

# High order extensions of Roe schemes for two dimensional nonconservative hyperbolic systems

M. J. Castro <sup>a</sup>, E.D. Fernández-Nieto <sup>b</sup>, A. M. Ferreiro <sup>c</sup>,  
J. A. García-Rodríguez <sup>c</sup>, C. Parés <sup>a</sup>

<sup>a</sup>*Dpto. de Análisis Matemático, Universidad de Málaga, Campus Teatinos s/n, 29080 Málaga. Spain.*

<sup>b</sup>*Dpto. de Matemática Aplicada I, Universidad de Sevilla, E.T.S. Arquitectura, 41012 Sevilla. Spain.*

<sup>c</sup>*Dpto. de Matemáticas, Universidad de A Coruña, Campus de Elviña s/n, 15071 A Coruña. Spain.*

---

## Abstract

*This paper is concerned with the development of well-balanced high order Roe methods for two-dimensional nonconservative hyperbolic systems. In particular, we are interested in extending the methods introduced in [3] to the two-dimensional case. We also investigate the well-balance properties and the consistency of the resulting schemes. We focus in applications to one and two layer shallow water systems.*

*Key words:* Generalized Roe Schemes, 2d nonconservative hyperbolic systems, nonconservative products, finite volume schemes, conservation laws, source terms, Shallow Water systems, two-layer problems, geophysical flows.

---

*Email addresses:* `castro@anamat.cie.uma.es` (M. J. Castro), `edofer@us.es` (E.D. Fernández-Nieto), `aferreiro@udc.es` (A. M. Ferreiro), `jagrodriguez@udc.es` (J. A. García-Rodríguez), `pares@anamat.cie.uma.es` (C. Parés).

## 1 Introduction

The motivating question in this paper is how to develop well-balanced high order numerical schemes for PDE systems of the form

$$\frac{\partial U}{\partial t} + \frac{\partial F_1}{\partial x_1}(U) + \frac{\partial F_2}{\partial x_2}(U) = B_1(U) \frac{\partial U}{\partial x_1} + B_2(U) \frac{\partial U}{\partial x_2} + S_1(U) \frac{\partial H}{\partial x_1} + S_2(U) \frac{\partial H}{\partial x_2}, \quad (1)$$

where the unknown  $U(\mathbf{x}, t)$  is defined in  $D \times (0, T)$ ,  $D$  being a domain of  $\mathbb{R}^2$ , and takes values on an open convex subset  $\Omega$  of  $\mathbb{R}^N$ ;  $F_i, i = 1, 2$  are two regular functions from  $\Omega$  to  $\mathbb{R}^N$ ;  $B_i, i = 1, 2$  are two regular matrix-valued function from  $\Omega$  to  $\mathcal{M}_{N \times N}(\mathbb{R})$ ;  $S_i, i = 1, 2$  are two functions from  $D$  to  $\mathbb{R}^N$ ; and finally  $H(\mathbf{x})$  is a known function from  $D$  to  $\mathbb{R}$ .

System (1) includes as particular cases: systems of conservation laws ( $B_i = 0, S_i = 0, i = 1, 2$ ); systems of conservation laws with source term or balance laws ( $B_i = 0, i = 1, 2$ ); and coupled systems of conservation laws.

In particular, the shallow water systems that govern the flow of one layer or two superposed layers of immiscible homogeneous fluids can be written in the form (1). Systems with similar characteristics also appear in other fluid models as two-phase flows.

The purpose of this paper is to extend to the two-dimensional case the high-order numerical methods introduced in [3], based on a generalized first order Roe scheme and a high order reconstruction operator. To do this, first the system is written in the form:

$$W_t + \mathcal{A}_1(W)W_{x_1} + \mathcal{A}_2(W)W_{x_2} = 0, \quad \mathbf{x} = (x_1, x_2) \in D, \quad t \in (0, T), \quad (2)$$

being  $\mathcal{A}_i, i = 1, 2$  two regular matrix-valued functions from  $D \times \mathbb{R}$  to  $\mathcal{M}_{(N+1) \times (N+1)}(\mathbb{R})$ . In effect, adding to (1) the equation

$$\frac{\partial H}{\partial t} = 0,$$

the system can be rewritten in this form (see [12], [6], [7], [9], [10]).

The nonconservative products involved in (2) do not make sense in general within the framework of distributions. Here, we follow the theory developed by Dal Maso, Le Floch and Murat in [4] to give a sense to these products as *Borel measures*. This theory is based on the choice of a family of paths.

Once the system has been rewritten, the first goal is to obtain a general writing of a Roe scheme for (2). To do this, first we extend to the two-dimensional case the notion of Roe linearization introduced in [20], which is also based on the choice of a family of paths. Next, a finite volume mesh of  $\Omega$  is constructed and piecewise constant approximations of the solution are considered. These approximations are updated by considering, at any time level, a family of projected Riemann Problem in the normal direction to each edge of the mesh. These projected Riemann problems are then linearized by using the Roe linearization. The approximated solutions of these 1d linear Riemann problems

are finally averaged in the cells to obtain the new piecewise constant approximation of the solution.

High order extensions are next obtained by extending to the 2d case the procedure developed in [3] for 1d problems: a reconstruction operator is considered, i.e. an operator that, given a family of constant values at the cells of the mesh, provides two functions at the edges, in such a manner that if the values at the cell are the averages of a regular function, then the functions at the edges provided by the operator are high order approximations of the value of the traces of that regular function.

This paper is organized as follows: in next section we discuss briefly the definition of weak solution of (2). In Section 3 we present the general expression of a Roe method for such a system and we give some general results concerning its consistency and well-balance properties. Section 4 is devoted to the high order extension of Roe methods based on a reconstruction operator. The well-balance properties of such a scheme are also discussed. The numerical schemes obtained in Sections 2 and 3 are particularized to systems (1) in Section 5. The application to the particular cases of the one and two-layer shallow water systems are discussed in Section 6 and 7, respectively. A number of numerical tests are presented in these Sections to verify the performance and well-balanced properties of the schemes.

## 2 Weak solutions

We consider the problem:

$$W_t + \mathcal{A}_1(W)W_{x_1} + \mathcal{A}_2(W)W_{x_2} = 0, \quad \mathbf{x} = (x_1, x_2) \in D \subset \mathbb{R}^2, t \in (0, T), \quad (3)$$

where  $W(\mathbf{x}, t)$  takes values on a convex domain  $\Omega$  of  $\mathbb{R}^N$  and  $\mathcal{A}_i$ ,  $i = 1, 2$  are two smooth and locally bounded matrix-valued functions from  $\Omega$  to  $\mathcal{M}_{N \times N}(\mathbb{R})$ .

Given an unitary vector  $\eta = (\eta_1, \eta_2) \in \mathbb{R}^2$ , we define the matrix

$$\mathcal{A}(W, \eta) = \mathcal{A}_1(W)\eta_1 + \mathcal{A}_2(W)\eta_2.$$

We assume that (3) is strictly hyperbolic, i.e. for all  $W \in \Omega$  and  $\forall \eta \in \mathbb{R}^2$ , the matrix  $\mathcal{A}(W, \eta)$  has  $N$  real and distinct eigenvalues

$$\lambda_1(W, \eta) < \dots < \lambda_N(W, \eta).$$

$\mathcal{A}(W, \eta)$  is thus diagonalizable:

$$\mathcal{A}(W, \eta) = \mathcal{K}(W, \eta)\mathcal{D}(W, \eta)\mathcal{K}^{-1}(W, \eta),$$

being  $\mathcal{D}(W, \eta)$  the diagonal matrix whose coefficients are the eigenvalues of  $\mathcal{A}(W, \eta)$  and  $\mathcal{K}(W, \eta)$  is a matrix whose  $j$ -th column is an eigenvector  $R_j(W, \eta)$  associated to the eigenvalue  $\lambda_j(W, \eta)$ ,  $j = 1, \dots, N$ .

For discontinuous solutions  $W$ , the nonconservative products  $\mathcal{A}_k(W)W_{x_k}$ ,  $k = 1, 2$  do not make sense as distributions. However, the theory developed by Dal Maso, LeFloch and Murat in [4] allows to give a rigorous definition of nonconservative products, associated to the choice of a family of paths in  $\Omega$ .

**Definition 1** A family of paths in  $\Omega \subset \mathbb{R}^N$  is a locally Lipschitz map

$$\Phi: [0, 1] \times \Omega \times \Omega \times \mathcal{S}^1 \rightarrow \Omega.$$

where  $\mathcal{S}^1 \subset \mathbb{R}^2$  denotes the unit sphere, that satisfies the following properties:

- (1)  $\Phi(0; W_L, W_R, \eta) = W_L$  and  $\Phi(1; W_L, W_R, \eta) = W_R$ , for any  $W_L, W_R \in \Omega$ ,  $\eta \in \mathcal{S}^1$ .
- (2)  $\Phi(s; W_L, W_R, \eta) = \Phi(1-s; W_R, W_L, -\eta)$ , for any  $W_L, W_R \in \Omega$ ,  $s \in [0, 1]$ ,  $\eta \in \mathcal{S}^1$ .
- (3) Given an arbitrary bounded set  $\mathcal{B} \subset \Omega$ , there exists a constant  $k$  such that

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

for any  $W_L, W_R \in \mathcal{B}$ ,  $s \in [0, 1]$ ,  $\eta \in \mathcal{S}^1$ .

- (4) For every bounded set  $\mathcal{B} \subset \Omega$ , there exists a constant  $K$  such that

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

for each  $W_L^1, W_R^1, W_L^2, W_R^2 \in \mathcal{B}$ ,  $s \in [0, 1]$ ,  $\eta \in \mathcal{S}^1$ .

**Remark 1** The dependency of the family of paths on  $\eta$  can be dropped for rotational invariant systems. In fact, in [4] the families of path introduced to define the nonconservative products in the multidimensional case do not depend on  $\eta$ .

Suppose that a family of paths  $\Phi$  in  $\Omega$  has been chosen. Then, the nonconservative products in (3) can be interpreted as a Borel measure and a rigorous definition of weak solution can be given (see [4] for details). According to this definition, a piecewise regular function  $W$  is a weak solution if and only if the two following conditions are satisfied:

- (i)  $W$  is a classical solution where it is smooth.
- (ii) At every point of a discontinuity  $W$  satisfies the jump condition

$$\int_0^1 (\sigma \mathcal{I} - \mathcal{A}(\Phi(s; W^-, W^+, \eta), \eta)) \frac{\partial \Phi}{\partial s}(s; W^-, W^+, \eta) ds = 0, \quad (4)$$

where  $\mathcal{I}$  is the identity matrix;  $\sigma$ , the speed of propagation of the discontinuity;  $\eta$  a unit vector normal to the discontinuity at the considered point; and  $W^-, W^+$ , the lateral limits of the solution at the discontinuity.

Together with the definition of weak solutions, a notion of entropy has to be chosen, as the usual Lax's concept or one related to an entropy pair.

The choice of the family of paths is important because it determines the speed of propagation of discontinuities. The simplest choice is given by the family of segments:

$$\Phi(s; W_L, W_R, \eta) = W_L + s(W_R - W_L), \quad (5)$$

that corresponds to the definition of nonconservative products proposed by Volpert (see [22]). In practical applications, it has to be based on the physical background of the problem. In [12] a clear motivation for the selection of the family of paths is provided when a physical regularization by diffusion, dispersion, etc is available. Nevertheless, it is natural from the mathematical point of view to require this family to satisfy some hypotheses concerning the relation of the paths with the integral curves of the characteristic fields. Following [15], in this article we shall assume that the following hypotheses are fulfilled:

- (H1) Given  $\eta \in S^1$  and two states  $W_L$  and  $W_R$  belonging to the same integral curve  $\gamma$  of a linearly degenerate field of  $\mathcal{A}(W, \eta)$ , the path  $\Phi(\cdot; W_L, W_R, \eta)$  is a parameterization of the arc of  $\gamma$  linking  $W_L$  and  $W_R$ .
- (H2) Given  $\eta \in S^1$  and two states  $W_L$  and  $W_R$  belonging to the same integral curve  $\gamma$  of a genuinely nonlinear field and such that  $\lambda_i(W_L, \eta) < \lambda_i(W_R, \eta)$ , the path  $\Phi(\cdot; W_L, W_R, \eta)$  is a parameterization of the arc of  $\gamma$  linking  $W_L$  and  $W_R$ .
- (H3) Given  $\eta \in S^1$ , let us denote by  $\mathcal{RP}_\eta \subset \Omega \times \Omega$  the set of pairs  $(W_L, W_R)$  such that the Riemann problem

$$\begin{cases} U_t + \mathcal{A}(U, \eta)U_\xi = 0, \\ U(\xi, 0) = \begin{cases} W_L & \text{if } \xi < 0, \\ W_R & \text{if } \xi > 0, \end{cases} \end{cases} \quad (6)$$

has a unique self-similar weak solution composed by at most  $N$  simple waves (i.e. entropy shocks, contact discontinuities or rarefaction waves) connecting  $J + 1$  intermediate constant states

$$U_0 = W_L, U_1, \dots, U_{J-1}, U_J = W_R,$$

with  $J \leq N$ . Then, given  $(W_L, W_R) \in \mathcal{RP}_\eta$ , the curve described by the path  $\Phi(\cdot; W_L, W_R, \eta)$  in  $\Omega$  is equal to the union of those corresponding to the paths  $\Phi(\cdot; U_{j-1}, U_j, \eta)$ ,  $j = 1, \dots, J$ .

The reason to set these hypotheses is that they allow us to prove the three following natural properties (see [15]):

**Proposition 2** *Let us assume that the concept of weak solutions of (3) is defined on the basis of a family of paths satisfying hypotheses (H1)-(H3). Then:*  
*(i) Given two states  $W_L$  and  $W_R$  belonging to the same integral curve of a*

linearly degenerate field of  $\mathcal{A}(W, \eta)$ , the contact discontinuity given by

$$W(\mathbf{x}, t) = \begin{cases} W_L & \text{if } x_1\eta_1 + x_2\eta_2 < \sigma t, \\ W_R & \text{if } x_1\eta_1 + x_2\eta_2 > \sigma t, \end{cases}$$

where  $\sigma$  is the (constant) value of the corresponding eigenvalue through the integral curve, is an entropy weak solution of (3).

(ii) Let  $(W_L, W_R)$  be a pair belonging to  $\mathcal{RP}_\eta$  and let  $U$  be the solution of the Riemann problem (6). The following equality holds for every  $t > 0$ :

$$\langle \mathcal{A}(U(\cdot, t), \eta) U_\xi(\cdot, t), 1 \rangle = \int_0^1 \mathcal{A}(\Phi(s; W_L, W_R, \eta)) \frac{\partial \Phi}{\partial s}(s; W_L, W_R, \eta) ds.$$

(iii) Let  $(W_L, W_R)$  be a pair belonging to  $\mathcal{RP}_\eta$  and  $U_j$  any of the intermediate states involved by the solution of the Riemann problem (6). Then:

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

### 3 Roe methods for two-dimensional nonconservative systems

In order to construct a first order numerical scheme for (3), we first extend the concept of Roe linearization introduced in [20] for 1d problems, which is also based on the use of a family of paths:

**Definition 3** Given a family of paths  $\Psi$ , a function  $\mathcal{A}_\Psi: \Omega \times \Omega \times \mathcal{S}^1 \rightarrow \mathcal{M}_{N \times N}(\mathbb{R})$  is called a Roe linearization of (3), if it verifies the following properties:

(1) For each  $W_L, W_R \in \Omega$  and  $\eta \in \mathcal{S}^1$ ,  $\mathcal{A}_\Psi(W_L, W_R, \eta)$  has  $N$  distinct real eigenvalues:

$$\lambda_1(W_L, W_R, \eta) < \lambda_2(W_L, W_R, \eta) < \cdots < \lambda_N(W_L, W_R, \eta).$$

(2)  $\mathcal{A}_\Psi(W, W, \eta) = \mathcal{A}(W, \eta)$ , for every  $W \in \Omega$ ,  $\eta \in \mathcal{S}^1$ .

(3) For any  $W_L, W_R \in \Omega$ ,  $\eta \in \mathcal{S}^1$ :

$$\mathcal{A}_\Psi(W_L, W_R, \eta)(W_R - W_L) = \int_0^1 \mathcal{A}(\Psi(s; W_L, W_R, \eta), \eta) \frac{\partial \Psi}{\partial s}(s; W_L, W_R, \eta) ds. \quad (7)$$

Notice that if  $\mathcal{A}_k(W)$ ,  $k = 1, 2$  are the Jacobian matrices of two smooth flux functions  $F_k(W)$ ,  $k = 1, 2$ , (7) is independent of the family of paths and it reduces to the usual Roe property:

$$\mathcal{A}_\Psi(W_L, W_R, \eta) \cdot (W_R - W_L) = F_\eta(W_R) - F_\eta(W_L), \quad (8)$$

for any  $\eta = (\eta_1, \eta_2) \in \mathcal{S}^1$ , where

$$F_\eta(W) = \eta_1 F_1(W) + \eta_2 F_2(W), \quad (9)$$

represents the flux along the  $\eta$  direction.

Once a Roe linearization  $\mathcal{A}_\Psi$  has been chosen, in order to discretize the system, the domain  $D$  is decomposed into subsets with an easy geometry, called cells or finite volumes,  $V_i \subset \mathbb{R}^2$ . We assume here that the cells are closed polygons whose intersections are either empty, a complete edge or a vertex. We will denote by  $\mathcal{T}$  the mesh, i.e. the set of cells, and by  $NV$  the number of cells.

The following notation is considered: given a finite volume  $V_i$ ,  $N_i \in \mathbb{R}^2$  represents its center;  $\mathcal{N}_i$  is the set of indexes  $j$  such that  $V_j$  is a neighbor of  $V_i$ ;  $E_{ij}$  is the common edge to two neighbor cells  $V_i$  and  $V_j$ , and  $|E_{ij}|$  represents its length;  $\eta_{ij} = (\eta_{ij,1}, \eta_{ij,2})$  is the normal unit vector of the edge  $E_{ij}$  pointing towards the cell  $V_j$  (see Figure 1). Each cell can be decomposed in triangular subcells  $\{V_{ij}\}_{j \in \mathcal{N}_i}$ :  $V_{ij}$  is the triangle defined by the center of the cell  $N_i$  and the edge  $E_{ij}$ .  $|V_i|$  and  $|V_{ij}|$  represent, respectively, the areas of  $V_i$  and  $V_{ij}$ .  $\Delta x$  is the maximum of the diameters of the cells. Finally,  $W_i^n$  will represent the constant approximation of the averaged solution in the cell  $V_i$  at time  $t^n$  provided by the numerical scheme:

$$W_i^n \cong \frac{1}{|V_i|} \int_{V_i} W(\mathbf{x}, t^n) d\mathbf{x}.$$

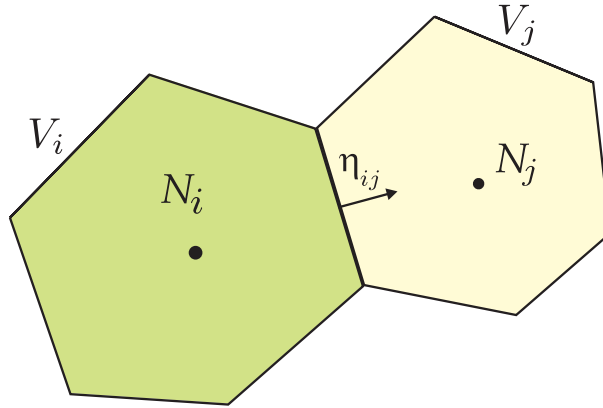


Fig. 1. Finite Volumes.

