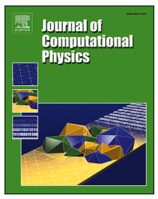
Journal of Computational Physics 535 (2025) 114047

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp



High-order WENO finite-difference methods for hyperbolic nonconservative systems of partial differential equations

Baifen Ren [a], Carlos Parés [b,*]

[a] School of Mathematical Sciences, Ocean University of China, Qingdao, 266100, China
[b] Departamento de Análisis Matemático, Estadística e Investigación Operativa, y Matemática aplicada, Universidad de Málaga, Bulevar Louis Pasteur, 31, 29010, Málaga, Spain

ARTICLE INFO

Keywords:
WENO finite difference scheme
High order accuracy
Well-balanced scheme
Nonconservative equations
Path-conservative method

ABSTRACT

This work aims to extend the well-known high-order WENO finite-difference methods for systems of conservation laws to nonconservative hyperbolic systems. The main difficulty of these systems both from the theoretical and the numerical points of view comes from the fact that the definition of weak solution is not unique: according to the theory developed by Dal Maso, LeFloch, and Murat in 1995, it depends on the choice of a family of paths. A new strategy is introduced here that allows non-conservative products to be written as the derivative of a generalized flux function that is defined locally on the basis of the selected family of paths. WENO reconstructions are then applied to this generalized flux. Moreover, if a Roe linearization is available, the generalized flux function can be evaluated through matrix-vector operations instead of path-integrals. Two different known techniques are used to extend the methods to problems with source terms and the well-balanced properties of the resulting schemes are studied. These numerical schemes are applied to a coupled Burgers' system and to the two-layer shallow water equations in one- and two- dimensions to obtain high-order methods that preserve water-at-rest steady states.

1. Introduction

We aim to construct well-balanced high-order WENO finite-difference schemes for hyperbolic nonconservative problems of the form

$$U_t + A_1(U)U_x + A_2(U)U_y = S_1(U)H_x + S_2(U)H_y, \tag{1}$$

where the unknown $U(x, y, t)$ takes value in an open convex set $\Omega \in \mathbb{R}^N$; $A_i(U)$, $i = 1, 2$ are smooth matrix-valued functions; $S_i(U)$, $i = 1, 2$ are vector-valued functions, and $H(x, y)$ is a known function from $\mathbb{R}^2$ to $\mathbb{R}$. The 1D case

$$U_t + A(U)U_x = S(U)H_x \tag{2}$$

will also be considered.

PDE systems of the form

$$U_t + F_1(U)_x + F_2(U)_y + B_1(U)U_x + B_2(U)U_y = S_1(U)H_x + S_2(U)H_y, \tag{3}$$

where $F_i(U)$, $i = 1, 2$ are the flux function and $B_i(U)$, $i = 1, 2$ are matrix-valued functions, can be considered as particular cases of (1). Systems of conservation laws ($B_i \equiv 0$, $S_i \equiv 0$, $i = 1, 2$) and systems of balance laws ($B_i \equiv 0$, $i = 1, 2$) are in turn particular cases of (3).

* Corresponding author.
E-mail addresses: renbaifen@stu.ouc.edu.cn (B. Ren), pares@uma.es (C. Parés).

https://doi.org/10.1016/j.jcp.2025.114047
Received 5 July 2024; Received in revised form 1 March 2025; Accepted 28 April 2025
Available online 1 May 2025
0021-9991/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

The numerical methods will be applied to the 1D and 2D hyperbolic two-layer shallow-water equations that govern the flow of two superposed layers of immiscible homogeneous fluids. This system is used in different ocean and coastal engineering simulations and there is a vast literature focusing on its numerical analysis: see for instance [1–7]. In order to illustrate the general procedure, we will also consider the 1D coupled Burgers' system introduced in [8]:

$$\begin{cases} u_t + uu_x + uv_x = 0, \\ v_t + vu_x + vv_x = 0, \end{cases} \tag{4}$$

that can be written in the form (2) with

$$A(U) = \begin{bmatrix} u & u \\ v & v \end{bmatrix}, \quad U \in \Omega,$$

with

$$\Omega = \left\{ \begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^2 \text{ s.t.} u + v > 0 \right\}.$$

The eigenvalues of $A(U)$ are

$$\lambda_1(U) = 0, \quad \lambda_2(U) = u + v,$$

and as so that the system is strictly hyperbolic in Ω.

The major difficulty of systems (1) or (3), from both the theoretical and the numerical points of view, comes from the fact that the presence of nonconservative products makes that, unlike for systems of conservation laws, the definition of weak solution is not unique. In the theory developed by Dal Maso, LeFloch, and Murat in [9], nonconservative products are defined as Borel measures based on the choice of a family of paths, i.e. a Lipschitz-continuous function $\Psi : [0,1] \times \Omega \times \Omega \to \Omega$ satisfying

$$\Psi(0; U_L, U_R) = U_L, \quad \Psi(1; U_L, U_R) = U_R$$

for all U_L, $U_R \in \Omega$, and

$$\Psi(s; U, U) = U$$

for all $U \in \Omega$ and $s \in [0,1]$. Once the family of paths has been used, the generalized Rankine-Hugoniot conditions satisfied by the admissible weak solutions at a jump are the following

$$\sigma(U^+ - U^-) = \int_0^1 \left(\sum_{i=1}^{2} n_i A_i(\Psi(s; U^-, U^+)) \right) \partial_s \Psi(s; U^-, U^+)\, ds,$$

where σ is the propagation speed; $U^\pm$ the lateral limits of the solution; and $\vec{n} = (n_1, n_2)$ a unit vector normal to the jump. For instance, in the particular case of system (4), it can be checked that the choice of the family of straight segments

$$\Psi_1(s; U^-, U^+) = U^- + s(U^+ - U^-), \quad s \in [0,1] \tag{5}$$

leads to the jump conditions

$$\begin{cases} \bar{u}(u^+ - u^-) + \bar{u}(v^+ - v^-) = \sigma(u^+ - u^-), \\ \bar{v}(u^+ - u^-) + \bar{v}(v^+ - v^-) = \sigma(v^+ - v^-), \end{cases} \tag{6}$$

with

$$\bar{u} = \frac{u^+ + u^-}{2}, \quad \bar{v} = \frac{v^+ + v^-}{2},$$

while the choice

$$\Psi_2(s; U^-, U^+) = \begin{bmatrix} u^- + s(4 - 3s)(u^+ - u^-) \\ v^- + s(v^+ - v^-) \end{bmatrix}, \quad s \in [0,1], \tag{7}$$

leads to

$$\begin{cases} \bar{u}(u^+ - u^-) + u^+(v^+ - v^-) = \sigma(u^+ - u^-), \\ v^-(u^+ - u^-) + \bar{v}(v^+ - v^-) = \sigma(v^+ - v^-). \end{cases} \tag{8}$$

Any choice of family paths leads to a consistent definition of weak solutions from the mathematical point of view. Therefore, given a particular application, the choice of the adequate family of paths has to be based on the consistency with the physics of the problem: for instance, the family of paths can be given by the viscous profiles corresponding to the neglected viscous terms (see [10] for a general discussion and [11] for the particular case of System (4)).

Based on this theory, a framework for designing finite-volume methods for nonconservative systems was introduced in [12] based on the concept of path-conservative methods. These methods have been extensively applied to solve nonconservative systems: see, for instance, [13–17].

In this paper, we focus on WENO finite-difference schemes. In the past decades, these methods have been widely applied to systems of conservation and balance laws: see for instance [18–23]. In [4] a fifth-order A-WENO finite-difference scheme for 1D and

2D systems of nonconservative hyperbolic systems was introduced. This scheme is based on the path-conservative central-upwind method and the global flux approach in which the integral of the nonconservative term is considered as a new flux function of the system. A high-order conservative method is then applied to the formal system of conservation laws. The evaluation of the new flux function requires the computation of the cell integrals of the nonconservative products using a high-order quadrature formula. Therefore, in addition to the flux reconstruction of the flux, it also requires high-order reconstructions of U at the quadrature points. If this approach is used, the method is reduced to the conservative one obviously when $A(U)$ is the Jacobian of a flux function F. Recently, in [24] a finite-difference WENO method has been introduced based on state reconstruction that does not require the computation of cell integrals of the conservative products. In contrast to global flux-based methods, they reduce to the conservative method only when $A(U)$ is linear.

The goal of this paper is to introduce a *local flux* approach to design high-order finite-difference methods for nonconservative systems. The key idea is to apply a standard high-order WENO reconstruction operator to the nonconservative products computed using the selected family of paths: more precisely, in the 1D case, the selected WENO operator will be applied to reconstruct quantities of the form

$$\int_0^1 A^{\pm}(\Psi(s;U_i,U_j))\partial_s\Psi(s;U_i,U_j)\,ds, \quad j\in S_i,$$

where S_i represents the stencil of the ith point and $A^{\pm}(U)$ represents a splitting of the matrix system $A(U)$. It will be seen that, if a path-consistent Roe linearization is available (see [25,26]), the quantities to be reconstructed can be computed by using matrix-vector products instead of path-integrals. The main advantages of this new one are the following:

- the accuracy in space of the methods only depends on the order of the selected WENO reconstruction provided that the selected family of paths satisfies a symmetry property to be described;
- no integrals involving quadrature points in the cells have to be computed which avoids having to use an additional reconstruction operator with uniform accuracy in the entire cells;
- the methods reduce to the standard finite-difference conservative WENO schemes when $A(U)$ is the Jacobian matrix of a flux function $F(U)$.

Two different matrix splittings will be considered here based on the standard Lax-Friedrichs and Upwind approaches to illustrate the strategy. Nevertheless it can also be applied to more general splittings or even to WENO reconstructions that are not based on splitting technique, as in the case of A-WENO. The application of the local flux approach combined with A-WENO reconstructions will be discussed in a forthcoming work.

A relevant property to be satisfied by the numerical methods solving systems of the form (1) or (3) is the preservation of some or all the steady-state solutions of the system, i.e. the well-balanced property. For instance, in the context of the one or the two-layer shallow-water equations, a minimal requirement to the numerical methods is to exactly preserve the steady states corresponding to water-at-rest, i.e. to satisfy the *C-property* according to [27]. Different techniques have been proposed to design well-balanced schemes including hydrostatic reconstruction related [28,29], relaxation methods [30], consistent discretization of the flux and source term [31,32], etc. See also [33–35].

In the context of finite-volume methods, a general strategy to design well-balanced methods has been described in [36]. In this strategy, a stationary solution whose average is the numerical approximation at every cell has to be computed at every time step. Then, a standard reconstruction operator is applied to the differences in the cell values at the stencil and the cell averages of the local stationary solution. In [37] this strategy has been extended to WENO finite-difference methods for systems of balance laws. Two different strategies will be followed here to obtain well-balanced numerical methods: one of them is the extension to nonconservative systems of the strategy introduced in this last reference, while the other consists of combining the Upwind splitting scheme with an adequate choice of family of paths.

As it is well known (see [38]), in the case of nonconservative systems, the numerical solutions obtained with finite-difference or similar methods that are formally consistent with the definition of weak solution related to a given family of paths may converge to functions that are not weak solutions according to that family. Nevertheless, it will be seen in Section 6 that the numerical results obtained for the two-layer shallow-water equations are similar to those obtained with other methods. Nevertheless, in order to ensure the convergence to functions that are weak solutions according to the selected family of paths, the numerical dissipation close to shocks has to be controlled: see for instance [39]. In [40,41] high-order finite-volume numerical methods that are able to correctly capture isolated shock waves have been designed based on the use of a discontinuous reconstruction operator in cells where a shock is detected: similar strategies can be adapted to the numerical methods introduced here, what will be done in future works.

The rest of the paper is organized as follows. In Section 2, firstly, the new WENO path-conservative schemes for 1D homogeneous (i.e. without source terms) nonconservative systems are introduced and their high-order accuracy property is proved. In this section the general problem is considered, so that the family of paths is, in principle, arbitrary. Nevertheless, it will be shown that a symmetry property has to be satisfied to ensure the high-order accuracy of the method. In Section 3, source terms are included and two strategies to obtain well-balanced methods are described. In Section 4, the proposed schemes are extended to 2D nonconservative systems. In Section 5, we apply the proposed scheme to 1D and 2D two-layer shallow water equations: the numerical results are presented in Section 6. Finally, some conclusions are drawn in Section 7.

2. Path-Conservative WENO finite-difference reconstruction methods

2.1. Conservative WENO finite difference schemes: a brief overview

The goal of this paper is to extend high-order finite-difference methods based on flux reconstructions for conservation laws system

$$U_t + F(U)_x = 0, \tag{9}$$

to nonconservative systems

$$U_t + A(U)U_x = 0. \tag{10}$$

It will be assumed here that the system is strictly hyperbolic, i.e. for every U, the matrix $A(U)$ has N different real eigenvalues

$$\lambda_1(U), \dots, \lambda_N(U).$$

In particular, systems of the form

$$U_t + F(U)_x + B(U)U_x = 0 \tag{11}$$

will be considered that can be written in the form (10) with

$$A(U) = J(F(U)) + B(U),$$

where $J(F(U))$ represents the Jacobian of the flux function $F(U)$.

Semi-discrete high-order finite-difference methods for systems of conservation laws (9) have the form:

$$\frac{dU_i}{dt} + \frac{1}{\Delta x}\left(F_{i+1/2} - F_{i-1/2}\right) = 0, \tag{12}$$

where $F_{i+1/2}$ is a high-order reconstruction of the flux function. The computational domain is $[a, b]$. Uniform meshes with a constant step size Δx will be considered, with cell centers denoted as x_i. The following notation is used for the cell interface:

$$x_{i+\frac{1}{2}} = x_i + \frac{\Delta x}{2}.$$

In the particular case of the WENO reconstruction of order $p = 2k + 1$ for systems of balance laws, two flux reconstructions are computed using the values at the points $x_{i-k}, \dots, x_{i+k}$:

$$F^L_{i+1/2} = \mathcal{R}^L(F(U_{i-k}), \dots, F(U_{i+k})), \tag{13}$$

$$F^R_{i-1/2} = \mathcal{R}^R(F(U_{i-k}), \dots, F(U_{i+k})). \tag{14}$$

These are the so-called left- and right-biased reconstructions, related by the equality:

$$\mathcal{R}^L(F(U_{i-k}), \dots, F(U_{i+k})) = \mathcal{R}^R(F(U_{i+k}), \dots, F(U_{i-k})).$$

In order to compute the numerical flux $F_{i+1/2}$, first a splitting of the flux function is considered

$$F(U) = F^+(U) + F^-(U),$$

in such a way that the eigenvalues of the Jacobian $J^+(U)$ (resp. $J^-(U)$) of $F^+(U)$ (resp. $F^-(U)$) are positive (resp. negative). A standard choice is the Lax-Friedrichs flux-splitting:

$$F^\pm(U) = \frac{1}{2}(F(U) \pm \alpha U),$$

where α is the maximum of the absolute value of the eigenvalues of $\{J(U_i)\}$, this maximum being taken over either local (WENO-LLF) or global (WENO-LF): see [21,22].

Then, the reconstruction operator is applied to $F^\pm$:

$$F^+_{i+1/2} = \mathcal{R}^L(F^+(U_{i-k}), \dots, F^+(U_{i+k})), \tag{15}$$

$$F^-_{i+1/2} = \mathcal{R}^R(F^-(U_{i+1-k}), \dots, F^-(U_{i+1+k})), \tag{16}$$

and finally,

$$F_{i+1/2} = F^+_{i+1/2} + F^-_{i+1/2}. \tag{17}$$

The reconstruction then satisfies

$$\frac{1}{\Delta x}\left(F_{i+1/2} - F_{i-1/2}\right) = F(U)_x + O(\Delta x^{2k+1}), \quad \forall i.$$

Any version of WENO reconstructions can be selected. In particular, in the numerical tests shown in Section 6, WENOZ is used (see [19]): for the sake of completeness, the expression of the fifth-order WENOZ reconstruction is recalled in A.

2.2. Extension to nonconservative systems

In order to extend these numerical methods to (10), let us first rewrite (12) as follows:

$$\frac{dU_i}{dt} + \frac{1}{\Delta x}\left(F_{i+1/2} - F\left(U_i\right) + F\left(U_i\right) - F_{i-1/2}\right) = 0, \tag{18}$$

or, equivalently

$$\frac{dU_i}{dt} + \frac{1}{\Delta x}\left(\widehat{D}^-_{i+1/2} + \widehat{D}^+_{i-1/2}\right) = 0, \tag{19}$$

with

$$\widehat{D}^-_{i+1/2} = F_{i+1/2} - F\left(U_i\right), \quad \widehat{D}^+_{i-1/2} = F\left(U_i\right) - F_{i-1/2}.$$

One has then

$$\begin{aligned}
\widehat{D}^-_{i+1/2} &= F_{i+1/2} - F(U_i) \\
&= F^+_{i+1/2} - F^+(U_i) + F^-_{i+1/2} - F^-(U_i) \\
&= \mathcal{R}^L(F^+(U_{i-k}), \dots, F^+(U_{i+k})) - F^+(U_i) \\
&+ \mathcal{R}^R(F^-(U_{i+1-k}), \dots, F^-(U_{i+1+k})) - F^-(U_i) \\
&= \mathcal{R}^L(F^+(U_{i-k}) - F^+(U_i), \dots, F^+(U_{i+k}) - F^+(U_i)) \\
&+ \mathcal{R}^R(F^-(U_{i+1-k}) - F^-(U_i), \dots, F^-(U_{i+1+k}) - F^-(U_i)) \\
&= \mathcal{R}^L(D^+_{i,i-k}, \dots, D^+_{i,i+k}) + \mathcal{R}^R(D^-_{i,i+1-k}, \dots, D^-_{i,i+1+k}),
\end{aligned}$$

where the following notation has been used

$$D^\pm_{j,k} = F^\pm(U_k) - F^\pm(U_j), \quad \forall j,k.$$

Analogously

$$\begin{aligned}
\widehat{D}^+_{i-1/2} &= F(U_i) - F_{i-1/2} \\
&= \mathcal{R}^L(D^+_{i-1-k,i}, \dots, D^+_{i-1+k,i}) + \mathcal{R}^R(D^-_{i-k,i}, \dots, D^-_{i+k,i}).
\end{aligned}$$

Finally, the last ingredient required to extend the numerical methods to nonconservative systems is a family of paths. Let us consider, in principle, an arbitrary family $\Psi : [0,1]\times\Omega\times\Omega \to \Omega$. Using Ψ we have:

$$D^\pm_{j,k} = F^\pm(U_k) - F^\pm(U_j) = \int_0^1 JF^\pm\left(\Psi(s;U_j,U_k)\right)\partial_s\Psi(s;U_j,U_k)\,ds,$$

where $JF^\pm$ represent the Jacobian matrices of $F^\pm$. The natural extension of the numerical method (12) to the system (10) is then given by (19) with

$$\widehat{D}^-_{i+1/2} = \mathcal{R}^L(D^+_{i,i-k}, \dots, D^+_{i,i+k}) + \mathcal{R}^R(D^-_{i,i+1-k}, \dots, D^-_{i,i+1+k}), \tag{20}$$

$$\widehat{D}^+_{i-1/2} = \mathcal{R}^L(D^+_{i-1-k,i}, \dots, D^+_{i-1+k,i}) + \mathcal{R}^R(D^-_{i-k,i}, \dots, D^-_{i+k,i}). \tag{21}$$

and

$$D^\pm_{j,k} = \int_0^1 A^\pm(\Psi(s;U_j,U_k))\partial_s\Psi(s;U_j,U_k)\,ds, \tag{22}$$

where $A^\pm(\Psi(s;U_j,U_k)$ is a matrix-splitting to be adequately chosen. For instance, for system (4), since the two eigenvalues are positive, a natural choice is given by

$$A^+(\Psi(s;U_j,U_k) = A(\Psi(s;U_j,U_k), \quad A^-(\Psi(s;U_j,U_k) = 0. \tag{23}$$

Please note that, while for systems of conservation laws only two reconstructions per intercell are required, here 4 reconstructions are needed in $x_{i+1/2}$: two reconstructions to compute $\widehat{D}^+_{i+1/2}$ and two others to compute $\widehat{D}^-_{i+1/2}$.

