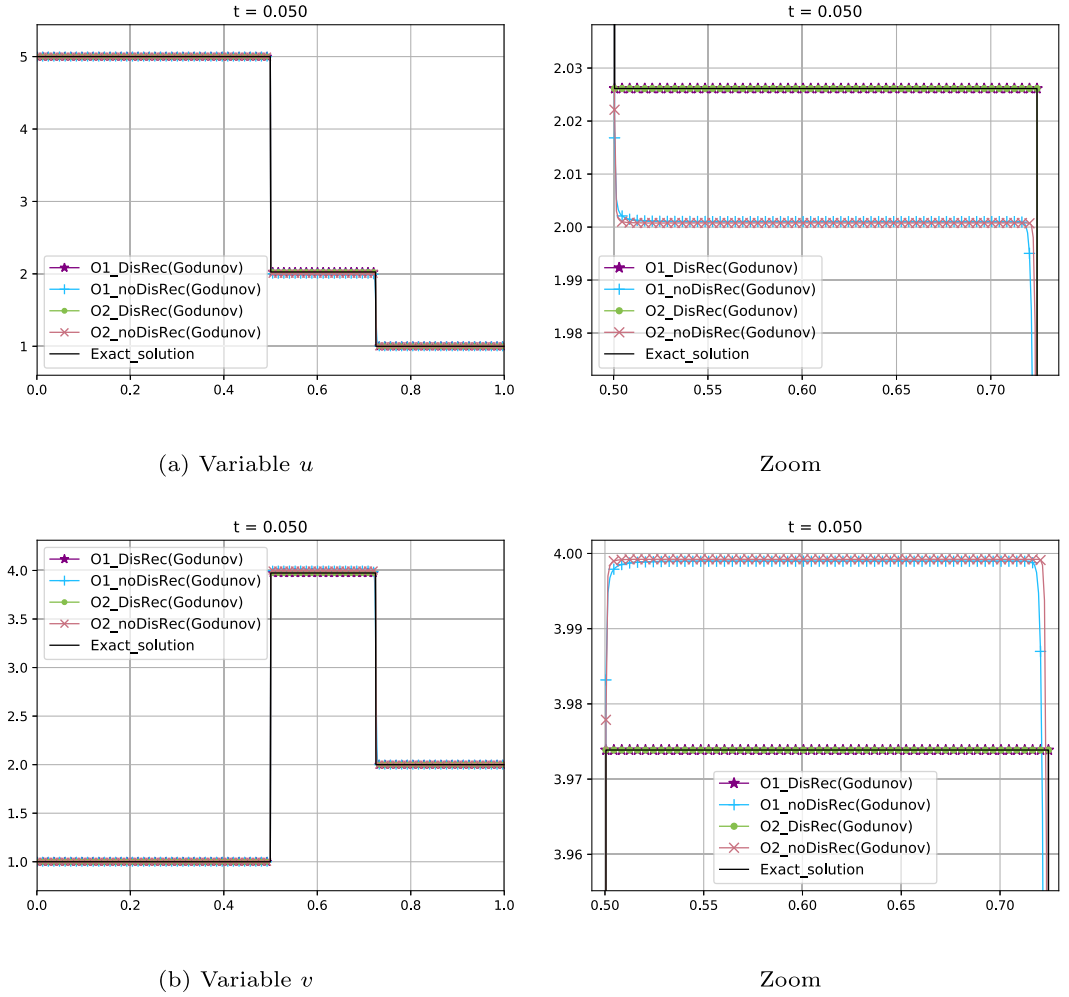
that is a stationary solution of the system (4.1). We show in Fig. 6 the numerical solution obtained with the first- and second-order discontinuous in-cell reconstruction using a 1000-mesh. The results in Fig. 7 and Table 1 show that the both schemes are well-balanced.

*Test 6: Perturbed stationary solution*

We consider finally the initial condition

$$\mathbf{u}_0(x) = [u_0(x), v_0(x)]^T = [\sin(x) + 0.2e^{-2000(r-0.5)^2}, 1 - \sin(x)]^T, \tag{4.6}$$

that is the stationary solution (4.5) with a perturbation in the variable $u$. Figs. 8 and 9 show the numerical solutions obtained at time $t = 0.2$ and $t = 1$ using a 1000-cell mesh together with a reference solution obtained with the first order in-cell discontinuous reconstruction Godunov scheme using a 10000-cell mesh. As it can be seen the second order methods




(a) Variable $u$ Zoom

(b) Variable $v$ Zoom

**Fig. 4.** Coupled Burgers system. Test 3: exact solution and numerical solutions obtained at time $t = 0.05$ with 1000 cells. Top: variable $u$ (left), zoom middle state (right). Down: variable $v$ (left), zoom middle state (right).

capture better the smooth parts of the solution and the ones with the in-cell reconstruction capture better the shock appearing in the perturbation. Observe that, in this case, the stationary solution (4.5) is not restored: a different equilibrium with a stationary bump placed at the initial location of the perturbation is obtained once the waves generated by the perturbation leaves the computational domain.

### *4.2. Gas dynamics equations in Lagrangian coordinates*

#### *4.2.1. Equations*

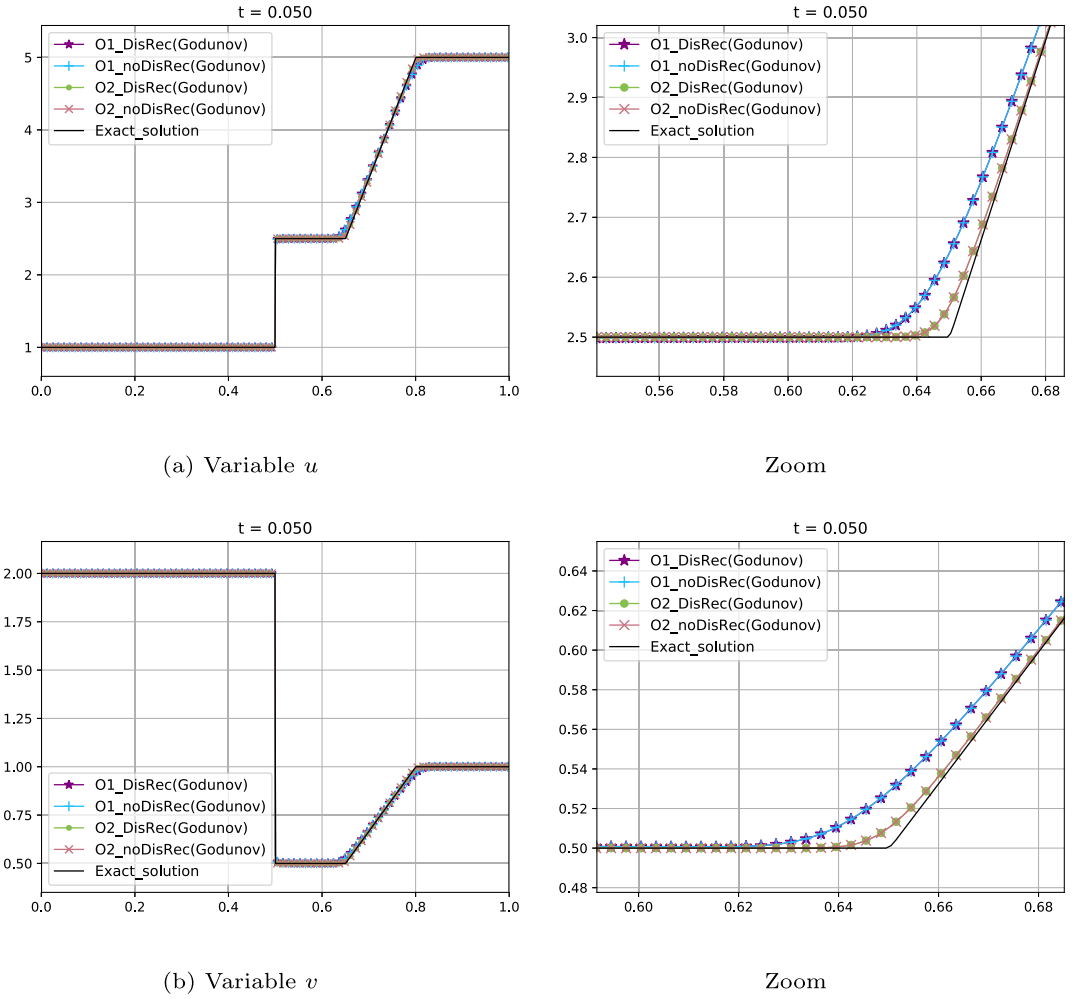
Let us consider the gas dynamics equations in Lagrangian coordinates:

$$\begin{cases} \partial_t \tau - \partial_x u = 0, \\ \partial_t u + \partial_x p = 0, \\ \partial_t E + \partial_x (pu) = 0, \end{cases} \tag{4.7}$$

where $\tau > 0$ represents the inverse of the density, $u$ is the velocity, $p = p(\tau, e) > 0$ is the pressure, $e$ is the specific internal energy, and $E = e + u^2/2$ the specific total energy. For the sake of simplicity, we consider a perfect gas equation of state $p(\tau, e) = (\gamma - 1)e/\tau$ where $\gamma > 1$. System (4.7) can be rewritten in nonconservative form as follows

$$\begin{cases} \partial_t \tau - \partial_x u = 0, \\ \partial_t u + \partial_x p = 0, \\ \partial_t e + p \partial_x u = 0, \end{cases} \tag{4.8}$$

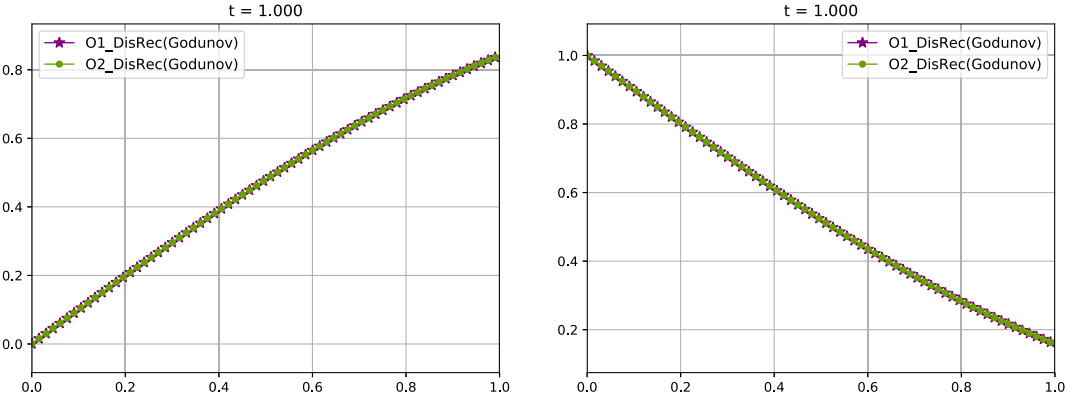
that can be written in the form (1.1) with




(a) Variable $u$ Zoom

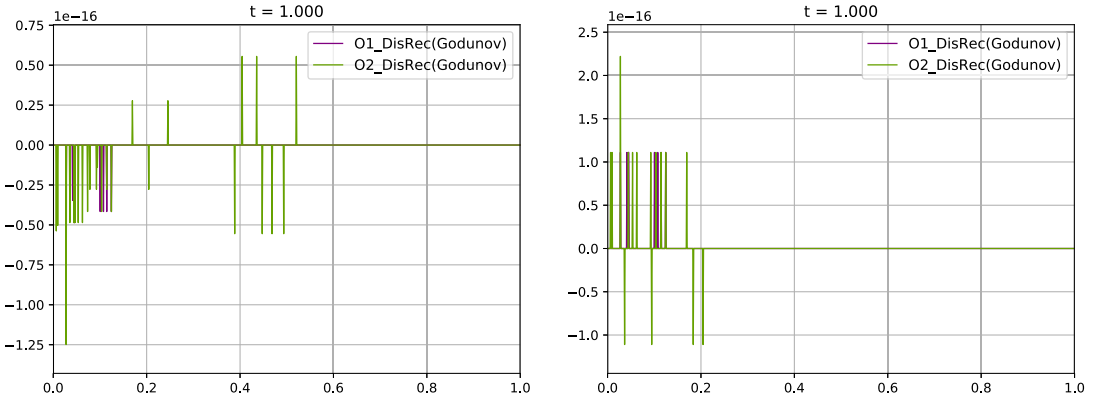
(b) Variable $v$ Zoom

**Fig. 5.** Coupled Burgers system. Test 4: exact solution and numerical solutions obtained at time $t = 0.05$ with 1000 cells. Top: variable $u$ (left), zoom rarefaction (right). Down: variable $v$ (left), zoom rarefaction (right).