Let us suppose that the approximations in the cells at time  $n$ ,  $\{W_i^n\}$ , are already known. To progress in time, we consider at each edge  $E_{ij}$  the following Riemann problem:

$$\begin{cases} W_t + \mathcal{A}_1(W)W_{x_1} + \mathcal{A}_2(W)W_{x_2} = 0, \\ W(x_1, x_2, t^n) = \begin{cases} W_i^n & \text{if } \eta_{ij,1}x_1 + \eta_{ij,2}x_2 < k_{ij}, \\ W_j^n & \text{if } \eta_{ij,1}x_1 + \eta_{ij,2}x_2 > k_{ij}, \end{cases} \end{cases} \quad (10)$$

being  $k_{ij} \in \mathbb{R}$  such that  $E_{ij}$  is contained in the straight line of equation:

$$\eta_{ij,1}x_1 + \eta_{ij,2}x_2 = k_{ij}. \quad (11)$$

It can be easily verified that the solution of this Riemann problem is given by:

$$W(x_1, x_2, t) = U(\eta_{ij,1}x_1 + \eta_{ij,2}x_2, t) \quad (12)$$

where  $U$  is the solution of the 1d Riemann problem:

$$\begin{cases} U_t + \mathcal{A}(U, \eta_{ij})U_\xi = 0, \\ U(\xi, t^n) = \begin{cases} W_i^n & \text{if } \xi < k_{ij}, \\ W_j^n & \text{if } \xi > k_{ij}, \end{cases} \end{cases} \quad (13)$$

We consider now its linear approximation:

$$\begin{cases} \bar{U}_t + \mathcal{A}_{ij}\bar{U}_\xi = 0, \\ \bar{U}(\xi, t^n) = \begin{cases} W_i^n & \text{if } \xi < k_{ij}, \\ W_j^n & \text{if } \xi > k_{ij}, \end{cases} \end{cases} \quad (14)$$

where  $\mathcal{A}_{ij}$  is the Roe matrix associated to the states  $W_i^n, W_j^n$ , and the direction  $\eta_{ij}$ ,

$$\mathcal{A}_{ij} = \mathcal{A}_\Psi(W_i^n, W_j^n, \eta_{ij}).$$

Now, the average of the solution at time  $t^{n+1}$  at the subcell  $V_{ij}$  is approximated by the average of  $\bar{U}$  at time  $t^{n+1}$  at the interval  $[k_{ij} - d_{ij}/2, k_{ij}]$ , being  $d_{ij} = d(N_i, E_{ij})$ , i.e.

$$\frac{1}{|V_{ij}|} \int_{V_{ij}} W(x, t^{n+1}) dx \cong U_{ij}^{n+1} = \frac{2}{d_{ij}} \int_{k_{ij}-d_{ij}/2}^{k_{ij}} \bar{U}(\xi, t^{n+1}) d\xi.$$

Finally,  $W_i^{n+1}$  is computed by averaging the approximations of the subcells:

$$W_i^{n+1} = \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |V_{ij}| U_{ij}^{n+1}. \quad (15)$$

Assuming the *CFL* condition

$$\max\{|\lambda_{ij,k}| : k = 1, \dots, N\} \cdot \Delta t \leq \frac{d_{ij}}{2}, \quad (16)$$

where  $\{\lambda_{ij,k}\}$  represent the eigenvalues of  $\mathcal{A}_{ij}$ , some straightforward calculations allow us to write the numerical scheme as follows:



$$W_i^{n+1} = W_i^n - \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \mathcal{A}_{ij}^-(W_j^n - W_i^n), \quad (17)$$

where

$$\mathcal{A}_{ij}^- = \mathcal{K}_{ij} \mathcal{D}_{ij}^- \mathcal{K}_{ij}^{-1},$$

being  $\mathcal{D}_{ij}^-$  the diagonal matrix whose coefficients are the negative parts of the eigenvalues of  $\mathcal{A}_{ij}$  and  $\mathcal{K}_{ij}$  is a  $N \times N$  matrix whose columns are associated eigenvectors. (17) is the general expression of a Roe scheme for problem (3).

The best choice of the family of paths  $\Psi$  appearing in the definition of the Roe linearization seems to be the family  $\Phi$  selected for the definition of weak solutions. In effect, Roe methods based on the family of paths  $\Phi$  can handle correctly with discontinuities in the following sense: let us suppose that the approximations at two neighbor cells,  $W_i^n$  and  $W_j^n$ , can be linked by an entropy discontinuity located at the straight line containing the edge  $E_{ij}$  and propagating at speed  $\sigma$ ; then, from (7) and (4) we deduce:

$$\mathcal{A}_{ij}(W_j^n - W_i^n) = \sigma(W_j^n - W_i^n),$$

i.e.  $\sigma$  is an eigenvalue of the intermediate matrix and  $W_j^n - W_i^n$  is an associated eigenvector. As a consequence, the solution of the linear Riemann problem (14) corresponding to the intercell  $E_{ij}$  coincides with the solution of the Riemann problem (10). Nevertheless, the construction of a Roe scheme with  $\Psi = \Phi$  can be difficult or very costly in practice. In this case, a simpler family of paths  $\Psi$  has to be chosen, as the family of segments. In [16] it was remarked that, in this case, the convergence of the numerical scheme can fail when the weak solution to approach involves discontinuities connecting states  $W^-$  and  $W^+$  such that the paths of the families  $\Phi$  and  $\Psi$  linking them are different.

**Remark 2** *Observe that in the deduction of the schemes a CFL-like requirement (16) has been imposed. In practice, the following condition:*

$$\max \left\{ \frac{|\lambda_{ij,k}|}{d_{ij}} : i = 1, \dots, NV, j \in \mathcal{N}_i, k = 1, \dots, N \right\} \cdot \Delta t \leq \gamma, \quad (18)$$

with  $0 < \gamma \leq 1$ , ensures the linear stability.

**Remark 3** *As in the case of systems of conservation laws, when sonic rarefaction waves appear it is necessary to modify the approximate Riemann problem solver in order to obtain entropy-satisfying solutions. The Harten-Hyman Entropy Fix technique (see [11]), for instance, can be easily adapted to this case.*

### 3.1 Consistency

The following result of consistency for smooth solutions can be proved:

**Theorem 4** *Let us suppose that  $\mathcal{A}_1(W)$ ,  $\mathcal{A}_2(W)$  are  $\mathcal{C}^1$  matrices with bounded derivatives. Let us also suppose that the chosen Roe linearizations  $\mathcal{A}_\Psi(\cdot, \cdot; \eta_{ij})$*

at each edge  $E_{ij}$  as well as the functions  $|\mathcal{A}(\cdot, \eta_{ij})|$  are  $\mathcal{C}^1$  with bounded derivatives. Let us assume that the finite volume mesh consists of regular polygons with an even number of edges and with the same diameter  $\Delta x$ . Then, the scheme (17) is consistent for smooth solutions.

**Remark 4** The regularity assumption for  $|\mathcal{A}_\Psi(\cdot, \cdot; \eta_{ij})|$  is equivalent, in practice, to assume that the eigenvalues of the intermediate matrices do not vanish, i.e. that the solutions to be approached do not have transitions. However, in practice the entropy fix applied to the scheme prevents the eigenvalues from vanishing and thus the hypothesis are satisfied even when transitions occur.

*Proof:*

First, the following notation is introduced:  $N_i = (N_{i,1}, N_{i,2})$  and  $N_{ij} = (N_{ij,1}, N_{ij,2})$  represent, respectively, the center of the cell  $V_i$  and the mid-point of the edge  $E_{ij}$ , so that

$$d_{ij} = d(N_i, N_{ij}).$$

Some easy applications of the divergence theorem allow us to prove the equalities:

$$\sum_{j \in \mathcal{N}_i} |E_{ij}| \eta_{ij} = 0; \quad (19)$$

$$\sum_{j \in \mathcal{N}_i} d_{ij} |E_{ij}| \eta_{ij,1}^2 = |V_i|; \quad (20)$$

$$\sum_{j \in \mathcal{N}_i} d_{ij} |E_{ij}| \eta_{ij,2}^2 = |V_i|; \quad (21)$$

$$\sum_{j \in \mathcal{N}_i} d_{ij} |E_{ij}| \eta_{ij,1} \eta_{ij,2} = 0. \quad (22)$$

More precisely, the divergence theorem has to be applied to the fields:  $(1, 0)$ ,  $(0, 1)$ ,  $(x_1 - N_{i,1}, 0)$ ,  $(0, x_2 - N_{i,2})$ , and  $(x_2 - N_{i,2}, 0)$  in  $V_i$ .

Let us consider  $W$  a regular solution of (3) and  $W_i^n = W(N_i, t^n)$ . We want to prove that:

$$\begin{aligned} \frac{W_i^{n+1} - W_i^n}{\Delta t} + \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \mathcal{A}_{ij}^-(W_j^n - W_i^n) \\ = (W_t + \mathcal{A}_1(W)W_{x_1} + \mathcal{A}_2(W)W_{x_2})(N_i, t^n) \\ + O(\Delta x, \Delta t). \end{aligned} \quad (23)$$

Clearly the first term is equal to  $W_t(N_i, t^n) + O(\Delta t)$ .

Let us analyze the second term. First, we use the equality

$$\mathcal{A}_{ij}^- = \frac{1}{2}(\mathcal{A}_{ij} - |\mathcal{A}_{ij}|), \quad (24)$$

where

$$|\mathcal{A}_{ij}| = \mathcal{K}_{ij} |\mathcal{D}_{ij}| \mathcal{K}_{ij}^{-1},$$

being  $|\mathcal{D}_{ij}|$  the diagonal matrix whose coefficients are the absolute value of the eigenvalues of  $\mathcal{A}_{ij}$ . We obtain:

$$\begin{aligned} \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \mathcal{A}_{ij}^-(W_j^n - W_i^n) &= \frac{1}{2|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \mathcal{A}_{ij}(W_j^n - W_i^n) \\ &\quad - \frac{1}{2|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| |\mathcal{A}_{ij}| (W_j^n - W_i^n). \end{aligned} \tag{25}$$

For the first summand in the right-hand side of (25) we have the equalities:

$$\begin{aligned} &\frac{1}{2|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \mathcal{A}_{ij}(W_j^n - W_i^n) \\ &= \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \mathcal{A}(W_i^n, \eta_{ij}) d_{ij} (\eta_{ij,1} W_{x_1}(N_i, t^n) + \eta_{ij,2} W_{x_2}(N_i, t^n)) + O(\Delta x) \\ &= \mathcal{A}_1(W_i^n) W_{x_1}(N_i, t^n) + \mathcal{A}_2(W_i^n) W_{x_2}(N_i, t^n) + O(\Delta x), \end{aligned} \tag{26}$$

where we have used the equalities

$$\mathcal{A}_\Psi(W_i^n, W_i^n, \eta_{ij}) = \mathcal{A}(W_i^n, \eta_{ij}) = \sum_{k=1}^2 \eta_{ij,k} \mathcal{A}_k(W_i^n),$$

the relations (19)-(22), and

$$\frac{1}{2|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| d_{ij} = 1, \quad \left| \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| d_{ij}^2 \right| \leq \Delta x.$$

Let us see finally that the second summand in the right-hand side of (25) is  $O(\Delta x)$ . To do this, we first consider a partition  $\{J_i, K_i\}$  of the set of indexes  $\mathcal{N}_i$  such that:

- $\text{card}(J_i) = \text{card}(K_i)$ ;
- given an index  $j \in J_i$ , there exists a unique index  $j^* \in K_i$  such that

$$\eta_{ij^*} = -\eta_{ij}.$$

Using this partition we have:

$$\begin{aligned}
& \frac{1}{2|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| |\mathcal{A}_{ij}| (W_j^n - W_i^n) \\
&= \frac{1}{2|V_i|} \sum_{j \in J_i} |E_{ij}| |\mathcal{A}_{ij}| (W_j^n - W_i^n) \\
&+ \frac{1}{2|V_i|} \sum_{j \in J_i} |E_{ij^*}| |\mathcal{A}_{ij^*}| (W_{j^*}^n - W_i^n) \\
&= \frac{1}{2|V_i|} \sum_{j \in J_i} |E_{ij}| \left( |\mathcal{A}(W(N_{ij}, t^n), \eta_{ij})| (W_j^n - W_i^n) \right. \\
&+ \left. |\mathcal{A}(W(N_{ij^*}, t^n), \eta_{ij})| (W_{j^*}^n - W_i^n) \right) + O(\Delta x) \\
&= \frac{1}{2|V_i|} \sum_{j \in J_i} |E_{ij}| \left( |\mathcal{A}(W(N_{ij}, t^n), \eta_{ij})| W_{\eta_{ij}}(N_{ij}, t^n) d_{ij} \right. \\
&- \left. |\mathcal{A}(W(N_{ij^*}, t^n), \eta_{ij})| W_{\eta_{ij}}(N_{ij^*}, t^n) d_{ij} \right) + O(\Delta x) \\
&= \frac{1}{2|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| d_{ij}^2 \frac{\partial}{\partial \eta_{ij}} \left( |\mathcal{A}(W, \eta_{ij})| W_{\eta_{ij}} \right) (N_i, t^n) + O(\Delta x) \\
&= O(\Delta x),
\end{aligned}$$

where  $W_\eta = \eta_{ij,1} W_{x_1} + \eta_{ij,2} W_{x_2}$ . □

**Remark 5** *The hypothesis on the finite volume mesh can be relaxed. The previous theorem still holds under the following hypothesis:*

- *the cells of  $\mathcal{T}$  have an even number of edges;*
- *for every cell  $V_i$ :*

$$\overrightarrow{N_i N_{ij}} = d_{ij} \eta_{ij} + O(\Delta x), \quad \forall j \in \mathcal{N}_i,$$

*being  $\Delta x = \max\{\text{diam}(V_i)\}$ ;*

- *the edges of every cell  $V_i$  can be taken in pairs  $(E_{ij}, E_{ij^*})$  verifying:*

$$\eta_{ij^*} = -\eta_{ij} + O(\Delta x),$$

$$|E_{ij^*}| = |E_{ij}| + O(\Delta x);$$

- *given two neighbor cells  $V_i, V_j$ :*

$$d_{ij} = d_{ji} + O(\Delta x).$$

- *$|\mathcal{A}(\cdot, \cdot)|$  is  $\mathcal{C}^1$  with bounded derivatives.*

For meshes consisting of polygons with an odd number of edges, the consistency may fail. Let us consider, for instance, the 1d linear transport equation interpreted as a 2d problem:

$$u_t + au_{x_1} = 0, \quad \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2, t \geq 0, \quad (27)$$

where  $a$  is, say, positive. This equation is a particular case of (3) with  $N = 1$ ,  $W = u$ ,  $\mathcal{A}_1(W) = a$ ;  $\mathcal{A}_1(W) = 0$ .

We fix the time step  $\Delta t$  and the space step  $\Delta x$  and consider points:

$$\mathbf{x}_j^k = (j\Delta x, k\Delta x), \quad j, k \in \mathbb{Z}.$$

We consider the mesh composed by the triangles whose vertices are the points of coordinates:

$$\mathbf{x}_j^k, \mathbf{x}_{j+1}^k, \mathbf{x}_j^{k+1}, \quad (28)$$

and those whose vertices are:

$$\mathbf{x}_j^k, \mathbf{x}_{j-1}^{k+1}, \mathbf{x}_j^{k+1}. \quad (29)$$

Let us consider the strip composed by the triangles of both families corresponding to a fixed value of  $k$ . Let  $V_i$  the triangle whose vertices are (28) with  $j = (i - 1)/2$  if  $i$  is odd, and the triangle whose vertices are (29) with  $j = i/2$  if  $i$  is even. Some straightforward calculations allow us to rewrite the scheme as follows:

$$u_i^{n+1} = u_i^n - 2 \frac{\Delta t}{\Delta x} a(u_i^n - u_{i-1}^n).$$

Clearly, the local truncation error is not first order. Nevertheless, the errors at the cells may compensate and the numerical scheme converge: notice that, even in the 1d case, the classical requirement of consistency fails for conservative schemes with consistent numerical fluxes when the grid is not uniform.

### 3.2 Well-balancing

The well-balance properties of Roe schemes for 1d nonconservative systems have been studied in [16]. In general, these properties are not inherited by their two dimensional extensions. Nevertheless some partial results can still be given.

**Definition 5** *Given an edge  $E_{ij}$ , we will denote by  $\Gamma_{ij}$  the set of integral curves of linearly degenerated fields of  $\mathcal{A}(W, \eta_{ij})$  such that the corresponding eigenvalue vanishes along the curve.*

**Theorem 6** *Let  $W$  be a regular stationary solution of (3) satisfying the following property: given two neighbor cells  $V_i, V_j$  the path*

$$s \in [0, 1] \rightarrow \Psi(s, W(N_i), W(N_j), \eta_{ij})$$

*is a parametrization of an arc  $\gamma$  of  $\Gamma_{ij}$ . Then the numerical scheme computes exactly the solution  $W$ .*

*Proof:*

Defining  $W_i^0 = W(N_i)$ , for all  $i$ . We obtain the following equalities:

$$\begin{aligned} \mathcal{A}_{ij}(W_j^0 - W_i^0) &= \mathcal{A}_\Psi(W_i^0, W_j^0, \eta_{ij})(W_j^0 - W_i^0) = \\ &= \int_0^1 \mathcal{A}(\Psi(s, W_i^0, W_j^0, \eta_{ij}), \eta_{ij}) \cdot \frac{\partial \Psi}{\partial s}(s, W_i^0, W_j^0, \eta_{ij}) ds = 0 \end{aligned}$$

where we have used the definition of  $\mathcal{A}_{ij}$ , the property (7) of the linearization and the fact that  $\Psi_{ij}(s, W_i^0, W_j^0, \eta_{ij})$  is a parametrization of a curve of  $\Gamma_{ij}$ . Then, 0 is an eigenvalue of  $\mathcal{A}_{ij}$  and  $W_j^0 - W_i^0$  an associated eigenvector. By definition of  $\mathcal{A}_{ij}^-$  we have:

$$\mathcal{A}_{ij}^-(W_j^0 - W_i^0) = 0, \text{ for all } j \in \mathcal{N}_i, \quad (30)$$

and thus:

$$W_i^n = W_i^0, \quad \forall i, n.$$

□

**Corollary 7** *If  $\Psi$  is the family of segments (5), then the numerical scheme solves exactly any regular stationary solution  $W$  such that, for any pair of neighbor cells,  $V_i, V_j$  the segment linking  $W(N_i)$  and  $W(N_j)$  belongs to a curve  $\gamma$  of  $\Gamma_{ij}$ .*

**Theorem 8** *Let us assume that  $\mathcal{A}_1, \mathcal{A}_2$  are of class  $\mathcal{C}^1$  with bounded derivatives and that the finite volume mesh  $\mathcal{T}$  satisfies the regularity conditions of Theorem 4. Let  $W$  be a regular stationary solution of (3) with the following property: given two neighbor cells  $V_i, V_j$  the states  $W(N_i), W(N_j)$  belong to a same curve  $\gamma$  of  $\Gamma_{ij}$ , and it is possible to find a parametrization*

$$U : [0, 1] \longrightarrow \Omega,$$

*of class  $\mathcal{C}^{k+1}$  of the arch of  $\gamma$  linking them, such that:*

$$\int_0^1 |U'(s) - \frac{\partial \Psi}{\partial s}(s, W(N_i), W(N_j), \eta_{ij})| ds = O(d(N_i, N_j)^{k+1}). \quad (31)$$

*Then, the scheme approximates the solutions  $W$  with order  $k$ .*

*Proof:*

We define again:

$$W_i^0 = W(N_i), \text{ for all } i.$$

Following the proof of Theorem 2 in [16], we obtain from (31) the estimate:

$$\mathcal{A}_{ij}^-(W_j^0 - W_i^0) = O(d(N_i, N_j)^{k+1}), \text{ for all } j \in \mathcal{N}_i.$$

As a result:

$$\begin{aligned}
W_i^1 &= W_i^0 - \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \mathcal{A}_{ij}^-(W_j^0 - W_i^0) \\
&= W_i^0 - \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} \frac{2|V_{ij}|}{d_{ij}} \mathcal{A}_{ij}^-(W_j^0 - W_i^0) \\
&= W_i^0 + O(\Delta x^k).
\end{aligned}$$

□

**Theorem 9** *Under the same hypothesis of Theorem 8, let us suppose that  $\Psi$  is the family of segments (5). Let  $W$  be a  $C^3$  stationary solution of (3) such that, given two neighbor cells  $V_i, V_j$ , the function*

$$s \in [0, 1] \rightarrow W(N_i + s(N_j - N_i))$$

*is a parametrization of an arc of some curve of  $\Gamma_{ij}$ . Then, the scheme approximates the solution  $W$  with order 2.*

*Proof:*

Defining:

$$W_i^0 = W(N_i), \text{ for all } i,$$

and proceeding as in the proof of Theorem 3 in [16], the following estimate can be obtained:

$$\mathcal{A}_{ij}^-(W_j^0 - W_i^0) = O(d(N_i, N_j)^3), \text{ for all } j \in \mathcal{N}_i.$$

From this estimate, the proof is concluded as in Theorem 8.