Observe that, while in the case of a system of conservation laws, the resulting numerical method is independent of the chosen family of paths (since it is equivalent to (12)), for nonconservative systems the numerical method depends on the chosen family of paths. For instance, in the particular case of System (4), if the matrix-splitting is given by (23), the choice of the family of paths (5), whose corresponding jump condition is (6), leads to

$$D^+_{j,k} = \begin{bmatrix} \bar{u}_{j,k}(u_k - u_j) + \bar{u}_{j,k}(v_k - v_j) \\ \bar{v}_{j,k}(u_k - u_j) + \bar{v}_{j,k}(v_k - v_j) \end{bmatrix}, \quad D^-_{j,k} = 0, \tag{24}$$

with

$$\bar{u}_{j,k} = \frac{u_j + u_k}{2}, \quad \bar{v}_{j,k} = \frac{v_j + v_k}{2},$$

while the choice (7), whose corresponding jump condition is (8), leads to

$$D^+_{j,k} = \begin{bmatrix} \bar{u}_{j,k}(u_k - u_j) + u_k(v_k - v_j) \\ v_j(u_k - u_j) + \bar{v}_{j,k}(v_k - v_j) \end{bmatrix}, \quad D^-_{j,k} = 0. \tag{25}$$

Definitions (5) and (7) lead to different numerical results: both of them are convergent for smooth solutions but, as it will be seen in Section 2.3, only (24) leads to a high-order accurate method. On the other hand, the two methods are expected to give different results for discontinuous solutions, since they are formally consistent with the different jump conditions, (6) or (8), corresponding to the selected paths. According to [38] this formal consistency does not ensure that the limits of the numerical solutions satisfy the expected jump conditions. In fact, the methods introduced here can fail in capturing correctly the discontinuities, as any other standard finite-difference type method. Nevertheless, they can be combined with techniques like the ones recently developed in [40,41] to improve their convergence to the sought weak solutions.

While for (4) the fluctuations (22) can be easily computed, this may be more difficult in other cases where numerical quadrature can be used to compute the integrals. Nevertheless, an alternative form of the method can be given if a Roe linearization is available in which the path-integrals are replaced by matrix-vector products. Remember that a Roe linearization (see [25,26]) is a matrix-valued function $A_\Psi : \Omega\times\Omega \mapsto \mathbb{R}^N \times \mathbb{R}^N$ that satisfies the following properties:

1. For each $U, V \in \Omega$, $A_\Psi(U, V)$ has N distinct real eigenvalues:
$$\lambda_1(U, V) < \lambda_2(U, V) < \cdots < \lambda_N(U, V).$$
2. $A_\Psi(U, U) = A(U)$, for every $U \in \Omega$.
3. For any $U, V \in \Omega$,
$$A_\Psi(U, V)(V - U) = \int_0^1 A(\Psi(s; U, V))\frac{\partial \Psi}{\partial s}(s; U, V)ds. \tag{26}$$

For instance, it can be easily checked that the matrices

$$A_{\Psi_1}(U^-, U^+) = \begin{bmatrix} \bar{u} & \bar{u} \\ \bar{v} & \bar{v} \end{bmatrix}, \quad A_{\Psi_2}(U^-, U^+) = \begin{bmatrix} \bar{u} & u^+ \\ v^- & \bar{v} \end{bmatrix}, \tag{27}$$

are Roe linearizations for system (4) related to the family of paths Ψ_1 and Ψ_2 given by (5) and (7) respectively. If a Roe linearization is available (as it is the case for the two-layer shallow-water system if the family of straight segments is selected) the fluctuations $D^\pm_{j,k}$ can be computed as in [16]:

$$D^\pm_{j,k} = A^\pm_\Psi(U_j, U_k)(U_k - U_j), \tag{28}$$

where

$$A^\pm_\Psi(U_j, U_k) = \frac{1}{2}\left(A_\Psi(U_j, U_k) \pm Q_\Psi(U_j, U_k)\right) \tag{29}$$

represents a splitting of the Roe linearization. Two different splittings will be considered here:

- Upwind splitting:
$$Q_\Psi(U_j, U_k) = \left|A_{j,k}\right|, \tag{30}$$
where
$$\left|A_{j,k}\right| = R_{j,k}\left|\Lambda_{j,k}\right|L_{j,k}.$$
Here, $\left|\Lambda_{j,k}\right|$ is the N-dimensional diagonal matrix whose coefficients are the absolute values of the eigenvalues of $A_{j,k}$:
$$|\lambda_{j,k;1}|, \ldots, |\lambda_{j,k;N}|;$$
$R_{j,k}$ is a matrix whose lth column $\vec{r}_{j,k;l}$ is an eigenvector associated to $\lambda_{j,k;l}$; and $L_{j,k} = R^{-1}_{j,k}$ is a matrix whose arrows are left-eigenvalues. A standard entropy-fix can be used to avoid the appearance of non-entropy discontinuities, like considering a regularization $|\cdot|_\epsilon$ of the absolute value function like in [42,43].
- Lax-Friedrichs(LF) splitting:
$$Q_\Psi(U_j, U_k) = \alpha I, \tag{31}$$
where I is the identity matrix and α is the global maximum of the absolute value of the eigenvalues, $\alpha \geq |\lambda_{j,k;l}|$, $l = 1, \ldots, N$.

The matrices involved in the Upwind splitting can be equivalently written as follows

$$A^\pm_\Psi(U, V) = P^\pm_\Psi(U, V)A_\Psi(U, V), \tag{32}$$

where

$$P^\pm_\Psi(U, V) = R_\Psi(U, V)M^\pm_\Psi(U, V)R^{-1}_\Psi(U, V). \tag{33}$$

Here $M^{\pm}_{\Psi}(U,V)$ represents the diagonal matrix whose coefficients are

$$\frac{1}{2}\big(1 \pm \operatorname{sign}\big(\lambda_l(U,V)\big)\big), \quad l = 1, \dots, N,$$

and $R_\Psi(U,V)$ is a matrix whose lth columns is an eigenvector $\vec{r}_l(U,V)$ associated to $\lambda_l(U,V)$. The fluctuations corresponding to this splitting can be then computed as follows: given two indices j,k, first the coordinates $\{\alpha_{j,k;l}\}_{l=1}^N$ of $U_k - U_j$ in the basis of eigenvectors of the Roe matrix $A_{j,k}$, i.e.

$$U_k - U_j = \sum_{l=1}^{N} \alpha_{j,k;l} \vec{r}_{j,k;l},$$

are computed by solving a linear system

$$R_{j,k}\vec{\alpha}_{j,k} = U_k - U_j. \tag{34}$$

Then one has

$$D^{\pm}_{j,k} = \sum_{l=1}^{N} \alpha_{j,k;l} \lambda^{\pm}_{j,k;l} \vec{r}_{j,k;l},$$

where, given $\lambda \in \mathbb{R}$, $\lambda^{\pm}$ represent the positive and negative part of λ, i.e.

$$\lambda^+ = \frac{\lambda + |\lambda|}{2}, \quad \lambda^- = \frac{\lambda - |\lambda|}{2}.$$

On the other hand, the method based on the LF splitting may be oscillatory if the reconstructions are not performed in characteristic fields. To avoid this, the reconstructions are computed in practice as follows:

$$\begin{aligned}\widehat{D}^-_{i+1/2} &= R_{i,i+1}\mathcal{R}^L(L_{i,i+1}D^+_{i,i-k}, \dots, L_{i,i+1}D^+_{i,i+k}) \\ &\quad + R_{i,i+1}\mathcal{R}^R(L_{i,i+1}D^-_{i,i+1-k}, \dots, L_{i,i+1}D^-_{i,i+1+k}),\end{aligned} \tag{35}$$

$$\begin{aligned}\widehat{D}^+_{i-1/2} &= R_{i-1,i}\mathcal{R}^L(L_{i-1,i}D^+_{i-1-k,i}, \dots, L_{i-1,i}D^+_{i-1+k,i}) \\ &\quad + R_{i-1,i}\mathcal{R}^R(L_{i-1,i}D^-_{i-k,i}, \dots, L_{i-1,i}D^-_{i+k,i}).\end{aligned} \tag{36}$$

Observe that, if the LF splitting is chosen and reconstructions in characteristic variables are performed, the right and left eigenvectors of the Roe matrices $A_{i,i+1}$ have to be computed. On the other hand, if the Upwind reconstruction is selected, the eigenvectors and eigenvalues of all the Roe matrices $A_{k,j}$ are required. Moreover, the linear system (34) has to be solved. Therefore, it is more computationally expensive for homogeneous problems. Nevertheless, it will be seen in Section 3 that the numerical treatment of the source term can compensate for this disadvantage.

Remark 1. Since the expression of WENO reconstructions is a linear combination of the fluxes whose coefficients depend nonlinearly on the data through the smoothness indicators, it can be shown that, for problems of the form (11), the numerical method (12) can be written in the form

$$\frac{dU_i}{dt} = -\frac{1}{\Delta x}\Big(F_{i+1/2} - F_{i-1/2} + \widehat{B}^-_{i+1/2} + \widehat{B}^+_{i-1/2}\Big), \tag{37}$$

where $F_{i+1/2}$ and $\widehat{B}_{i+1/2}$ are, respectively, standard WENO reconstructions of the flux function and the nonconservative terms, in which the nonlinear coefficients are the same.

2.3. Accuracy of the methods

Let us check that (19)–(22) is a high-order numerical method for (10).

Proposition 1. *Let us consider a smooth solution $U(x,t)$ of (10) and assume that $A(U)$ and Ψ are smooth. We also assume that Ψ satisfies*

$$\int_0^1 A(\Psi(s;V,U))\partial_s\Psi(s;V,U)\,ds = -\int_0^1 A(\Psi(s;U,V))\partial_s\Psi(s;U,V)\,ds \tag{38}$$

for all $U, V \in \Omega$. Then we have

$$\partial_t U(x_i,t) + \frac{1}{\Delta x}\Big(\widehat{D}^-_{i+1/2} + \widehat{D}^+_{i-1/2}\Big) = O(\Delta x^{2k+1}), \tag{39}$$

where $p = 2k+1$ is the order of the reconstruction operator.

Proof. Given an index i and a time t, let us define the function $G_i^t(x)$ as follows:

$$G_i^t(x) = \int_0^1 A(\Psi(s;U(x_i,t),U(x,t))\partial_s\Psi(s;U(x_i,t),U(x,t))\,ds. \tag{40}$$

This function satisfies

$$G_i^t(x_i) = 0, \quad \partial_x G_i^t(x_i) = A(U(x_i,t))U_x(x_i,t).$$

In effect,

$$G_i^t(x_i) = \int_0^1 A(\Psi(s;U(x_i,t),U(x_i,t))\partial_s\Psi(s;U(x_i,t),U(x_i,t))\,ds = 0,$$

since

$$\Psi(s;U(x_i,t),U(x_i,t)) = U(x_i,t), \quad \forall s.$$

On the other hand:

$$\begin{aligned}
\partial_x G_i^t(x_i) &= \lim_{h\to 0}\frac{G_i^t(x_i+h)-G_i^t(x_i)}{h}\\
&= \lim_{h\to 0}\frac{1}{h}\int_0^1 A(\Psi(s;U(x_i,t),U(x_i+h,t)))\partial_s\Psi(s;U(x_i,t),U(x_i+h,t))\,ds\\
&= \lim_{h\to 0}\int_0^1 \frac{1}{h}\big(A(\Psi(s;U(x_i,t),U(x_i+h,t))) - A(U(x_i,t))\big)\partial_s\Psi(s;U(x_i,t),U(x_i+h,t))\,ds\\
&\quad + \lim_{h\to 0} A(U(x_i,t))\frac{1}{h}\int_0^1 \partial_s\Psi(s;U(x_i,t),U(x_i+h,t))\,ds\\
&= \lim_{h\to 0}\int_0^1 \frac{1}{h}\big(A(\Psi(s;U(x_i,t),U(x_i+h,t))) - A(U(x_i,t))\big)\partial_s\Psi(s;U(x_i,t),U(x_i+h,t))\,ds\\
&\quad + \lim_{h\to 0} A(U(x_i,t))\frac{U(x_i+h,t)-U(x_i,t)}{h}\\
&= A(U(x_i,t))U_x(x_i,t),
\end{aligned}$$

where, in the first term, it has been used again that

$$\partial_s\Psi(s;U(x_i,t),U(x_i+h,t)) \to \partial_s\Psi(s;U(x_i,t),U(x_i,t)) = 0.$$

Observe that, for all j:

$$D_{i,j} = G_i^t(x_j), \quad D_{j,i} = -G_i^t(U)(x_j).$$

Therefore

$$\begin{aligned}
\widehat{D}^-_{i+1/2} &= \mathcal{R}^L(D^+_{i,i-k},\dots,D^+_{i,i+k}) + \mathcal{R}^R(D^-_{i,i+1-k},\dots,D^-_{i,i+1+k})\\
&= \mathcal{R}^L(G_i^{t,+}(x_{i-k}),\dots,G_i^{t,+}(x_{i+k})) + \mathcal{R}^R(G_i^{t,-}(x_{i+1-k}),\dots,G_i^{t,-}(x_{i+1+k})) = \hat{G}^t_{i,i+1/2},\\
\widehat{D}^+_{i-1/2} &= \mathcal{R}^L(D^+_{i-1-k,i},\dots,D^+_{i-1+k,i}) + \mathcal{R}^R(D^-_{i-k,i},\dots,D^-_{i+k,i})\\
&= -\mathcal{R}^L(G_i^{t,+}(x_{i-1-k}),\dots,G_i^{t,+}(x_{i-1+k})) - \mathcal{R}^R(G_i^{t,-}(x_{i-k}),\dots,G_i^{t,-}(x_{i+k})) = -\hat{G}_{i,i-1/2},
\end{aligned}$$

where $G_i^{t,\pm}$ represents the splitting of the function G_i^t and $\hat{G}^t_{i,i\pm 1/2}$ is its WENO reconstruction. Therefore we have:

$$\begin{aligned}
\frac{1}{\Delta x}(\widehat{D}^+_{i-1/2} + \widehat{D}^-_{i+1/2}) &= \frac{1}{\Delta x}(\hat{G}_{i,i+1/2} - \hat{G}_{i,i-1/2})\\
&= \partial_x G_i^t(x_i) + O(\Delta x^{2k+1})\\
&= A(U(x_i,t))U_x(x_i,t) + O(\Delta x^{2k+1}),
\end{aligned}$$

which leads to (39). □

The symmetry condition (38) is satisfied by the family of straight segments

$$\Psi(s;U,V) = U + s(V-U).$$

In effect

$$\begin{aligned}
\int_0^1 A(\Psi(s;V,U))\partial_s\Psi(s;V,U)\,ds &= \left(\int_0^1 A(V+s(U-V))\,ds\right)(U-V)\\
&= -\left(\int_0^1 A(U+s(V-U))\,ds\right)(V-U)\\
&= -\int_0^1 A(\Psi(s;U,V))\partial_s\Psi(s;U,V)\,ds.
\end{aligned}$$

Therefore, in the particular case of system (4), the definition (24), based on the choice of straight segments, leads to a high-order method. On the other hand, (38) is not satisfied for (7) and, it will be seen in Section 6.1 that the method corresponding to (25) is only first-order accurate, what shows that this condition is necessary as well.

According to the proof of Proposition 1, the numerical method can be interpreted as follows: the PDE system is first formally rewritten as the system of balance laws

$$\partial_t U_i + \partial_x \mathcal{F}_i^L = 0, \tag{41}$$

with

$$\mathcal{F}_i^L(x,t) = G_i^t(x),$$

where G_i is the function given by (40); then WENO reconstructions are applied to the generalized flux function $\mathcal{F}_i^L$. In the particular case of a system of the form (11) it can be easily checked that this is equivalent to reconstructing the generalized flux function

$$\mathcal{F}^L = F + \mathcal{B}_i^L,$$

where

$$\mathcal{B}_i^L(x,t) = \int_0^1 B(\Psi(s;U(x_i,t),U(x,t))\partial_s\Psi(s;U(x_i,t),U(x,t))\,ds,$$

which is defined for every i, while in the global-flux approach the generalized flux function to be reconstructed is

$$\mathcal{F}^G = F + \mathcal{B}^G,$$

where

$$\mathcal{B}^G(x,t) = \int_a^x B(U)U_x\,dx$$

is globally defined. This is why it was said above that a *local flux* approach is followed here. Following this approach, $\mathcal{B}_i^L$ is approximated at the node points of the stencil as follows

$$\mathcal{B}_i^L(x_j,t) \approx B_\Psi(U_i,U_j)(U_j - U_i), \quad j \in S_i,$$

where B_Ψ is the linearization of B used in the Roe matrix, While in the case of the global flux approach (similar to the approach in the finite volume method as [44]), $\mathcal{B}^G$ is numerically approximated at the nodes using a recursive formula such that

$$\mathcal{B}_0^G = 0; \quad \mathcal{B}_{i+1}^G = \mathcal{B}_i^G + \Delta x \sum_{l=0}^{M} \alpha_l B(U_i^l) D_x U_i^l, \quad i = 0,\dots,NP-1,$$

where α^l, $l = 0,\dots,M$ are the weights of the selected quadrature form and U_i^l, $D_x U_i^l$, $l = 0,\dots,M$ are high-order approximations of U and U_x at the quadrature points, NP is the total number of discrete points. Therefore $M+1$ additional reconstructions (of state in this case) are necessary. Summing up, while in the local flux approach two flux WENO reconstructions per point are needed (to compute $\widehat{D}^+_{i-1/2}$ and $\widehat{D}^-_{i+1/2}$), in the global flux approach one flux WENO reconstruction per intercell and $M+1$ state reconstructions per cell are necessary. It means that the local flux approach requires $2NP$ flux reconstructions per stage of the ODE solver used for the temporal discretization while the global flux approach requires $(M+2)NP$. On the other hand, if the order of the WENO reconstruction is $2k+1$, then M has to be greater or equal than k so that the numerical method preserves the order of the WENO reconstructions. Therefore, the number of reconstructions in the global flux approach is greater than $(k+2)NP$ compared to the $2NP$ reconstructions in the local flux approach.