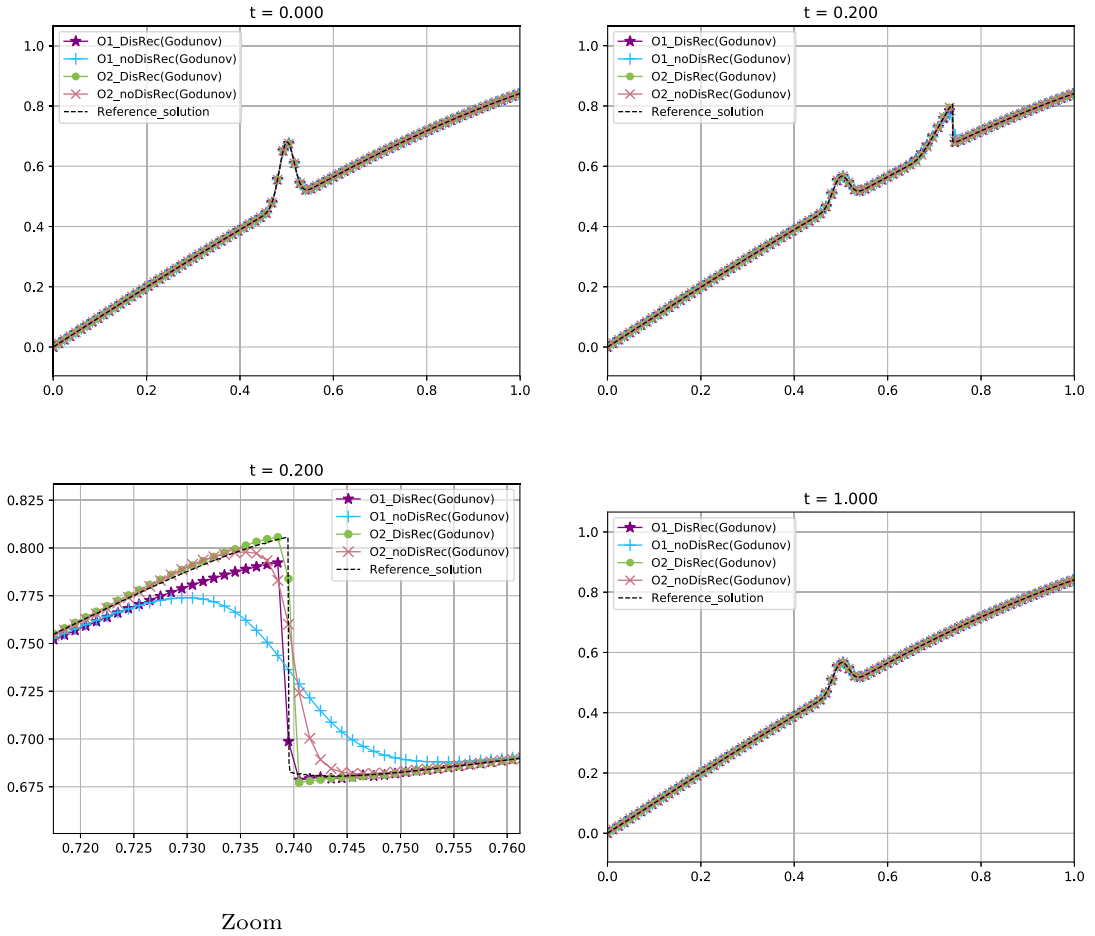

(a) Variable $u$ (b) Variable $v$

**Fig. 6.** Coupled Burgers system. Test 5: numerical solution of (4.5) at time $t = 1.00$ with 1000 cells. Left: variable $u$. Right: variable $v$.




(a) Variable $u$ (b) Variable $v$

**Fig. 7.** Coupled Burgers system. Test 5: difference between the numerical solution at $t = 1.00$ and the stationary solution. Left: variable $u$. Right: variable $v$.




Zoom

**Fig. 8.** Coupled Burgers system. Test 6: variable $u$. Top: initial condition (left), reference and numerical solutions obtained at time $t = 0.2$ with 1000 cells (right). Down: zoom of the perturbation area at time $t = 0.2$ (left), reference and numerical solutions obtained at time $t = 1$ (right).

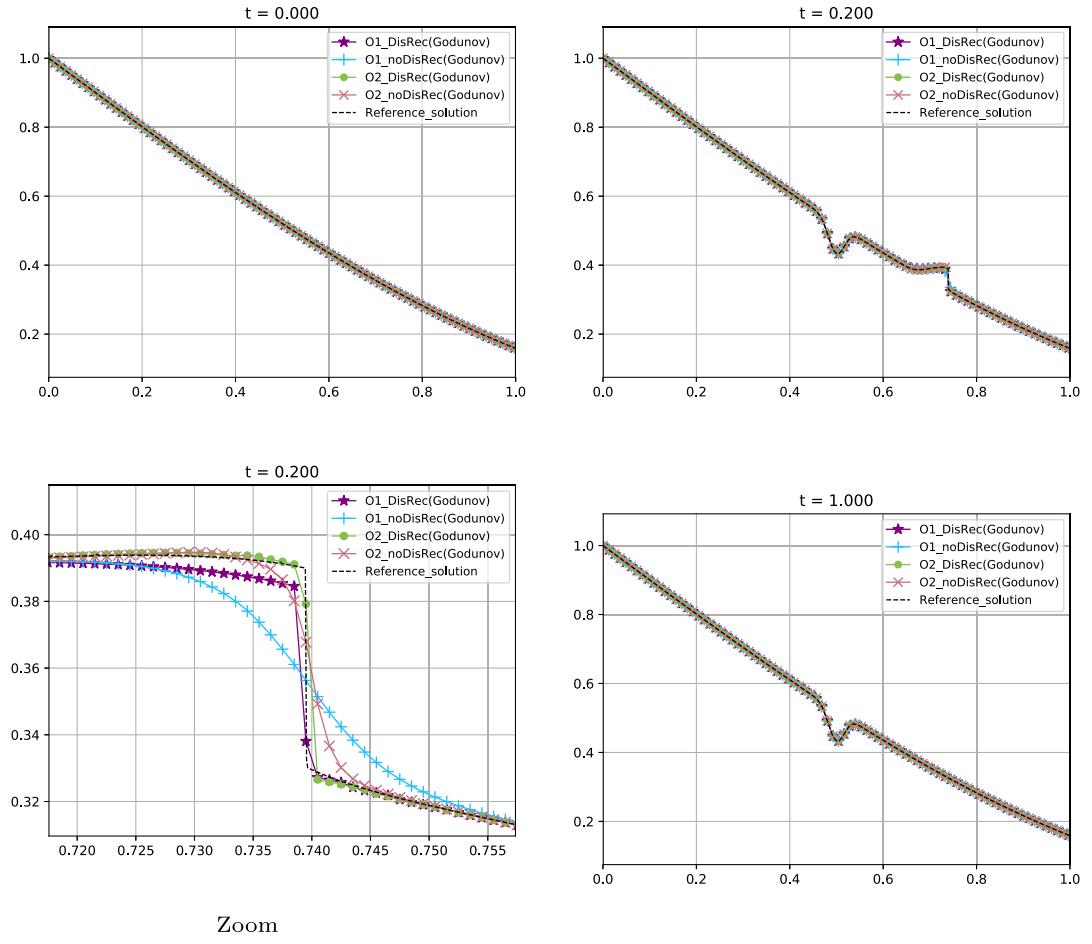



**Fig. 9.** Coupled Burgers system. Test 6: variable $v$. Top: initial condition (left), reference and numerical solutions obtained at time $t = 0.2$ with 1000 cells (right). Down: zoom of the perturbation area at time $t = 0.2$ (left), reference and numerical solutions obtained at time $t = 1$ (right).

$$\mathbf{u} = \begin{pmatrix} \tau \\ u \\ e \end{pmatrix}, \quad \mathcal{A}(\mathbf{u}) = \begin{pmatrix} 0 & -1 & 0 \\ -\dfrac{(\gamma-1)e}{\tau^2} & 0 & \dfrac{\gamma-1}{\tau} \\ 0 & \dfrac{(\gamma-1)e}{\tau} & 0 \end{pmatrix}.$$

The system is strictly hyperbolic with eigenvalues

$$\lambda_1(\mathbf{u}) = -\sqrt{\gamma p/\tau}, \quad \lambda_2(\mathbf{u}) = 0, \quad \lambda_3(\mathbf{u}) = \sqrt{\gamma p/\tau},$$

whose characteristic fields are given by the eigenvectors

$$R_1(\mathbf{u}) = [1, \sqrt{\gamma p/\tau}, -p]^T, \quad R_2(\mathbf{u}) = [1, 0, p/(\gamma-1)], \quad R_3(\mathbf{u}) = [1, -\sqrt{\gamma p/\tau}, -p]^T.$$

$R_2(\mathbf{u})$ is linearly degenerate and $R_i(\mathbf{u})$, $i = 1, 3$ are genuinely nonlinear: see [16]. On the other hand, the admissible solutions of (4.7) are selected by Lax entropy inequalities, which here are equivalent to:

$$\sigma(\tau_+ - \tau_-) \geq 0, \tag{4.9}$$

where $\tau_-$ and $\tau_+$ are the values of $\tau$ at both sides of the discontinuity and $\sigma$ its speed of propagation.