□

**Remark 6** *In the proofs of the two previous theorems it is possible to weaken the hypothesis for the mesh: it is enough to assume that there exists  $C^* > 0$  independent of  $\Delta x$  such that*

$$\frac{d(N_i, N_j)}{2d_{ij}} \leq C^*.$$

Notice that the hypotheses concerning the stationary solutions in the previous theorems are much more restrictive than in the 1d case: while in 1d problems the values taken by a stationary solution at the centers of two neighbor cells always belong to the same integral curve of a linearly degenerate field whose corresponding eigenvalue vanishes along the curve (see [16]), this is not the case for general two dimensional stationary solutions. Nevertheless, this happens when the stationary solution is essentially 1d and the mesh is rectangular and properly oriented. More precisely, let  $\eta \in \mathbb{R}^2$  be a fixed unitary vector and  $U(\xi)$  a regular stationary solution of the one dimensional problem:

$$U_t + \mathcal{A}(U, \eta)U_\xi = 0. \quad (32)$$

Let  $W$  be the stationary solution of problem (3) given by:

$$W(x_1, x_2) = U(x_1\eta_1 + x_2\eta_2).$$

If we have a rectangular mesh such that for every edge  $E_{ij}$ , the vector  $\eta_{ij}$  is either parallel or orthogonal to  $\eta$ , then it can be easily verified that the restriction of the solution to the segment linking the centers of two neighbor cells is a solution of the projected problem:

$$U_t + \mathcal{A}(U, \eta_{ij})U_\xi = 0,$$

and, as a result, its values at those centers are connected by a curve of  $\Gamma_{ij}$ .

For this kind of solutions and meshes, the 2d schemes have the same well balanced properties that the one dimensional schemes from which they have been derived.

Nevertheless, we will see that Corollary 7 is enough to ensure the  $\mathcal{C}$ -Property of the schemes applied to shallow water systems of one or two layers, i.e. the stationary solutions corresponding to water at rest situations are preserved.

#### 4 High order schemes based on reconstruction of states

In this section, we present a high order extension by state reconstructions of the previously proposed scheme.

Let us consider first the case of systems of conservation laws

$$W_t + F_1(W)_{x_1} + F_2(W)_{x_2} = 0. \quad (33)$$

These systems can be considered as a particular case of (3) in which the matrices  $\mathcal{A}_i(W)$ ,  $i = 1, 2$  are the Jacobians:

$$\mathcal{A}_i(W) = \frac{\partial F_i}{\partial W}, \quad i = 1, 2.$$

High order methods based on the reconstruction of states can be built for (33) using the following procedure: given a first order conservative scheme with numerical flux function  $G(U, V; \eta)$ , a reconstruction operator of order  $p$  is considered, that is, an operator that associates to a given family  $\{W_i\}_{i=1}^{NV}$  of values at the cells two families of functions defined at the edges:

$$\gamma \in E_{ij} \rightarrow W_{ij}^\pm(\gamma),$$

in such a way that, whenever

$$W_i = \frac{1}{|V_i|} \int_{V_i} W(\mathbf{x}) d\mathbf{x} \quad (34)$$

for some smooth function  $W$ , then

$$W_{ij}^\pm(\gamma) = W(\gamma) + O(\Delta x^p), \quad \forall \gamma \in E_{ij}.$$



Once the first order method and the reconstruction operator have been chosen, the method of lines can be used to develop high order methods for (33): the idea is to discretize only in space, leaving the problem continuous in time. This procedure leads to a system of ordinary differential equations which is solved using a standard numerical method.

Let  $\overline{W}_i(t)$  denotes the cell average of a regular solution  $W$  of (33) over the cell  $V_i$  at time  $t$ :

$$\overline{W}_i(t) = \frac{1}{|V_i|} \int_{V_i} W(x, t) dx.$$

The following equation can be easily obtained for the cell averages:

$$\overline{W}'_i(t) = -\frac{1}{|V_i|} \left( \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} F(W(\gamma, t)) \cdot \eta_{ij} d\gamma \right). \quad (35)$$

The first order method and the reconstructions are now used to approach the values of the fluxes at the edges:

$$W'_i(t) = -\frac{1}{|V_i|} \left( \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} G(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}) d\gamma \right), \quad (36)$$

being  $W_i(t)$  the approximation to  $\overline{W}_i(t)$  provided by the scheme and  $W_{ij}^\pm(\gamma, t)$  the reconstruction at  $\gamma \in E_{ij}$  corresponding to the family  $\{W_i(t)\}_{i=1}^{NV}$ . It can be shown that (36) is an approximation of order  $p$  of (35).

In practice, the integral terms in (36) are approached by means of a numerical quadrature of order  $\bar{r} \geq p$  at least:

$$\int_a^b f(s) ds = (b-a) \left( \sum_{l=1}^{n(\bar{r})} \omega_l f(x_l) \right) + O(\Delta x^{\bar{r}}), \quad (37)$$

where  $n(\bar{r})$  denotes the number of points,  $\omega_l$  are the weights, and  $x_l = a + s_l(b-a)$  with  $s_l \in [0, 1]$ , represent the quadrature points. The expression of the corresponding semi-discrete numerical scheme is then as follows:

$$W'_i(t) = -\frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \left( \sum_{l=1}^{n(\bar{r})} \omega_l G(W_{ij,l}^-(t), W_{ij,l}^+(t), \eta_{ij}) \right). \quad (38)$$

where

$$W_{ij,l}^\pm(t) = W_{ij}^\pm(a_{ij} + s_l(b_{ij} - a_{ij}), t), \quad (39)$$

$a_{ij}$  and  $b_{ij}$  being the extremes of the edge  $E_{ij}$ .

We assume here that the first order scheme is a Roe method, i.e.:

$$G(U, V, \eta) = \frac{F_\eta(U) + F_\eta(V)}{2} - \frac{1}{2} |\mathcal{A}(U, V, \eta)| (V - U), \quad (40)$$

where  $F_\eta$  is given by (9) and  $\mathcal{A}(U, V, \eta)$  is the Roe matrix associated to the states  $U, V$ , and to the unit vector  $\eta$ , i.e. an intermediate matrix that verifies

the Roe property:

$$F_\eta(V) - F_\eta(U) = \mathcal{A}(U, V, \eta) \cdot (V - U). \quad (41)$$

Let us now generalize the semi-discrete methods (36) or (38) to the nonconservative system (3). We will assume that the reconstructions are calculated as follows: given the family  $\{W_i\}_{i=1}^{NV}$  of values at the cells, first an approximation function is constructed at every cell  $V_i$ , based on the values of  $W_i$  at some of the cells close to  $V_i$  (the *stencil*):

$$P_i(\mathbf{x}) = P_i(\mathbf{x}; \{W_j\}_{j \in \mathcal{B}_i}),$$

for some set of indexes  $\mathcal{B}_i$ . If, for instance, the reconstruction only depends on the neighbor cells of  $V_i$ , then  $\mathcal{B}_i = \mathcal{N}_i \cup \{i\}$ . These approximations functions are calculated usually by means of an interpolation or approximation procedure. Once these functions have been constructed, the reconstruction at  $\gamma \in E_{ij}$  are defined as follows:

$$W_{ij}^-(\gamma) = \lim_{\mathbf{x} \rightarrow \gamma} P_i(\mathbf{x}), \quad W_{ij}^+(\gamma) = \lim_{\mathbf{x} \rightarrow \gamma} P_j(\mathbf{x}). \quad (42)$$

Clearly, for any  $\gamma \in E_{ij}$  the following equalities are satisfied:

$$W_{ij}^-(\gamma) = W_{ji}^+(\gamma); \quad W_{ij}^+(\gamma) = W_{ji}^-(\gamma).$$

We suppose that the reconstruction operator satisfies the following properties:

(HP1) It is conservative, i.e. the following equality holds for any cell  $V_i$ :

$$W_i = \frac{1}{|V_i|} \int_{V_i} P_i(\mathbf{x}) d\mathbf{x}. \quad (43)$$

(HP2) It is of order  $p$ .

(HP3) It is of order  $q$  in the interior of the cells, i.e. if the operator is applied to a sequence  $\{W_i\}$  satisfying (34) for some smooth function  $W(\mathbf{x})$ , then:

$$P_i(\mathbf{x}) = W(\mathbf{x}) + O(\Delta x^q), \quad \forall \mathbf{x} \in \text{int}(V_i). \quad (44)$$

(HP4) Under the assumption of the previous property, the gradient of  $P_i$  provides an approximation of order  $m$  of the gradient of  $W$ :

$$\nabla P_i(\mathbf{x}) = \nabla W(\mathbf{x}) + O(\Delta x^m), \quad \forall \mathbf{x} \in \text{int}(V_i). \quad (45)$$

Let us denote by  $P_i^t$  the approximation functions corresponding to the approximations of the cell averages  $W_i(t)$ , i.e.

$$P_i^t(\mathbf{x}) = P_i(\mathbf{x}; \{W_j(t)\}_{j \in \mathcal{B}_i}).$$

$W_{ij}^-(\gamma, t)$  (resp.  $W_{ij}^+(\gamma, t)$ ) is then defined by

$$W_{ij}^-(\gamma, t) = \lim_{\mathbf{x} \rightarrow \gamma} P_i^t(\mathbf{x}), \quad W_{ij}^+(\gamma, t) = \lim_{\mathbf{x} \rightarrow \gamma} P_j^t(\mathbf{x}). \quad (46)$$

Using (24) and (41), (36) can be rewritten as follows:

$$\begin{aligned}
W_i'(t) &= -\frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( \mathcal{A}_{ij}^-(\gamma, t) (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) \right) d\gamma \\
&\quad - \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( F_{\eta_{ij}}(W_{ij}^-(\gamma, t)) \right) d\gamma \\
&= -\frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( \mathcal{A}_{ij}^-(\gamma, t) (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) \right) d\gamma \\
&\quad - \frac{1}{|V_i|} \int_{V_i} \nabla \cdot (F \circ P_i^t)(\mathbf{x}) d\mathbf{x} \\
&= -\frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( \mathcal{A}_{ij}^-(\gamma, t) (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) \right) d\gamma \\
&\quad - \frac{1}{|V_i|} \int_{V_i} \left( \mathcal{A}_1(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_1}(\mathbf{x}) + \mathcal{A}_2(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_2}(\mathbf{x}) \right) dx,
\end{aligned} \tag{47}$$

where the following notation has been used

$$\mathcal{A}_{ij}(\gamma, t) = \mathcal{A} \left( W_{ij}^+(\gamma, t), W_{ij}^-(\gamma, t), \eta_{ij} \right),$$

together with the definition of the reconstructions and the divergence theorem.

Notice now that (36) can be easily generalized to obtain a numerical scheme for solving (3):

$$\begin{aligned}
W_i'(t) &= -\frac{1}{|V_i|} \left[ \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( \mathcal{A}_{ij}^-(\gamma, t) (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) \right) d\gamma \right. \\
&\quad \left. + \int_{V_i} \left( \mathcal{A}_1(P_i^t(x)) \frac{\partial P_i^t}{\partial x_1}(x) + \mathcal{A}_2(P_i^t(x)) \frac{\partial P_i^t}{\partial x_2}(x) \right) dx \right],
\end{aligned} \tag{48}$$

where now

$$\mathcal{A}_{ij}(\gamma, t) = \mathcal{A}_\Psi \left( W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij} \right),$$

being  $\mathcal{A}_\Psi$  the chosen Roe linearization.

#### 4.1 Order of accuracy

The cell averages of a smooth solution of (3),  $\overline{W}_i(t)$ , satisfy:

$$\overline{W}_i'(t) = -\frac{1}{|V_i|} \int_{V_i} \left( \mathcal{A}_1(W(\mathbf{x})) W_{x_1}(\mathbf{x}) + \mathcal{A}_2(W(\mathbf{x})) W_{x_2}(\mathbf{x}) \right) dx. \tag{49}$$

Thus, (48) is expected to be an accurate approximation of (49). This fact is stated in the following result, whose proof is similar to the corresponding result for 1d problems (see [3]):

**Theorem 10** *Let us assume that  $\mathcal{A}_1$  and  $\mathcal{A}_2$  are of class  $\mathcal{C}^2$  with bounded derivatives and  $\mathcal{A}_\Psi$  is bounded for all  $i, j$ . Let us also suppose that the reconstruction operator satisfies the hypothesis (HP1)-(HP4). Then (48) is an approximation of order at least  $\alpha = \min(p, q + 1, m + 1)$  to the system (49) in the following sense:*

$$\begin{aligned} & \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( \mathcal{A}_{ij}^-(\gamma, t) (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) \right) d\gamma \\ & + \int_{V_i} \left( \mathcal{A}_1(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_1}(\mathbf{x}) + \mathcal{A}_2(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_2}(\mathbf{x}) \right) d\mathbf{x} \\ & = \int_{V_i} (\mathcal{A}_1(W(\mathbf{x}, t)) W_{x_1}(\mathbf{x}, t) + \mathcal{A}_2(W(\mathbf{x}, t)) W_{x_2}(\mathbf{x}, t)) d\mathbf{x} + O(\Delta x^{\alpha+1}), \end{aligned} \quad (50)$$

for every solution  $W$  smooth enough, being  $W_{ij}^\pm(\gamma, t)$  the associated reconstructions and  $P_i^t$  the approximation functions corresponding to the family

$$\bar{W}_i(t) = \frac{1}{|V_i|} \int_{V_i} W(\mathbf{x}, t) d\mathbf{x}.$$

#### 4.2 Approximation of the integral terms

In practice, the integral terms in (48) are numerically approached. In this case, together with a 1d formula (37) for the integrals on the edges, it can be also necessary to choose a quadrature formula of order  $\bar{s}$  for the integrals in the cells:

$$\int_{V_i} f(\mathbf{x}) d\mathbf{x} = |V_i| \sum_{l=1}^{n(\bar{s})} \alpha_l^i f(\mathbf{x}_l^i) + \mathcal{O}(|V_i|^{\bar{s}}). \quad (51)$$

In order to preserve the order of the numerical scheme, it is necessary to have  $\bar{r} \geq \alpha$  and  $\bar{s} \geq \alpha$ . The numerical scheme writes then as follows:

$$\begin{aligned} W_i'(t) = & -\frac{1}{|V_i|} \left[ \sum_{j \in \mathcal{N}_i} |E_{ij}| \sum_{l=1}^{n(\bar{r})} w_l \mathcal{A}_{ij,l}^-(t) (W_{ij,l}^+(t) - W_{ij,l}^-(t)) \right. \\ & \left. + |V_i| \sum_{l=1}^{n(\bar{s})} \alpha_l^i \left( \mathcal{A}_1(P_i^t(\mathbf{x}_l^i)) \frac{\partial P_i^t}{\partial x_1}(\mathbf{x}_l^i) + \mathcal{A}_2(P_i^t(\mathbf{x}_l^i)) \frac{\partial P_i^t}{\partial x_2}(\mathbf{x}_l^i) \right) \right], \end{aligned} \quad (52)$$

where

$$\begin{aligned} W_{ij,l}^\pm(t) &= W_{ij}^\pm(a_{ij} + s_l(b_{ij} - a_{ij}), t), \\ \mathcal{A}_{ij,l}(t) &= \mathcal{A}_\Psi(W_{ij,l}^-(t), W_{ij,l}^+(t), \eta_{ij}). \end{aligned}$$

#### 4.3 Well-balance properties.

In this section we study the well-balanced properties of the schemes (48) or (52).

**Definition 11** We consider a semi-discrete method to approximate (3):

$$\begin{cases} W'_i(t) = \frac{1}{|V_i|} \mathcal{H}(\mathbf{W}_j(t), j \in \mathcal{B}_i), \\ \mathbf{W}(0) = \mathbf{W}_0, \end{cases} \quad (53)$$

where  $\mathbf{W}(t) = \{W_i(t)\}_{i=1}^{NV}$  represents the vector of the approximations to the averaged values of the exact solution;  $\mathbf{W}_0 = \{W_i^0\}$  is the vector of the initial conditions; and  $\mathcal{B}_i$  are the stencils. Given a smooth stationary solution  $W$  of the system, the numerical scheme is said to be exactly well-balanced for  $W$  if the vector or its cell averages is a critical point of (53), i.e.

$$\mathcal{H}(\mathbf{W}_j, j \in \mathcal{B}_i) = 0, \quad (54)$$

and it is said to be well-balanced with order  $p$  if

$$\mathcal{H}(\mathbf{W}_j, j \in \mathcal{B}_i) = \mathcal{O}(\Delta x^{p+1}). \quad (55)$$

Let us also introduce the concept of well-balance reconstruction operator:

**Definition 12** Given a smooth stationary solution of (3), a reconstruction operator is said to be well-balanced for  $W(\mathbf{x})$  if the approximation functions  $P_i(\mathbf{x})$  associated to the averaged values of  $W$  are also stationary solutions of the system, i.e.

$$\mathcal{A}_1(P_i(\mathbf{x}))P_i(\mathbf{x})_{x_1} + \mathcal{A}_2(P_i(\mathbf{x}))P_i(\mathbf{x})_{x_2} = 0, \quad \forall \mathbf{x} \in V_i, \quad i = 1, \dots, NV.$$

The two following results can be easily proved:

**Theorem 13** Let  $W$  be a smooth stationary solution of (3). Let us suppose that both the first order Roe method and the reconstruction operator chosen are exactly well-balanced for  $W$ . Then the numerical schemes (48) and (52) are also exactly well-balanced for  $W$ .