3. Problems with source terms and well-balanced property

3.1. Well-balanced property

Let us first consider a system of the form (10) in which $\lambda = 0$ is an eigenvalue of $A(U)$ for every $U \in \Omega$. The well-balanced property of the methods is related to the preservation of the stationary solutions U^* of the system, which satisfy the equation

$$A(U^*)U_x^* = 0.$$

Observe that, U_x^* is an eigenvector associated with the null eigenvalue for all x such that $U^*(x)_x \neq 0$. As an example, it can be easily checked that the stationary solutions of (4) are the set of functions

$$U^*(x) = \begin{bmatrix} u^*(x) \\ v^*(x) \end{bmatrix} \text{ s.t. } u^*(x) + v^*(x) = constant. \tag{42}$$

The method described in Section 2.2 has then the well-balanced property given by the following results the proof of which is trivial:

Proposition 2. *Let $U^*(x)$ be a stationary solution of system (10). If, for every $x_L < x_R$ one has*

$$A_\Psi(U^*(x_L), U^*(x_R))(U^*(x_R) - U^*(x_L)) = 0 \tag{43}$$

and the selected matrix-splitting is such that

$$A_\Psi(U, V)(V - U) = 0 \implies A^\pm_\Psi(U, V)(V - U) = 0, \tag{44}$$

then the numerical method (19)–(28) is well-balanced for U^*, i.e. $\{U^*(x_i)\}$ is an equilibrium of the ODE system (19).

Observe that (44) is satisfied for the Upwind splitting approach, as can be easily deduced from (32), but not for the LF splitting: in effect, in this case one has

$$A_\Psi(U, V)(V - U) = 0 \implies A^\pm_\Psi(U, V)(V - U) = \pm\frac{\alpha}{2}(V - U).$$

Nevertheless, the modification of the identity matrix technique introduced in [16] can be applied to modify the splitting so that (43) is satisfied.

As an application of Proposition 2, it can be easily checked that the property (43) is satisfied for the Roe matrix A_{Ψ_1} defined in (27) for every stationary solution (42) of System (4). Therefore, the choices of the family of straight segments and the Upwind splitting lead to a numerical method for (4) that is fully well-balanced (and high-order accurate). On the other hand, (43) is not satisfied for the Roe matrix A_{Ψ_2}.

Property (43) is discussed in [45] in relation with the well-balanced property of Roe methods. Let us only recall that, given a stationary solution U^*, this property is satisfied if the family of paths is such that, for all $x_L, x_R \in \mathbb{R}$ with $x_L < x_R$, the functions

$$s \in [0, 1] \to \Psi\big(s; U^*(x_L), U^*(x_R)\big) \in \Omega$$

and

$$x \in [x_L, x_R] \to U^*(x) \in \Omega$$

define the same curve in Ω. In effect, if this is the case one has:

$$\begin{aligned} A_\Psi(U, V)(V - U) &= \int_0^1 A(\Psi(s; U, V))\partial_s\Psi(s; U, V)\, ds \\ &= \int_{x_L}^{x_R} A(U^*(x))U^*_x(x)\, dx \\ &= 0, \end{aligned}$$

where a change of parameter has been applied to obtain the second equality and the fact that U^* is a stationary solution has been used in the third one. In particular, if the family of straight segments is chosen, the property (43) is satisfied for all stationary solutions such that the curve defined by $x \to U^*(x)$ lies in a straight line: this is the case of (4) whose stationary solutions (42) lie in a straight line of equation $u + v = constant$ in the u, v plane.

3.2. Source terms

Let us consider now problems with source term

$$U_t + A(U)U_x = S(U)H_x, \tag{45}$$

where $H(x)$ is a known function, whose stationary solutions satisfy

$$A(U^*)U^*_x = S(U^*)H_x. \tag{46}$$

3.2.1. Strategy 1

The first strategy consists in writing (45) in the form (10) as follows (see [45,46]):

$$W_t + \mathcal{A}(W)W_x = 0, \tag{47}$$

with

$$W = \begin{bmatrix} U \\ H \end{bmatrix} \in \Omega \times \mathbb{R}, \quad \mathcal{A}(W) = \left[\begin{array}{c|c} A(U) & -S(U) \\ 0 & 0 \end{array}\right],$$

and then the strategy described in Section 2.2 is applied to (47). To do this, a family of paths

$$\widetilde{\Psi}(s; W_L, W_R) = \begin{bmatrix} \Psi_U(s; W_L, W_R) \\ \Psi_H(s; W_L, W_R) \end{bmatrix}$$

satisfying (38) (as, for instance, the family of straight segments) and a Roe linearization have to be chosen first. As in [45], let us assume that a Roe matrix of the form

$$\mathcal{A}_{\widetilde{\Psi}}(W_L, W_R) = \left[\begin{array}{c|c} A_{\widetilde{\Psi}}(W_L, W_R) & -S_{\widetilde{\Psi}}(W_L, W_R) \\ 0 & 0 \end{array}\right]$$

is available, where

- $A_{\widetilde{\Psi}}(W_L, W_R)$ has N real different eigenvalues $\lambda_i(W_L, W_R)$, $i = 1, \dots, N$;
- $A_{\widetilde{\Psi}}(W, W) = A(W)$ for all $W = [U, H]^T$;
- $S_{\widetilde{\Psi}}(W, W) = S(W)$ for all $W = [U, H]^T$;
- for all $W_L, W_R \in \Omega \times \mathbb{R}$

$$A_{\widetilde{\Psi}}(W_L, W_R)(U_R - U_L) = \int_0^1 A(\Psi_U(s; W_L, W_R))\partial_s \Psi_U(s; W_L, W_R)\, ds;$$

$$S_{\widetilde{\Psi}}(W_L, W_R)(H_R - H_L) = \int_0^1 S(\Psi_U(s; W_L, W_R))\partial_s \Psi_H(s; W_L, W_R)\, ds.$$

In this paragraph, the Upwind splitting is considered. Some algebraic computations show that the corresponding splitting is given by the matrices

$$\mathcal{A}^{\pm}_{\widetilde{\Psi}}(W_L, W_R) = \left[\begin{array}{c|c} P^{\pm}_{\widetilde{\Psi}}(W_L, W_R)A_{\widetilde{\Psi}}(W_L, W_R) & -P^{\pm}_{\widetilde{\Psi}}(W_L, W_R)S_{\widetilde{\Psi}}(W_L, W_R) \\ 0 & 0 \end{array}\right]$$

where $P^{\pm}_{\widetilde{\Psi}}$ are the projection matrices defined as in (33).

If the trivial equation for the artificial unknown H is removed, the numerical method can be written again as (19)-(20)-(21) where now

$$D^{\pm}_{j,k} = P^{\pm}_{j,k}\big(A_{j,k}(U_k - U_j) - S_{j,k}(H(x_k) - H(x_j))\big), \tag{48}$$

with

$$P^{\pm}_{j,k} = P^{\pm}_{\widetilde{\Psi}}(W_j, W_k), \quad A_{j,k} = A_{\widetilde{\Psi}}(W_j, W_k), \quad S_{j,k} = S_{\widetilde{\Psi}}(W_j, W_k).$$

Accordingly, the fluctuations can be computed as follows: given two indices j, k, first the coordinates $\{\alpha_{j,k;l}\}_{l=1}^N$ of $U_k - U_j - A^{-1}_{j,k}S_{j,k}(H(x_k) - H(x_j))$ in the basis of eigenvectors of the Roe matrix $A_{j,k}$ are computed:

$$U_k - U_j - A^{-1}_{j,k}S_{j,k}(H(x_k) - H(x_j)) = \sum_{l=1}^N \alpha_{j,k;l}\vec{r}_{j,k;l}. \tag{49}$$

Then, one has:

$$A_{j,k}(U_k - U_j) - S_{j,k}(H(x_k) - H(x_j)) = \sum_{l=1}^N \alpha_{j,k;l}\lambda_{j,k;l}\vec{r}_{j,k;l},$$

and then

$$D^{\pm}_{j,k} = \sum_{l=1}^N \alpha_{j,k;l}\lambda^{\pm}_{j,k;l}\vec{r}_{j,k;l}.$$

The reconstruction is then performed as in the case of homogeneous problems.

Proposition 2 can be then applied to this particular case to show that, given $H(x)$, the numerical method is well-balanced for a stationary solution U^* provided that (43) holds, i.e. if

$$A_{\widetilde{\Psi}}(U^*(x_L), U^*(x_R))(U^*(x_R) - U^*(x_L)) = S_{\widetilde{\Psi}}(U^*(x_L), U^*(x_R))(H(x_R) - H(x_L)) \tag{50}$$

for all $x_L < x_R$. This is the case if

$$s \in [0, 1] \to \widetilde{\Psi}\left(s; \begin{bmatrix} U^*(x_L) \\ H(x_L) \end{bmatrix}; \begin{bmatrix} U^*(x_R) \\ H(x_R) \end{bmatrix}\right) \tag{51}$$

and

$$x \in [x_L, x_R] \to \begin{bmatrix} U^*(x) \\ H(x) \end{bmatrix} \tag{52}$$

define the same curve for all $x_L < x_R$. In particular, if the family of straight segments is chosen, then the numerical method is well-balanced for every stationary solution such that (52) lies in a straight line for every $x_L < x_R$. This property will be used in Section 5 to define numerical methods that preserve water-at-rest solutions for the two-layer shallow-water system. More sophisticated families of paths could be considered to preserve more general stationary solutions, as the ones based on the Generalized Hydrostatic Reconstruction introduced in [29] which will be done in a forthcoming paper.

The numerical treatment of the source term in this strategy can be interpreted as it was done in Section 2.2 for the nonconservative products $B(U)U_x$: the source term is first written as the derivative of a new flux function

$$S(U)H_x = \partial_x S_i^L$$

with

$$S_i^L(x, t) = \int_0^1 S(\Psi_U(s; W(x_i, t), W(x, t))\partial_s \Psi_H(s; W(x_i, t), W(x, t))\, ds,$$

while in the global-flux approach, it is rewritten as

$$S(U)H_x = \partial_x S^G$$

with

$$S^G(x,t) = \int_{x_0}^{x} S(U)H_x\, dx.$$

Again, S_i^L is approximated at the node points of the stencil as follows

$$S_i^L(x_j,t) \approx S_{\widetilde{\Psi}}(W_i, W_j)(H(x_j) - H(x_i)), \quad j \in S_i,$$

and no integrals at the cells have to be approximated thus avoiding the need to calculate new reconstructions at the quadrature points.

3.2.2. Strategy 2

Strategy 2 extends to the nonconservative system the technique proposed in [37] for systems of balance laws. Unlike Strategy 1 the well-balanced property of the methods based on this strategy will not depend on either the choice of the family of paths or the matrix splitting.

Let us consider first the numerical method for (45) given by

$$\frac{dU_i}{dt} + \frac{1}{\Delta x}\left(\widehat{D}^-_{i+1/2} + \widehat{D}^+_{i-1/2}\right) = S(U_i)H_x(x_i), \tag{53}$$

where the fluctuations $\widehat{D}^+_{i-1/2}$ are defined by (28) with any choice of family of paths, Roe matrix and matrix-splitting. Under the hypothesis of Proposition 2, this method is highly accurate but in principle does not preserve any stationary solution. Let us modify the method so that a given m-parameter family of stationary solutions

$$U^*(x; c_1, \dots, c_m), \tag{54}$$

with $m \le N$, is preserved. To do this, let us assume that there exists a $m \times N$ matrix C with rank m such that, given any point $\bar{x}$ and any state $\bar{U}$, there exists a unique stationary solution of the family satisfying

$$CU^*(\bar{x}) = C\bar{U}, \tag{55}$$

i.e., this system of equations determines the value of the m parameters. The idea is then to rewrite the equation at x_i equivalently as follows:

$$\frac{dU_i}{dt} + A(U(x_i,t))U_x(x_i,t) - A(U_i^*)U^*_{i,x} = (S(U(x_i,t)) - S(U_i^*(x_i,t)))H_x(x_i),$$

where U_i^* is the unique stationary solution of the m-parameter family satisfying

$$CU_i^*(x_i) = CU(x_i,t). \tag{56}$$

The idea is then to discretize $A(U)U_x$ and $A(U_i^*)U^*_{i,x}$ together by applying the strategy introduced in Section 2.2 what leads to the numerical method:

$$\frac{dU_i}{dt} + \frac{1}{\Delta x}\left(\widehat{D}^-_{i+1/2} + \widehat{D}^+_{i-1/2}\right) = (S(U_i) - S(U_i^*(x_i)))H_x(x_i), \tag{57}$$

with

$$\begin{aligned}\widehat{D}^-_{i+1/2} &= \mathcal{R}^L(D^+_{i,i-k} - D^{*,+}_{i,i-k}, \dots, D^+_{i,i+k} - D^{*,+}_{i,i+k}) \\ &\quad + \mathcal{R}^R(D^-_{i,i+1-k} - D^{*,-}_{i,i+1-k}, \dots, D^-_{i,i+1+k} - D^{*,-}_{i,i+1+k}),\end{aligned} \tag{58}$$

$$\begin{aligned}\widehat{D}^+_{i-1/2} &= \mathcal{R}^L(D^+_{i-1-k,i} - D^{*,+}_{i-1-k,i}, \dots, D^+_{i-1+k,i} - D^{*,+}_{i-1+k,i}) \\ &\quad + \mathcal{R}^R(D^-_{i-k,i} - D^{*,-}_{i-k,i}, \dots, D^-_{i+k,i} - D^{*,-}_{i+k,i}),\end{aligned} \tag{59}$$

where the starred fluctuations are given by

$$D^{*,\pm}_{j,k} = A^{\pm}_{\Psi}(U_i^*(x_j), U_i^*(x_k))(U_i^*(x_k) - U_i^*(x_j)), \tag{60}$$

The following result then holds.

Proposition 3. *The numerical method (57)–(59) is well-balanced for all the stationary solutions of the family (54), i.e. $\{U^*(x_i; c_1, \dots, c_m)\}$ is an equilibrium of the ODE system (19) for every stationary solution U^*.*

The proof is straightforward. As an application, let us consider the family of stationary solutions

$$U^*(x;c) = \begin{bmatrix} u^*(x;c) \\ v^*(x;c) \end{bmatrix} = \begin{bmatrix} \frac{c}{2} + \sin(x) \\ \frac{c}{2} - \sin(x) \end{bmatrix}, \quad c \in \mathbb{R}$$

of System (4). Given $\bar{x}$ and $\bar{U} = [\bar{u}, \bar{v}]^T$ the equation

$$u^*(x; c) + v^*(x; c) = \bar{u} + \bar{v}$$

determines the value of the parameter

$$c = \bar{u} + \bar{v},$$

i.e. $m = 1$ and $C = [1, 1]$ in this case. Therefore, if the family of straight segments and the Roe matrix A_{Ψ_1} in (27) are chosen, the numerical method (57) with

$$U_i^*(x) = \begin{bmatrix} \frac{u_i+v_i}{2} + \sin(x) \\ \frac{u_i+v_i}{2} - \sin(x) \end{bmatrix}$$

is high-order accurate and well-balanced for the given family of stationary solution. This technique will be used in next section to design again a numerical method that preserves water-at-rest solutions for the two-layer shallow-water system that is based on the LF splitting.

For some particular problems, this strategy can be extended to design fully well-balanced methods, i.e. methods that preserve all the stationary solutions. In effect, let us suppose that, given $\bar{x}$ and $\bar{U}$, there is only one stationary solution $U^*(x)$ such that

$$U^*(\bar{x}) = \bar{U}$$

(i.e. $m = N$ and C is the identity matrix) or, there are several but it is possible to use a criterion to select one of them (like the flow regime, for instance). Then, the strategy can be applied by taking U_i^* as the unique or the selected stationary solution such that

$$U_i^*(x_i) = U(x_i, t).$$

In this case, the numerical method reduces to (19), with the fluctuations given by (58)–(59).

In particular, this technique allows the design of fully well-balanced numerical methods for the shallow-water system: see [37]. In fact, the numerical methods introduced here based on Strategy 2 reduce to the ones in this reference when they are applied to systems of balance laws, as the shallow-water system. In the reference, it has been shown that theses methods deal correctly with discontinuous bottom functions H: the fully-well balanced methods are able to correctly capture the stationary contact discontinuities standing on the points of discontinuity of H: see [37] for details. This technique can be extended to derive numerical methods that are fully well-balanced for the 1D two-layer shallow-water systems but the computation of moving stationary solution is more involved: this will be the object of future work.

3.3. Implementation

We summarize here the implementation of two well-balanced methods: Method 1 in which Strategy 1 is combined with the Upwind splitting, and Method 2 corresponding to Strategy 2, LF splitting, and reconstruction in characteristic variables. Let us assume that approximations U_j to the solution are available at the cell points; the fluctuations and the source terms at a point x_i are then computed as follows.