*4.2.2. Simple waves*

Once the family of paths has been chosen, the simple waves of this system are:

- Stationary contact discontinuities linking states $\mathbf{u}_l$, $\mathbf{u}_r$ such that

$$u_l = u_r, \quad p_l = p_r.$$

- Rarefactions waves joining states $\mathbf{u}_l$, $\mathbf{u}_r$ that satisfy

$$u_l < u_r,$$

and the relations given by the Riemann invariants:
  - 1-rarefactions:

$$2\sqrt{\frac{\gamma e_l}{\gamma-1}} + u_l = 2\sqrt{\frac{\gamma e_r}{\gamma-1}} + u_r, \quad \frac{e_l}{\tau_l^{\gamma-1}} = \frac{e_r}{\tau_r^{\gamma-1}}.$$

  - 2-rarefactions:

$$2\sqrt{\frac{\gamma e_l}{\gamma-1}} - u_l = 2\sqrt{\frac{\gamma e_r}{\gamma-1}} - u_r, \quad \frac{e_l}{\tau_l^{\gamma-1}} = \frac{e_r}{\tau_r^{\gamma-1}}.$$

- Shock waves joining states $\mathbf{u}_l$ and $\mathbf{u}_r$ that satisfy

$$u_l > u_r$$

and the jump conditions:

$$\begin{aligned}
&\sigma[\tau] = -[u],\\
&\sigma[u] = [p],\\
&\sigma[e] = \int_0^1 \phi_p(s;\mathbf{u}_l,\mathbf{u}_r)\partial_s\phi_u(s;\mathbf{u}_l,\mathbf{u}_r)\,ds.
\end{aligned}$$

If, for instance, the family of straight segments is chosen for the variables $\tau, u, p$

$$\phi_\tau(s;\mathbf{u}_l,\mathbf{u}_r) = \tau_l + s(\tau_r-\tau_l); \quad \phi_u(s;\mathbf{u}_l,\mathbf{u}_r) = u_l + s(u_r-u_l); \quad \phi_p(s;\mathbf{u}_l,\mathbf{u}_r) = p_l + s(p_r-p_l),$$

the jump conditions reduce to:

$$\begin{aligned}
&\sigma[\tau] = (u_l-u_r),\\
&\sigma[u] = p_r - p_l,\\
&\sigma[e] = \frac{1}{2}(p_r+p_l)(u_r-u_l).
\end{aligned}$$

It can be easily checked that these jump conditions are equivalent to the standard Rankine-Hugoniot conditions corresponding to the conservative formulation (4.7) and thus, the weak solutions are the same.

A Roe matrix is given in this case by:

$$\mathcal{A}(\mathbf{u}_l,\mathbf{u}_r) = \mathcal{A}(\bar{\mathbf{u}}), \quad \bar{\mathbf{u}}(\mathbf{u}_l,\mathbf{u}_r) = (\bar\tau,\bar u,\bar p),$$

with

$$\bar\tau = \frac{\tau_l+\tau_r}{2}, \quad \bar u = \frac{u_l+u_r}{2}, \quad \bar e = \frac{\bar p\,\bar\tau}{\gamma-1}, \quad \bar p = \frac{p_l+p_r}{2},$$

see [21].

*4.2.3. Cell-marking criterion and in-cell discontinuous reconstruction*

In [11] the in-cell discontinuous reconstruction technique has been used to correct the results that are obtained with the standard Roe path-conservative scheme. To apply this technique, a cell is marked if

$$u_{j-1}^n \geq u_{j+1}^n.$$

The second strategy to select the speed, and the left and right states of the discontinuous reconstruction based on the Roe matrix is used here (see Subsection 3.2). More precisely:

- If $u_{j-1}^n = u_{j+1}^n$ then

$$\sigma_j^n = 0, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n, \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}^n.$$

- If $u^n_{j-1} > u^n_{j+1}$ and $\tau^n_{j+1} - \tau^n_{j-1} < 0$ then

$$\sigma^n_j = -\sqrt{\gamma \bar{p}/\bar{\tau}}, \quad \mathbf{u}^n_{j,l} = \mathbf{u}^n_{j-1}, \quad \mathbf{u}^n_{j,r} = \mathbf{u}^n_{j-1} + \alpha_1 R_1(\mathbf{u}^n_{j-1}, \mathbf{u}^n_{j+1}).$$

- If $u^n_{j-1} > u^n_{j+1}$ and $\tau^n_{j+1} - \tau^n_{j-1} > 0$ then

$$\sigma^n_j = \sqrt{\gamma \bar{p}/\bar{\tau}}, \quad \mathbf{u}^n_{j,l} = \mathbf{u}^n_{j+1} - \alpha_3 R_3(\mathbf{u}^n_{j-1}, \mathbf{u}^n_{j+1}), \quad \mathbf{u}^n_{j,r} = \mathbf{u}_{j+1}.$$

Here $\bar{p}$ and $\bar{\tau}$ represent the Roe intermediate values of $p$ and $\tau$, and $\alpha_k$, $k = 1, 2, 3$ the coordinates of $\mathbf{u}^n_{j+1} - \mathbf{u}^n_{j-1}$ in the basis of eigenvectors of the Roe matrix, i.e. $\mathbf{u}^n_{j+1} - \mathbf{u}^n_{j-1} = \sum_{k=1}^{3} \alpha_k R_k(\mathbf{u}^n_{j-1}, \mathbf{u}^n_{j+1})$. This method is extended here to second order by following the procedure described in Section 3.

#### 4.2.4. Numerical tests

*Test 1: Isolated 1-shock*

Let us consider the following initial condition taken from [11]

$$\mathbf{u}_0(x) = [\tau_0(x), u_0(x), p_0(x)]^T = \begin{cases} [2.09836065573770281, 2.3046638387921279, 1]^T & \text{if } x < 0.5, \\ [8, 0, 0.1]^T & \text{otherwise.} \end{cases}$$

The solution of the Riemann problem consists of a 1-shock wave joining the left and right states. Fig. 10 compares the exact solution with the numerical approximations at time $t = 0.5$ obtained with Roe method, its second order extension based on the standard MUSCL reconstruction, and the first- and second-order discontinuous in-cell reconstruction schemes using 300-cell mesh and CFL = 0.5: as it can be seen Roe methods does not capture the discontinuities properly (as it was noted in [1]) what is not the case for the two other methods.

*Test 2: 1-shock + contact discontinuity + 3-shock*

Let us consider the following initial condition taken from [11]

$$\mathbf{u}_0(x) = [\tau_0(x), u_0(x), p_0(x)]^T = \begin{cases} [5, 3.323013993227, 0.481481481481]^T & \text{if } x < 0.5, \\ [8, 0, 0.1]^T & \text{otherwise.} \end{cases}$$

The solution of the Riemann problem consists of a 1-shock wave with negative speed, a stationary contact discontinuity, and a 3-shock that coincides with the one in the first test problem. Fig. 11 shows the numerical solutions at time $t = 0.5$ using a mesh of 300 cells and CFL = 0.5 and the conclusions are the same: the in-cell discontinuous reconstruction methods of order 1 and 2 get good approximations of the exact solution while Roe method and its second-order extension based on the standard MUSCL reconstruction do not. Fig. 12 shows the numerical solutions obtained with O1_DisRec at time $t = 0.505$ using different meshes: it can be observed that the intermediate state between the shocks and the contact discontinuity are not exactly captured as it happens with isolated shocks. Nevertheless the numerical solutions seem to converge to the exact solution when $\Delta x \to 0$.

*Test 3: 1-rarefaction + contact discontinuity + 3-shock*

Let us consider now the initial condition

$$\mathbf{u}_0(x) = [\tau_0(x), u_0(x), p_0(x)]^T = \begin{cases} [2.09836065573770281, 3.323013993227, 1]^T & \text{if } x < 0.5, \\ [8, 4, 0.1]^T & \text{otherwise.} \end{cases}$$

The solution of the Riemann problem consists of a 1-rarefaction wave whose head and tail have negative speeds, a stationary contact discontinuity, and a 3-shock with positive speed. Fig. 13 shows the numerical solutions at time $t = 0.5$ using a mesh of 300 cells and CFL = 0.5. Although all the methods capture correctly the rarefaction wave, second order methods do it better, as expected; concerning the stationary contact discontinuity and the shock wave, only the first- and second-order in-cell discontinuous reconstruction methods capture the exact solution.

*Test 4: well-balanced property*

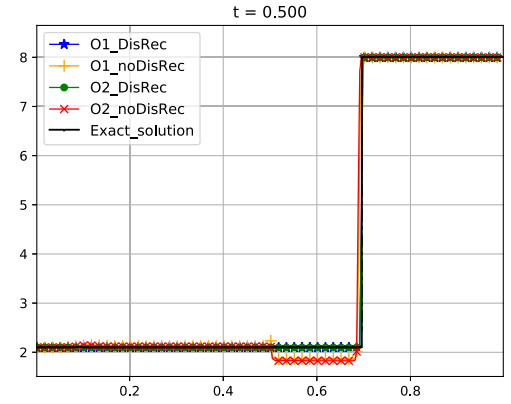
In order to test the well-balanced property, let us add a gravitational source term to system (4.8):

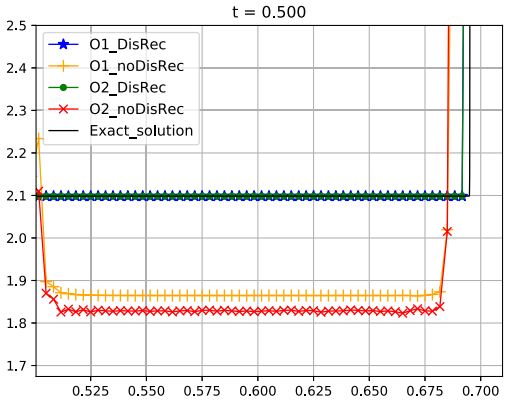
$$\begin{cases} \partial_t \tau - \partial_x u = 0, \\ \partial_t u + \partial_x p = g, \\ \partial_t e + p \partial_x u = 0, \end{cases} \tag{4.10}$$

where $g$ is the intensity of the gravitational field.

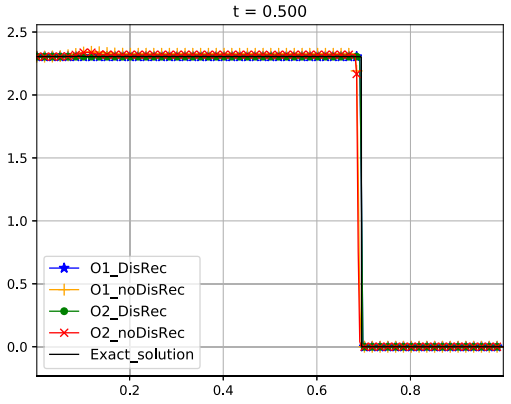
The stationary solutions of (4.10) verify:

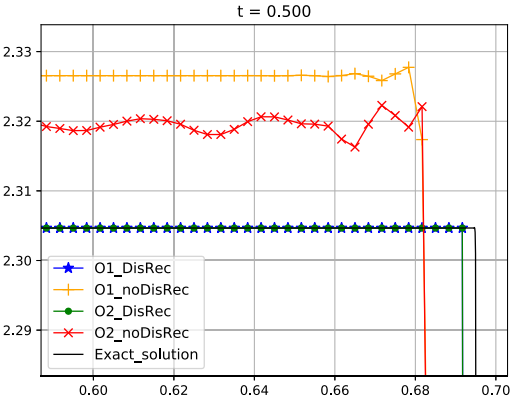
$$u = \text{constant}, \quad p - gx = \text{constant}. \tag{4.11}$$