**Theorem 14** Under the hypotheses of Theorem 10, the schemes (48) and (52) are well-balanced with order at least  $\alpha = \min(p, q + 1, r + 1)$ .

## 5 Systems of conservation laws with non conservative products and source terms

We consider in this section systems of the form:

$$W_t + F_1(W)_{x_1} + F_2(W)_{x_2} = B_1(W)W_{x_1} + B_2(W)W_{x_2} + S_1(W)H_{x_1} + S_2(W)H_{x_2}, \quad (56)$$

where  $W(\mathbf{x}, t): D \times (0, T) \mapsto \Omega \subset \mathbb{R}^N$ ,  $D$  being a bounded domain of  $\mathbb{R}^2$ ;  $\Omega$ , a convex subset of  $\mathbb{R}^N$ ;  $F_i: \Omega \mapsto \mathbb{R}^N$ ,  $B_i: \Omega \mapsto \mathcal{M}_{N \times N}(\mathbb{R})$  and  $S_i: \Omega \mapsto \mathbb{R}^N$ ,  $i = 1, 2$ , are regular and locally bounded functions. Finally  $H: D \subset \mathbb{R}^2 \mapsto \mathbb{R}$  is a known function.

If the equation

$$H_t = 0,$$

is added to the system and  $H$  is considered like a new unknown of the problem (whose value is determined by the initial condition), (56) can be rewritten in the form:

$$\widetilde{W}_t + \widetilde{\mathcal{A}}_1(\widetilde{W})\widetilde{W}_{x_1} + \widetilde{\mathcal{A}}_2(\widetilde{W})\widetilde{W}_{x_2} = 0, \quad (57)$$

where  $\widetilde{W}$  is the augmented vector:

$$\widetilde{W} = \begin{bmatrix} W \\ H \end{bmatrix},$$

and the block structure of the matrices  $\widetilde{\mathcal{A}}_k(\widetilde{W}) \in \mathcal{M}_{(N+1) \times (N+1)}(\mathbb{R})$ ,  $k = 1, 2$ , are given by

$$\widetilde{\mathcal{A}}_k(\widetilde{W}) = \left[ \begin{array}{c|c} \mathcal{A}_k(W) & -S_k(W) \\ \hline 0 & 0 \end{array} \right], \quad k = 1, 2. \quad (58)$$

Here,

$$\mathcal{A}_k(W) = \mathcal{J}_k(W) - B_k(W), \quad k = 1, 2,$$

being  $\mathcal{J}_k$ ,  $k = 1, 2$  the Jacobian matrix of the flux functions  $F_k$ ,  $k = 1, 2$ :

$$\mathcal{J}_k(W) = \frac{\partial F_k}{\partial W}, \quad k = 1, 2.$$

The following notation will be used:

$$\begin{aligned} \mathcal{A}(W, \eta) &= \eta_1 \mathcal{A}_1(W) + \eta_2 \mathcal{A}_2(W), \\ \mathcal{J}(W, \eta) &= \eta_1 \mathcal{J}_1(W) + \eta_2 \mathcal{J}_2(W), \\ B(W, \eta) &= \eta_1 B_1(W) + \eta_2 B_2(W), \\ S(W, \eta) &= \eta_1 S_1(W) + \eta_2 S_2(W). \end{aligned}$$

We assume that (56) is hyperbolic: for any  $\eta$  and  $W \in \Omega$ , the matrix  $\mathcal{A}(W, \eta)$  has  $N$  real distinct eigenvalues

$$\lambda_1(W, \eta) < \dots < \lambda_N(W, \eta),$$

and associated eigenvectors  $R_j(W, \eta)$ ,  $j = 1, \dots, N$ . If these eigenvalues do not vanish, (57) is a strictly hyperbolic system: given a unit vector  $\eta$  and a state  $\widetilde{W}$ , the eigenvalues of the matrix

$$\widetilde{\mathcal{A}}(\widetilde{W}, \eta) = \widetilde{\mathcal{A}}_1(\widetilde{W}_1)\eta_1 + \widetilde{\mathcal{A}}_2(\widetilde{W}_2)\eta_2,$$

are:

$$\lambda_1(W, \eta), \dots, \lambda_N(W, \eta), 0$$

with associated eigenvectors:

$$\widetilde{R}_1(\widetilde{W}, \eta), \dots, \widetilde{R}_{N+1}(\widetilde{W}, \eta),$$

given by

$$\tilde{R}_i(\tilde{W}, \eta) = \begin{bmatrix} R_i(W, \eta) \\ 0 \end{bmatrix}, \quad i = 1, \dots, N; \quad \tilde{R}_{N+1}(\tilde{W}, \eta) = \begin{bmatrix} \mathcal{A}(W, \eta)^{-1} \cdot S(W, \eta) \\ 1 \end{bmatrix}.$$

In order to construct Roe matrices for (57), first of all a family of paths  $\tilde{\Psi}$  has to be chosen. The following notation will be used to describe, given a vector  $\eta$ , the path linking two states  $\tilde{W}_0, \tilde{W}_1$ :

$$\tilde{\Psi}(s; \tilde{W}_0, \tilde{W}_1, \eta) = \begin{bmatrix} \Psi(s; \tilde{W}_0, \tilde{W}_1, \eta) \\ \Psi_{N+1}(s; \tilde{W}_0, \tilde{W}_1, \eta) \end{bmatrix} = \begin{bmatrix} \Psi_1(s; \tilde{W}_0, \tilde{W}_1, \eta) \\ \Psi_2(s; \tilde{W}_0, \tilde{W}_1, \eta) \\ \vdots \\ \Psi_{N+1}(s; \tilde{W}_0, \tilde{W}_1, \eta) \end{bmatrix}.$$

Let us suppose that, given any unit vector  $\eta$  and two states  $\tilde{W}_j = [W_j, H_j]^T$ ,  $j = 0, 1$  it is possible to calculate:

- A matrix  $\mathcal{J}(W_0, W_1, \eta)$  such that:

$$\mathcal{J}(W_0, W_1, \eta)(W_1 - W_0) = F_\eta(W_1) - F_\eta(W_0),$$

i.e. a Roe matrix for the flux function  $F_\eta$ .

- A matrix  $B_{\tilde{\Psi}}(W_0, W_1, \eta)$  satisfying:

$$B_{\tilde{\Psi}}(W_0, W_1, \eta)(W_1 - W_0) = \int_0^1 B(\Psi(s; \tilde{W}_0, \tilde{W}_1, \eta), \eta) \frac{\partial \Psi}{\partial s}(s; \tilde{W}_0, \tilde{W}_1, \eta) ds; \quad (59)$$

- A vector  $S_{\tilde{\Psi}}(W_0, W_1, \eta)$  satisfying:

$$S_{\tilde{\Psi}}(W_0, W_1, \eta)(H_1 - H_0) = \int_0^1 S(\Psi(s; \tilde{W}_0, \tilde{W}_1, \eta), \eta) \cdot \frac{\partial \Psi_{N+1}}{\partial s}(s; \tilde{W}_0, \tilde{W}_1, \eta) ds, \quad (60)$$

Then, it can be easily verified that the matrix:

$$\tilde{\mathcal{A}}_{\tilde{\Psi}}(\tilde{W}_0, \tilde{W}_1, \eta) = \left[ \begin{array}{c|c} \mathcal{A}_{\tilde{\Psi}}(W_0, W_1, \eta) & -S_{\tilde{\Psi}}(W_0, W_1, \eta) \\ \hline 0 & 0 \end{array} \right], \quad (61)$$

with

$$\mathcal{A}_{\tilde{\Psi}}(W_0, W_1, \eta) = \mathcal{J}(W_0, W_1, \eta) - B_{\tilde{\Psi}}(W_0, W_1, \eta),$$

is a Roe linearization provided that it has  $N + 1$  real different eigenvalues.

Let us suppose that the approximations at time  $t^n$

$$\tilde{W}_i^n = \begin{bmatrix} W_i^n \\ H_i \end{bmatrix},$$

have already been obtained. The following notation will be used:

$$\begin{aligned}\mathcal{J}_{ij} &= \mathcal{J}(W_i^n, W_j^n, \eta_{ij}), \\ B_{ij} &= B_{\tilde{\Psi}}(W_i^n, W_j^n, \eta_{ij}), \\ S_{ij} &= S_{\tilde{\Psi}}(W_i^n, W_j^n, \eta_{ij}), \\ \mathcal{A}_{ij} &= \mathcal{J}_{ij} - B_{ij},\end{aligned}$$

$$\tilde{\mathcal{A}}_{ij} = \tilde{\mathcal{A}}_{\tilde{\Psi}}(\tilde{W}_i^n, \tilde{W}_j^n, \eta_{ij}) = \left[ \begin{array}{c|c} \mathcal{A}_{ij} & -S_{ij} \\ \hline 0 & 0 \end{array} \right].$$

The corresponding Roe scheme reads then as follows:

$$\tilde{W}_i^{n+1} = \tilde{W}_i^n - \frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \tilde{\mathcal{A}}_{ij}^- (\tilde{W}_j^n - \tilde{W}_i^n). \quad (62)$$

Dropping the  $(N+1)$ -th components (which are not relevant as  $H$  is a known function), some straightforward calculations allow us to rewrite the scheme as follows:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \left( G_{ij} - B_{ij} \cdot (W_j^n - W_i^n) - \mathcal{P}_{ij}^- S_{ij} (H_j - H_i) \right) \quad (63)$$

where

$$G_{ij} = \frac{1}{2} \left( F_{\eta_{ij}}(W_i^n) + F_{\eta_{ij}}(W_j^n) \right) - \frac{1}{2} |\mathcal{A}_{ij}| (W_j^n - W_i^n)$$

is the usual Roe flux and

$$\mathcal{P}_{ij}^- = \frac{1}{2} (\mathcal{I} - |\mathcal{A}_{ij}| \mathcal{A}_{ij}^{-1}).$$

This latter matrix can also be written in the form

$$\mathcal{P}_{ij}^- = \frac{1}{2} \mathcal{K}_{ij} (\mathcal{I} - \text{sgn}(\mathcal{D}_{ij})) \mathcal{K}_{ij}^{-1},$$

where  $\text{sgn}(\mathcal{D}_{ij})$  is the diagonal matrix whose coefficients are the signs of the eigenvalues  $\lambda_{ij,1}, \dots, \lambda_{ij,N}$  of  $\mathcal{A}_{ij}$  and  $\mathcal{K}_{ij}$  is an  $N \times N$  matrix whose columns are associated eigenvectors.

Using the equality  $\mathcal{A}_{ij}^- = \mathcal{P}_{ij}^- \mathcal{A}_{ij}$  the numerical scheme can also be written in this way:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| F_{ij}^-, \quad (64)$$

where

$$F_{ij}^- = \mathcal{P}_{ij}^- (\mathcal{A}_{ij} (W_j^n - W_i^n) - S_{ij} (H_j - H_i)). \quad (65)$$



In a similar way, the semi-discrete high order extension of the Roe scheme (63) based on a reconstruction operator (48) can be expressed as follows:

$$\begin{aligned}
W'_i = & -\frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( G_{ij}(\gamma, t) - B_{ij}(\gamma, t)(W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) \right. \\
& \left. - \mathcal{P}_{ij}^-(\gamma, t) S_{ij}(\gamma, t)(H_{ij}^+(\gamma) - H_{ij}^-(\gamma)) \right) d\gamma \\
& + \frac{1}{|V_i|} \int_{V_i} \left( \sum_{k=1}^2 B_k(P_i^t(\mathbf{x})) \frac{\partial P_i^t}{\partial x_k}(\mathbf{x}) + \sum_{k=1}^2 S_k(P_i^t(\mathbf{x})) \frac{\partial P_{N+1,i}^t}{\partial x_k}(\mathbf{x}) \right) d\mathbf{x},
\end{aligned} \tag{66}$$

where the following notation has been used:

$$\mathbf{x} \in V_i \rightarrow \tilde{P}_i^t(\mathbf{x}) = \begin{bmatrix} P_i^t(\mathbf{x}) \\ P_{N+1,i}^t(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} P_{1,i}^t(\mathbf{x}) \\ P_{2,i}^t(\mathbf{x}) \\ \vdots \\ P_{N,i}^t(\mathbf{x}) \\ P_{N+1,i}^t(\mathbf{x}) \end{bmatrix},$$

represents the approximation function at the cell  $V_i$  at time  $t$ ;

$$\gamma \in E_{ij} \rightarrow \tilde{W}_{ij}^\pm(\gamma, t) = \begin{bmatrix} W_{ij}^\pm(\gamma, t) \\ H_{ij}^\pm(\gamma) \end{bmatrix},$$

the reconstructions at the edge  $\Gamma_{ij}$  at time  $t$ ;

$$\begin{aligned}
\mathcal{J}_{ij}(\gamma, t) &= \mathcal{J}(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}), \\
B_{ij}(\gamma, t) &= B_{\tilde{\Psi}}(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}), \\
S_{ij}(\gamma, t) &= S_{\tilde{\Psi}}(W_{ij}^-(\gamma, t), W_{ij}^+(\gamma, t), \eta_{ij}), \\
\mathcal{A}_{ij}(\gamma, t) &= \mathcal{J}_{ij}(\gamma, t) - \mathcal{B}_{ij}(\gamma, t), \\
G_{ij}(\gamma, t) &= \frac{1}{2} \left( F_{\eta_{ij}}(W_{ij}^-(\gamma, t)) + F_{\eta_{ij}}(W_{ij}^+(\gamma, t)) \right) \\
&\quad - \frac{1}{2} |\mathcal{A}_{ij}(\gamma, t)| (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)), \\
\mathcal{P}_{ij}^-(\gamma, t) &= \frac{1}{2} \mathcal{K}_{ij}(\gamma, t) (\mathcal{I} - \text{sgn}(\mathcal{D}_{ij}(\gamma, t))) \mathcal{K}_{ij}^{-1}(\gamma, t),
\end{aligned}$$

where  $\text{sgn}(\mathcal{D}_{ij}(\gamma, t))$  is the diagonal matrix whose coefficients are the signs of the eigenvalues  $\lambda_{ij,1}(\gamma, t), \dots, \lambda_{ij,N}(\gamma, t)$  of  $\mathcal{A}_{ij}(\gamma, t)$  and  $\mathcal{K}_{ij}(\gamma, t)$  is a  $N \times N$  matrix whose columns are associated eigenvectors.

Finally, the numerical scheme (52) can be rewritten in this particular case

as follows:

$$\begin{aligned}
W'_i = & -\frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \sum_{l=1}^{n(\bar{r})} w_l \left( G_{ij,l}(t) - B_{ij,l}(t)(W_{ij,l}^+(t) - W_{ij,l}^-(t)) \right. \\
& \left. - \mathcal{P}_{ij,l}(t)^- S_{ij,l}(t)(H_{ij,l}^+(t) - H_{ij,l}^-(t)) \right) \\
& + \frac{1}{|V_i|} \sum_{l=1}^{n(\bar{s})} \alpha_l^i \left( \sum_{k=1}^2 B_k(P_i^t(\mathbf{x}_l)) \frac{\partial P_i^t}{\partial x_k}(\mathbf{x}_l) + \sum_{k=1}^2 S_k(P_i^t(\mathbf{x}_l)) \frac{\partial P_{N+1,i}^t}{\partial x_k}(\mathbf{x}_l) \right)
\end{aligned} \tag{67}$$

using the obvious notation.

## 6 Applications to the shallow water system

### 6.1 Equations

Let us consider the one layer shallow water system without friction terms:

$$\begin{cases} \frac{\partial h}{\partial t} + \frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} = 0, \\ \frac{\partial q_1}{\partial t} + \frac{\partial}{\partial x_1} \left( \frac{q_1^2}{h} + \frac{g}{2} h^2 \right) + \frac{\partial}{\partial x_2} \left( \frac{q_1 q_2}{h} \right) = gh \frac{\partial H}{\partial x_1}, \\ \frac{\partial q_2}{\partial t} + \frac{\partial}{\partial x_1} \left( \frac{q_1 q_2}{h} \right) + \frac{\partial}{\partial x_2} \left( \frac{q_2^2}{h} + \frac{g}{2} h^2 \right) = gh \frac{\partial H}{\partial x_2}. \end{cases} \tag{68}$$

which are the equations governing the flow of a shallow layer of homogeneous inviscid fluid in a two dimensional domain  $D \subset \mathbb{R}^2$ . In the equations,  $H(\mathbf{x})$  represents the depth function measured from a fixed level of reference;  $g$  is the gravity;  $q_j(\mathbf{x}, t)$  represents the mass-flow in the direction  $j$ ; and  $h(\mathbf{x}, t)$ , the thickness of the layer. These quantities are related to the vertical averaged velocity  $(u_1(\mathbf{x}, t), u_2(\mathbf{x}, t))$  by the relations:

$$q_j(\mathbf{x}, t) = u_j(\mathbf{x}, t)h(\mathbf{x}, t), \quad j = 1, 2.$$

Problem (68) can be written in the form (56) with:

$$W = \begin{bmatrix} h \\ q_1 \\ q_2 \end{bmatrix}, \tag{69}$$

$$F_1(W) = \begin{bmatrix} q_1 \\ \frac{q_1^2}{h} + \frac{1}{2}gh^2 \\ \frac{q_1q_2}{h} \end{bmatrix}, \quad F_2(W) = \begin{bmatrix} q_2 \\ \frac{q_1q_2}{h} \\ \frac{q_2^2}{h} + \frac{1}{2}gh^2 \end{bmatrix},$$

$$S_1(W) = \begin{bmatrix} 0 \\ gh \\ 0 \end{bmatrix}, \quad S_2(W) = \begin{bmatrix} 0 \\ 0 \\ gh \end{bmatrix},$$

and  $B_1(W) = B_2(W) = 0$ .

## 6.2 Numerical schemes

We consider here a Roe linearization based on the family of segments:

$$\tilde{\Psi}(s; \tilde{W}_L, \tilde{W}_R, \eta) = \tilde{W}_L + s(\tilde{W}_R - \tilde{W}_L).$$

Following the procedure described in Section 5 we obtain the following numerical scheme:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \left( \mathcal{P}_{ij}^-(\mathcal{A}_{ij}(W_j^n - W_i^n) - S_{ij}(H_j - H_i)) \right), \quad (70)$$

where:

$$W_i^n = \begin{bmatrix} h_i^n \\ q_{1,i}^n \\ q_{2,i}^n \end{bmatrix}, \quad u_{l,i}^n = \frac{q_{l,i}^n}{h_i^n}, \quad l = 1, 2,$$

$$\mathcal{A}_{ij} = \begin{bmatrix} 0 & \eta_{ij,1} & \eta_{ij,2} \\ (-u_{1,ij}^2 + c_{ij}^2)\eta_{ij,1} - u_{1,ij}u_{2,ij}\eta_{ij,2} & 2u_{1,ij}\eta_{ij,1} + u_{2,ij}\eta_{ij,2} & u_{1,ij}\eta_{ij,2} \\ -u_{1,ij}u_{2,ij}\eta_{ij,1} + (-u_{2,ij}^2 + c_{ij}^2)\eta_{ij,2} & u_{2,ij}\eta_{ij,1} & u_{1,ij}\eta_{ij,1} + 2u_{2,ij}\eta_{ij,2} \end{bmatrix}, \quad (71)$$

$$S_{ij} = \begin{bmatrix} 0 \\ gh_{ij}\eta_{ij,1} \\ gh_{ij}\eta_{ij,2} \end{bmatrix}. \quad (72)$$

The following notation has been used:

$$c_{ij} = \sqrt{gh_{ij}}, \quad u_{l,ij} = \frac{\sqrt{h_i}u_{l,i} + \sqrt{h_j}u_{l,j}}{\sqrt{h_i} + \sqrt{h_j}}, \quad l = 1, 2 \text{ y } h_{ij} = \frac{h_i + h_j}{2}. \quad (73)$$

As the family of paths doesn't verify the hypotheses (H1)-(H3) required to the family of paths related to the definition of weak solutions, the numerical scheme is not expected to correctly capture every discontinuity. As it is the case for the corresponding 1d scheme, while shocks related to the genuinely nonlinear fields evolving in regions where  $H$  is continuous are expected to be correctly approached, this is not the case for contact discontinuities related to bottom jumps, unless if the states linked for such a discontinuity correspond to water at rest solutions (see [16] for the discussion of these aspects in the 1d case). Nevertheless, this is not a limiting factor when the function  $H$  is continuous, as stationary contact discontinuities cannot appear.