1. Compute the Roe matrices $A_{i,j} = A_\Psi(U_i, U_j)$, $j = i-k, \dots, i+k$ and their eigenvalues $\{\lambda_{i,j;l}\}_{l=1}^N$.
2. Compute the fluctuations at the stencil points as follows:
   (a) Method 1:
   Compute the eigenvector matrices $R_{i,j}$, $j = i-k, \dots, i+k$.
   Solve the linear systems

   $$R_{i,j}\vec{\alpha} = U_j - U_i - A_{i,j}^{-1} S_{i,j}(H(x_j) - H(x_i))$$

   to obtain $\{\alpha_{i,j;l}\}_{l=1}^N$ such that

   $$U_j - U_i - A_{i,j}^{-1} S_{i,j}(H(x_j) - H(x_i)) = \sum_{l=1}^N \alpha_{i,j;l}\vec{r}_{i,j;l},$$

   Compute then

   $$D_{i,j}^\pm = \sum_{i=1}^N \alpha_{i,j;l}\lambda_{i,j;l}^\pm \vec{r}_{i,j;l}.$$

   (b) Method 2:
   Compute first the stationary solution U_i^* of the family (54) satisfying

   $$CU_i^*(x_i) = CU_i$$

   and evaluate it at x_j, $j = i-k, \dots, i+k$.
   Compute the Roe matrices $A_{i,j}^* = A_\Psi(U_i^*(x_i), U_i^*(x_j))$.

Compute then

$$D^{\pm}_{i,j} = \frac{1}{2}\big(A_{i,j}\big(U_j - U_i\big) \pm \alpha\big(U_j - U_i\big)\big),$$
$$D^{*,\pm}_{i,j} = \frac{1}{2}\Big(A^*_{i,j}\big(U^*_i(x_j) - U^*_i(x_i)\big) \pm \alpha\big(U^*_i(x_j) - U^*_i(x_i)\big)\Big),$$

where α is chosen so that $\alpha > |\lambda_{i,j;l}|$ for all i, j, l.

3. Compute the WENO reconstructions of the fluctuations at the cell interfaces as follows:

(a) Method 1:

$$\widehat{D}^-_{i+1/2} = \mathcal{R}^L(D^+_{i,i-k}, \dots, D^+_{i,i+k}) + \mathcal{R}^R(D^-_{i,i+1-k}, \dots, D^-_{i,i+1+k}),$$
$$\widehat{D}^+_{i-1/2} = \mathcal{R}^L(D^+_{i-1-k,i}, \dots, D^+_{i-1+k,i}) + \mathcal{R}^R(D^-_{i-k,i}, \dots, D^-_{i+k,i}).$$

(b) Method 2:

Compute the matrices of right and left eigenvectors, $R_{i-1,i}$, $L_{i-1,i}$, (resp. $R_{i,i+1}$, $L_{i,i+1}$) of $A_{i-1,i}$ (resp. $A_{i,i+1}$)
Then compute:

$$\begin{aligned}\widehat{D}^-_{i+1/2} = {} & R_{i,i+1}\mathcal{R}^L(L_{i,i+1}\Big(D^+_{i,i-k} - D^{*,+}_{i,i-k}\Big), \dots, L_{i,i+1}\Big(D^+_{i,i+k} - D^{*,+}_{i,i+k}\Big)) \\ & + R_{i,i+1}\mathcal{R}^R(L_{i,i+1}\Big(D^-_{i,i+1-k} - D^{*,-}_{i,i+1-k}\Big), \dots, L_{i,i+1}\Big(D^-_{i,i+1+k} - D^{*,-}_{i,i+1+k}\Big)), \\ \widehat{D}^+_{i-1/2} = {} & R_{i-1,i}\mathcal{R}^L(L_{i-1,i}\Big(D^+_{i-1-k,i} - D^{*,+}_{i-1-k,i}\Big), \dots, L_{i-1,i}\Big(D^+_{i-1+k,i} - D^{*,+}_{i-1+k,i}\Big)) \\ & + R_{i-1,i}\mathcal{R}^R(L_{i-1,i}\Big(D^-_{i-k,i} - D^{*,-}_{i-k,i}\Big), \dots, L_{i-1,i}\Big(D^-_{i+k,i} - D^{*,-}_{i+k,i}\Big)).\end{aligned}$$

Compute finally the source term:

$$\big(S(U_i) - S(U^*_i(x_i))\big)H_x(x_i).$$

Remark 2. Although in Section 2.2 it was said that the numerical methods based on the Upwind splitting are more computationally expensive than those based on the LF splitting, observe that the numerical treatment of the source term in Method 2 requires the computation of the fluctuations corresponding to the chosen stationary solutions so that, as it will be seen in Section 6, the computational costs of both methods are comparable.

4. Extension to 2D systems

4.1. Homogeneous problems

This section extends the 1D path-conservative fifth-order WENO scheme to 2D nonconservative systems of the form

$$U_t + A_1(U)U_x + A_2(U)U_y = 0. \tag{61}$$

The system is supposed again to be strictly hyperbolic, i.e. for all U and all $\theta \in [0, 2\pi)$, the matrix

$$\cos(\theta)A_1(U) + \sin(\theta)A_2(U)$$

has N different real eigenvalues. Systems with flux terms and nonconservative products

$$U_t + F(U)_x + G(U)_y + C(U)U_x + D(U)U_y = 0, \tag{62}$$

can be considered as particular cases in which $A_1(U) = J(F(U)) + C(U), A_2(U) = J(G(U)) + D(U)$.

Let us assume again that Roe linearizations $A_{i,\Psi}(U, V)$ of $A_i(U)$, $i = 1, 2$ are available for the selected family of paths Ψ. We consider uniform Cartesian meshes with points (x_i, y_j) with steps Δx and Δy in the x and y direction. WENO methods can be extended dimension by dimension:

$$U'_{i,j} = -\frac{1}{\Delta x}\Big(\widehat{D}^-_{i+1/2,j} + \widehat{D}^+_{i-1/2,j}\Big) - \frac{1}{\Delta y}\Big(\widehat{D}^-_{i,j+1/2} + \widehat{D}^+_{i,j-1/2}\Big), \tag{63}$$

where $U_{i,j} \approx U(x_i, y_j)$ and

$$\widehat{D}^-_{i+1/2,j} = \mathcal{R}^L(D^+_{i,i-k;j}, \dots, D^+_{i,i+k;j}) + \mathcal{R}^R(D^-_{i,i+1-k;j}, \dots, D^-_{i,i+1+k;j}), \tag{64}$$
$$\widehat{D}^+_{i-1/2,j} = \mathcal{R}^L(D^+_{i-1-k,i;j}, \dots, D^+_{i-1+k,i;j}) + \mathcal{R}^R(D^-_{i-k,i;j}, \dots, D^-_{i+k,i;j}), \tag{65}$$
$$\widehat{D}^-_{i,j+1/2} = \mathcal{R}^L(D^+_{i;j,j-k}, \dots, D^+_{i;j,j+k}) + \mathcal{R}^R(D^-_{i;j,j+1-k}, \dots, D^-_{i;j,j+1+k}), \tag{66}$$
$$\widehat{D}^+_{i,j-1/2} = \mathcal{R}^L(D^+_{i;j-1-k,j}, \dots, D^+_{i;j-1+k,j}) + \mathcal{R}^R(D^-_{i;j-k,j}, \dots, D^-_{i;j+k,j}). \tag{67}$$

Here, the following notation has been used:

$$D^{\pm}_{i,l;j} = A^{\pm}_{1;i,l;j}(U_{l,j} - U_{i,j}), \quad l = i-k, \dots, i+k, \tag{68}$$
$$D^{\pm}_{i;j,l} = A^{\pm}_{2;i;j,l}(U_{i,l} - U_{i,j}), \quad l = j-k, \dots, j+k, \tag{69}$$

where

$$A_{1;i,l;j} = A_{1,\Psi}(U_{i,j}, U_{l,j}), \quad A_{2;i;j,l} = A_{2,\Psi}(U_{i,j}, U_{i,l})$$

and the super-indices $\pm$ represent their splitting matrices: both the Upwind and the LF splittings can be readily extended to 2D.

Proposition 1 can be easily extended to prove that (63) is a numerical method of order $p = 2k + 1$ provided that Property (38) is satisfied for A_1 and A_2.

4.2. Problems with source terms

We now consider systems of the form

$$U_t + A_1(U)U_x + A_2(U)U_y = S_1(U)H_x + S_2(U)H_y, \tag{70}$$

where $H(x, y)$ is again a known function. Strategy 1 described in Section 3 can be readily extended to problem (70): it is enough to redefine the fluctuations as follows:

$$D^{\pm}_{i,l;j} = A^{\pm}_{1;i,l;j}(U_{l,j} - U_{i,j}) - S^{\pm}_{1;i,l;j}(H(x_l, y_j) - H(x_i, y_j)), \tag{71}$$

$$D^{\pm}_{i;j,l} = A^{\pm}_{2;i;j,l}(U_{i,l} - U_{i,j}) - S^{\pm}_{2;i;j,l}(H(x_i, y_l) - H(x_i, y_j)), \tag{72}$$

where $A_{1;i,k;j}$, $S_{1;i,k;j}$, represent respectively the intermediate matrix and source term given by the Roe linearization for the states $U_{i,j}$ and $U_{l,j}$ and $A_{2;i;j,l}$, $S_{2;i;j,l}$ the corresponding to the states $U_{i,j}$, $U_{i,l}$.

It can be shown as in Proposition 2 that the numerical method is well-balanced for stationary solutions that satisfy that given $(x_L, y), (x_R, y), (x, y_D), (x, y_U) \in \mathbb{R}^2$ one has

$$\begin{aligned} &A_{1,\widetilde{\Psi}}(U^*(x_L, y), U^*(x_R, y))(U^*(x_R, y) - U^*(x_L, y)) \\ &\qquad = S_{1,\widetilde{\Psi}}(U^*(x_L, y), U^*(x_R, y))(H(x_R, y) - H(x_L, y)), \end{aligned} \tag{73}$$

$$\begin{aligned} &A_{2,\widetilde{\Psi}}(U^*(x, y_D), U^*(x, y_U))(U^*(x, y_U) - U^*(x, y_D)) \\ &\qquad = S_{2,\widetilde{\Psi}}(U^*(x, y_D), U^*(x, y_U))(H(x, y_U) - H(x, y_D)) \end{aligned} \tag{74}$$

These equalities are satisfied if the functions

$$s \in [0, 1] \to \Psi_U\left(s; \begin{bmatrix} U^*(x_L, y) \\ H(x_L, y) \end{bmatrix}; \begin{bmatrix} U^*(x_R, y) \\ H(x_R, y) \end{bmatrix}\right), \tag{75}$$

$$s \in [0, 1] \to \Psi_U\left(s; \begin{bmatrix} U^*(x, y_D) \\ H(x, y_D) \end{bmatrix}; \begin{bmatrix} U^*(x, y_U) \\ H(x, y_U) \end{bmatrix}\right), \tag{76}$$

define respectively the same curves in Ω as

$$x \in [x_L, x_R] \to U^*(x, y), \tag{77}$$

$$y \in [y_D, y_U] \to U^*(x, y). \tag{78}$$

This geometrical property is much more restrictive in 2D than in 1D and, in general, only stationary solutions that are essentially 1D or some particular families of stationary solutions satisfy them, as will be seen in the two-layer shallow-water case.

Strategy 2 can be extended easily to 2D problems to design numerical methods that preserve a given family of m-parameter stationary solutions

$$U^*(x, y; c_1, \dots, c_m)$$

with $m < N$, assuming again that, given $\bar{x}, \bar{y}, \bar{U}$, there exists a unique choice of the parameters such that

$$CU^*_i(\bar{x}, \bar{y}) = C\bar{U},$$

Once an element of the family $U^*_{i,j}$ has been determined, the numerical solution is written as follows:

$$\begin{aligned} &U'_{i,j} + \frac{1}{\Delta x}\left(\widehat{D}^-_{i+1/2,j} + \widehat{D}^+_{i-1/2,j}\right) + \frac{1}{\Delta y}\left(\widehat{D}^-_{i,j+1/2} + \widehat{D}^+_{i,j-1/2}\right) \\ &\qquad = (S_1(U_{i,j}) - S_1(U^*_{i,j}(x_i, y_j)))H_x(x_i, y_j) + (S_2(U_{i,j}) - S_2(U^*_{i,j}(x_i, y_j)))H_y(x_i, y_j), \end{aligned} \tag{79}$$

with

$$\begin{aligned} \widehat{D}^-_{i+1/2,j} &= \mathcal{R}^L(D^+_{i,i-k;j} - D^{*,+}_{i,i-k;j}, \dots, D^+_{i,i+k;j} - D^{*,+}_{i,i+k;j}) \\ &\quad + \mathcal{R}^R(D^-_{i,i+1-k;j} - D^{*,-}_{i,i+1-k;j}, \dots, D^-_{i,i+1+k;j} - D^{*,-}_{i,i+1+k;j}), \end{aligned} \tag{80}$$

$$\begin{aligned} \widehat{D}^+_{i-1/2,j} &= \mathcal{R}^L(D^+_{i-1-k,i;j} - D^{*,+}_{i-1-k,i;j}, \dots, D^+_{i-1+k,i;j} - D^{*,+}_{i-1+k,i;j}) \\ &\quad + \mathcal{R}^R(D^-_{i-k,i;j} - D^{*,-}_{i-k,i;j}, \dots, D^-_{i+k,i;j} - D^{*,-}_{i+k,i;j}), \end{aligned} \tag{81}$$

$$\widehat{D}^-_{i,j+1/2} = \mathcal{R}^L(D^+_{i;j,j-k} - D^{*,+}_{i;j,j-k}, \dots, D^+_{i;j,j+k} - D^{*,+}_{i;j,j+k})$$

$$+\mathcal{R}^R(D^-_{i;j,j+1-k}-D^{*,-}_{i;j,j+1-k},\cdots,D^-_{i;j,j+1+k}-D^{*,-}_{i;j,j+1+k}), \tag{82}$$

$$\widehat{D}^+_{i,j-1/2}=\mathcal{R}^L(D^+_{i;j-1-k,j}-D^{*,+}_{i;j-1-k,j},\cdots,D^+_{i;j-1+k,j}-D^{*,+}_{i;j-1+k,j})$$
$$+\mathcal{R}^R(D^-_{i;j-k,j}-D^{*,-}_{i;j-k,j},\cdots,D^-_{i;j+k,j}-D^{*,-}_{i;j+k,j}). \tag{83}$$

Here, as in the 1D case, in the fluctuations $D^{*,\pm}_{k,l;m}$, $U_{k,m}$ and $U_{l,m}$ are replaced by $U^*_{i,j}(x_k,y_m)$ and $U^*_{i,j}(x_l,y_m)$, and in the fluctuations $D^{*,\pm}_{k;l,m}$, $U_{k,l}$ and $U_{k,m}$ are replaced by $U^*_{i,j}(x_k,y_l)$ and $U^*_{i,j}(x_k,y_m)$.

5. Application to the two-layer shallow water model

In this Section, we apply the proposed well-balanced schemes in Section 4 to 2D two-layer shallow water system that governs the flow of two superposed layers of immiscible fluids with different constant densities:

$$\begin{aligned}
&(h_1)_t+(h_1u_{1,1})_x+(h_1u_{1,2})_y=0,\\
&(h_1u_{1,1})_t+\left(h_1u_{1,1}^2+\frac{g}{2}h_1^2\right)_x+(h_1u_{1,1}u_{1,2})_y=-gh_1Z_x-gh_1(h_2)_x,\\
&(h_1u_{1,2})_t+(h_1u_{1,1}u_{1,2})_x+\left(h_1u_{1,2}^2+\frac{g}{2}h_1^2\right)_y=-gh_1Z_y-gh_1(h_2)_y,\\
&(h_2)_t+(h_2u_{2,1})_x+(h_2u_{2,2})_y=0,\\
&(h_2u_{2,1})_t+\left(h_2u_{2,1}^2+\frac{g}{2}h_2^2\right)_x+(h_2u_{2,1}u_{2,2})_y=-gh_2Z_x-grh_2(h_1)_x,\\
&(h_2u_{2,2})_t+(h_2u_{2,1}u_{2,2})_x+\left(h_2u_{2,2}^2+\frac{g}{2}h_2^2\right)_y=-gh_2Z_y-grh_2(h_1)_y,
\end{aligned} \tag{84}$$

where

- h_k, $k=1,2$ is the thickness of the kth layer;
- $\vec{u}_k=(u_{k,1},u_{k,2})$, $k=1,2$ is the velocity of the kth layer;
- $r=\dfrac{\rho_1}{\rho_2}$, where ρ_k, $k=1,2$ is the density of the kth layer ($\rho_1<\rho_2$). (index 1 corresponds to the upper layer);
- $c_k=\sqrt{gh_k}$, $k=1,2$;
- $Z(x,y)$ is the bottom function.

The equations can be written in the form (70) with $N=6$, $H\equiv -Z$,

$$U=\left[h_1,h_1u_{1,1},h_1u_{1,2},h_2,h_2u_{2,1},h_2u_{2,2}\right]^T,$$

$$A_1(U)=\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0\\
c_1^2-u_{1,1}^2 & 2u_{1,1} & 0 & c_1^2 & 0 & 0\\
-u_{1,1}u_{1,2} & u_{1,2} & u_{1,1} & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 0\\
rc_2^2 & 0 & 0 & c_2^2-u_{2,1}^2 & 2u_{2,1} & 0\\
0 & 0 & 0 & -u_{2,1}u_{2,2} & u_{2,2} & u_{2,1}
\end{bmatrix},\quad S_1(U)=\begin{bmatrix}0\\gh_1\\gh_1\\0\\0\\0\end{bmatrix},$$

$$A_2(U)=\begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 0\\
-u_{1,1}u_{1,2} & u_{1,2} & u_{1,1} & 0 & 0 & 0\\
c_1^2-u_{1,2}^2 & 0 & 2u_{1,2} & c_1^2 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1\\
0 & 0 & 0 & -u_{2,1}u_{2,2} & u_{2,2} & u_{2,1}\\
rc_2^2 & 0 & 0 & c_2^2-u_{2,2}^2 & 0 & 2u_{2,2}
\end{bmatrix},\quad S_2(U)=\begin{bmatrix}0\\0\\0\\0\\gh_2\\gh_2\end{bmatrix}. \tag{85}$$

The characteristic equation of $A_1(U)$ is

$$(\lambda-u_{1,1})(\lambda-u_{2,1})\Big(\big((\lambda-u_{1,1})^2-gh_1\big)\big((\lambda-u_{2,1})^2-gh_2\big)-rg^2h_1h_2\Big)=0.$$

The eigenvalues are then the four roots λ_k, $k=1,.,4$ of the equation

$$\big((\lambda-u_{1,1})^2-gh_1\big)\big((\lambda-u_{2,1})^2-gh_2\big)=rg^2h_1h_2,$$

and

$$\lambda_5=u_{1,1},\quad \lambda_6=u_{2,1}.$$

The corresponding eigenvectors are

$$\vec{R}_k = \begin{bmatrix} 1 \\ \lambda_k \\ u_{1,2} \\ \mu_k \\ \mu_k\lambda_k \\ \mu_k u_{2,2} \end{bmatrix}, \quad k = 1, \dots, 4, \quad \vec{R}_5 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{R}_6 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \tag{86}$$

with

$$\mu_k = \frac{(\lambda_k - u_{1,1})^2}{c_1^2} - 1.$$

The eigenvalues and eigenvectors of $A_2(U)$ can be computed similarly.