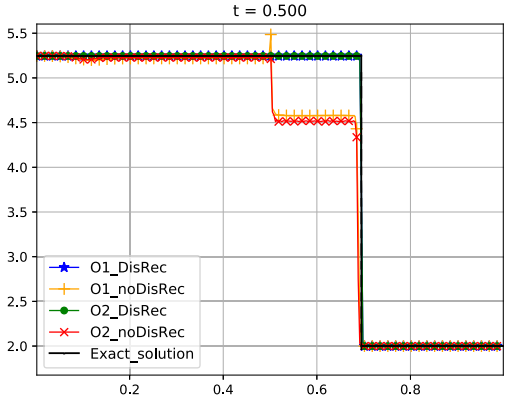



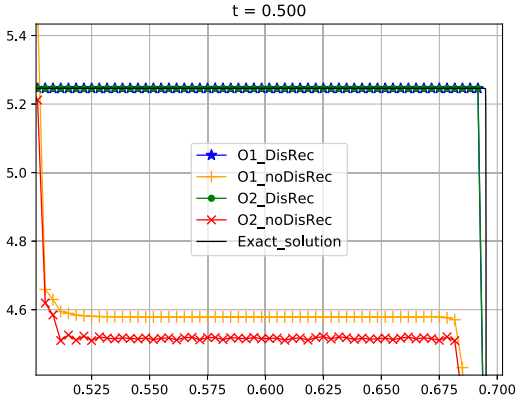

(a) Variable $\tau$ Zoom



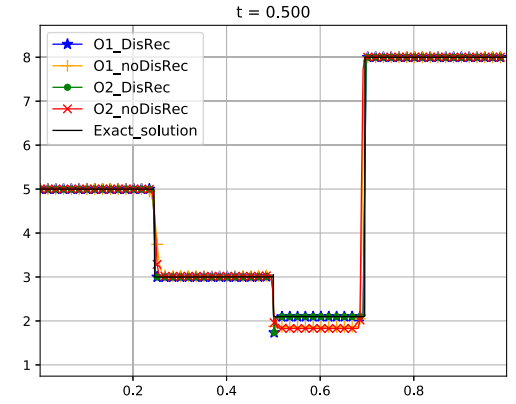

(b) Variable $u$ Zoom



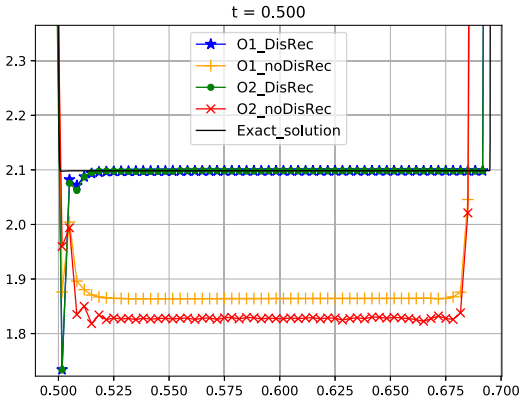




(c) Variable $e$ Zoom

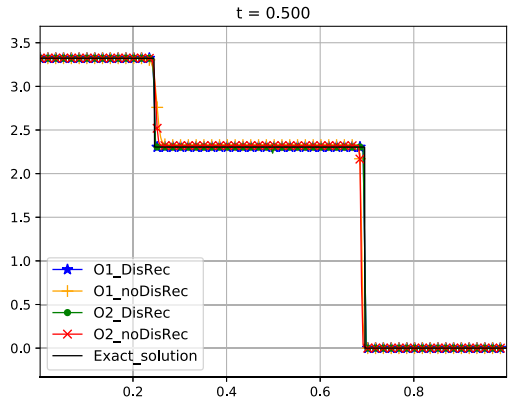
**Fig. 10.** Gas dynamics equations in Lagrangian coordinates. Test 1: exact solution and numerical solutions obtained at time $t = 0.5$ with 300 cells. Top: variable $\tau$ (left), zoom right middle state (right). Middle: variable $u$ (left), zoom right middle state (right). Down: variable $e$ (left), zoom right middle state (right).



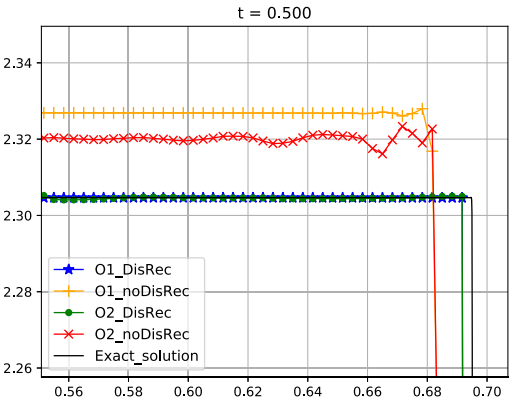

(a) Variable $\tau$



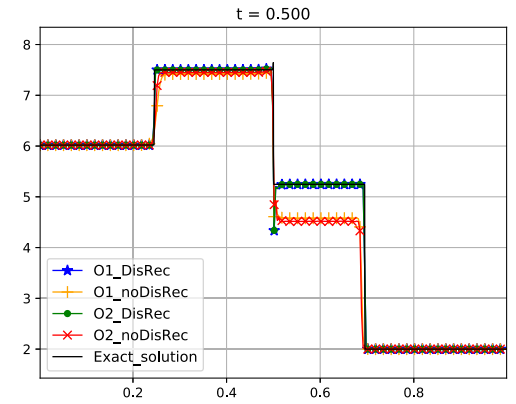

Zoom



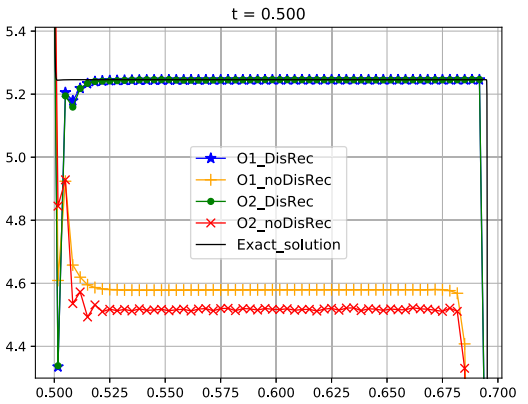

(b) Variable $u$




Zoom

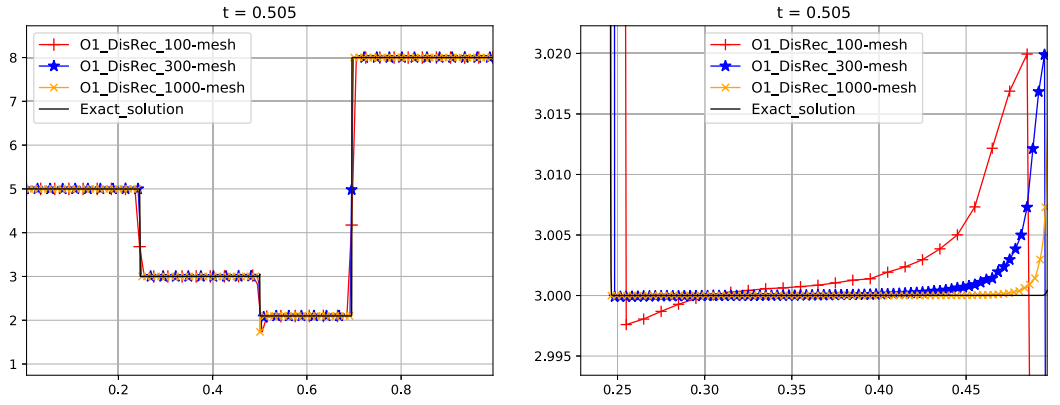



(c) Variable $e$




Zoom

**Fig. 11.** Gas dynamics equations in Lagrangian coordinates. Test 2: exact solution and numerical solutions obtained at time $t = 0.5$ with 300 cells. Top: variable $\tau$ (left), zoom right middle state (right). Middle: variable $u$ (left), zoom right middle state (right). Down: variable $e$ (left), zoom right middle state (right).




(a) Variable $\tau$ Zoom

**Fig. 12.** Gas dynamics equations in Lagrangian coordinates. Test 2: variable $\tau$. Left: exact solution and numerical solutions obtained with O1_Disrec at time $t = 0.505$ with 100, 300 and 1000 cells. Right: zoom left middle state.

**Table 2**
$L^1$ errors $\|\Delta \cdot\|_1$ at time $t = 10$ for the Gas dynamics equations in Lagrangian coordinates with initial conditions (4.12).

| $\|\Delta\tau\|_1$ (1st) | $\|\Delta u\|_1$ (1st) | $\|\Delta e\|_1$ (1st) | $\|\Delta\tau\|_1$ (2nd) | $\|\Delta u\|_1$ (2nd) | $\|\Delta e\|_1$ (2nd) |
|---|---|---|---|---|---|
| 7.40E-19 | 7.33E-17 | 0.00 | 7.40E-19 | 7.33E-17 | 0.00 |

Therefore, the curves of the family $\Gamma$ are the straight lines of the space $u, p, x$ defined by (4.11). Since the selected family of paths is linear in these variables, property (P) is satisfied and thus Roe method is well-balanced. Let us consider the following initial condition:

$$\mathbf{u}_0(x) = [\tau_0(x), u_0(x), p_0(x)]^T = \begin{cases} [1, 1, gx + 1]^T & \text{if } x < 0.5, \\ [2, 1, gx + 1]^T & \text{otherwise.} \end{cases} \tag{4.12}$$

We observe in Fig. 14 and Table 2 that the first- and second-order in-cell discontinuous reconstruction are well-balanced.

### 4.3. Modified Shallow Water system

#### 4.3.1. Equations

Let us consider the modified Shallow Water system introduced in [9]:

$$\begin{cases} \partial_t h + \partial_x q = 0, \\ \partial_t q + \partial_x \left( \dfrac{q^2}{h} \right) + qh\partial_x h = 0, \end{cases} \tag{4.13}$$

where $\mathbf{u} = [h, q]^t$ belongs to $\Omega = \{\mathbf{u} \in \mathbb{R}^2 | \quad 0 < q, \ 0 < h < (16q)^{1/3}\}$. This system can be written in the form (1.1) with

$$\mathcal{A}(\mathbf{u}) = \begin{bmatrix} 0 & 1 \\ -u^2 + uh^2 & 2u \end{bmatrix},$$

being $u = q/h$. The system is strictly hyperbolic over $\Omega$ with eigenvalues

$$\lambda_1(\mathbf{u}) = u - h\sqrt{u}, \quad \lambda_2(\mathbf{u}) = u + h\sqrt{u},$$

whose characteristic fields, given by the eigenvectors

$$R_1(\mathbf{u}) = [1, u - h\sqrt{u}]^T, \quad R_2(\mathbf{u}) = [1, u + h\sqrt{u}]^T,$$

are genuinely nonlinear.