Concerning the well-balance properties of the schemes, the following results can be stated:

**Proposition 15** *The scheme (70) solves exactly the stationary solutions corresponding to water at rest or vacuum.*

*Proof:*

The expression of the matrix  $\tilde{\mathcal{A}}(\tilde{W}, \eta)$  for (68) is:

$$\tilde{\mathcal{A}}(\tilde{W}, \eta) = \begin{bmatrix} 0 & \eta_1 & \eta_2 & 0 \\ (-u_1^2 + c^2)\eta_1 - u_1u_2\eta_2 & 2u_1\eta_1 + u_2\eta_2 & u_1\eta_2 & -gh\eta_1 \\ -u_1u_2\eta_1 + (-u_2^2 + c^2)\eta_2 & u_2\eta_1 & u_1\eta_1 + 2u_2\eta_2 & -gh\eta_2 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where  $u_l = q_l/h$ ,  $l = 1, 2$ , and  $c = \sqrt{gh}$ . Some straightforward calculations allow us to show that the straight lines of equations:

$$q_1 = 0, \quad q_2 = 0, \quad h - H = cst, \quad (74)$$

are integral curves of the linearly degenerate field of  $\tilde{\mathcal{A}}(\tilde{W}, \eta)$ .

Let us consider a stationary solution corresponding to water at rest:

$$\tilde{W}(x_1, x_2) = \begin{bmatrix} C + H(x_1, x_2) \\ 0 \\ 0 \\ H(x_1, x_2) \end{bmatrix}, \quad (75)$$

being  $C$  a positive constant. Given two neighbor cells  $V_i, V_j$  the segment linking

$\widetilde{W}(N_i)$  and  $\widetilde{W}(N_j)$  is clearly contained in a straight line of the family (74). Applying Corollary 7, we deduce that the scheme solves exactly this solution. The proof is similar for the stationary solutions corresponding to vacuum.  $\square$

Let us now turn to the semi-discrete high order extension of (70). Following Section 5 its general expression is as follows:

$$\begin{aligned}
W'_i = & -\frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( F_{\eta_{ij}}(W_{ij}^-(\gamma, t)) \right. \\
& \left. + \mathcal{P}_{ij}(\gamma, t)^- \left( \mathcal{A}_{ij}(\gamma, t)(W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) - S_{ij}(\gamma, t)(H_{ij}^+ - H_{ij}^-) \right) \right) d\gamma \\
& + \frac{1}{|V_i|} \int_{V_i} \begin{pmatrix} 0 \\ gP_{h,i}^t(\mathbf{x}) \frac{\partial P_{H,i}^t(\mathbf{x})}{\partial x_1} \\ gP_{h,i}^t(\mathbf{x}) \frac{\partial P_{H,i}^t(\mathbf{x})}{\partial x_2} \end{pmatrix} d\mathbf{x},
\end{aligned} \tag{76}$$

where the following notation has been used:

$$\widetilde{P}_i^t(\mathbf{x}) = \begin{pmatrix} P_{h,i}^t(\mathbf{x}) \\ P_{q_1,i}^t(\mathbf{x}) \\ P_{q_2,i}^t(\mathbf{x}) \\ P_{H,i}^t(\mathbf{x}) \end{pmatrix}$$

represents the approximation function at time  $t$  at the  $i$ -th cell;  $W_{ij}^\pm(\gamma, t)$ ,  $H_{ij}^\pm(\gamma, t)$  the reconstructed states at  $\gamma \in E_{ij}$  at time  $t$ ; and  $\mathcal{A}_{ij}(\gamma, t)$ ,  $\mathcal{P}_{ij}(\gamma, t)^-$ , the Roe and projection matrices associated to the reconstructed states  $W_{ij}^\pm(\gamma, t)$ .

Concerning the well-balance properties of this high order extension, the following result can be stated:

**Proposition 16** *Let us suppose that the approximation functions are exact for constant functions. Moreover, let us suppose that the reconstructions of the variables  $h$  and  $H$  satisfy the equality:*

$$P_{\eta,i}^t = P_{h,i}^t - P_{H,i}^t, \quad \forall i, \tag{77}$$

being  $\eta = h - H$ . Then, the reconstruction operator, and thus the semi-discrete scheme (76), is well-balanced for the stationary solutions corresponding to water at rest and vacuum.

*Proof:*

Let us consider again a stationary solution corresponding to water at rest (75). We have:

$$\eta(\mathbf{x}) = h(\mathbf{x}) - H(\mathbf{x}) = C, \forall \mathbf{x}.$$

As a consequence:

$$P_{\eta,i}(\mathbf{x}) = C, \quad \forall \mathbf{x} \in V_i, \forall i,$$

and thus:

$$P_{h,i}^0(\mathbf{x}) - P_{H,i}^0(\mathbf{x}) = C, \quad \forall \mathbf{x} \in V_i, \forall i.$$

Using this last equality, it is trivial to verify that:

$$\tilde{P}_i^0(\mathbf{x}) = \begin{pmatrix} P_{h,i}^0(\mathbf{x}) \\ 0 \\ 0 \\ P_{H,i}^0(\mathbf{x}) \end{pmatrix}$$

is also a stationary solution corresponding to water at rest. □

### 6.3 Numerical experiments

Some numerical tests are presented here in order to validate the performances of the Roe scheme and its high order extension. We have considered structured meshes. The high order extension is based on the third order bi-hyperbolic reconstruction introduced in [18] that generalizes the 1d reconstruction presented in [14] (see also [19]). This reconstruction operator satisfies hypotheses (HP1)-(HP3) with  $p = q = 3$ ,  $m = 2$ . The time-stepping used for the third order scheme is based on an optimal TVD Runge-Kutta method (see [8], [17]). In order to obtain the equality (77), the reconstruction procedure is applied to  $\eta$  and the variable  $H$ , and then (77) is used to define the reconstruction of the remaining variable. The integral terms have been approximated by means of a Gaussian quadrature of order three. In the sequel, this high order extension of Roe scheme will be referred to as BHRoe.

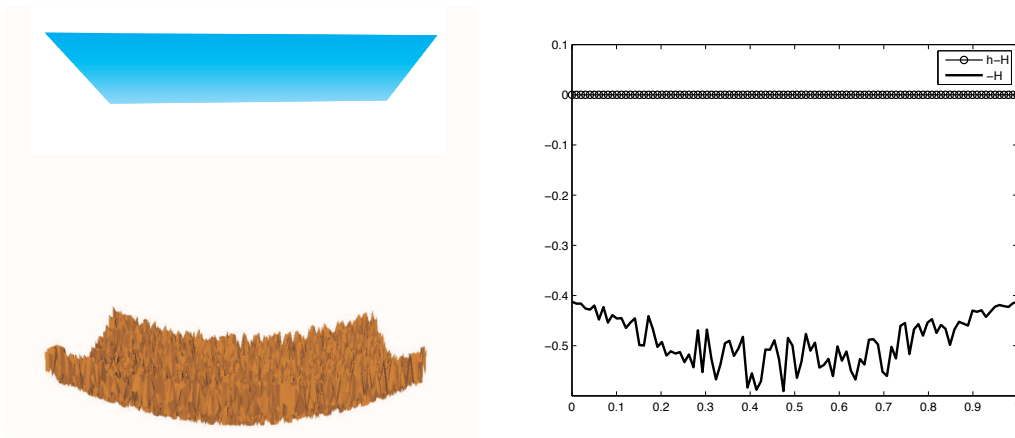


Fig. 2. Test case 6.3.1 (water at rest). Free surface and bottom topography ( $-H$ ).

### 6.3.1 Verification of the $\mathcal{C}$ -Property

This first test has been designed to verify numerically the  $\mathcal{C}$ -Property for both the Roe and BHRoe schemes.  $H(\mathbf{x})$  is given by a smooth function with a random perturbation (see Figure 2). As initial conditions we have taken  $h(\mathbf{x}, 0) = H(\mathbf{x})$  and  $q_1(\mathbf{x}, 0) = q_2(\mathbf{x}, 0) = 0$ ,  $\mathbf{x} \in [0, 1] \times [0, 1]$ . For both schemes, we consider a rectangular mesh with  $\Delta x = \Delta y = 0.01$ , and CFL=0.9. Periodic boundary conditions are also considered.

Table 1

Verification of the  $\mathcal{C}$ -Property.

Unknowns	$h$	$q_1$	$q_2$
$L^1$ error (Roe)	$6.55E^{-17}$	$4.04E^{-16}$	$4.16E^{-16}$
$L^1$ error (BHRoe)	$5.59E^{-17}$	$6.29E^{-16}$	$5.91E^{-16}$

As expected both numerical schemes preserve the steady state solution exactly up to machine accuracy. This fact can be observed in Table 1.

### 6.3.2 Well-balance properties for 1d stationary solutions

In this numerical experiment we test the well-balance property for smooth stationary solutions for both numerical schemes Roe and BHRoe. According to Theorem 9 and to the discussion at the end of Subsection 3.2, Roe scheme should approximate with order 2 any smooth essentially 1d stationary solutions if a properly oriented rectangular mesh is used. On the other hand, according to Theorem 14, BHRoe should approximate any smooth stationary solution with order 3.

We consider here the depth function given by:

$$H(\mathbf{x}) = 2 - 0.2 e^{-0.16(x_1-10)^2}, \quad \mathbf{x} \in [0, 20] \times [0, 20],$$

and as initial condition, the extension to 2d of the 1d subcritical stationary solution, is given (see Figure 3). The initial solution is computed as follows: first  $q_1$  and  $q_2$  are fixed

$$q_1(\mathbf{x}, 0) = 0.15, \quad q_2(\mathbf{x}, 0) = 0.$$

The value of  $h$  at the boundary  $x_1 = 20$  is also fixed:

$$h(20, x_2, 0) = 0.5.$$

Then, the value of  $h(\mathbf{x}, 0)$  at every point  $(x_1, x_2)$  is given by the greatest real root of the polynomial:

$$gh^3(\mathbf{x}, 0) - (gH(\mathbf{x}) + K)h^2(\mathbf{x}, 0) + \frac{1}{2}q_1^2(\mathbf{x}, 0) = 0,$$

Table 2

Test case 6.3.2:  $L^1$  errors and order. Roe scheme.

N.Cells	error $h$	order $h$	error $q_1$	order $q_1$
$10 \times 10$	$1.28E^{-1}$	-	$2.97E^{-5}$	-
$20 \times 20$	$3.20E^{-2}$	1.98	$1.74E^{-6}$	4.09
$40 \times 40$	$8.19E^{-3}$	1.98	$9.42E^{-8}$	4.21
$80 \times 80$	$1.96E^{-3}$	2.06	$4.72E^{-9}$	4.32
$160 \times 160$	$3.93E^{-4}$	2.32	$2.96E^{-10}$	4.00

where  $K$  is given by:

$$K = \frac{1}{2} \frac{q_1^2(20, 0, 0)}{h^2(20, 0, 0)} + g(h(20, 0, 0) - H(20, 0)).$$

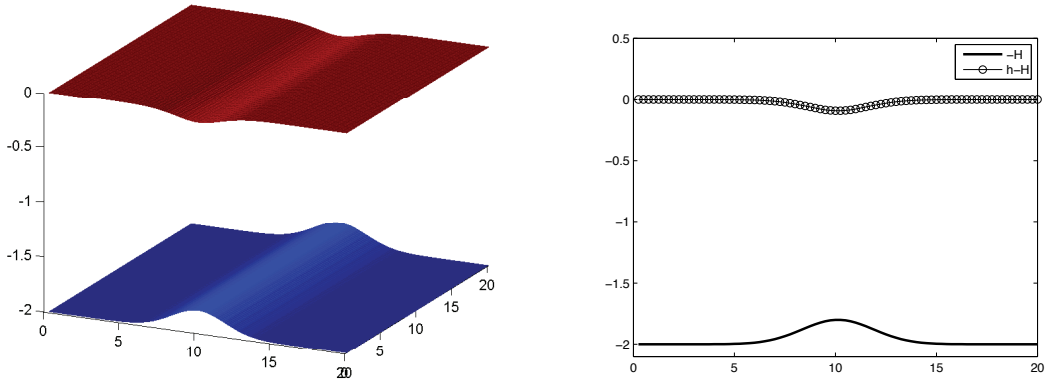


Fig. 3. Subcritical steady state solution: free surface elevation and bottom topography.

As boundary conditions,  $q_1(0, x_2, t) = 0.15$  is fixed at the boundary  $x_1 = 0$ , while,  $h(20, x_2, t) = 0.5$  is fixed at the boundary  $x_2 = 20$ . Wall boundary conditions are fixed at boundaries  $x_2 = 0$  and  $x_2 = 20$ . The CFL is set to 0.9.

Table 2 shows the  $L^1$  errors obtained with the Roe scheme, while Table 3 summaries the  $L^1$  errors and order obtained with the BHRoe scheme. As expected, Roe achieves second order while BHRoe achieves third order.

### 6.3.3 Accuracy test.

We consider a test proposed in [23] in order to measure the accuracy of the BHRoe scheme. Specifically, the bottom topography is defined as

$$H(\mathbf{x}) = 2 - \sin(2\pi x_1) - \cos(2\pi x_2),$$



Table 3

Test case 6.3.2:  $L^1$  errors and order. BHRoe scheme.

N.Cells	error $h$	order $h$	error $q_1$	order $q_1$
$10 \times 10$	$1.90E^{-5}$	-	$7.88E^{-4}$	-
$20 \times 20$	$1.89E^{-6}$	3.38	$1.49E^{-4}$	2.40
$40 \times 40$	$1.33E^{-6}$	0.51	$6.08E^{-5}$	1.29
$80 \times 80$	$2.57E^{-7}$	2.37	$1.01E^{-5}$	2.59
$160 \times 160$	$3.64E^{-8}$	2.82	$1.19E^{-6}$	3.08

Table 4

Accuracy test.  $L^1$  errors and orders.

N. cells	error $h$	order $h$	error $q_1$	order $q_1$	error $q_2$	order $q_2$
$25 \times 25$	$1.52E^{-02}$	-	$4.13E^{-02}$	-	$9.10E^{-02}$	-
$50 \times 50$	$3.47E^{-03}$	2.13	$8.33E^{-03}$	2.31	$2.19E^{-02}$	2.05
$100 \times 100$	$5.45E^{-04}$	2.67	$1.26E^{-03}$	2.72	$3.50E^{-03}$	2.65
$200 \times 200$	$7.56E^{-05}$	2.85	$1.77E^{-04}$	2.84	$4.79E^{-04}$	2.87
$400 \times 400$	$9.79E^{-06}$	2.95	$2.25E^{-05}$	2.97	$6.28E^{-05}$	2.93

the initial water height is

$$h(\mathbf{x}, 0) = 10 + e^{\sin(2\pi x_1)} \cos(2\pi x_2),$$

while the initial discharges are given by

$$q_1(\mathbf{x}, 0) = \sin(\cos(2\pi x_1))\sin(2\pi x_2), \quad q_2(\mathbf{x}, 0) = \cos(2\pi x_1) \cos(\sin(2\pi x_2)).$$

The computational domain is the unit square and periodic boundary conditions have been imposed.

Table 4 shows the results obtained at time  $t = 0.05$ , as shocks developed later for this problem. The solution computed with the BHRoe scheme on a mesh with  $800 \times 800$  grid points has been taken as reference solution. The CFL number has been taken as 0.5. As it can be seen, third order of accuracy is achieved.

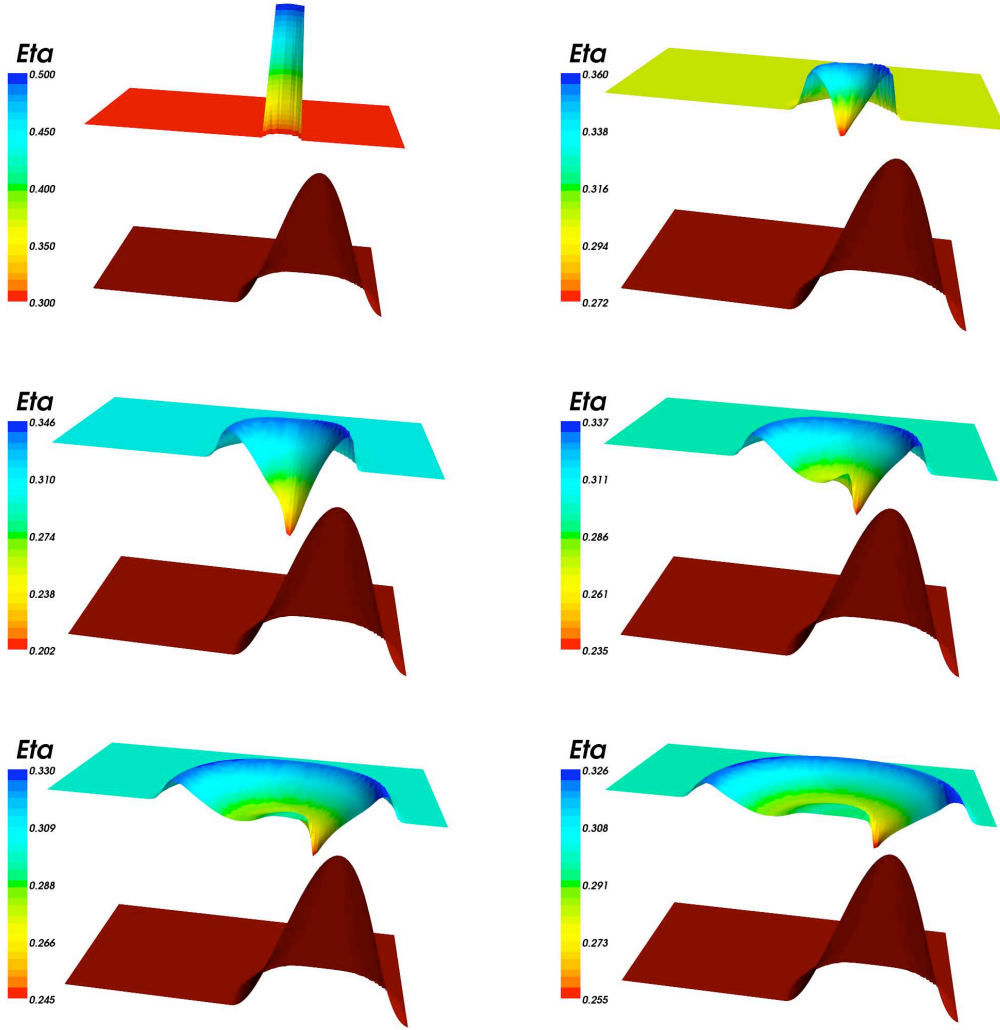


Fig. 4. Circular Dam-break on an closed channel (free surface, slice of the channel at  $y = 1$ ): BHRoe. From top to bottom and left to right, results at times  $t = 0.00, 0.05, 0.10, 0.15, 0.20, 0.25$ .