For $r \approx 1$, first-order approximation of $\lambda_1, \dots, \lambda_4$ is given similar as [47]:

$$\lambda^{\pm}_{ext} = U_m \pm \sqrt{g(h_1 + h_2)}, \quad \lambda^{\pm}_{\text{int}} = U_c \pm \sqrt{g' \frac{h_1 h_2}{h_1 + h_2} \left[1 - \frac{(u_1 - u_2)^2}{g'(h_1 + h_2)} \right]}, \tag{87}$$

where

$$U_m = \frac{h_1 u_1 + h_2 u_2}{h_1 + h_2}, \quad U_c = \frac{h_1 u_2 + h_2 u_1}{h_1 + h_2},$$

and $g' = (1 - r)g$ is the reduced gravity. Observe that λ_{int} become complex when

$$\frac{(u_1 - u_2)^2}{g'(h_1 + h_2)} > 1,$$

so that the system is expected to lose hyperbolicity in these cases. Numerical techniques to overcome sporadic episodes of loss of hyperbolicity can be found in [1,48].

The 1D two-layer shallow-water model will be also considered: it can be written in the form (45) with $N = 4$, $H \equiv -Z$,

$$U = \begin{bmatrix} h_1 \\ h_1 u_1 \\ h_2 \\ h_2 u_2 \end{bmatrix}, \quad A(U) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ gh_1 - u_1^2 & 2u_1 & gh_1 & 0 \\ 0 & 0 & 0 & 1 \\ rgh_2 & 0 & gh_2 - u_2^2 & 2u_2 \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ gh_1 \\ 0 \\ gh_2 \end{bmatrix}, \tag{88}$$

where now u_k, h_k, $k = 1, 2$ are respectively the velocity and thickness of the layers. The family of straight segments will be considered here to compute nonconservative products. The following Roe linearization corresponding to the family of straight segments is available (see [49]):

$$A(U_L, U_R) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\bar{u}_1^2 + \bar{c}_1^2 & 2\bar{u}_1 & \bar{c}_1^2 & 0 \\ 0 & 0 & 0 & 1 \\ r\bar{c}_2^2 & 0 & -\bar{u}_2^2 + \bar{c}_2^2 & 2\bar{u}_2 \end{bmatrix}, \quad S(U_L, U_R) = \begin{bmatrix} 0 \\ g\bar{h}_1 \\ 0 \\ g\bar{h}_2 \end{bmatrix},$$

where

$$\bar{u}_k = \frac{\sqrt{h_{L,k}}u_{L,k} + \sqrt{h_{R,k}}u_{R,k}}{\sqrt{h_{L,k}} + \sqrt{h_{R,k}}}, \quad \bar{h}_k = \frac{h_{L,k} + h_{R,k}}{2}, \quad \bar{c}_k = \sqrt{g\bar{h}_k}, \quad k = 1, 2.$$

This Roe matrix and its natural extension to 2D will be used to implement WENO methods. Steady-states solutions that correspond to water-at-rest equilibria constitute a 2-parameter family:

$$u_{1,1} = u_{1,2} = u_{2,1} = u_{2,2} \equiv 0, \; h_1 = c_1, \; h_2 = -Z + c_2,$$

where c_1, c_2 are arbitrary parameters ($c_1 > 0$, $c_2 > \max(Z)$). We will focus here on methods that preserve this family. Both Strategies 1 and 2 described in Section 3 for 1D problems and 4.2 for 2D problems can be followed to design numerical methods that preserve water-at-rest solutions. In effect, for Strategy 1, the equalities (44), (73), (74) can be easily checked for these stationary solutions: the equality of the curves given by (75) and (77) or those given by (76) and (78) can be easily checked if the family of straight segments is chosen: it derives from the linear nature of the relationships between variables that characterize water-at-rest solutions.

For Strategy 2, observe that, given $\bar{x}, \bar{y}$ and a state $\bar{U} = [\bar{h}_1, \bar{h}_1\bar{u}_{1,1}, \bar{h}_1\bar{u}_{1,2}, \bar{h}_2, \bar{h}_2\bar{u}_{2,1}, \bar{h}_2\bar{u}_{2,2}]^T$, there exists a unique water-at-rest stationary solution such as

$$h_1^*(\bar{x}, \bar{y}) = \bar{h}_1, \quad h_2^*(\bar{x}, \bar{y}) = \bar{h}_2,$$

which is the one given by:

$$h_1^*(x, y) = \bar{h}_1, \quad \bar{h}_2^*(x, y) = -Z(x, y) + Z(\bar{x}, \bar{y}) + \bar{h}_2, \quad u^*_{1,1} = u^*_{1,2} = u^*_{2,1} = u^*_{2,2} = 0,$$

i.e. in this case

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

so that the stationary solution $U^*_{i,j}$ used to implement the method is given by

$$U^*_{i,j}(x,y) = \left[h_{1;i,j}, 0, 0, -Z(x,y) + Z(x_i,y_j) + h_{2;i,j}, 0, 0\right]^T. \tag{89}$$

When the Upwind splitting scheme is used, systems of the form

$$R \cdot \vec{\alpha} = U_R - U_L - A_l^{-1} S_k (H_R - H_L),$$

have to be solved to compute α_i, $i = 1, \dots, 6$ such that (49) is satisfied. Here, $A_l = A_l(U_L, U_R)$, $S_l = S_l(U_L, U_R)$, $l = 1, 2$ represent the Roe linearization in the x or y direction, and

$$R = \left[\begin{array}{c|c|c|c} & & & \\ \vec{R}_1 & \dots & \vec{R}_6 & \\ & & & \end{array} \right], \quad \vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_6 \end{bmatrix}.$$

Let us suppose that $l = 1$. In this case, $\vec{R}_i$, $i = 1, \dots, 6$ are given by (86). It can be checked that the system can be solved as follows:

1. Solve system

$$R \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = V_R - V_L - A^{-1} S (H_R - H_L),$$

where

$$V_L = \begin{bmatrix} h_{1,L} \\ h_{1,L} u_{1,1,L} \\ h_{2,L} \\ h_{2,L} u_{2,1,L} \end{bmatrix}, \quad V_R = \begin{bmatrix} h_{1,R} \\ h_{1,R} u_{1,1,R} \\ h_{2,R} \\ h_{2,R} u_{2,1,R} \end{bmatrix},$$

$$R = \begin{bmatrix} 1, & \dots, & 1 \\ \lambda_1, & \dots, & \lambda_4 \\ \mu_1, & \dots, & \mu_4 \\ \mu_1 \lambda_1, & \dots, & \mu_4 \lambda_4 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \tilde{c}_1^2 - \tilde{u}_{1,1}^2 & 2\tilde{u}_{1,1} & \tilde{c}_1^2 & 0 \\ 0 & 0 & 0 & 1 \\ r\tilde{c}_2^2 & 0 & \tilde{c}_2^2 - \tilde{u}_{2,1}^2 & 2\tilde{u}_{2,1} \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ g\bar{h}_1 \\ 0 \\ g\bar{h}_2 \end{bmatrix}.$$

2. If $\lambda_i \neq 0$, $i = 5, 6$ compute

$$\alpha_5 = \frac{F_3 - u_{1,2} \sum_{j=1}^4 \alpha_j}{\lambda_5},$$
$$\alpha_6 = \frac{F_6 - u_{2,2} \sum_{j=1}^4 \mu_j \alpha_j}{\lambda_6}.$$

Otherwise (which is the case in water-at-rest stationary solutions) define

$$\alpha_5 = \alpha_6 = 0.$$

In other words, the system to be solved in the 2D case reduces to those arising in 1D problems. Observe that, following this algorithm, the cases in which eigenvalues λ_i, $i = 5, 6$, vanish, and thus A_k cannot be inverted, the coefficients α_j can be still computed.

6. Numerical solutions

In this section, we present some numerical results for 1D coupled Burgers, the 1D and 2D two-layer shallow water equations. Inheriting the notation from Section 3.3, two different WENO methods have been implemented and tested:

- WENO methods with Upwind splitting and Strategy 1 in Section 3.2.1 for the treatment of the source terms: these family of schemes will be called Method 1.
- WENO methods with LF splitting and Strategy 2 in Section 3.2.2 for the treatment of the source terms: these family of schemes will be called Method 2.

For all the methods:

- fifth-order WENO is used in all of the numerical tests but in accuracy test;
- the third order TVD Runge-Kutta method in [43] is used for time stepping, which is a convex combination of forward Euler steps;

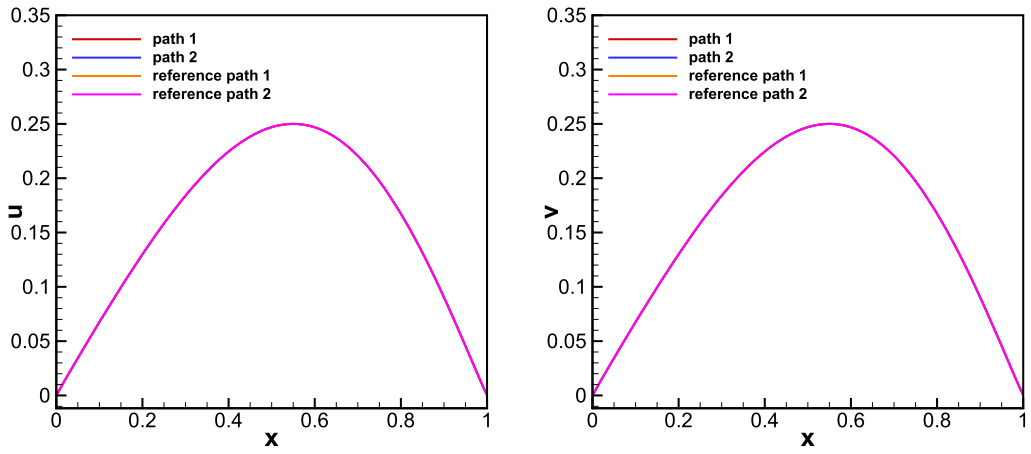


Fig. 1. Smooth solution of the Coupled Burgers' system. Numerical solutions for u (left) and v (right) at $t = 0.1$ obtained with the numerical methods consistent with the families of paths Ψ_1 and Ψ_2 using a mesh of 200 points.

Table 1
Accuracy test with different paths, L^∞ norm error for the coupled Burgers equations.

N	path 1		path 2	
	L^∞ error	Order	L^∞ error	Order
25	2.13e-6		5.76e-4	
50	5.01e-8	5.41	2.86e-4	1.01
100	1.29e-9	5.28	1.42e-4	1.01
200	3.65e-11	5.14	6.98e-5	1.02
400	9.95e-13	5.20	3.38e-5	1.05

- the computation of eigenvalues and eigenvectors are computed in analytic form according to (87) and (86), respectively, and the LAPACK library is used to solve linear systems.
- CFL $= 0.45$ is used in all cases.
- the free boundary conditions are imposed except in the steady-state solutions and accuracy test.

The numerical experiments have been implemented using FORTRAN 90 compiled with the INTEL ifort compiler run in a Dell Precision workstation with 24 CPU cores and 128 gigabytes of memory.

In many numerical tests the results are comparable and, when this is the case, we only show the results obtained with Method 2.

6.1. 1D coupled burgers equation

In this section, we consider the coupled Burgers' system (4) and compare the numerical methods obtained with the methods based on the fluctuations (24) and (25) based respectively on the choice of the family of paths Ψ_1 given by (5) and Ψ_2 given by (7).

6.1.1. Smooth solution

The initial condition

$$[u(x,0), v(x,0)]^T = [u(x,0), v(x,0)]^T = [0.25\sin(\pi x), 0.25\sin(\pi x)]^T \tag{90}$$

is first consider in the interval $[0,1]$ and periodic boundary conditions are imposed. Fig. 1 shows the numerical solutions obtained at time $t = 0.1$ with a mesh of 200 points, when the solution is still smooth. Reference solutions have been computed in a mesh of 6400 points and Table 1 shows the errors and the accuracy of the methods: as it can be seen, both methods seem to converge to the same solution, but only the one consistent with the family (5) achieves the optimal accuracy while the one consistent with the family (7) in only first-order accurate. We note that the high-order accuracy condition (38) for family of paths is then necessary.

6.1.2. Shock waves

The goal of this test is to show that numerical methods based on different families of paths give different results in presence of discontinuous waves. We consider the Riemann problem given by the initial condition

$$[u(x,0), v(x,0)]^T = \begin{cases} [2,2]^T, & \text{if } x \le 0.5, \\ [1,1]^T, & \text{otherwise.} \end{cases} \tag{91}$$

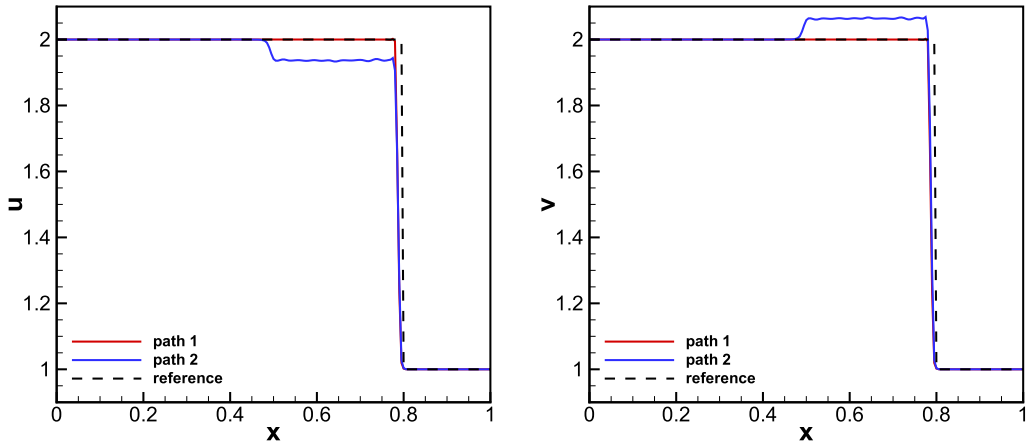


Fig. 2. Riemann problem for the coupled Burgers' system. Numerical solutions for u (left) and v (right) at $t = 0.1$ are obtained consistent with the families of paths Ψ_1 and Ψ_2 using a mesh of 200 points.

The analytical solution is used as the reference solution. If the family of straight segments Ψ_1 is chosen, the solution is an isolated shock wave traveling at speed 3. After numerically simulating the solution up to $t = 0.1$ using 200 uniform grids, the results are presented in Fig. 2. As it can be seen, in this particular case the numerical method consistent with the family of paths (5) captures the correct solution, while the one consistent with the family (7) gives a solution with a stationary contact discontinuity at $x = 0.5$ and a different shock wave traveling at the same speed.

6.2. 1D two-layer shallow water model

6.2.1. Small perturbations of a steady-state solution

We use this test, taken from [31], to verify the well-balanced property. The density ratio is $r = 0.98$ and the gravitational constant is $g = 10$. We consider smooth and discontinuous bottom topography. The smooth bottom topography is given by

$$Z(x) = \begin{cases} 0.25(\cos(10\pi(x-0.5))+1) - 2, & \text{if } 0.4 < x < 0.6, \\ -2, & \text{otherwise,} \end{cases}$$

and the discontinuous one by

$$Z(x) = \begin{cases} -1.5, & \text{if } x > 0.5, \\ -2, & \text{otherwise.} \end{cases}$$

The initial data is given by:

$$\big(h_1, h_1u_1, h_2, h_2u_2\big)(x,0) = \begin{cases} (1+\xi, 0, -1 - Z(x), 0), & \text{if } 0.1 < x < 0.2, \\ (1, 0, -1 - Z(x), 0), & \text{otherwise.} \end{cases}$$

The computational domain is $[-0.2, 1]$ with extrapolation boundary conditions. We have first checked that for $\xi = 0$ the initial condition is preserved to machine accuracy, i.e. that the C-property is satisfied. Next, we set $\xi = 0.00001$. The numerical solutions computed with 200 points are compared with a reference solution computed with 2000 points. The final time is $t = 0.15$. The numerical water surface, i.e. $\eta_1 = -Z + h_1 + h_2$, corresponding to the smooth and the discontinuous topographies are reported in Fig. 3 (left and right respectively): it can be seen that small waves are captured without spurious oscillations as expected.

6.2.2. Riemann problems

We consider here two Riemann problems with flat bottom topography. This test is taken from [8]. In the first test, the computational domain is $[0, 1]$ and the final time is $t = 0.1$. The density ratio is $r = 0.98$ and the gravitational constant is $g = 10$. The initial condition is given by

$$\big(h_1, h_1u_1, h_2, h_2u_2\big)(x,0) = \begin{cases} (0.50, 1.250, 0.50, 1.250), & \text{if } x < 0.3, \\ (0.45, 1.125, 0.55, 1.375), & \text{otherwise.} \end{cases}$$

The flat bottom is placed at $Z(x) = -1$. The water surface $E = -1 + h_1 + h_2$, a zoom of h_1, and the velocity u_1 obtained at the final time with a mesh of 200 cells are compared in Fig. 4 with a reference solution obtained with 10,000 points.