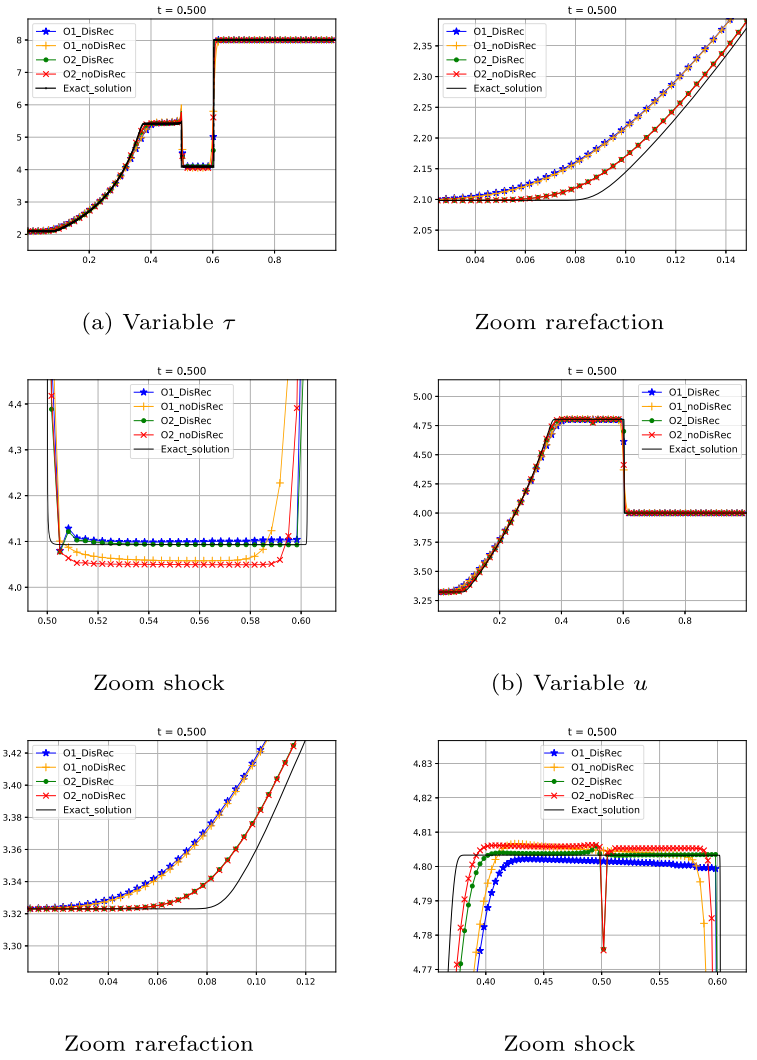



**Fig. 13.** Gas dynamics equations in Lagrangian coordinates. Test 3: exact solution and numerical solutions obtained at time $t = 0.5$ with 300 cells. Top: variable $\tau$ (left), zoom rarefaction (center), zoom shock (right). Middle: variable $u$ (left), zoom rarefaction (center), zoom shock (right). Top: variable $e$ (left), zoom rarefaction (center), zoom shock (right).

### *4.3.2. Simple waves*

Once the family of paths has been chosen, the simple waves of this system are:

- 1-rarefaction waves joining states $\mathbf{u}_l$, $\mathbf{u}_r$ such that

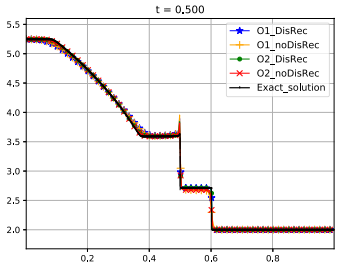
$$h_r < h_l, \quad \sqrt{u_l} + h_l/2 = \sqrt{u_r} + h_r/2,$$

and 2-rarefaction waves joining states $\mathbf{u}_l$, $\mathbf{u}_r$ such that

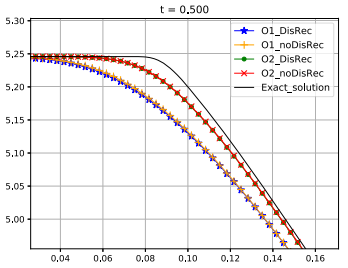
$$h_l < h_r, \quad \sqrt{u_l} - h_l/2 = \sqrt{u_r} - h_r/2.$$

- 1-shock and 2-shock waves joining states $\mathbf{u}_l$ and $\mathbf{u}_r$ such that $h_l < h_r$ or $h_r < h_l$ respectively, that satisfy the jump conditions:

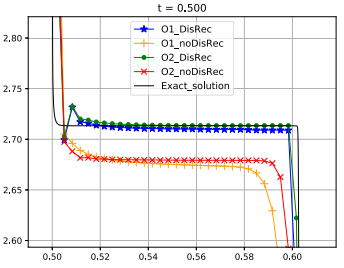
$$\sigma[h] = [q],$$
$$\sigma[q] = \left[\frac{q^2}{h}\right] + \int_0^1 \phi_q(s;\mathbf{u}_l,\mathbf{u}_r)\phi_h(s;\mathbf{u}_l,\mathbf{u}_r)\partial_s\phi_h(s;\mathbf{u}_l,\mathbf{u}_r)\,ds.$$




(c) Variable $e$

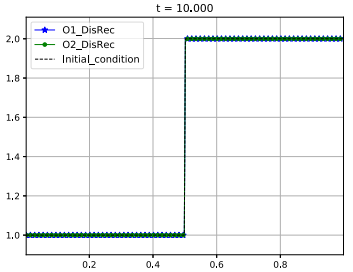



Zoom rarefaction

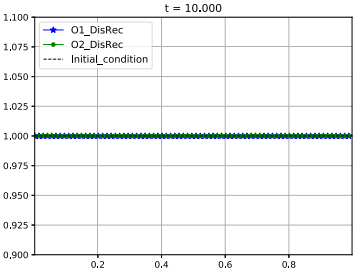



Zoom shock

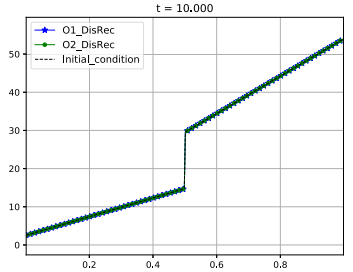
**Fig. 13.** (*continued*)




(a) Variable $\tau$




(b) Variable $u$




(c) Variable $e$

**Fig. 14.** Gas dynamics equations in Lagrangian coordinates. Test 4: numerical solution at $t = 10.00$ and the initial stationary solution. Left: variable $\tau$. Center: variable $u$. Right: variable $e$.

If, for instance, the following family of path is chosen:

$$\phi(s;\mathbf{u}_l,\mathbf{u}_r) = \begin{bmatrix} \phi_h(s;\mathbf{u}_l,\mathbf{u}_r) \\ \phi_q(s;\mathbf{u}_l,\mathbf{u}_r) \end{bmatrix} = \begin{cases} \begin{bmatrix} h_l + 2s(h_r - h_l) \\ q_l \end{bmatrix} & \text{if } 0 \le s \le \frac{1}{2}, \\ \begin{bmatrix} h_r \\ q_l + (2s-1)(q_r - q_l) \end{bmatrix} & \text{if } \frac{1}{2} \le s \le 1, \end{cases}$$

the jump conditions reduce to:

$$\sigma[h] = [q],$$
$$\sigma[q] = \left[\frac{q^2}{h}\right] + q_l\left[\frac{h^2}{2}\right].$$

If this family of paths has been selected and Lax's entropy criterion is used, the simple waves of the system are as follows:

- Given a left-hand state $\mathbf{u}_l$, the 1-shock $\mathcal{S}_1(\mathbf{u}_l)$ and the 2-shock $\mathcal{S}_2(\mathbf{u}_l)$ curves consisting of all the right-hand states that can be connected with $\mathbf{u}_l$ through a 1-shock and a 2-shock wave respectively, are:

$$\mathcal{S}_1(\mathbf{u}_l) : u = u_l - \sqrt{\frac{u_l(h+h_l)}{2h}}(h-h_l), \quad h > h_l, \tag{4.14}$$
$$\mathcal{S}_2(\mathbf{u}_l) : u = u_l - \sqrt{\frac{u_l(h+h_l)}{2h}}(h-h_l), \quad h < h_l. \tag{4.15}$$

  Moreover, given two states $\mathbf{u}_l$ and $\mathbf{u}_r$ connected by a 1-shock wave or a 2-shock wave, the speed of the shock is given by:

$$\sigma_1(\mathbf{u}_l, \mathbf{u}_r) = u_l - \sqrt{h_r u_l \frac{h_l + h_r}{2}}, \tag{4.16}$$
$$\sigma_2(\mathbf{u}_l, \mathbf{u}_r) = u_l + \sqrt{h_r u_l \frac{h_l + h_r}{2}}, \tag{4.17}$$

  respectively.
- Given a left-hand state $\mathbf{u}_l$, the 1-rarefaction $\mathcal{R}_1(\mathbf{u}_l)$ and the 2-rarefaction $\mathcal{R}_2(\mathbf{u}_l)$ consisting of all the right-hand states that can be connected with $\mathbf{u}_l$ through a 1-rarefaction and a 2-rarefaction wave, respectively, are:

$$\mathcal{R}_1(\mathbf{u}_l) : u = \left(\frac{h_l - h}{2} + \sqrt{u_l}\right)^2, \quad h < h_l, \tag{4.18}$$
$$\mathcal{R}_2(\mathbf{u}_l) : u = \left(\frac{h - h_l}{2} + \sqrt{u_l}\right)^2, \quad h > h_l. \tag{4.19}$$

*4.3.3. Cell-marking criterion*

The criterion to mark the cells is the following:

1. If $h^n_{j+1} > h^n_{j-1}$ and

$$u^n_{j-1} - \sqrt{\frac{u^n_{j-1}(h^n_{j+1} + h^n_{j-1})}{2h^n_{j+1}}}(h^n_{j+1} - h^n_{j-1}) < u^n_{j+1} < \left(\frac{h^n_{j+1} - h^n_{j-1}}{2} + \sqrt{u^n_{j-1}}\right)^2,$$

   the solution of the Riemann problem consists of a 1-shock and a 2-rarefaction waves: the cell is marked.
2. If $h^n_{j+1} < h^n_{j-1}$ and

$$u^n_{j-1} + \sqrt{\frac{u^n_{j-1}(h^n_{j+1} + h^n_{j-1})}{2h^n_{j+1}}}(h^n_{j+1} - h^n_{j-1}) < u^n_{j+1} < \left(\frac{h^n_{j-1} - h^n_{j+1}}{2} + \sqrt{u^n_{j-1}}\right)^2,$$

   the solution of the Riemann problem consists of a 1-rarefaction and a 2-shock waves: the cell is marked.
3. If $h^n_{j+1} > h^n_{j-1}$ and

$$u^n_{j+1} < u^n_{j-1} - \sqrt{\frac{u^n_{j-1}(h^n_{j+1} + h^n_{j-1})}{2h^n_{j+1}}}(h^n_{j+1} - h^n_{j-1}),$$

   or $h^n_{j+1} < h^n_{j-1}$ and

$$u^n_{j+1} < u^n_{j-1} + \sqrt{\frac{u^n_{j-1}(h^n_{j+1} + h^n_{j-1})}{2h^n_{j+1}}}(h^n_{j+1} - h^n_{j-1}),$$

   the solution of the Riemann problem consists of a 1-shock and a 2-shock waves: the cell is marked.
4. Otherwise the solution of the Riemann problem consists of two rarefactions and the cell is not marked.