#### 6.3.4 Circular dam-break problem

We consider the following test problem: the domain is the square  $D = [0, 2] \times [0, 2]$ . The depth function is given by:

$$H(\mathbf{x}) = \begin{cases} 0.6 - b(\mathbf{x}), & \text{if } (x_1 - 1.5)^2 + (x_2 - 1)^2 \leq (0.1)^2 \\ 0.6 & \text{otherwise,} \end{cases}$$

with  $b(\mathbf{x}) = \frac{1}{8} (\cos(2\pi(x_1 - 0.5)) + 1) (\cos(2\pi x_2) + 1)$ .

The initial condition is:

$$h(\mathbf{x}, 0) = \begin{cases} H(\mathbf{x}) + 0.5 & \text{if } (x_1 - 1.25)^2 + (x_2 - 1)^2 \leq (0.1)^2, \\ H(\mathbf{x}) & \text{otherwise,} \end{cases}$$

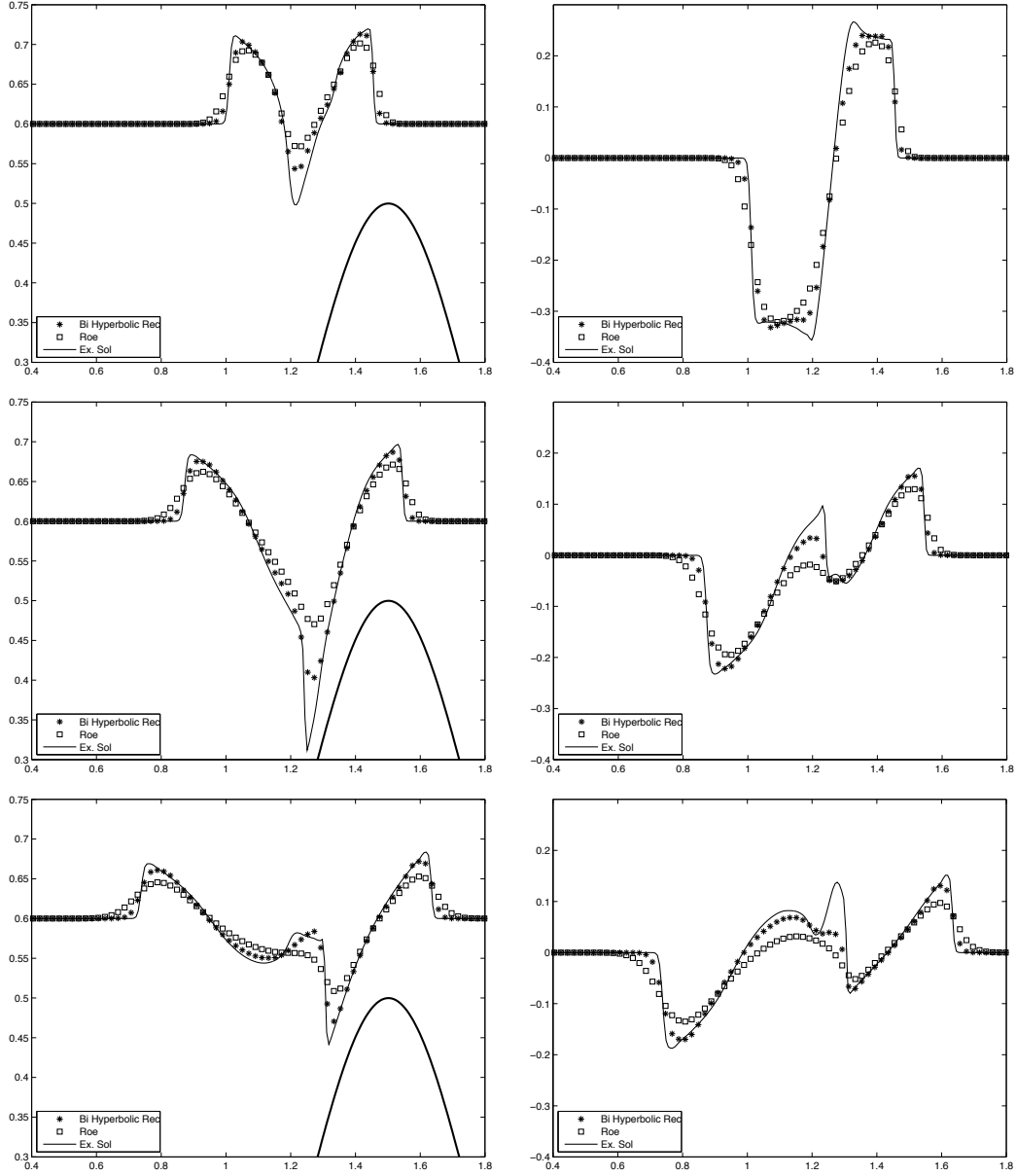


Fig. 5. Circular dam-break on an closed channel. Comparison between Roe and BHRoe at times  $t = 0.05, 0.1, 0.15$  at the longitudinal section  $x_2 = 10$ . Free surface (left column) and variable  $q_{1,1}$  (right column).

and  $q_1(\mathbf{x}, 0) = q_2(\mathbf{x}, 0) = 0$  (see Figure 4). We consider wall boundary conditions at  $x_2 = 0$  and  $x_2 = 2$  and free boundary conditions at  $x_1 = 0$  and  $x_1 = 2$ . The CFL is set to 0.9 and  $\Delta x = \Delta y = 0.02$ .

In Figure 4, the computed free surface and the bottom topography obtained with BHRoe at different times ( $t = 0, 0.05, 0.1, 0.15, 0.2, 0.25$  s) are shown.

Figure 5 shows a comparison between the solutions obtained with Roe, BHRoe and a reference solution computed over a finer mesh composed with  $800 \times 800$  cells, in the longitudinal section  $x_2 = 10$ . It can be noticed that BHRoe scheme is less diffusive than Roe scheme.

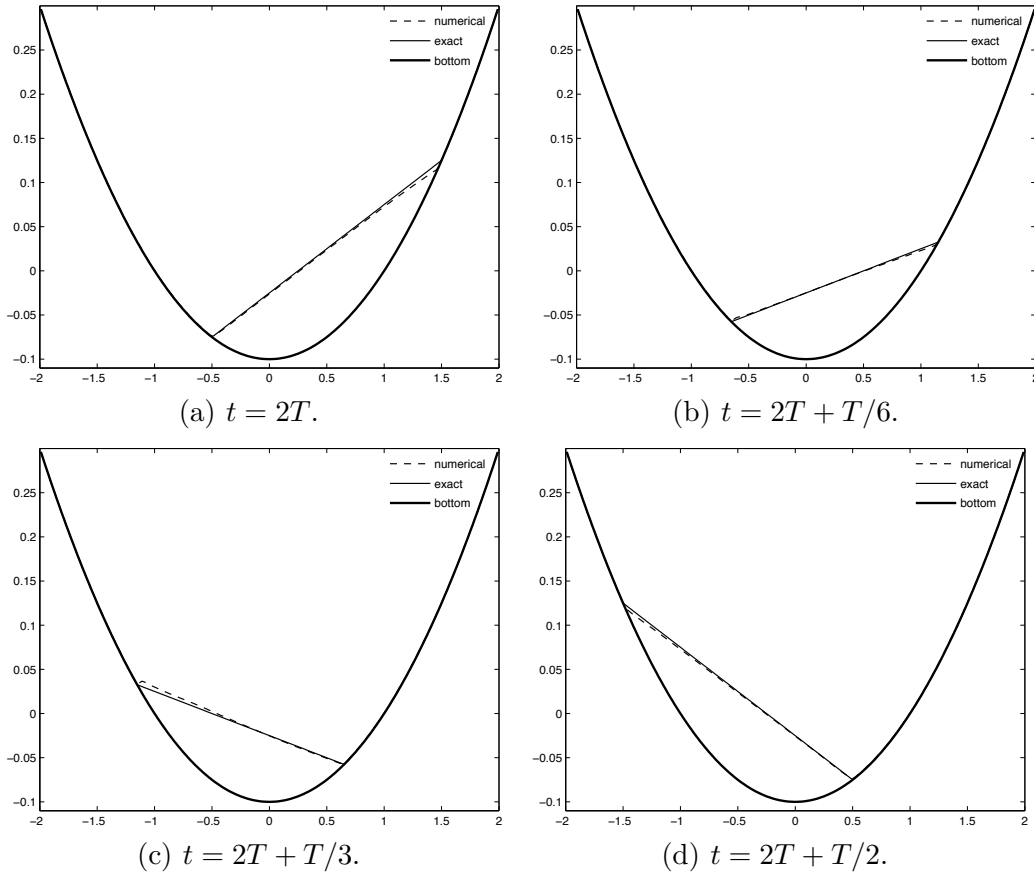


Fig. 6. 2-d oscillating lake: surface elevation vs  $x_1$ -coordinate, for  $x_2 = 0$ . Results obtained with BHM Roe scheme.

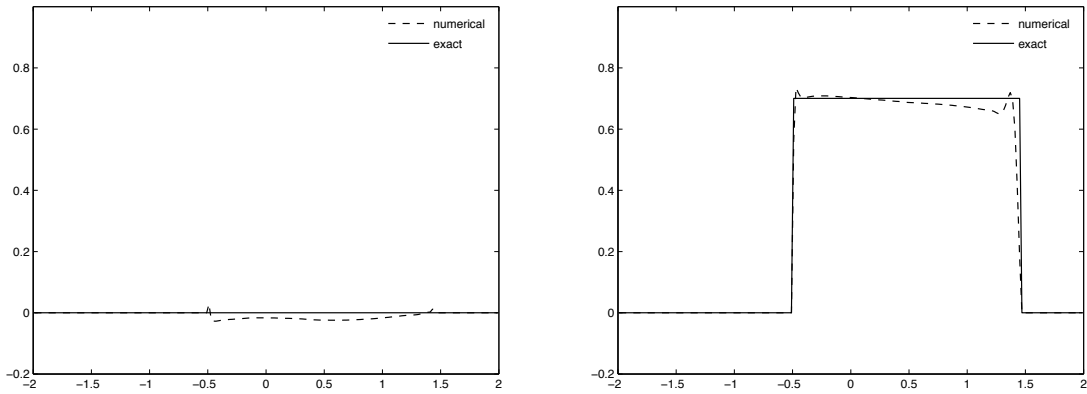


Fig. 7. 2-d oscillating lake: velocities at time  $t = 2T$ . Left:  $u_1$ . Right:  $u_2$ . Results obtained with BHM Roe scheme.

### 6.3.5 A two-dimensional oscillating lake

In [2], a modified Roe scheme is proposed in order to deal with wet/dry fronts. Combining the cited modified Roe scheme together with the bi-hyperbolic

reconstruction, a high order numerical scheme to deal with wet/dry fronts can be constructed and we denote by BHM Roe. This numerical test is design to show its performance in solutions where wet/dry fronts appear.

Let us consider the paraboloidal topography defined by the depth function

$$H(x_1, x_2) = h_0 \left( 1 - \frac{x_1^2 + x_2^2}{a^2} \right), \quad (x_1, x_2) \in [-2, 2] \times [-2, 2],$$

together with the periodic analytical solution of the two-dimensional shallow water equations stated in [21]:

$$\begin{aligned} h(x_1, x_2, t) &= \max \left( 0, \frac{\sigma h_0}{a^2} (2x_1 \cos(\omega t) + x_2 \sin(\omega t) - \sigma) + H(x_1, x_2) \right), \\ u_1(x_1, x_2, t) &= -\sigma \omega \sin(\omega t), \quad u_2(x_1, x_2, t) = \sigma \omega \cos(\omega t), \end{aligned}$$

where  $u_1$  and  $u_2$  are the velocities in the  $x_1$  and  $x_2$  directions, and  $\omega = \sqrt{2gh_0}/a$ . The values  $a = 1$ ,  $\sigma = 0.5$  and  $h_0 = 0.1$  have been considered for this test.

The computations have been performed using a quadrilateral mesh with  $\Delta x = \Delta y = 0.02$  and CFL number 0.7. Comparisons between the numerical and the analytical free surfaces at different times are shown in Figure 6, where  $T$  represents the oscillation period. Although a small distortion near the shorelines can be observed in some cases, they can be reduced using a finer spatial discretization. On the other hand, the planar form of the free surface is maintained throughout the computation.

To obtain accurate approximations of the velocity is a much more difficult issue. In Figure 7 are shown comparisons for both the  $u_1$  and  $u_2$  velocities at time  $t = 2T$ . As it can be observed, the position of the wet/dry fronts have been accurately captured, despite the small perturbations appearing in the wet zone.

## 7 Application to the two-layer shallow water system

### 7.1 Equations

We consider the two-layer shallow water system without friction terms:

$$\left\{ \begin{array}{l}
\frac{\partial h_1}{\partial t} + \frac{\partial q_{1,1}}{\partial x_1} + \frac{\partial q_{1,2}}{\partial x_2} = 0, \\
\frac{\partial q_{1,1}}{\partial t} + \frac{\partial}{\partial x_1} \left( \frac{q_{1,1}^2}{h_1} + \frac{g}{2} h_1^2 \right) + \frac{\partial}{\partial x_2} \left( \frac{q_{1,1} q_{1,2}}{h_1} \right) = -gh_1 \frac{\partial h_2}{\partial x_1} + gh_1 \frac{dH}{dx_1}, \\
\frac{\partial q_{1,2}}{\partial t} + \frac{\partial}{\partial x_1} \left( \frac{q_{1,1} q_{1,2}}{h_1} \right) + \frac{\partial}{\partial x_2} \left( \frac{q_{1,2}^2}{h_1} + \frac{g}{2} h_1^2 \right) = -gh_1 \frac{\partial h_2}{\partial x_2} + gh_1 \frac{\partial H}{\partial x_2}, \\
\frac{\partial h_2}{\partial t} + \frac{\partial q_{2,1}}{\partial x_1} + \frac{\partial q_{2,2}}{\partial x_2} = 0, \\
\frac{\partial q_{2,1}}{\partial t} + \frac{\partial}{\partial x_1} \left( \frac{q_{2,1}^2}{h_2} + \frac{g}{2} h_2^2 \right) + \frac{\partial}{\partial x_2} \left( \frac{q_{2,1} q_{2,2}}{h_2} \right) = -grh_2 \frac{\partial h_1}{\partial x_1} + gh_2 \frac{dH}{dx_1}, \\
\frac{\partial q_{2,2}}{\partial t} + \frac{\partial}{\partial x_1} \left( \frac{q_{2,1} q_{2,2}}{h_2} \right) + \frac{\partial}{\partial x_2} \left( \frac{q_{2,2}^2}{h_2} + \frac{g}{2} h_2^2 \right) = -grh_2 \frac{\partial h_1}{\partial x_2} + gh_2 \frac{\partial H}{\partial x_2}.
\end{array} \right. \quad (78)$$

The fluid is assumed to occupy a two dimensional domain  $D \subset \mathbb{R}^2$ . Again  $H$  is the depth function and  $g$  is the gravity. In the equations, index 1 makes reference to the upper layer and index 2 to the lower one. Each layer is assumed to have a constant density,  $\rho_i$ ,  $i = 1, 2$  ( $\rho_1 < \rho_2$ ).  $r = \frac{\rho_1}{\rho_2}$  is the density ratio. The unknowns  $q_{i,j}(\mathbf{x}, t)$  and  $h_i(\mathbf{x}, t)$  represent respectively the mass-flow in the  $x_j$  direction and the thickness of the  $i$ -th layer at point  $\mathbf{x}$  at time  $t$ .

This problem can also be written in the form (56) with:

$$W = [h_1; q_{1,1}; q_{1,2}; h_2; q_{2,1}; q_{2,2}]^T, \quad (79)$$

$$F_1(W) = \begin{bmatrix} q_{1,1} \\ \frac{q_{1,1}^2}{h_1} + \frac{1}{2}gh_1^2 \\ \frac{q_{1,1}q_{1,2}}{h_1} \\ q_{2,1} \\ \frac{q_{2,1}^2}{h_2} + \frac{1}{2}gh_2^2 \\ \frac{q_{2,1}q_{2,2}}{h_2} \end{bmatrix}, \quad F_2(W) = \begin{bmatrix} q_{1,2} \\ \frac{q_{1,1}q_{1,2}}{h_1} \\ \frac{q_{1,2}^2}{h_1} + \frac{1}{2}gh_1^2 \\ q_{2,2} \\ \frac{q_{2,1}q_{2,2}}{h_2} \\ \frac{q_{2,2}^2}{h_2} + \frac{1}{2}gh_2^2 \end{bmatrix}, \quad (80)$$

$$B_1(W) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -gh_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -rgh_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad B_2(W) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -gh_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -rgh_2 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (81)$$

$$S_1(\mathbf{x}, W) = \left[ 0; gh_1; 0; 0; gh_2; 0 \right]^T, \quad (82)$$

$$S_2(\mathbf{x}, W) = \left[ 0; 0; gh_1; 0; 0; gh_2 \right]^T. \quad (83)$$

## 7.2 Numerical schemes

We consider again a Roe linearization based on the family of segments:

$$\tilde{\Psi}(s; \tilde{W}_L, \tilde{W}_R, \eta) = \tilde{W}_L + s(\tilde{W}_R - \tilde{W}_L).$$

The procedure described in Section 5 is followed again to obtain the numerical scheme:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{|V_i|} \sum_{j \in \mathcal{N}_i} |E_{ij}| \left( \mathcal{P}_{ij}^-(\mathcal{A}_{ij}(W_j^n - W_i^n) - S_{ij}(H_j - H_i)) \right), \quad (84)$$

where:

$$W_i^n = \left[ h_{1,i}^n; q_{1,1,i}^n; q_{1,2,i}^n; h_{2,i}^n; q_{2,1,i}^n; q_{2,2,i}^n \right]^T, \quad u_{k,l,i}^n = \frac{q_{k,l,i}^n}{h_i^n}, \quad l = 1, 2, k = 1, 2. \quad (85)$$

$$\mathcal{A}_{ij} = \left[ \begin{array}{c|c} J_{ij}^1 & -B_{ij}^{1,2} \\ \hline -B_{ij}^{2,1} & J_{ij}^2 \end{array} \right]. \quad (86)$$

Here,

$$J_{ij}^k = \begin{bmatrix} 0 & \eta_{ij,1} & \eta_{ij,2} \\ (-u_{k,1,ij}^2 + c_{k,ij}^2)\eta_{ij,1} - u_{k,1,ij}u_{k,2,ij}\eta_{ij,2} & 2u_{k,1,ij}\eta_{ij,1} + u_{k,2,ij}\eta_{ij,2} & u_{k,1,ij}\eta_{ij,1} \\ u_{k,1,ij}u_{k,2,ij}\eta_{ij,1} + (-u_{k,2,ij}^2 + c_{k,ij}^2)\eta_{ij,2} & -u_{k,2,ij}\eta_{ij,2} & u_{k,1,ij}\eta_{ij,1} + 2u_{k,2,ij}\eta_{ij,2} \end{bmatrix}, \quad (87)$$

$$B_{ij}^{1,2} = \begin{bmatrix} 0 & 0 & 0 \\ -c_{1,ij}^2\eta_{ij,1} & 0 & 0 \\ -c_{1,ij}^2\eta_{ij,2} & 0 & 0 \end{bmatrix}, \quad B_{ij}^{2,1} = \begin{bmatrix} 0 & 0 & 0 \\ -rc_{2,ij}^2\eta_{ij,1} & 0 & 0 \\ -rc_{2,ij}^2\eta_{ij,2} & 0 & 0 \end{bmatrix}, \quad (88)$$

and

$$S_{ij} = \begin{bmatrix} 0 \\ gh_{1,ij}\eta_{ij,1} \\ gh_{1,ij}\eta_{ij,2} \\ 0 \\ gh_{2,ij}\eta_{ij,1} \\ gh_{2,ij}\eta_{ij,2} \end{bmatrix}, \quad (89)$$

being

$$c_{k,ij} = \sqrt{gh_{k,ij}}, \quad u_{k,l,ij} = \frac{\sqrt{h_{k,i}}u_{k,l,i} + \sqrt{h_{k,j}}u_{k,l,j}}{\sqrt{h_{k,i}} + \sqrt{h_{k,j}}}, \quad k = 1, 2; l = 1, 2, \quad (90)$$

$$h_{k,ij} = \frac{h_{k,i} + h_{k,j}}{2}, \quad k = 1, 2. \quad (91)$$

Two eigenvalues of  $\mathcal{A}_{ij}$  are given by:

$$\lambda_{ij,1} = (u_{1,1,ij}, u_{1,2,ij}) \cdot \eta_{ij}, \quad \lambda_{ij,2} = (u_{2,1,ij}, u_{2,2,ij}) \cdot \eta_{ij},$$

and the four other are the roots of the fourth order equation:

$$\left((\lambda - \lambda_{ij,1})^2 - gh_1\right) \left((\lambda - \lambda_{ij,2})^2 - gh_2\right) = rg^2h_1h_2.$$

As it happens in the 1d case, there is not an easy analytical expression for these 4 eigenvalues.