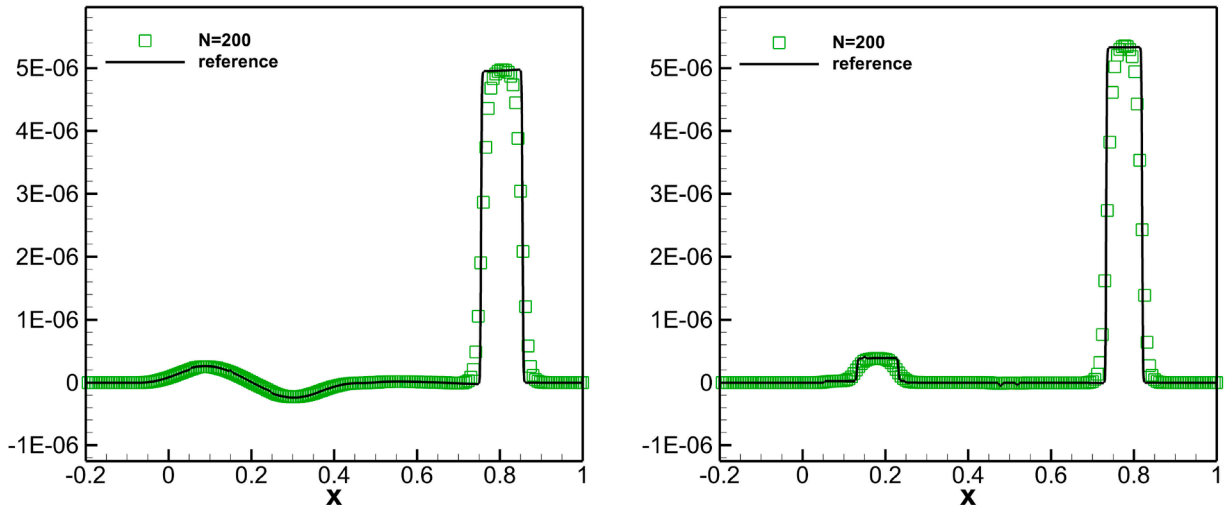


Fig. 3. Small perturbation of a water-at-rest stationary solution with smooth (left) and discontinuous (right) bottom topographies. The numerical solution obtained with 200 points is compared with the reference solution obtained with 2000 points: water surface at $t = 0.15$.

In the second test case, taken from [50], the computational domain is $[-5, 5]$ and the final time is $t = 1$. The density ratio is again $r = 0.98$ and the gravitational constant is $g = 9.81$. The initial data is given by

$$\big(h_1, h_1u_1, h_2, h_2u_2\big)(x,0) = \begin{cases} (1.8, 0, 0.2, 0), & \text{if } x < 0, \\ (0.2, 0, 1.8, 0), & \text{otherwise.} \end{cases}$$

The bottom is placed now at $Z(x) = -2$. The water surface $E = -2 + h_1 + h_2$ and the interface $\eta = -2 + h_2$ obtained with a 500-point mesh are shown in Fig. 5, together with a reference solution obtained with a fine mesh with 5000 points.

6.2.3. An internal dam-break with flat bottom

In this test, taken from [49], an internal dam-break over a flat bottom is simulated. The computational domain is $[0, 10]$ with free boundary conditions. The final time is $t = 10$. The gravitational constant $g = 10$ and the density ratio $r = 0.98$. The initial condition is given by:

$$\big(h_1, h_1u_1, h_2, h_2u_2\big)(x,0) = \begin{cases} (0.2, 0, 0.8, 0), & \text{if } x < 5, \\ (0.8, 0, 0.2, 0), & \text{otherwise.} \end{cases}$$

The results computed with a 200-point mesh are compared in Fig. 6 with a reference solution computed using 3200 points: that solutions agree well with those in [48,49].

6.2.4. Internal dam-break with non-flat bottom

In this test, taken from [4], a discontinuous stationary solution is reached starting from an internal dam-break initial condition given by

$$\big(h_1, h_1u_1, h_2, h_2u_2\big)(x,0) = \begin{cases} (1.6, 0, -1.6 - Z(x), 0) & \text{if } x < 0, \\ (0.7, 0, -0.7 - Z(x), 0) & \text{otherwise,} \end{cases}$$

and the bottom topography is given by

$$Z(x) = 0.25e^{-x^2} - 2.$$

The constant gravitational acceleration is $g = 9.81$ and the density ratio is $r = 0.998$. The computational domain is $[-5, 5]$. The steady state obtained with 500 points is shown in Fig. 7. The converged results agree well with reference solutions computed with 2000 grids and with those in [4].

6.2.5. Accuracy test

In this test, taken from [4], we check the empirical order of accuracy of the methods using a smooth solution. The bottom is flat, $g = 9.81, r = 0.98$, and the initial condition is given by:

$$\big(h_1, h_1u_1, h_2, h_2u_2\big)(x,0) = \Big(1 - \frac{1}{2}\sin(8x), 0, 0.6 + \frac{1}{2}\sin(8x), 0\Big).$$

The computational domain is $[-\pi/2, 3\pi/2]$ and periodic boundary conditions are considered. We compute the numerical solutions until $t = 0.1$. We use 5th, and 7th order WENO-Z for this smooth problem to check the numerical accuracy. The reference solutions

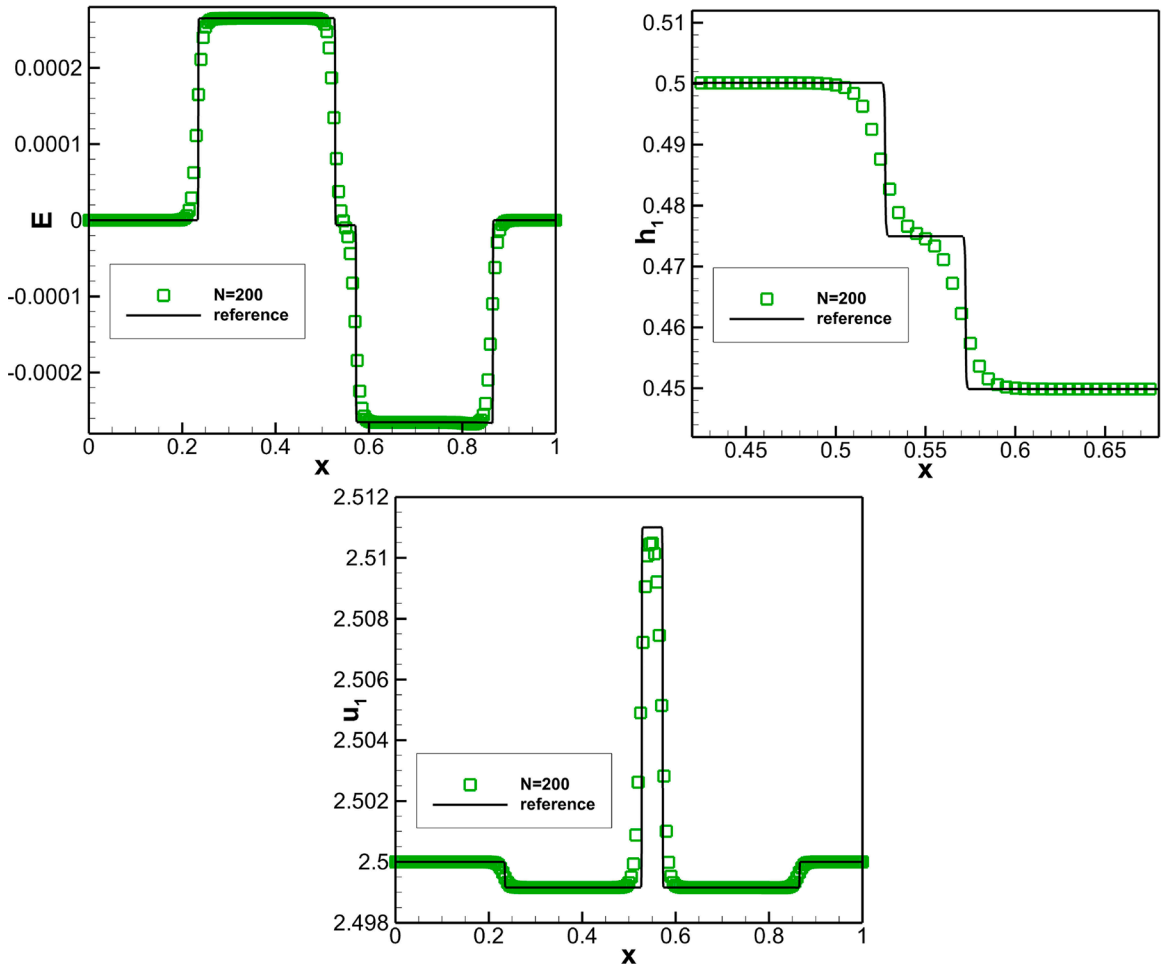


Fig. 4. Riemann problem 1. The numerical solution computed with 200 points is compared to a reference solution computed with 10,000 points. Top left: water surface; top right: h_1 (zoom); bottom: u_1.

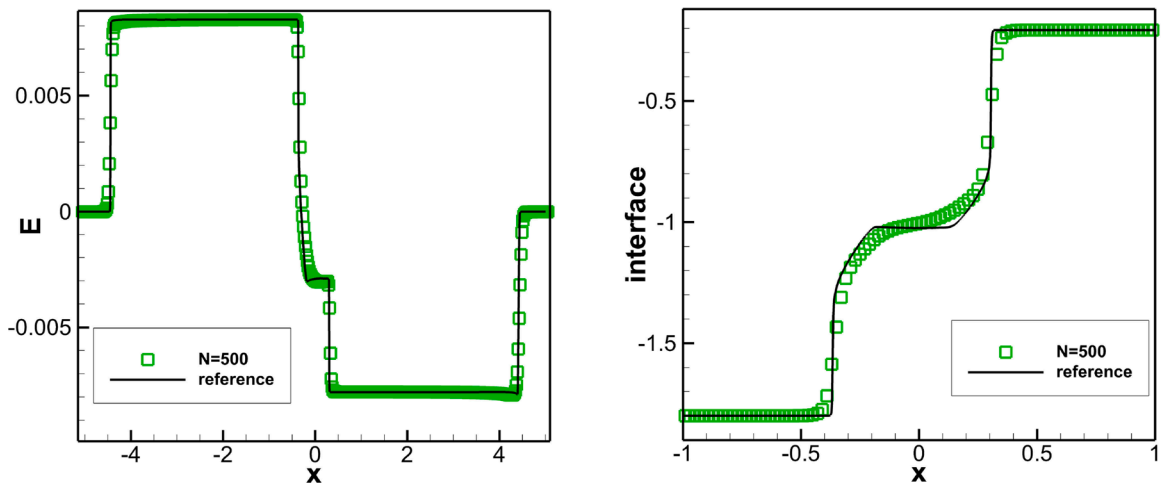


Fig. 5. Riemann problem 2. The numerical solution obtained with 500 points is compared to a reference solution computed with 5000 points. Left: water surface; right: interface.

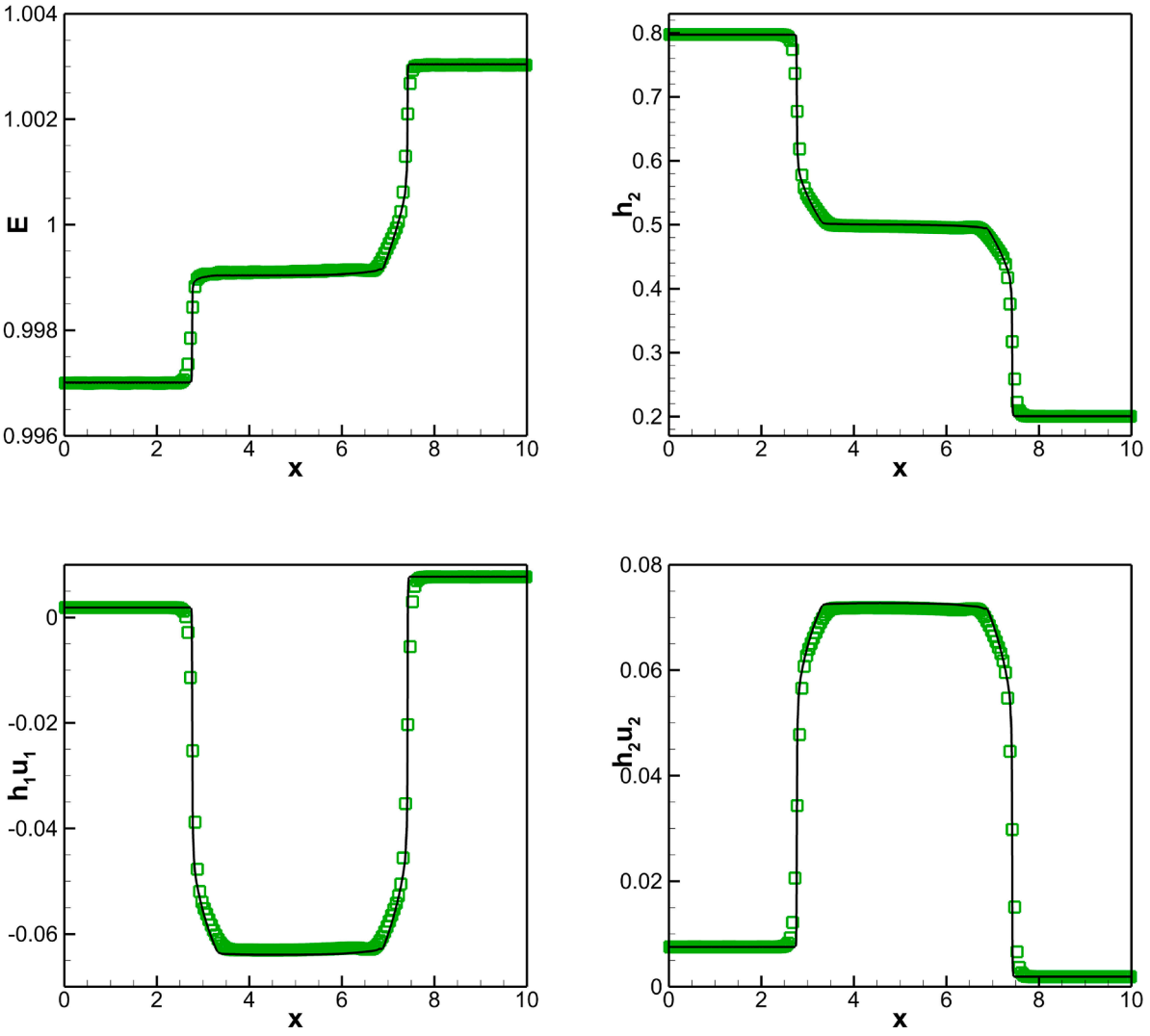


Fig. 6. Internal dam-break with flat bottom. The numerical solution computed with 200 points is compared to a reference solution computed using 3200 points: free surface (top-left), h_2 (top-right), h_1u_1 (down-left), and h_2u_2 (down-right) at time $t = 10$.

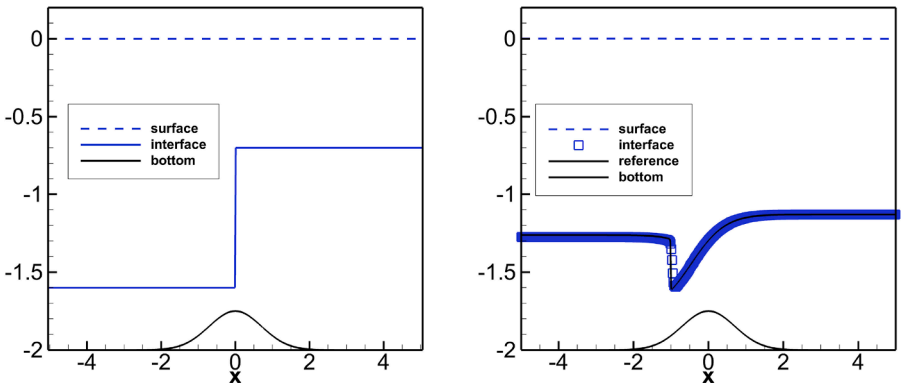


Fig. 7. Internal dam-break with non-flat bottom. The numerical solution obtained with 500 points is compared to a reference solution computed with 2000 points. Free surface, interface, and bottom topography: initial condition (left), reached steady state (right).

Table 2
L^∞ error and convergence rates using WENO3, WENO5, and WENO7.

N	3th order test		5th order test		7th order test	
	h_1	h_1+h_2	h_1	h_1+h_2	h_1	h_1+h_2
25	3.32e-1 —	5.84e-3 —	5.38e-2 —	1.42e-3 —	1.62e-1 —	3.37e-3 —
50	1.10e-2 (1.60)	2.10e-3 (1.48)	4.65e-3 (3.53)	4.50e-4 (1.66)	1.20e-2 (3.76)	6.26e-4 (2.43)
100	2.08e-2 (2.40)	4.96e-4 (2.08)	4.84e-4 (3.26)	1.42e-5 (4.98)	9.58e-4 (3.65)	1.82e-5 (5.10)
200	3.19e-3 (2.71)	7.43e-5 (2.74)	1.95e-5 (4.63)	5.52e-7 (4.69)	1.94e-5 (5.63)	1.82e-7 (6.64)
400	4.24e-4 (2.91)	9.75e-6 (2.93)	5.92e-7 (5.04)	1.74e-8 (4.99)	2.04e-7 (6.57)	1.49e-9 (6.94)
800	5.31e-5 (3.00)	1.23e-6 (2.99)	1.76e-8 (5.07)	4.33e-10 (5.33)	1.64e-9 (6.96)	1.01e-11(7.20)

Table 3
L^1 errors at time 0.1 using two meshes of 50×50 and 100×100 points.

	h_1	h_1u_1	h_1v_1	h_2	h_2u_1	h_2v_2
50×50	7.77e-16	1.36e-16	2.79e-16	9.81e-16	9.78e-15	1.36e-16
100×100	1.98e-15	1.88e-16	5.71e-16	1.44e-15	1.32e-14	1.88e-16

Table 4
Methods 1 and 2: CPU times and speedup.

	CPU time	Speedup ratio
Method 1	68.38 s	
Method 2	63.54 s	1.08

for 5th are computed by each order WENO scheme with 6400 uniform grids. We use 12,800 grids to obtain the reference solution for the 7th order accuracy test to get a more accurate reference. The third-order TVD Runge Kutta is used in all cases for time stepping and, in order to reach the same order of accuracy in time and space, the time-step is set to $\Delta t^{\frac{k}{3}}$, where k is the WENO order and Δt is the step given by the CFL condition.

Table 2 show the L^∞ error and the empirical order of convergence: in all cases, the optimal order is reached.

6.3. 2D two-layer shallow water model

6.3.1. 2-D steady-state solution

This test is used to check the well-balanced property of the methods for water-at-rest solutions. The constant density ratio is $r = 0.98$ and the gravitational constant is $g = 10$. The bottom topography is given by a smooth function

$$Z(x,y) = 0.05e^{-100(x^2+y^2)} - 1,$$

and the initial condition is given by

$$h_1 = 0.5,\ h_2 = 1 - Z(x,y),\ u_{1,1} = u_{1,2} = u_{2,1} = u_{2,2} = 0.$$

The computational domain is $[-1,1]\times[-1,1]$. The final time is $t = 0.1$. This initial boundary value problem is a modification of the exact C-property test, proposed in [32] for the shallow water equations. The numerical L^1 errors corresponding to two meshes of 50×50 and 100×100 points are shown in Table 3, as it can be seen the initial condition is preserved to machine accuracy.