*4.3.4. In-cell discontinuous reconstruction*

A Roe matrix is given in this case by

$$\mathcal{A}(\mathbf{u}_l, \mathbf{u}_r) = \begin{bmatrix} 0 & 1 \\ -\bar{u}^2 + q_l \bar{h} & 2\bar{u} \end{bmatrix},$$

where

$$\bar{u} = \frac{\sqrt{h_l}u_l + \sqrt{h_r}u_r}{\sqrt{h_l} + \sqrt{h_r}}, \quad \bar{h} = \frac{h_l + h_r}{2}.$$

The following strategy based on this Roe matrix (see Subsection 3.2) is used to select the speed, and the left and right states of the discontinuous reconstruction:

- If the solution of the Riemann problem consists of a 1-shock and a 2-rarefaction waves (case 1):

$$\sigma_j^n = \bar{u} - h_{j-1}^n \sqrt{\bar{u}}, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n, \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j-1}^n + \alpha_1 R_1(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n),$$

  where $\bar{u}$ is the Roe average of $u_{j-1}^n$ and $u_{j+1}^n$, and $\alpha_k$, $k = 1, 2$ represent the coordinates of $\mathbf{u}_{j+1}^n - \mathbf{u}_{j-1}^n$ in the basis of eigenvectors of the Roe matrix, i.e. $\mathbf{u}_{j+1}^n - \mathbf{u}_{j-1}^n = \sum_{k=1}^2 \alpha_k R_k(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n)$.
- If the solution of the Riemann problem consists of a 1-rarefaction and a 2-shock waves (case 2):

$$\sigma_j^n = \bar{u} + h_{j-1}^n \sqrt{\bar{u}}, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_{j+1}^n - \alpha_2 R_2(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n), \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}^n.$$

- If the solution of the Riemann problem consists of a 1-shock and a 2-shock waves (case 3) we select one of them depending on the amplitude of the $\alpha_1$ and $\alpha_2$ coefficients in order to choose the 'dominant' one:
  - If $|\alpha_1| \leq |\alpha_2|$ then:

$$\sigma_j^n = \bar{u} + h_{j-1}^n \sqrt{\bar{u}}, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_{j+1}^n - \alpha_2 R_2(\mathbf{u}_{j-1}^n, \mathbf{u}_{j+1}^n), \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}.$$

  - If $|\alpha_1| > |\alpha_2|$ then:

$$\sigma_j^n = \bar{u} - h_{j-1}^n \sqrt{\bar{u}}, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n, \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j-1}^n + \alpha_1 R_1(\mathbf{u}_{j-1}, \mathbf{u}_{j+1}).$$

The variable $h$ is selected in (3.4).

*4.3.5. Non-isolated shocks capturing*

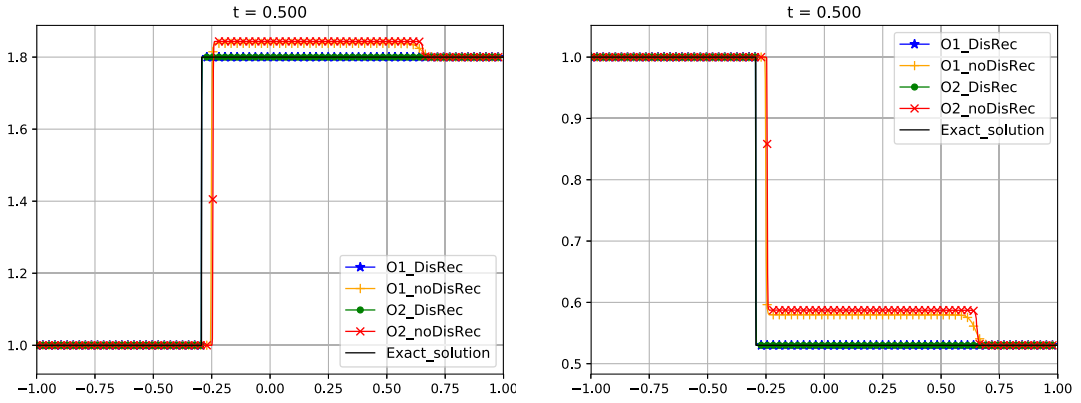
Although, according to Theorem 3.2, the first- and second-order in-cell discontinuous reconstruction methods capture exactly isolated shock waves, this is not the case for non-isolated shocks, as it has been seen in Fig. 12 for the Gas dynamics equations in Lagrangian coordinates. Nevertheless they clearly improve the results provided by standard methods and, in particular, the numerical results seem to converge to the right weak solution as $\Delta x$ tends to 0.

This fact is also observed for the Modified Shallow Water system, specially in the case of two non-isolated shock waves traveling in opposite directions, as it will be seen in Test 2. In order to improve the numerical results, we have developed a more sophisticated in-cell discontinuous reconstruction based on the exact solution of the Riemann problems (see Subsection 3.2) that allows one to capture better the intermediate states. The key ingredients are:

- The solution of the Riemann problem with initial data $\mathbf{u}_{j-1,r}^{n-1}$ and $\mathbf{u}_{j+1,l}^{n-1}$ is used to mark the cells instead of the one corresponding to the initial data $\mathbf{u}_{j-1}^n$ and $\mathbf{u}_{j+1}^n$, where $\mathbf{u}_{j-1,r}^{n-1}$ and $\mathbf{u}_{j+1,l}^{n-1}$ are the states selected in the discontinuous reconstruction in the previous time step.
- The exact intermediate state is used when the solution of the Riemann problem involves two shock waves.
- If the solution of this Riemann problem involves two shock waves traveling in the same direction, a reconstruction with two discontinuities (one for each of the shock waves) is considered, so that the complete structure of the Riemann solution is imposed.

The details of the reconstruction are given in Appendix A.

The numerical methods using the first strategy for the discontinuous reconstruction (based on the Roe matrix) will be labeled again by O$p$_DisRec and those using the second one (based on the exact solutions of the Riemann problems) by O$p$_ExactDisRec.




(a) Variable $h$ (b) Variable $q$

**Fig. 15.** Modified Shallow Water system. Test 1: Numerical solutions obtained withe the first- and second-order methods with and without discontinuous reconstruction based on the Roe matrix at time $t = 0.5$ with 1000 cells. Left: variable $h$. Right: variable $q$.

#### 4.3.6. Numerical tests

*Test 1: Isolated 1-shock*

Let us consider the following initial condition taken from [9]

$$\mathbf{u}_0(x) = [h_0(x), q_0(x)]^T = \begin{cases} [1, 1]^T & \text{if } x < 0, \\ [1.8, 0.530039370688997]^T & \text{otherwise.} \end{cases}$$

The solution of the Riemann problem consists of a 1-shock wave joining the left and right states. Fig. 15 compares the exact solution and the numerical approximations at time $t = 0.15$ obtained with Roe method, its second order extension based on the standard MUSCL-Hancock reconstruction, and the first- and second-order discontinuous in-cell reconstruction schemes based on the Roe matrix using 1000-cell mesh and CFL = 0.5: as it can be seen the standard Roe methods does not capture the discontinuities properly what is not the case for the in-cell discontinuous reconstruction methods based on the Roe structure. The results obtained with O$p$_ExactDisRec are similar.

*Test 2: left-moving 1-shock + right-moving 2-shock*

Let us consider the following initial condition

$$\mathbf{u}_0(x) = [h_0(x), q_0(x)]^T = \begin{cases} [1, 1]^T & \text{if } x < 0, \\ [1.5, 0.1855893974385]^T & \text{otherwise.} \end{cases} \tag{4.20}$$

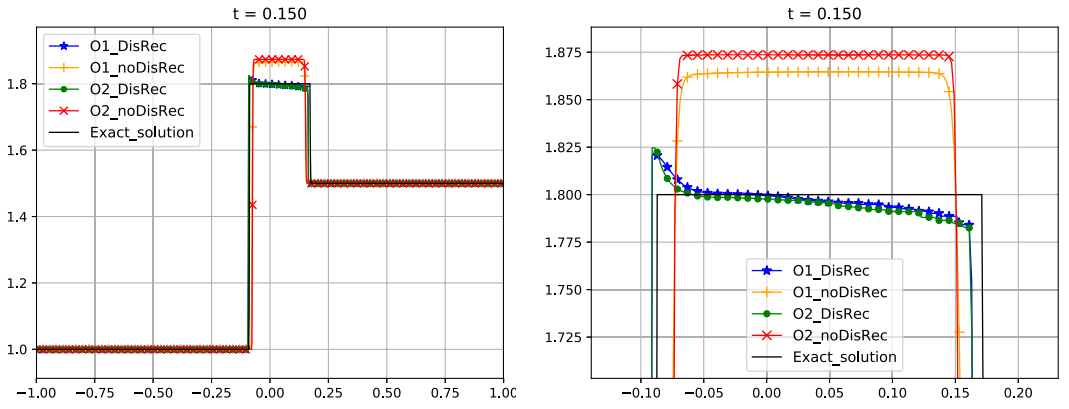
The solution of the Riemann problem consists of a 1-shock wave with negative speed and a 2-shock with positive speed with intermediate state $\mathbf{u}_* = [1.8, 0.530039370688997]^T$. Fig. 16 compares the exact solution with the numerical approximations at time $t = 0.15$ obtained with Roe method, its second order extension based on the standard MUSCL-Hancock reconstruction, and the first- and second-order discontinuous in-cell reconstruction schemes based on the Roe matrix using 1000-cell mesh and CFL = 0.5: as it can be seen none of them capture the discontinuities exactly, although the ones using in-cell discontinuous reconstruction do it better. Fig. 17 shows the numerical solutions obtained with the first-order method with discontinuous reconstruction based on the Roe matrix at time $t = 0.15$ using different cell meshes: as we can see the numerical solutions seem to converge to the exact solution as $\Delta x \to 0$. In Fig. 18 the results given by O$p$_ExactDisRec, $p = 1, 2$ are shown: observe that both of them capture exactly the two shocks.

*Test 3: right-moving 1-shock + right-moving 2-shock*

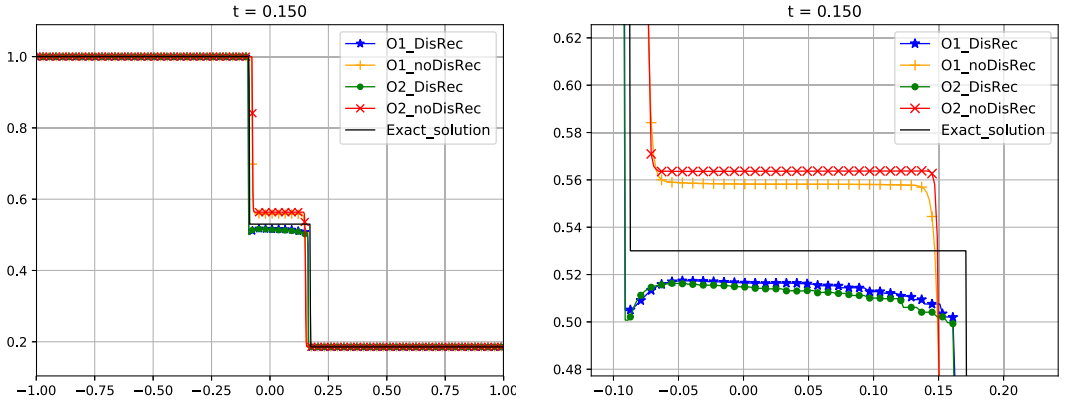
Let us consider the following initial condition

$$\mathbf{u}_0(x) = [h_0(x), q_0(x)]^T = \begin{cases} [1, 1]^T & \text{if } x < 0, \\ [5, 2.86423084288]^T & \text{otherwise.} \end{cases} \tag{4.21}$$

The solution of the Riemann problem consists of a 1-shock and a 2-shock waves with positive speed and intermediate state $\mathbf{u}_* = [1.5, 5.96906891076]^T$. Fig. 19 shows the exact solution and the numerical approximations at time $t = 0.06$ obtained with Roe method, its second-order extension based on the standard MUSCL-Hancock reconstruction, and the first- and second-order discontinuous in-cell reconstruction schemes based on the Roe structure using 1000-cell mesh and CFL = 0.5: as in the previous test case, the in-cell discontinuous reconstruction captures the shocks and intermediate state much better




(a) Variable $h$ Zoom



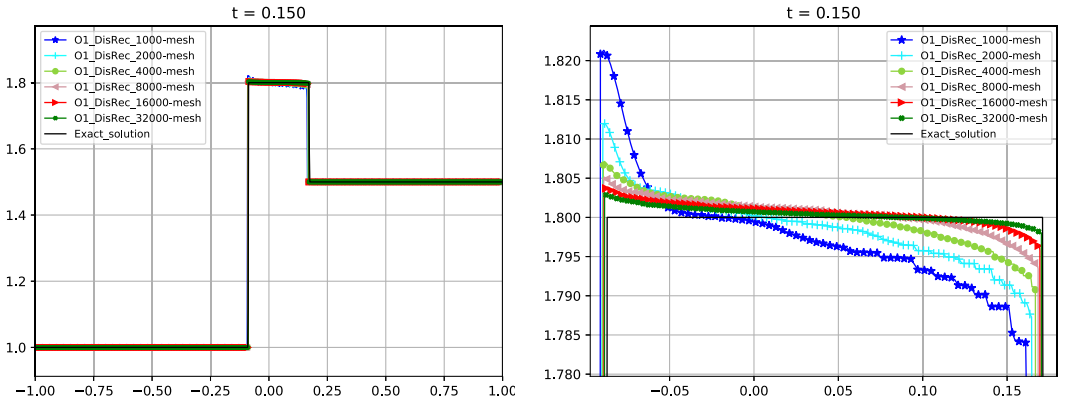

(b) Variable $q$ Zoom

**Fig. 16.** Modified Shallow Water system. Test 2: Numerical solutions obtained with the first- and second-order methods with and without discontinuous reconstruction based on the Roe matrix at time $t = 0.15$ with 1000 cells. Top: variable $h$ (left), zoom middle state (right). Down: variable $q$ (left), zoom middle state (right).