In this case, the choice of an appropriate family of paths to define the weak solutions of the problem based on a regularization of the problem is a difficult task. With this numerical scheme, which is based on the family of segments, the speeds of the shocks related to the genuinely nonlinear fields given by the scheme are expected to fit to Volpert's definition of nonconservative products



(see [22]) which is equivalent to the definition corresponding to the family of segments, i.e. they are expected to fit to the Rankine-Hugoniot condition (4) corresponding to the family of segments.

Concerning the approximation of contact discontinuities related to the linearly degenerate field, again only those related to water at rest solutions are exactly captured.

The numerical scheme satisfies again the  $\mathcal{C}$ -Property:

**Proposition 17** *The scheme (84) solves exactly the solutions corresponding to water at rest or vacuum.*

The proof is analogous to that of Proposition 15.

The expression of a semi-discrete high order extension of (84) is now as follows:

$$\begin{aligned}
W'_i = & -\frac{1}{|V_i|} \sum_{j \in \mathcal{N}_i} \int_{E_{ij}} \left( F_{\eta_{ij}}(W_{ij}^-(\gamma, t)) + \mathcal{P}_{ij}(\gamma, t)^- \left( \mathcal{A}_{ij}(\gamma, t) (W_{ij}^+(\gamma, t) - W_{ij}^-(\gamma, t)) \right. \right. \\
& \left. \left. - S_{ij}(\gamma, t) (H_{ij}^+ - H_{ij}^-) \right) \right) d\gamma \\
& + \frac{1}{|V_i|} \int_{V_i} \begin{pmatrix} 0 \\ gP_{h_1,i}^t(\mathbf{x}) \frac{\partial}{\partial x_1} (P_{H,i}^t - P_{h_2,i}^t)(\mathbf{x}) \\ gP_{h_1,i}^t(\mathbf{x}) \frac{\partial}{\partial x_2} (P_{H,i}^t - P_{h_2,i}^t)(\mathbf{x}) \\ 0 \\ gP_{h_2,i}^t(\mathbf{x}) \frac{\partial}{\partial x_1} (P_{H,i}^t - rP_{h_1,i}^t)(\mathbf{x}) \\ gP_{h_2,i}^t(\mathbf{x}) \frac{\partial}{\partial x_2} (P_{H,i}^t - rP_{h_1,i}^t)(\mathbf{x}) \end{pmatrix} d\mathbf{x},
\end{aligned} \tag{92}$$

where the following notation has been used:

$$\tilde{P}_i^t(\mathbf{x}) = \begin{pmatrix} P_{h_1,i}^t(\mathbf{x}) \\ P_{q_{1,1},i}^t(\mathbf{x}) \\ P_{q_{1,2},i}^t(\mathbf{x}) \\ P_{h_2,i}^t(\mathbf{x}) \\ P_{q_{2,1},i}^t(\mathbf{x}) \\ P_{q_{2,2},i}^t(\mathbf{x}) \\ P_{H,i}^t(\mathbf{x}) \end{pmatrix}$$

represents the approximation function at time  $t$  at the  $i$ -th cell;

$$W_{ij}^\pm(\gamma, t), \quad H_{ij}^\pm(\gamma, t),$$

the reconstructed states and depths at  $\gamma \in E_{ij}$  at time  $t$ ; and

$$\mathcal{A}_{ij}(\gamma, t), \quad \mathcal{P}_{ij}(\gamma, t)^-,$$

the Roe and projection matrices associated to the reconstructed states  $W_{ij}^\pm(\gamma, t)$ .

Concerning the well-balance properties of this high order extension, the following result can be stated:

**Proposition 18** *Let us suppose that the approximation functions are exact for constant functions. Moreover, let us suppose that the reconstructions of the variables  $h_2$ , and  $H$  satisfy the equality:*

$$P_{\eta_2, i}^t = P_{h_2, i}^t - P_{H, i}^t, \quad \forall i, \quad (93)$$

being  $\eta_2 = h_2 - H$ . Then, the reconstruction operator, and thus the semi-discrete scheme (92), is well-balanced for the stationary solutions corresponding to water at rest.

*Proof:*

Let us consider again a stationary solution corresponding to water at rest for the two-layer shallow-water system. We have:

$$h_1(\mathbf{x}) = \bar{h}_1 \in \mathbb{R} \text{ and } \eta_2(\mathbf{x}) = h_2(\mathbf{x}) - H(\mathbf{x}) = C, \forall \mathbf{x}.$$

As a consequence:

$$P_{h_1, i}(\mathbf{x}) = \bar{h}_1, \quad P_{\eta_2, i}(\mathbf{x}) = C, \quad \forall \mathbf{x} \in V_i, \forall i,$$

and thus:

$$P_{h_1, i}^0(\mathbf{x}) = \bar{h}_1, \quad P_{h_2, i}^0(\mathbf{x}) - P_{H, i}^0(\mathbf{x}) = C, \quad \forall \mathbf{x} \in V_i, \forall i.$$

Using this last equality, it is trivial to verify that:

$$\tilde{P}_i^0(\mathbf{x}) = \begin{pmatrix} P_{h_1, i}^0(\mathbf{x}) \\ 0 \\ 0 \\ P_{h_2, i}^0(\mathbf{x}) \\ 0 \\ 0 \\ P_{H, i}^0(\mathbf{x}) \end{pmatrix}$$

is also a stationary solution corresponding to water at rest.

□

### 7.3 Numerical experiments

We present here some tests to validate the performance of the Roe method and its high order extension BHRoe, when they are applied to the bi-layer shallow-water system. Again, we consider structured meshes, the third order bi-hyperbolic reconstruction, an optimal TVD Runge-Kutta method for the time stepping, and a Gaussian quadrature of order three for the integral terms. As in the one-layer shallow-water system, to obtain the equality (93), the reconstruction procedure is applied to  $\eta_2$  and  $H$ , and then (93) is used to recover the reconstruction of the variable  $h_2$ .

#### 7.3.1 Verification of the $\mathcal{C}$ -Property

In order to numerically verify the  $\mathcal{C}$ -Property for both Roe and BHRoe schemes when they are applied to the bi-layer shallow-water system, the same domain and depth function of Test 6.3.1 is used (see Figure 8). The following initial condition is used:  $h_1(\mathbf{x}, 0) = 0.4$ ,  $h_2(\mathbf{x}, 0) = H(\mathbf{x})$ ,  $q_{1,1}(\mathbf{x}, 0) = q_{1,2}(\mathbf{x}, 0) = q_{2,1}(\mathbf{x}, 0) = q_{2,2}(\mathbf{x}, 0) = 0$ ,  $\mathbf{x} \in [0, 1] \times [0, 1]$ . The ratio of densities is set to  $r = 0.998$ ,  $\Delta x = \Delta y = 0.01$ , and CFL is equals to 0.9. Periodic boundary conditions are also considered.

Tables 5 and 6 show the obtained results for each layer using Roe and BHRoe schemes. As expected both numerical schemes preserve the steady state solution exactly up to machine accuracy.

Table 5

Verification of the  $\mathcal{C}$ -Property. Upper layer.

Unknowns	$h_1$	$q_{1,1}$	$q_{1,2}$
$L^1$ error (Roe)	$5.75E^{-17}$	$3.58E^{-17}$	$6.71E^{-17}$
$L^1$ error (BHRoe)	$1.62E^{-19}$	$5.58E^{-16}$	$3.77E^{-16}$

Table 6

Verification of the  $\mathcal{C}$ -Property. Lower layer.

Unknowns	$h_2$	$q_{2,1}$	$q_{2,2}$
$L^1$ error (Roe)	$5.23E^{-17}$	$1.14E^{-16}$	$1.49E^{-16}$
$L^1$ error (BHRoe)	$4.91E^{-17}$	$1.39E^{-17}$	$1.12E^{-15}$

#### 7.3.2 Accuracy test.

This test is design to verify numerically the accuracy of BHRoe scheme when it is applied to the bi-layer shallow-water system. We consider the computational domain  $D = [-1, 1] \times [-1, 1]$ . The depth function is given by:

$$H(\mathbf{x}) = 2 + 0.1e^{-4\|\mathbf{x}\|^2},$$

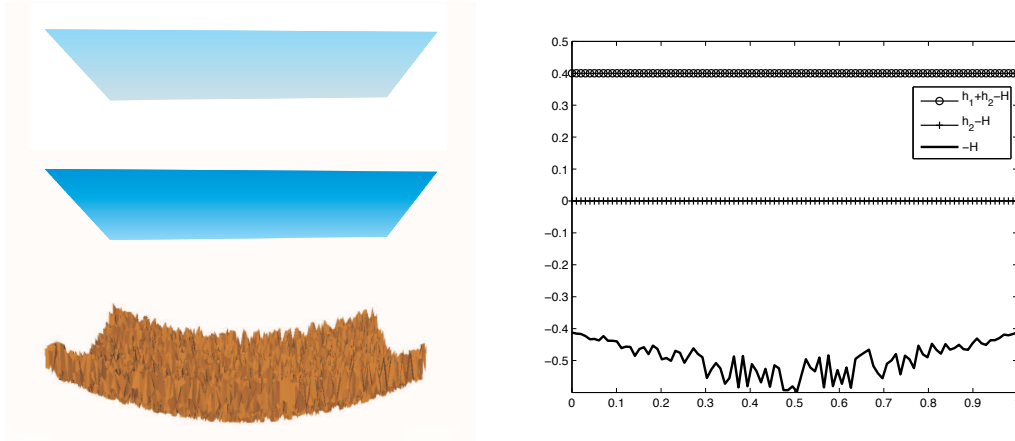


Fig. 8. Test case 7.3.1. Free surface, interface, and bottom topography.

Table 7

Test case 7.3.2:  $L^1$  errors and order. BHRoe (upper layer).

N.Cells	error $h_1$	order $h_1$	error $q_{1,1}$	order $q_{1,1}$	error $q_{1,2}$	order $q_{1,2}$
$8 \times 8$	$2.07E^{-1}$	-	$8.89E^{-1}$	-	$8.95E^{-1}$	-
$16 \times 16$	$8.05E^{-2}$	1.37	$2.64E^{-1}$	1.75	$2.64E^{-1}$	1.76
$32 \times 32$	$1.79E^{-2}$	2.17	$6.18E^{-2}$	2.12	$6.11E^{-2}$	2.12
$64 \times 64$	$3.22E^{-3}$	2.47	$1.58E^{-2}$	1.95	$1.58E^{-2}$	1.95
$128 \times 128$	$7.36E^{-4}$	2.13	$2.94E^{-3}$	2.42	$2.94E^{-3}$	2.42
$256 \times 256$	$1.14E^{-4}$	2.69	$4.25E^{-4}$	2.79	$4.25E^{-4}$	2.79

and the initial condition is:

$$h_1(\mathbf{x}, 0) = 1 - 0.01e^{-4\|\mathbf{x}\|^2}, \quad q_{1,1}(\mathbf{x}, 0) = q_{1,2}(\mathbf{x}, 0) = 0,$$

and

$$h_2(\mathbf{x}, 0) = 1 - 0.1e^{-6\|\mathbf{x}\|^2}, \quad q_{2,1}(\mathbf{x}, 0) = q_{2,2}(\mathbf{x}, 0) = 0.$$

Periodic boundary conditions are considered and  $r = 0.98$ . The CLF is set to 0.7. A reference solution over a mesh of  $512 \times 512$  cells is computed at  $t = 0.01$ . In Tables 7 and 8 the  $L^1$  errors and order obtained at time  $t = 0.01$  using BHRoe are shown.

### 7.3.3 Well-balance property for stationary 1d solutions

Let us consider the computational domain  $D = [-3, 3] \times [-3, 3]$  and the depth function

$$H(\mathbf{x}) = 2 - 0.5 e^{-0.5 x_1^2}, \quad \mathbf{x} \in [-3, 3] \times [-3, 3].$$

We construct a smooth stationary solution as follows: the values of  $q_{i,j}$  are

Table 8

Test case 7.3.2:  $L^1$  errors and order. BHRoe (lower layer).

N.Cells	error $h_2$	order $h_2$	error $q_{2,1}$	order $q_{2,1}$	error $q_{2,2}$	order $q_{2,2}$
$8 \times 8$	$1.99E^{-1}$	-	$9.02E^{-1}$	-	$9.09E^{-1}$	-
$16 \times 16$	$7.96E^{-2}$	1.33	$2.68E^{-1}$	1.75	$2.69E^{-1}$	1.76
$32 \times 32$	$1.81E^{-2}$	2.14	$6.21E^{-2}$	2.11	$6.21E^{-2}$	2.11
$64 \times 64$	$3.26E^{-3}$	2.47	$1.61E^{-2}$	1.95	$1.61E^{-2}$	1.95
$128 \times 128$	$6.98E^{-4}$	2.22	$2.99E^{-3}$	2.43	$2.99E^{-3}$	2.42
$256 \times 256$	$1.14E^{-4}$	2.62	$4.31E^{-4}$	2.79	$4.31E^{-4}$	2.79

fixed:

$$\begin{aligned} q_{1,1}(\mathbf{x}, 0) &= 0.15, & q_{1,2}(\mathbf{x}, 0) &= 0, \\ q_{2,1}(\mathbf{x}, 0) &= -0.15, & q_{2,2}(\mathbf{x}, 0) &= 0, \end{aligned}$$

as well as the values of  $h_1$  and  $h_2$  at  $x_1 = 3$ :

$$h_1(0, x_2, 0) = 0.5, \quad h_2(0, x_2, 0) = H(0, x_2, 0) - 0.5.$$

The density ratio is set to  $r = 0.98$ . The values of  $h_1(\mathbf{x}, 0)$  and  $h_2(\mathbf{x}, 0)$  are then obtained by solving the nonlinear system:

$$\begin{cases} \frac{1}{2}(q_{1,1})^2 + gh_1^3 = -gh_2h_1^2 + (gH + K_1)h_1^2, \\ \frac{1}{2}(q_{2,1})^2 + gh_2^3 = rgh_1h_2^2 + (gH + K_2)h_2^2, \end{cases}$$

where  $K_1$  and  $K_2$  are obtained by computing the mechanical energy of both layers at  $x_1 = 3$ . In Figure 9 the stationary solution obtained is depicted.

We have applied Roe and BHRoe taking this stationary solution as initial condition with CFL equals to 0.9. As boundary conditions, the free surface and the interface are fixed at  $x_1 = 3$  and  $q_{1,1}$  and  $q_{2,1}$  are fixed at  $x_1 = -3$ . Wall boundary conditions are considered at  $x_2 = -3$  and  $x_2 = 3$ . The results obtained are shown in Tables 9, 10, 11, and 12.

As expected, Roe achieves second order, while BHRoe achieves third order.

#### 7.3.4 Internal circular dam-break

Let us consider now the domain  $[-5, 5] \times [-5, 5]$ , and a flat bottom topography given by the function  $H(\mathbf{x}) = 2$ . As initial condition we set:

$$h_1(\mathbf{x}, 0) = \begin{cases} 1.8 & \text{if } \|\mathbf{x}\|^2 > 4, \\ 0.2 & \text{if } \|\mathbf{x}\|^2 \leq 4, \end{cases}$$

Table 9

Test case 7.3.3:  $L^1$  errors and order. Roe (upper layer).

N.Cells	error $h_1$	order $h_1$	error $q_{1,1}$	order $q_{1,1}$
$10 \times 10$	$4.01E^{-2}$	-	$8.86E^{-4}$	-
$20 \times 20$	$1.14E^{-2}$	1.81	$2.27E^{-4}$	1.96
$40 \times 40$	$2.92E^{-3}$	1.96	$5.62E^{-5}$	2.01
$80 \times 80$	$5.97E^{-4}$	2.29	$1.09E^{-5}$	2.36
$160 \times 160$	$1.25E^{-4}$	2.25	$2.32E^{-6}$	2.24

Table 10

Test case 7.3.3:  $L^1$  errors and order. Roe (lower layer).

N.Cells	error $h_2$	order $h_2$	error $q_{2,1}$	order $q_{2,1}$
$10 \times 10$	$3.96E^{-2}$	-	$8.69E^{-4}$	-
$20 \times 20$	$1.13E^{-2}$	1.81	$2.22E^{-4}$	1.97
$40 \times 40$	$2.89E^{-3}$	1.97	$5.50E^{-5}$	2.01
$80 \times 80$	$5.90E^{-4}$	2.29	$1.07E^{-5}$	2.36
$160 \times 160$	$1.22E^{-4}$	2.27	$2.21E^{-6}$	2.27

Table 11

Test case 7.3.3:  $L^1$  errors and order. BHRoe (upper layer).

N.Cells	error $h_1$	order $h_1$	error $q_{1,1}$	order $q_{1,1}$
$10 \times 10$	$3.30E^{-2}$	-	$6.71E^{-3}$	-
$20 \times 20$	$3.80E^{-3}$	3.12	$7.56E^{-4}$	3.15
$40 \times 40$	$6.90E^{-4}$	2.46	$1.34E^{-4}$	2.50
$80 \times 80$	$1.06E^{-4}$	2.70	$1.89E^{-5}$	2.82
$160 \times 160$	$1.40E^{-5}$	2.92	$2.38E^{-6}$	2.99

Table 12

Test case 7.3.3:  $L^1$  errors and order. BHRoe (lower layer).