We have used this test to compare the computational costs of Methods 1 and 2: the CPU times corresponding to a computation with 200×200 points are shown in Table 4. The computational costs of Strategy 1 and 2 are comparable while Strategy 1 is slightly more computationally expensive.

6.3.2. Interface propagation with flat bottom

This test, taken from [50], is aimed to verify the robustness of the numerical method. The initial condition is

$$\big(h_1, h_1u_{1,1}, h_1u_{1,2}, h_2, h_2u_{2,1}, h_2u_{2,2}\big)(x,y,0) = \begin{cases} (0.50, 1.250, 1.250, 0.50, 1.250, 10.250), & \text{if } (x,y)\in\Omega, \\ (0.45, 1.125, 1.125, 0.55, 1.375, 1.375), & \text{otherwise,} \end{cases}$$

where $\Omega = \{x<-0.5, y<0\}\cup\big\{(x+0.5)^2+(y+0.5)^2<0.25\big\}\cup\{x<0, y<-0.5\}$. The constant gravitational acceleration is $g = 10$ and the density ratio is $r = 0.98$. The flat bottom topography is given by $Z(x,y)\equiv -1$. The computational domain is $[-0.55, 0.7]\times[-0.55, 0.7]$ and the final time, $t = 0.1$. The computational results obtained with two meshes of 400×400 and 800×800 are shown in Fig. 8.

6.3.3. Interface propagation with non-flat bottom

This test is similar to the one in Section 6.3.2 but with a non-flat bottom given by

$$Z(x,y) = 0.05e^{-100(x^2+y^2)} - 1,$$

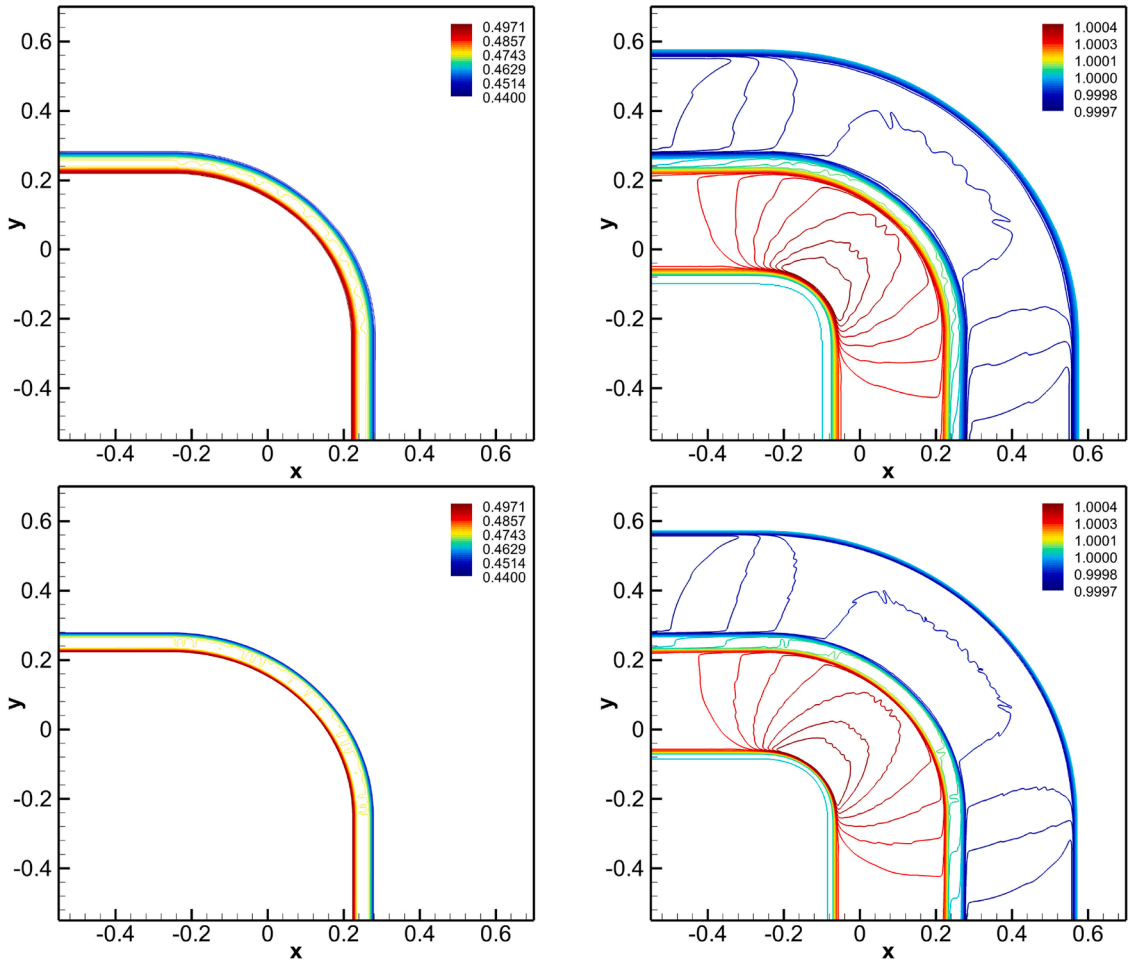


Fig. 8. Interface propagation in 2D with flat bottom. Numerical solutions obtained with 400×400 (top row) and 800×800 (bottom row) grids: upper layer thickness h_1 (left) and water surface $h_1 + h_2$ (right).

and the following initial data:

$$\big(h_1, h_1u_{1,1}, h_1u_{1,2}, h_2, h_2u_{2,1}, h_2u_{2,2}\big)(x, y, 0) = \begin{cases} (0.50, 1.250, 1.250, -0.50 - Z(x, y), 1.250, 1.250), & \text{if } (x, y) \in \Omega, \\ (0.45, 1.125, 1.125, -0.45 - Z(x, y), 1.375, 1.375), & \text{otherwise.} \end{cases}$$

The computational domain, the values of g and r, and the final time are the same. We show again the results computed with 400×400 and 800×800 points in Fig. 9. The numerical solutions are in good agreement with those presented in [6] but finer structures of the flow are actually captured here.

6.3.4. *Internal circular dam break over flat bottom topography*

This example is taken from [46]: an internal circular dam-break problem with a flat bottom is considered. The initial conditions are given by

$$\big(h_1, h_1u_{1,1}, h_1u_{1,2}, h_2, h_2u_{2,1}, h_2u_{2,2}\big)(x, y, 0) = \begin{cases} (1.8, 0, 0, -1.8 - Z(x, y), 0, 0), & \text{if } x^2 + y^2 > 4, \\ (0.2, 0, 0, -0.2 - Z(x, y), 0, 0), & \text{otherwise.} \end{cases}$$

In this test, we consider $g = 9.81$, $r = 0.998$, and $Z(x, y) \equiv -2$. The final time is $t = 20$. We show the contour lines of the water interface for times $t = 4, 20$ in Fig. 10. Diagonal slices $y = x$ of the numerical results computed with 200×200 points for times $t = 4, 6, 10, 14, 16, 20$ are shown in Fig. 11. It is worth mentioning that the results agree well with those in [4,46].

6.3.5. *Internal circular dam break over nonflat bottom topography*

In this test, we consider nonflat bottom

$$Z(x, y) = 0.5e^{(x^2+y^2)} - 2,$$

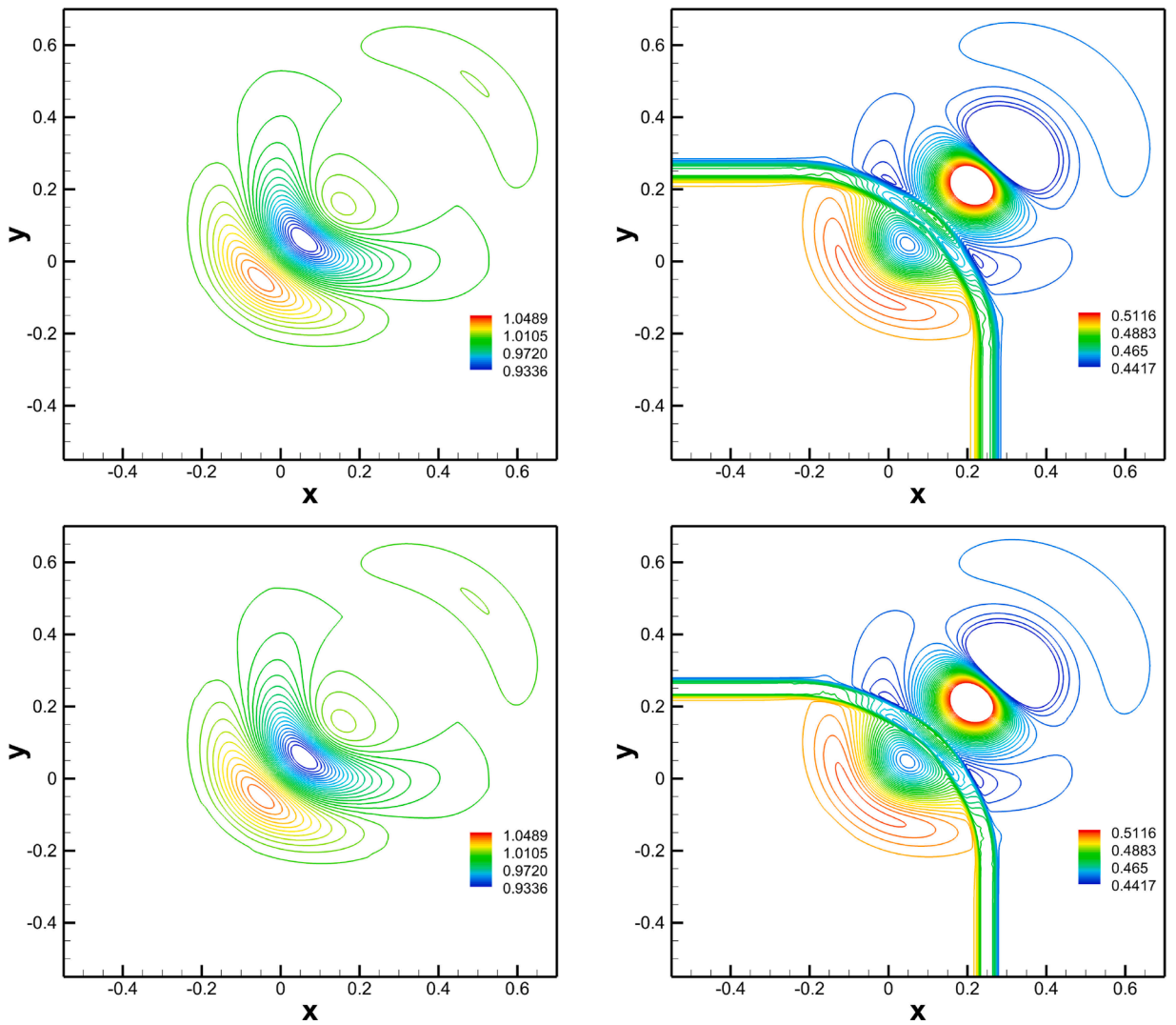


Fig. 9. Interface propagation in 2-D with nonflat bottom. Numerical solutions computed with 400×400 (top row) and 800×800 (bottom row) grids: water surface $h_1 + h_2 + Z$ (left) and upper layer thickness h_1 (right).

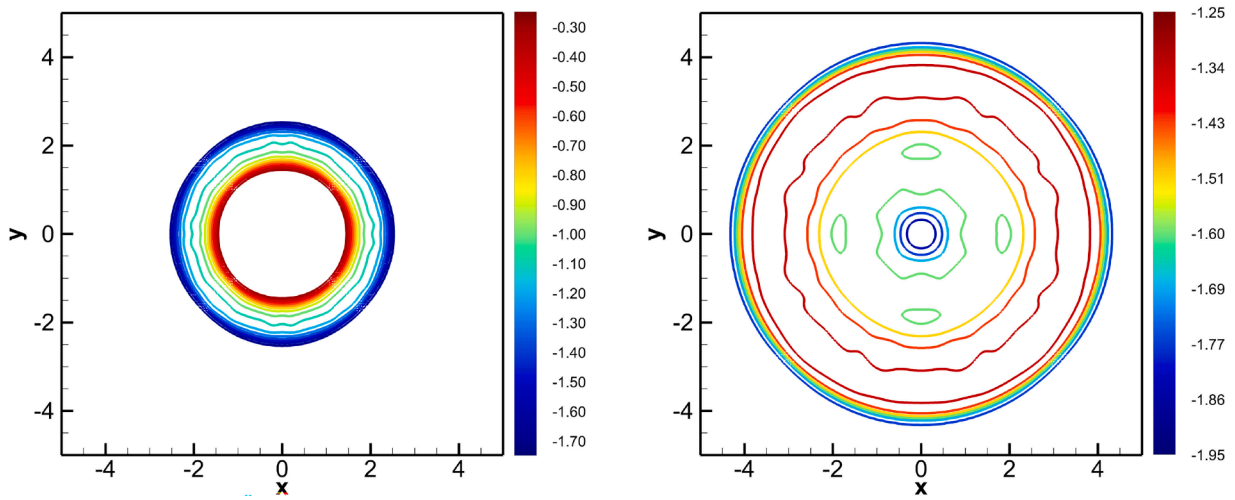


Fig. 10. Internal circular dam breaking in 2D with a flat bottom. Contour lines of the interface $h_2 + Z$ at times $t = 4$ (left), $t = 20$ (right).

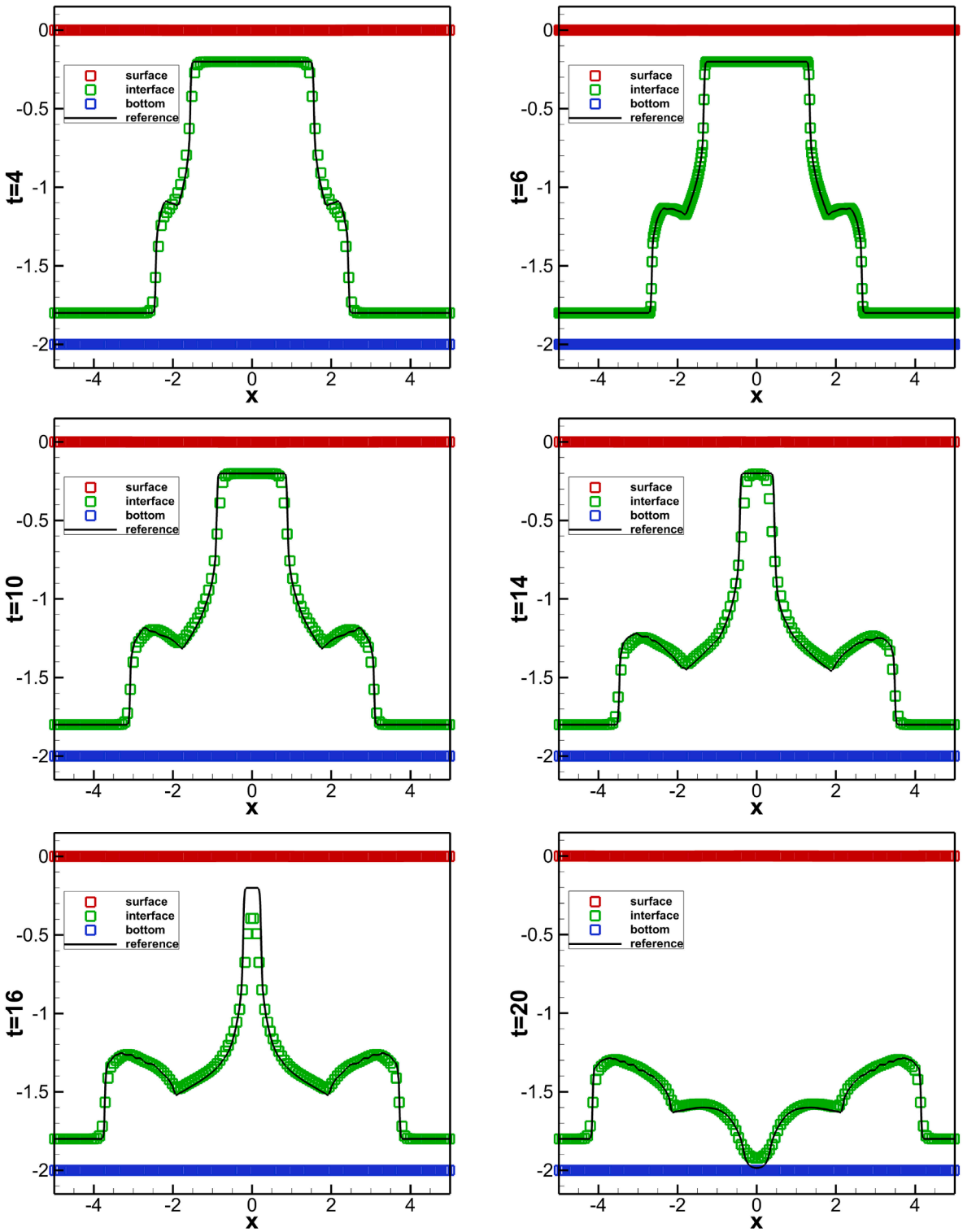


Fig. 11. Internal circular dam breaking in 2D with a flat bottom. Numerical solutions computed with 200×200 grids. Diagonal slices of the water surface $h_1 + h_2 + Z$, interface $h_2 + Z$ and bottom at times $t = 4, 6, 10, 14, 16, 20$ (from top to bottom and from left to right).