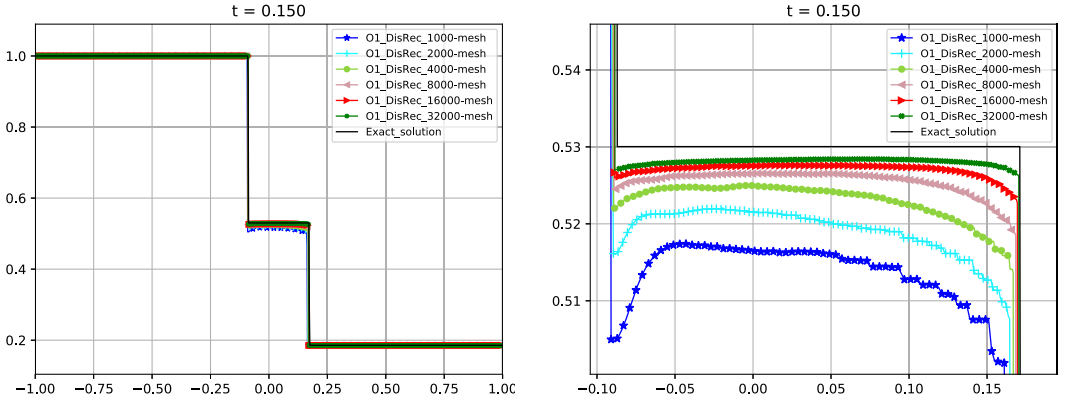
than the standard first- and second-order Roe methods. In Fig. 20 the results given by the first- and second-order in-cell discontinuous schemes based in the exact solution of the Riemann problems are shown: both of them capture exactly the exact solution.

## 5. Conclusions

In this paper, an extension to second-order accuracy of the in-cell discontinuous reconstruction methods introduced in [11] is presented: it has been compared with the first-order one using several numerical tests. We observe, as expected, an improvement in the smooth parts of the solutions. The isolated shock-capturing property is enunciated, proved and tested. In the presence of more than one shock we have used two different strategies: one based on the linearized Riemann problem when a Roe matrix is available, and another one based on the exact Riemann problems when the solution is explicitly known. We have observed that the strategy based on the Roe matrix can fail when the intermediate states appearing in the solution of the linearized Riemann problem do not coincide with the exact intermediate states appearing in the solution of the exact Riemann problem. The only important part in the in-cell reconstruction procedure is to know the exact intermediate states, so it is not necessary to know the entire structure of the exact Riemann problem. The well-balanced properties of the schemes are also studied. Future work will focus on the extension to high-order accuracy through the Taylor expansion and the application of the Cauchy-Kovalevski procedure and the extension of in-cell reconstruction for models with more waves appearing in their Riemann problems.

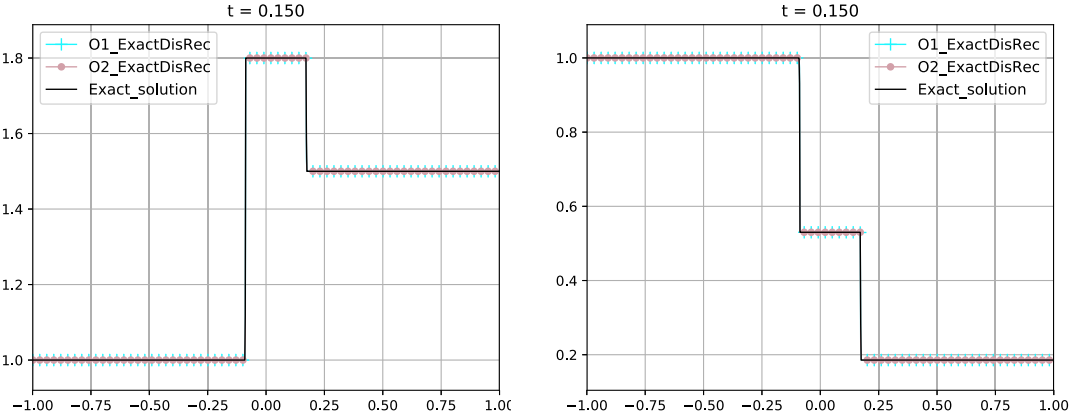



(a) Variable $h$ Zoom



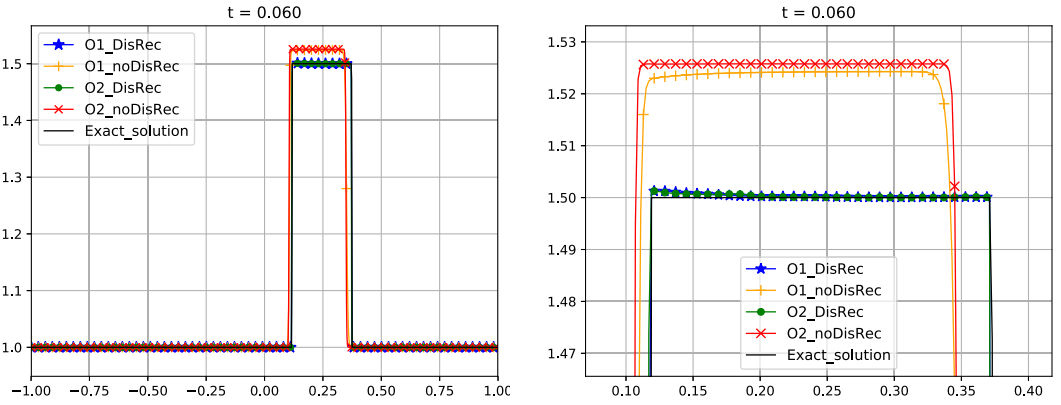

(b) Variable $q$ Zoom

**Fig. 17.** Modified Shallow Water system. Test 2: Numerical solutions obtained with the first-order methods with discontinuous reconstruction based on the Roe matrix at time $t = 0.15$ with different cell meshes. Top: variable $h$ (left), zoom middle state (right). Down: variable $q$ (left), zoom middle state (right).

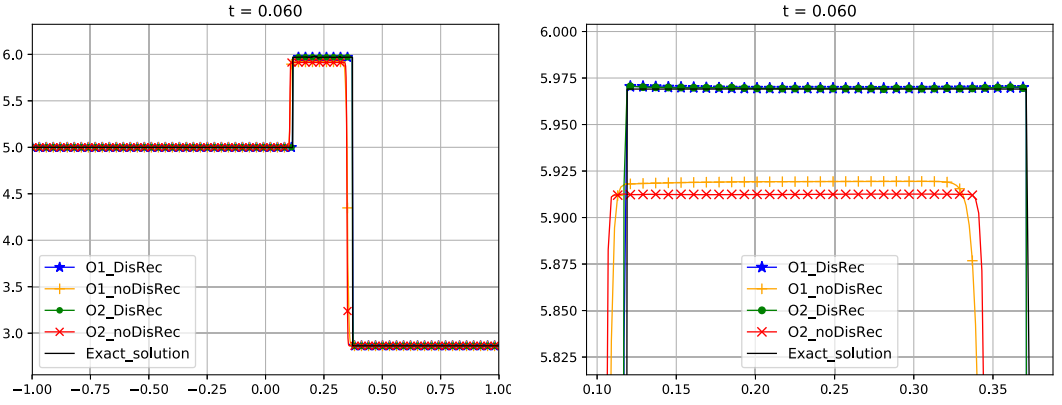



(a) Variable $h$ (b) Variable $q$

**Fig. 18.** Modified Shallow Water system. Test 2: Numerical solutions obtained with the first- and second-order methods with discontinuous reconstruction based on the exact solutions of the Riemann problems at time $t = 0.15$ with 1000 cells. Left : variable $h$. Right: variable $q$.

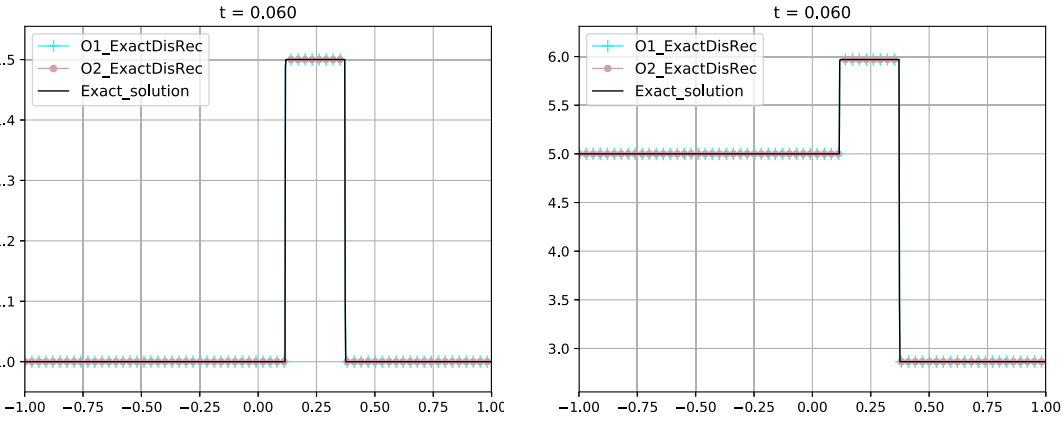



(a) Variable $h$ Zoom




(b) Variable $q$ Zoom

**Fig. 19.** Modified Shallow Water system. Test 3: Numerical solutions obtained with the first- and second-order methods with and without discontinuous reconstruction based on the Roe matrix at time $t = 0.06$ with 1000 cells. Top: variable $h$ (left), zoom middle state (right). Down: variable $q$ (left), zoom middle state (right).




(a) Variable $h$ (b) Variable $q$

**Fig. 20.** Modified Shallow Water system. Test 3: Numerical solutions obtained with the first- and second-order methods with discontinuous reconstruction based on the exact solutions of the Riemann problems at time $t = 0.06$ with 1000 cells. Left: variable $h$. Right: variable $q$.