N.Cells	error $h_2$	order $h_2$	error $q_{2,1}$	order $q_{2,1}$
$10 \times 10$	$3.70E^{-2}$	-	$6.57E^{-3}$	-
$20 \times 20$	$4.26E^{-3}$	3.12	$7.25E^{-4}$	3.18
$40 \times 40$	$7.42E^{-4}$	2.52	$1.20E^{-4}$	2.60
$80 \times 80$	$1.10E^{-4}$	2.75	$1.66E^{-5}$	2.85
$160 \times 160$	$1.48E^{-5}$	2.90	$2.07E^{-6}$	3.00

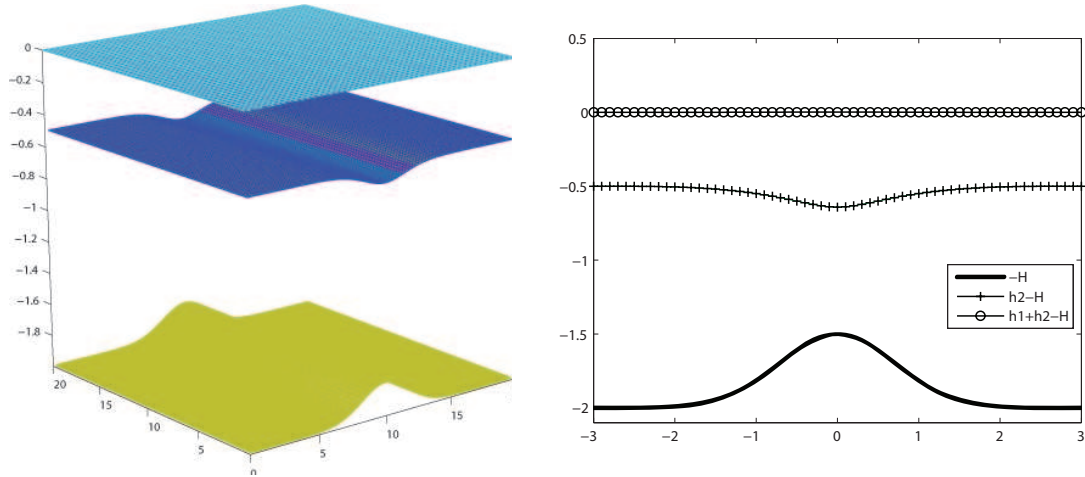


Fig. 9. Test 7.3.3: initial condition. 3D view and longitudinal section.

$h_2(\mathbf{x}, 0) = 2 - h_1(\mathbf{x}, 0)$  and  $q_{1,1}(\mathbf{x}, 0) = q_{1,2}(\mathbf{x}, 0) = q_{2,1}(\mathbf{x}, 0) = q_{2,2}(\mathbf{x}, 0) = 0$  (see Figure 10).

A rectangular mesh with  $\Delta x = \Delta y = 0.05$  is considered. The CFL is set to 0.9 and the density ratio is set to  $r = 0.998$ . Free boundary conditions are imposed at all boundaries.

In Figure 10 the evolution of the interface obtained with BHRoe scheme is depicted at times  $t = 0, 1, 2, 3, 4$  and 5s. Notice the internal circular wave expanding through the domain.

In Figures 11 and 12 the results obtained with both schemes Roe and BHRoe and a reference solution computed over a finer mesh composed with  $800 \times 800$  cells are compared at the longitudinal section  $x_2 = 0$ . Figure 11 shows the free surface and the interface, and Figure 12 the mass fluxes.

**Acknowledgments.** This research has been partially supported by the Spanish Government Research projects MTM2006-08075 and MTM2006-01275. The numerical computations have been performed at the Laboratory of Numerical Methods of the University of Málaga.

## References

- [1] L. ARMI, D. FARMER. *Maximal two-layer exchange through a contraction with barotropic net flow*. J. Fluid Mech. 164:27–51, 1986.
- [2] M.J. CASTRO, A. FERREIRO, J.A. GARCÍA, J. GONZÁLEZ-VIDA, J. MACÍAS, C. PARÉS, M.E. VÁZQUEZ-CENDÓN. *On the numerical treatment of wet/dry fronts in shallow flows: application to one-layer and two-layers systems*. Math. and Comp. model., 42(3-4):419–439, 2005
- [3] M.J. CASTRO, J.M. GALLARDO, C. PARÉS. *High order finite volume schemes based on reconstruction of states for solving hyperbolic systems with*

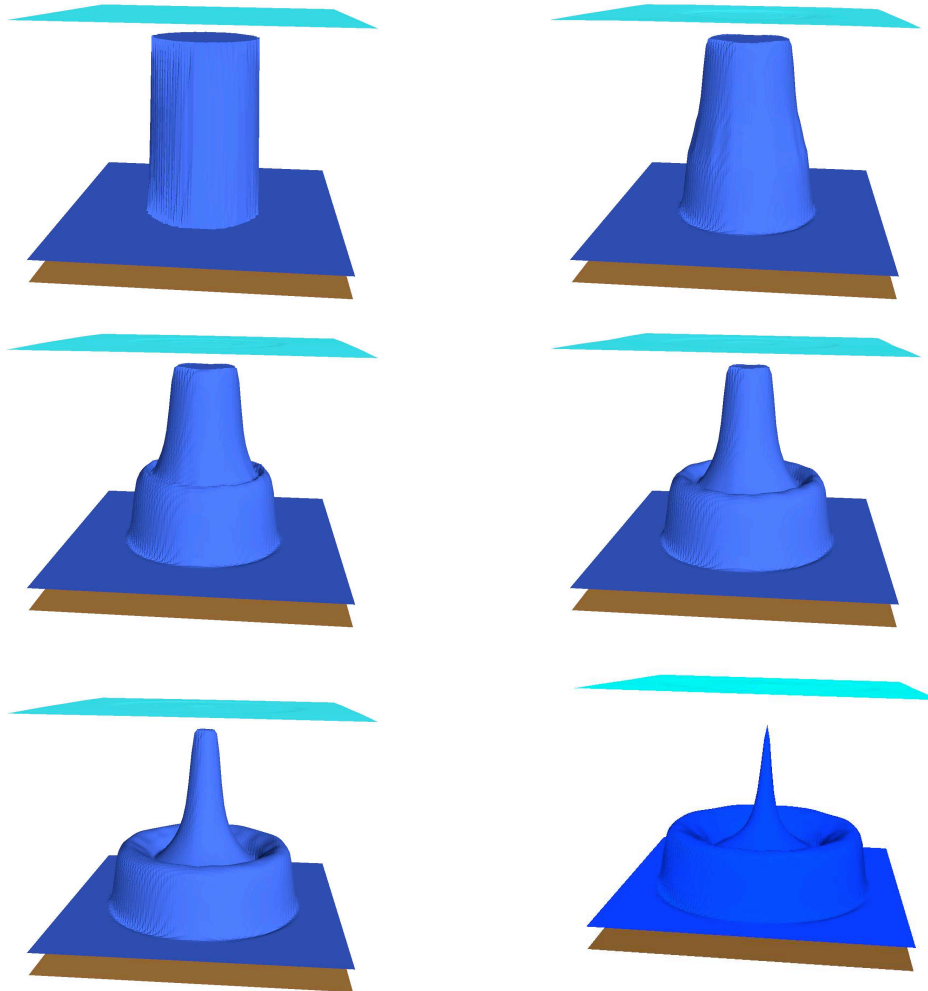


Fig. 10. Internal circular dam-break (interface, upper view). BHRoe. Free surface and interface at times  $t = 0, 1, 2, 3, 4, 5$  (from top to bottom and left to right).

*nonconservative products. Applications to shallow water systems.* Math. Comp., 75: 1103-1134, 2006.

- [4] G. DAL MASO, P.G. LEFLOCH AND F. MURAT. *Definition and weak stability of nonconservative products.* J. Math. Pures Appl. 74:483–548, 1995.
- [5] D. FARMER, L. ARMI. *Maximal two-layer exchange over a sill and through a combination of a sill and contraction with barotropic flow.* J. Fluid Mech. 164:53–76, 1986.
- [6] L. GOSSE *A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms.* Comp. Math. with Applic. 39:135–159, 2000.
- [7] L. GOSSE. *A well-balanced scheme using non-conservative products designed for hyperbolic system of conservation laws with source terms.* Mat. Mod. Meth. Appl. Sc. 11: 339–365, 2001.



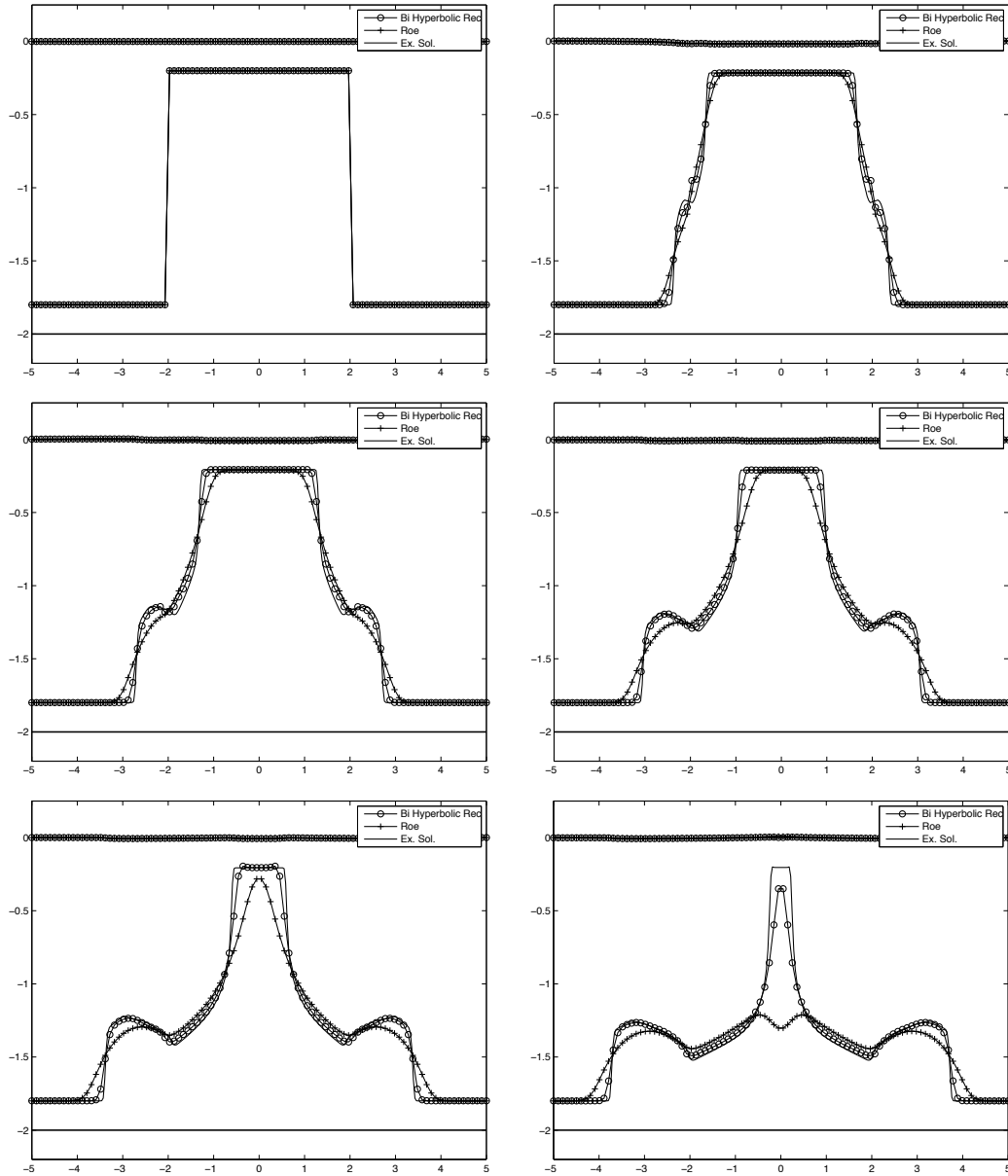


Fig. 11. Internal circular dam-break. Interface and free surface at section  $x_2 = 0$  at times  $t = 0, 1, 2, 3, 4, 5$  (from top to bottom and left to right).

- [8] S. GOTTLIEB AND C. W. SHU. *Total variation diminishing Runge-Kutta schemes*. *Mat. Comp.* 67: 73–85, 1998.
- [9] J.M. GREENBERG AND A.Y. LEROUX. *A well balanced scheme for the numerical processing of source terms in hyperbolic equations*. *SIAM J. Numer. Anal.* 33: 1–16, 1996.
- [10] J.M. GREENBERG, A.Y. LEROUX, R. BARAILLE, A. NOUSSAIR. *Analysis and approximation of conservation laws with source terms*. *SIAM J. Numer. Anal.* 34: 1980–2007, 1997.
- [11] A. HARTEN, J.M. HYMAN. *Self-adjusting grid methods for one-dimensional*

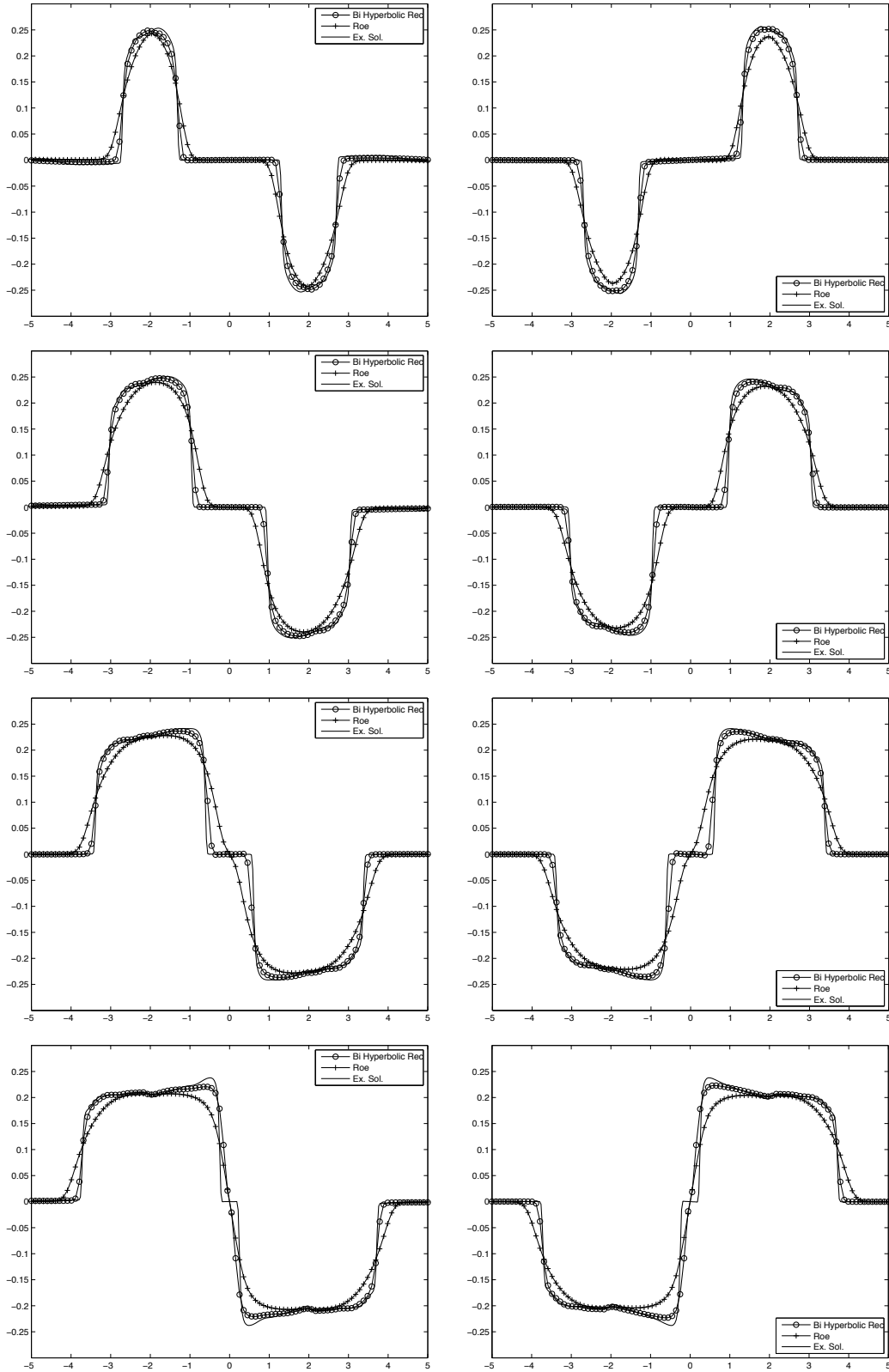


Fig. 12. Internal circular dam-break.  $q_{1,1}$  (left column)  $q_{2,1}$  (right column) at section  $x_2 = 0$  at times  $t = 2, 3, 4$ , and  $5$ .

- hyperbolic conservation laws*. J. Comp. Phys. 50:235–269, 1983.
- [12] P.G. LEFLOCH. *Shock waves for nonlinear hyperbolic systems in nonconservative form*, Institute for Math. and its Appl., Minneapolis, Preprint 593, 1989.
- [13] R. J. LEVEQUE. *Balancing source terms and flux gradient in high-resolution Godunov methods: the quasi-steady wave-propagation algorithm*. J. Comput. Phys. 146: 346–365, 1998.
- [14] A. MARQUINA. *Local piecewise hyperbolic reconstruction of numerical fluxes for non linear scalar conservation laws*. SIAM, Journal Sci. Comput. 15(4), 892–915, 1994.
- [15] M.L. MUÑOZ AND C. PARÉS. *Godunov’s method for nonconservative hyperbolic systems*. ESAIM: M2AN 41(1): 169-185, 2007.
- [16] C. PARÉS, M.J. CASTRO. *On the well-balance property of Roe’s method for nonconservative hyperbolic systems. Applications to shallow-water systems*. ESAIM: M2AN, 38(5):821–852, 2004.
- [17] C. W. SHU AND S. OSHER. *Efficient implementation of essentially non-oscillatory shock capturing schems*. J. Comp. Phys., 77:439–471, 1998.
- [18] H. J. SCHROLL AND F. SVENSSON. *A Bihyperbolic Fiinite Volume Method for Quadrilateral Meshes*. SIAM: J. Sci. Comput., 26(2): 237–260, 2006.
- [19] S. SERNA. *A class of extended limiters applied to piecewise hyperbolic methods*. SIAM: J. Sci. Comput., 28(1): 123–140, 2006.
- [20] I. TOUMI. *A weak formulation of Roe’s approximate Riemann Solver*. J. Comp. Phys. 102(2): 360–373, 1992.
- [21] W.C. THACKER. *Some exact solutions to the nonlinear shallow-water wave equations*. J. Fluid Mech. 107: 499–508, 1981.
- [22] A. I. VOLPERT, *Spaces BV and quasilinear equations*, Math. USSR Sbornik, 73 (1967), pp. 255–302.
- [23] Y. Xing, C.-W. Shu, *High order finite difference WENO schemes with the exact conservation property for the shallow water equations*, J. Comput. Phys. 208 (2005) 206–227.