## CRediT authorship contribution statement

All authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The research of CP, MC and EPG was partially supported by the Spanish Government (SG), the European Regional Development Fund (ERDF), the Regional Government of Andalusia (RGA), and the University of Málaga (UMA) through the projects of reference RTI2018-096064-B-C21 (SG-ERDF), UMA18-Federja-161 (RGA-ERDF-UMA), and P18-RT-3163 (RGA-ERDF). EPG was also financed by the Junior Scientific Visibility Program from the Foundation Mathématiques Jacques Hadamard for a stay of three month in the Laboratoire de Mathématiques de Versailles (LMV) with reference ANR-11-LABX-0056-LMH, LabEx LMH. TML was supported by the Spanish Government (SG) through the projects of reference RTI2018-096064-B-C22. The authors thank the anonymous reviewer whose comments helped to improve the paper.

## Appendix A. Non-isolated shock capturing in-cell discontinuous reconstruction for the modified Shallow Water system

In order to avoid an excess of indices the following notation will be used:

$$\mathbf{u}^{n-1}_{j-1,r} = \mathbf{u}_L = [h_L, q_L]^T, \quad \mathbf{u}^{n-1}_{j+1,l} = \mathbf{u}_R = [h_R, q_R]^T.$$

The discontinuous reconstruction is then as follows:

- If the solution of the Riemann problem consists of 1-shock and a 2-rarefaction (case 1) then

$$\sigma^n_j = \sigma_1(\mathbf{u}_l, \mathbf{u}_*), \quad \mathbf{u}^n_{j,l} = \mathbf{u}^n_{j-1}, \quad \mathbf{u}^n_{j,r} = \mathbf{u}_*,$$

  where $\mathbf{u}_* = [h_*, q_*]^T$ is the intermediate state in the solution of the Riemann problem: $h_*$ is the root of the function:

$$f_{s,r}(h) = \left(\frac{h-h_r}{2} + \sqrt{u_r}\right)^2 - u_l + \sqrt{\frac{u_l(h+h_l)}{2h}}(h-h_l),$$

  such that $h_l < h_* < h_r$. Once $h_*$ has been computed, $q_*$ is given by

$$q_* = h_*\left(\frac{h_*-h_r}{2} + \sqrt{u_r}\right)^2.$$

- If the solution of the Riemann problem consists of a 1-rarefaction and a 2-shock (case 2), then:

$$\sigma^n_j = \sigma_2(\mathbf{u}_*, \mathbf{u}_r), \quad \mathbf{u}^n_{j,l} = \mathbf{u}_*, \quad \mathbf{u}^n_{j,r} = \mathbf{u}^n_{j+1},$$

  where $\mathbf{u}_* = [h_*, q_*]^T$ is the intermediate state: $h_*$ is the root of the function:

$$f_{r,s}(h) = \left(\frac{h_l-h}{2} + \sqrt{u_l}\right)\left(\frac{h_l-h}{2} + \sqrt{u_l} + \sqrt{\frac{h_r+h}{2h_r}}(h_r-h)\right) - u_r,$$

  such that $h_r < h_* < h_l$. Once $h_*$ has been computed, $q_*$ is given by

$$q_* = h_*\left(\frac{h_l-h_*}{2} + \sqrt{u_l}\right)^2.$$

- If the solution of the Riemann problem consists of a 1-shock and a 2-shock (case 3), the intermediate state $\mathbf{u}_* = [h_*, q_*]^T$ can be computed as follows: $h_*$ is the root of the function

$$f_{s,s}(h) = u_*(h) + \sqrt{\frac{u_*(h)(h+h_r)}{2h_r}}(h_r-h) - u_r,$$

  where

$$u_*(h) = u_l - \sqrt{\frac{u_l(h+h_l)}{2h}}(h-h_l),$$

such that $h_* < h_l$ and $h_* < h_r$. Once $h_*$ has been computed, $q_*$ is obtained by:

$$q_* = h_* u_*(h_*).$$

Let us denote by $\sigma_1$ and $\sigma_2$ the speeds of the 1 and the 2 shock waves $\sigma_1(\mathbf{u}_l, \mathbf{u}_*)$ and $\sigma_2(\mathbf{u}_*, \mathbf{u}_r)$. The discontinuous reconstruction is then selected as follows:

– If $\sigma_1 < 0 < \sigma_2$: let $d_1$ and $d_2$ be given by

$$d_1 = \frac{h_* - h_j^n}{h_* - h_l}, \quad d_2 = \frac{h_r - h_j^n}{h_r - h_*}.$$

Then:

∗ If $|\sigma_1| \leq |\sigma_2|$:

· If $0 \leq d_2 \leq 1$, then

$$\sigma_j^n = \sigma_2, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_*, \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}^n.$$

· Otherwise, if $0 \leq d_1 \leq 1$, then

$$\sigma_j^n = \sigma_1, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n, \quad \mathbf{u}_{j,r}^n = \mathbf{u}_*.$$

∗ If $|\sigma_1| > |\sigma_2|$:

· If $0 \leq d_1 \leq 1$, then

$$\sigma_j^n = \sigma_1, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_{j-1}^n, \quad \mathbf{u}_{j,r}^n = \mathbf{u}_*.$$

· Otherwise, if $0 \leq d_2 \leq 1$, then

$$\sigma_j^n = \sigma_2, \quad \mathbf{u}_{j,l}^n = \mathbf{u}_*, \quad \mathbf{u}_{j,r}^n = \mathbf{u}_{j+1}^n.$$

– Otherwise (i.e. if $0 \leq \sigma_1 < \sigma_2$ or $\sigma_1 < \sigma_2 \leq 0$): let $d_1$ and $d_2$ be such that

$$\begin{cases} d_1 h_l + (d_2 - d_1) h_* + (1 - d_2) h_r = h_j^n, \\ d_1 q_l + (d_2 - d_1) q_* + (1 - d_2) q_r = q_j^n. \end{cases} \tag{A.1}$$

Then:

$$P_j^n(x,t) = \begin{cases} \mathbf{u}_l & \text{if } x \leq x_{j-1/2} + d_1 \Delta x + \sigma_1 (t - t_n), \\ \mathbf{u}_* & \text{if } x_{j-1/2} + d_1 \Delta x + \sigma_1 (t - t_n) \leq x \leq x_{j-1/2} + d_2 \Delta x + \sigma_2 (t - t_n), \\ \mathbf{u}_r & \text{otherwise.} \end{cases} \tag{A.2}$$

This in-cell discontinuous reconstruction can only be done if $0 \leq d_1, d_2 \leq 1$, otherwise the cell is unmarked. Moreover, if $d_1 = d_2 = 1$ and the speeds of the shocks are positive (resp. if $d_1 = d_2 = 0$ and the speeds of the shocks are negative) the cell is unmarked and the cell $I_{j+1}$ (resp. the cell $I_{j-1}$) is marked if necessary.

Observe that, when the speeds of the shocks have the same sign, the discontinuous reconstruction coincides with the solution of the Riemann problem.

## References


[1] R. Abgrall, S. Karni, A comment on the computation of non-conservative products, J. Comput. Phys. 229 (8) (2010) 2759–2763.

[2] B. Audebert, F. Coquel, Hybrid Godunov-Glimm method for a nonconservative hyperbolic system with kinetic relations, in: Numerical Mathematics and Advanced Applications, Springer, Berlin, Heidelberg, 2006, pp. 646–653.

[3] A. Beljadid, P.G. LeFloch, S. Mishra, C. Parés, Schemes with well-controlled dissipation. Hyperbolic systems in nonconservative form, Commun. Comput. Phys. 21 (4) (2017) 913–946.

[4] C. Berthon, Schéma nonlinéaire pour l'approximation numérique d'un système hyperbolique non conservatif, C. R. Math. 335 (12) (2002) 1069–1072.

[5] C. Berthon, F. Coquel, Nonlinear projection methods for multi-entropies Navier-Stokes systems, in: Innovative Methods for Numerical Solution of Partial Differential Equations, World Scientific, 2002, pp. 278–304.

[6] M. Castro, J.M. Gallardo, C. Parés, High order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. Applications to shallow-water systems, Math. Comput. 75 (255) (2006) 1103–1135.

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

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

[9] M.J. Castro, P.G. LeFloch, M.L. Muñoz-Ruiz, C. Parés, Why many theories of shock waves are necessary: convergence error in formally path-consistent schemes, J. Comput. Phys. 227 (17) (2008) 8107–8129.
[10] M.J. Castro, T. Morales de Luna, C. Parés, Well-balanced schemes and path-conservative numerical methods, in: Handbook of Numerical Analysis, in: Handbook of Numerical Methods for Hyperbolic Problems Applied and Modern Issues, vol. 18, Elsevier, 2017, pp. 131–175.
[11] C. Chalons, Path-conservative in-cell discontinuous reconstruction schemes for non conservative hyperbolic systems, Commun. Math. Sci. 18 (1) (2020) 1–30.
[12] C. Chalons, F. Coquel, Navier-Stokes equations with several independent pressure laws and explicit predictor-corrector schemes, Numer. Math. 101 (3) (2005) 451–478.
[13] C. Chalons, F. Coquel, A new comment on the computation of non-conservative products using Roe-type path conservative schemes, J. Comput. Phys. 335 (2017) 592–604.
[14] G. Dal Maso, P.G. LeFloch, F. Murat, Definition and weak stability of nonconservative products, J. Math. Pures Appl. 74 (6) (1995) 483–548.
[15] U.S. Fjordholm, S. Mishra, Accurate numerical discretizations of non-conservative hyperbolic systems, Modél. Math. Anal. Numér. 46 (1) (2012) 187–206.
[16] E. Godlewski, P.-A. Raviart, Numerical Approximation of Hyperbolic Systems of Conservation Laws, Springer, 1995.
[17] A. Harten, J.M. Hyman, Self adjusting grid methods for one-dimensional hyperbolic conservation laws, J. Comput. Phys. 50 (2) (1983) 235–269.
[18] A. Hiltebrand, S. Mishra, C. Parés, Entropy-stable space–time DG schemes for non-conservative hyperbolic systems, Modél. Math. Anal. Numér. 52 (3) (2018) 995–1022.
[19] P.G. LeFloch, S. Mishra, Numerical methods with controlled dissipation for small-scale dependent shocks, Acta Numer. 23 (2014) 743–816.
[20] M.L. Muñoz-Ruiz, C. Parés, Godunov method for nonconservative hyperbolic systems, Modél. Math. Anal. Numér. 41 (1) (2007) 169–185.
[21] C.D. Munz, On Godunov-type schemes for lagrangian gas dynamics, SIAM J. Numer. Anal. 31 (1) (1994) 17–42.
[22] C. Parés, Numerical methods for nonconservative hyperbolic systems: a theoretical framework, SIAM J. Numer. Anal. 44 (1) (2006) 300–321.
[23] C. Parés, M.J. Castro-Díaz, On the well-balance property of Roe's method for nonconservative hyperbolic systems. Applications to shallow-water systems, Modél. Math. Anal. Numér. 38 (5) (2004) 821–852.
[24] I. Toumi, A weak formulation of Roe's approximate Riemann solver, J. Comput. Phys. 102 (2) (1992) 360–373.
[25] B. Van Leer, Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme, J. Comput. Phys. 14 (4) (1974) 361–370.
[26] B. Van Leer, On the relation between the upwind-differencing schemes of Godunov, Engquist–Osher and Roe, SIAM J. Sci. Stat. Comput. 5 (1) (1984) 1–20.