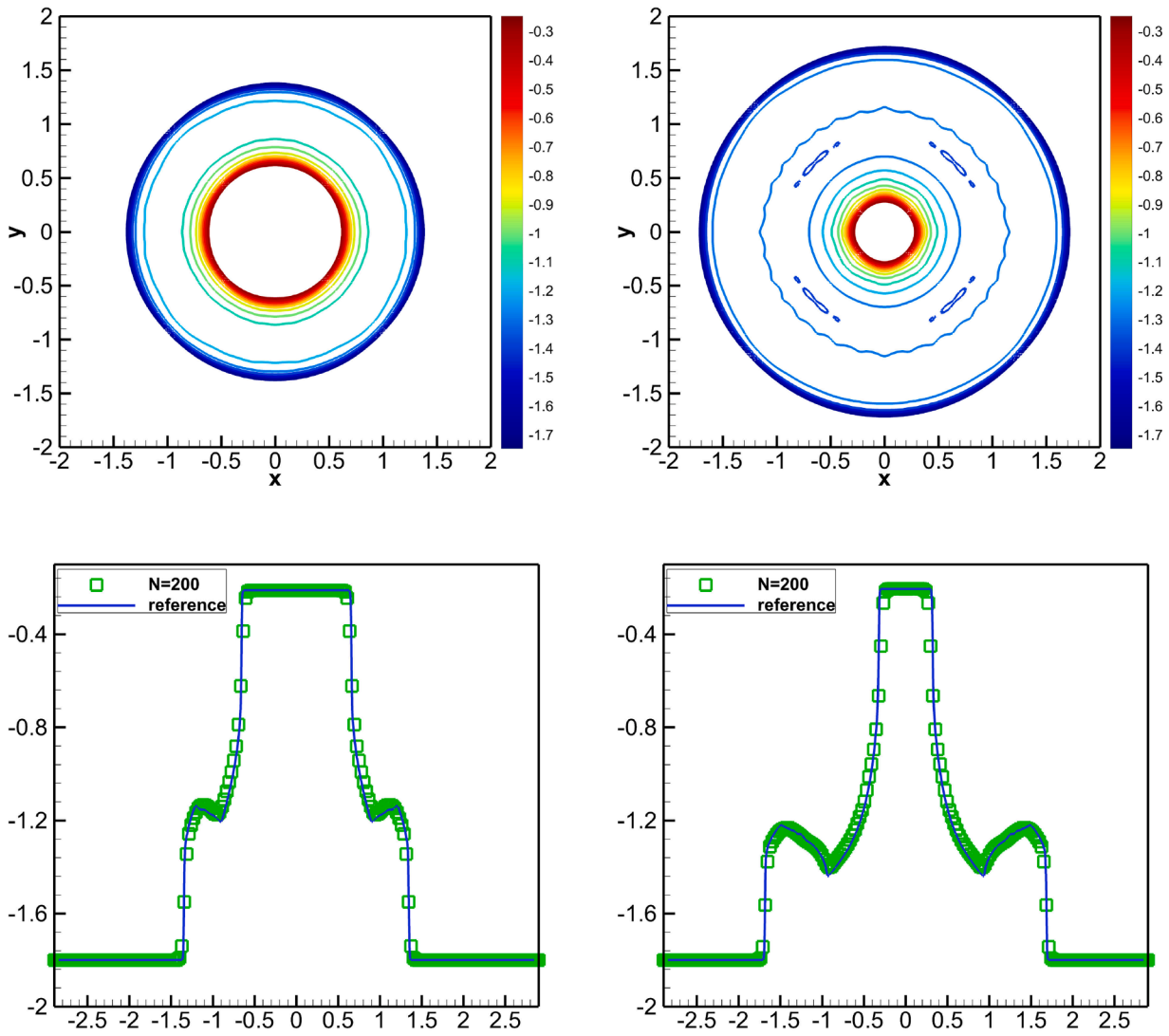


Fig. 12. Internal circular dam break over nonflat bottom topography. Numerical solutions computed with 200×200 grids. Contour lines of the interface $h_2 + Z$ at times $t = 1$ (top left) and $t = 2$ (top right); diagonal slices of the interface at time $t = 1$ (down left) and $t = 2$ (down right).

and the initial condition

$$\big(h_1, h_1u_{1,1}, h_1u_{1,2}, h_2, h_2u_{2,1}, h_2u_{2,2}\big)(x, y, 0) = \begin{cases} (1.8, 0, 0, -1.8 - Z(x, y), 0, 0), & \text{if } x^2 + y^2 > 1, \\ (0.2, 0, 0, -0.2 - Z(x, y), 0, 0), & \text{otherwise.} \end{cases}$$

The constant gravitational acceleration is $g = 9.81$ and the density ratio is $r = 0.98$. The computational domain is $[-2, 2] \times [-2, 2]$ and a uniform mesh with 200×200 grids is considered. The contour lines and diagonal slices of the water interface at $t = 1, t = 2$ are shown in Fig. 12.

7. Conclusion

A new family of high-order WENO finite-difference methods to solve hyperbolic nonconserative PDE systems have been proposed. These methods are based on a general strategy in which, instead of reconstructing fluxes using a WENO operator, what we reconstruct is the nonconservative products of the system which are computed using the selected family of paths. Moreover, if a Roe linearization is available, the nonconservative products can be computed through matrix-vector operations instead of path-integrals. The main advantages of this theory are the following:

- the high-order accuracy in space only depends on the order of the selected WENO operator, provided that the path satisfies the symmetry property;

- no integrals in the cells have to be computed so that reconstructions with uniform accuracy in the entire cells are required;
- a unified WENO framework that treats non-conservative equations consistently with conservative equations. The framework is compatible with state reconstruction techniques.

The methods have been extended to systems with source terms to design well-balanced methods. This methodology has been successfully applied to obtain high-order numerical methods for the 1D and 2D two-layer shallow water equations that preserve water-at-rest steady states. A number of numerical tests confirm the high-order accuracy of the methods as well as their shock-capturing and well-balanced properties.

The second strategy introduced here to design well-balanced methods can be combined with the technique developed in [51] to obtain numerical methods that preserve both water-at-rest and moving equilibria for the 1D shallow-water model or some particular families of stationary solutions in the 2D case: this will be done in an upcoming work.

Another further development concerns the convergence of the methods: as it happens for general finite-difference and related methods, the convergence of the numerical results to functions that are weak solutions of the system according to the selected family of paths is not ensured for the schemes introduced here. Some techniques recently developed in the context of high-order finite-volume methods in [40,41] can be adapted to the methods introduced here to ensure that isolated shocks that satisfy the generalized Rankine-Hugoniot associated to the selected family of paths are correctly captured.

In memoriam

This paper is dedicated to the memory of Prof. Arturo Hidalgo López (*July 03rd 1966 - †August 26th 2024) of the Universidad Politecnica de Madrid.

CRediT authorship contribution statement

Baifen Ren: Writing – original draft, Visualization, Validation, Software, Methodology; **Carlos Parés:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

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

Acknowledgements

The work of B. Ren is supported by the China Scholarship Council. The work of C. Parés is supported by Grant PID2022-137637NB-C21 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU, and Grant PDC2022-133663-C21, funded by MICIU/AEI/10.13039/501100011033 and by "European Union NextGenerationEU/PRTR".

Data availability

Data will be made available on request.

Appendix A. Fifth-order WENOZ reconstruction

Let us recall here the expression of the fifth-order WENOZ reconstruction used to compute $F_{i+1/2}$ as an example. The expression of $F_{i-1/2}$ can be obtained then using the mirror principle. Given $F_j = F(U_j)$, $j = i-2, i-1, i, i+1, i+2$, $F_{i+1/2}$ is computed as follows:

$$F_{i+1/2} = \sum_{k=0}^{2} \omega_k F^k_{i+1/2}, \tag{A.1}$$

where

$$F^0_{i+1/2} = \frac{1}{6}(2F_{i-2} - 7F_{i-1} + 11F_i),\ F^1_{i+1/2} = \frac{1}{6}(-F_{i-1} + 5F_i + 2F_{i+1}),\ F^2_{i+1/2} = \frac{1}{6}(2F_i + 5F_{i+1} - F_{i+2}),$$

are third-order interpolation formulas computed in 3 *substencils*, and ω_k, $k = 0, 1, 2$ are nonlinear weights to be computed. In the classical WENO-JS scheme [21] the nonlinear weights are given by

$$\omega_k^{(JS)} = \frac{\alpha_k}{\sum_{j=0}^{2} \alpha_j}, \quad \alpha_k = \frac{d_k}{(\beta_k + \varepsilon)^p}, \quad k = 0, 1, 2.$$

Here $d_0 = 1/10$, $d_1 = 6/10$, $d_2 = 3/10$, are the ideal weights leading to the global fifth-order interpolation formula

$$F_{i+1/2} = \frac{1}{60}(2F_{i-2} - 13F_{i-1} + 47F_i + 27F_{i+1} - 3F_{i+2});$$

β_k are the smoothness indicators

$$\beta_k = \sum_{l=1}^{2} \Delta x^{2l-1} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \left(\frac{d^l}{dx^l} F^k_{i+\frac{1}{2}}(x) \right)^2 dx, \quad k = 0, 1, 2,$$

that are defined so that the weights are close to the ideal ones in smooth regions but, when a discontinuity is detected in the stencil, the contribution of the sub-stencil containing it is close to 0 (non-oscillatory weights). WENO-Z scheme [19] propose the global smooth indicator $\tau_5 = |\beta_2 - \beta_0|$ to achieve the optimal accuracy at the critical points,

$$\omega_k^{(Z)} = \frac{\alpha_k}{\sum_{l=0}^{2} \alpha_l}, \quad \alpha_k = d_k \left(1 + \left(\frac{\tau_5}{\beta_k + \varepsilon} \right)^p \right), \quad k = 0, 1, 2,$$

$\varepsilon = 10^{-12}$, $p = 2$ are used in this study. There is a vast literature related to the computation of optimal order smoothness indicators: see for instance [52,53].

References

[1] M.J. Castro, E.D. Fernández-Nieto, J.M. González-Vida, C. Parés, Numerical treatment of the loss of hyperbolicity of the two-layer shallow-water system, J. Sci. Comput. 48 (1) (2011) 16–40. https://doi.org/10.1007/s10915-010-9427-5

[2] M.J. Castro, J.A. García-Rodríguez, J.M. González-Vida, J. Macías, C. Parés, M.E. Vázquez-Cendón, Numerical simulation of two-layer shallow water flows through channels with irregular geometry, J. Comput. Phys. 195 (1) (2004) 202–235.

[3] N. Chalmers, E. Lorin, On the numerical approximation of one-dimensional nonconservative hyperbolic systems, J. Comput. Sci. 4 (1) (2013) 111–124. Computational Methods for Hyperbolic Problems, https://doi.org/10.1016/j.jocs.2012.08.002

[4] S. Chu, A. Kurganov, M. Na, Fifth-order A-WENO schemes based on the path-conservative central-upwind method, J. Comput. Phys. 469 (2022) 111508. https://doi.org/10.1016/j.jcp.2022.111508

[5] J. Dong, X. Qian, A robust numerical scheme based on auxiliary interface variables and monotone-preserving reconstructions for two-layer shallow water equations with wet–dry fronts, Comput. Fluids 272 (2024) 106193.

[6] X. Liu, A new well-balanced finite-volume scheme on unstructured triangular grids for two-dimensional two-layer shallow water flows with wet-dry fronts, J. Comput. Phys. 438 (2021) 110380. https://doi.org/10.1016/j.jcp.2021.110380

[7] J. Murillo, S. Martinez-Aranda, A. Navas-Montilla, P. García-Navarro, Adaptation of flux-based solvers to 2D two-layer shallow flows with variable density including numerical treatment of the loss of hyperbolicity and drying/wetting fronts, J. Hydroinf. 22 (5) (2020) 972–1014. https://doi.org/10.2166/hydro.2020.207

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

[9] G.D. Maso, P.L. Floch, F. Murat, Definition and weak stability of nonconservative products, Journal de Mathématiques Pures et Appliquées 74 (1995) 483–548.

[10] M.J. Castro, U.S. Fjordholm, S. Mishra, C. Parés, Entropy conservative and entropy stable schemes for nonconservative hyperbolic systems, SIAM J. Numer. Anal. 51 (3) (2013) 1371–1391. https://doi.org/10.1137/110845379

[11] C. Berthon, Nonlinear scheme for approximating a non-conservative hyperbolic system, Comptes Rendus Mathématiques de l'Académi des Sciences 395 (2002) 1069–1072.

[12] C. Parés, Numerical methods for nonconservative hyperbolic systems: a theoretical framework, SIAM J. Numer. Anal. 44 (1) (2006) 300–321.

[13] M.J. Castro, Y. Cheng, A. Chertock, A. Kurganov, Solving two-mode shallow water equations using finite volume methods, Commun. Comput. Phys. 16 (5) (2014) 1323–1354. https://doi.org/10.4208/cicp.180513.230514a

[14] M.J. Castro, J.M. Gallardo, C. Parés, High order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. Applications to shallow-water systems, Math. Comput. 75 (255) (2006) 1103–1135. https://doi.org/10.1090/s0025-5718-06-01851-5

[15] M.J. Castro, T. Morales de Luna, C. Parés, Chapter 6 - Well-balanced schemes and path-conservative numerical methods, in: R. Abgrall, C.-W. Shu (Eds.), Handbook of Numerical Methods for Hyperbolic Problems, 18 of *Handbook of Numerical Analysis*, Elsevier, 2017, pp. 131–175.

[16] M.J. Castro, A. Pardo, C. Parés, E.F. Toro, On some fast well-balanced first order solvers for nonconservative systems, Math. Comput. 79 (2010) 1427–1472. https://doi.org/10.1090/S0025-5718-09-02317-5

[17] M.J. Castro, C. Parés, G. Puppo, G. Russo, Central schemes for nonconservative hyperbolic systems, SIAM J. Sci. Comput. 34 (5) (2012) B523–B558.

[18] R. Borges, M. Carmona, B. Costa, W.S. Don, An improved weighted essentially non-oscillatory scheme for hyperbolic conservation laws, J. Comput. Phys. 227 (6) (2008) 3191–3211.

[19] M. Castro, B. Costa, W.S. Don, High order weighted essentially non-oscillatory WENO-Z schemes for hyperbolic conservation laws, J. Comput. Phys. 230 (5) (2011) 1766–1792. https://doi.org/10.1016/j.jcp.2010.11.028

[20] A.K. Henrick, T.D. Aslam, J.M. Powers, Mapped weighted essentially non-oscillatory schemes: achieving optimal order near critical points, J. Comput. Phys. 207 (2) (2005) 542–567. https://doi.org/10.1016/j.jcp.2005.01.023

[21] G.-S. Jiang, C.-W. Shu, Efficient implementation of weighted ENO schemes, J. Comput. Phys. 126 (1) (1996) 202–228.

[22] C.-W. Shu, Essentially Non-Oscillatory and Weighted Essentially Non-Oscillatory Schemes for Hyperbolic Conservation Laws, Springer Berlin Heidelberg, 1998, p. 325–432. https://doi.org/10.1007/bfb0096355

[23] J. Zhu, J. Qiu, A new type of finite volume WENO schemes for hyperbolic conservation laws, J. Sci. Comput. 73 (2–3) (2017) 1338–1359. https://doi.org/10.1007/s10915-017-0486-8

[24] D.S. Balsara, D. Bhoriya, C.-W. Shu, H. Kumar, Efficient finite difference WENO scheme for hyperbolic systems with non-conservative products, Commun. Appl. Math. Comput. 6 (2) (2024) 907–962. https://doi.org/10.1007/s42967-023-00275-9

[25] P.L. Roe, Approximate Riemann solvers, parameter vectors, and difference schemes, J. Comput. Phys. 135 (2) (1997) 250–258. https://doi.org/10.1006/jcph.1997.5705

[26] I. Toumi, A weak formulation of Roe's approximate Riemann solver, J. Comput. Phys. 102 (2) (1992) 360–373. https://doi.org/10.1016/0021-9991(92)90378-C

[27] A. Bermúdez, M.E. Vázquez-Cendón, Upwind methods for hyperbolic conservation laws with source terms, Comput. Fluids 23 (8) (1994) 1049–1071.

[28] E. Audusse, F. Bouchut, M.-O. Bristeau, R. Klein, B. Perthame, A fast and stable well-Balanced scheme with hydrostatic reconstruction for shallow water flows, SIAM J. Sci. Comput. 25 (6) (2004) 2050–2065.

[29] M.J. Castro, A. Pardo, C. Parés, Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique, Math. Models Methods Appl. Sci. 17 (12) (2007) 2055–2113.

[30] R. Abgrall, S. Karni, Two-Layer shallow water system: a relaxation approach, SIAM J. Sci. Comput. 31 (3) (2009) 1603–1627. https://doi.org/10.1137/06067167X

[31] R.J. LeVeque, Balancing source terms and flux gradients in high-resolution godunov methods: the quasi-steady wave-propagation algorithm, J. Comput. Phys. 146 (1) (1998) 346–365.

[32] 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 (1) (2005) 206–227.

[33] F. Bouchut, V. Zeitlin, A robust well-balanced scheme for multi-layer shallow water equations, Discrete Contin. Dyn. Syst. B 13 (4) (2010) 739–758. https://doi.org/10.3934/dcdsb.2010.13.739
[34] J.M. Greenberg, A.Y. Leroux, A well-balanced scheme for the numerical processing of source terms in hyperbolic equations, SIAM J. Numer. Anal. 33 (1) (1996) 1–16.
[35] K.T. Mandli, A numerical method for the two layer shallow water equations with dry states, Ocean Modell. 72 (2013) 80–91. https://doi.org/10.1016/j.ocemod.2013.08.001
[36] M.J. Castro, C. Parés, Well-balanced high-order finite volume methods for systems of balance laws, J. Sci. Comput. 82(2), 48 (2020).
[37] C. Parés, C. Parés-Pulido, Well-balanced high-order finite difference methods for systems of balance laws, J. Comput. Phys. 425 (2021) 109880. https://doi.org/10.1016/j.jcp.2020.109880
[38] M.J. Castro, P.G. LeFloch, M.L. Muñoz-Ruiz, C. Parés, Why many theories of shock waves are necessary: convergence error in formally path-consistent schemes, J. Comput. Phys. 227 (17) (2008) 8107–8129. https://doi.org/10.1016/j.jcp.2008.05.012
[39] A. Beljadid, P. LeFloch, Siddhartha, C. Parés, Schemes with well-controlled dissipation. hyperbolic systems in nonconservative form, Commun. Comput. Phys. 21 (4) (2018) 913–946.
[40] E. Pimentel-García, M.J. Castro, C. Chalons, T. Morales de Luna, C. Parés, In-cell discontinuous reconstruction path-conservative methods for non conservative hyperbolic systems - second-order extension, J. Comput. Phys. 459 (2022) 111152. https://doi.org/10.1016/j.jcp.2022.111152
[41] E. Pimentel-García, M.J. Castro, C. Chalons, C. Parés, High-order in-cell discontinuous reconstruction path-conservative methods for nonconservative hyperbolic systems – DR.MOOD generalization, Numer. Methods Partial Differ. Equ. 40 (6) (2024) e23133.
[42] A. Harten, J.M. Hyman, Self adjusting grid methods for one-dimensional hyperbolic conservation laws, J. Comput. Phys. 50 (2) (1983) 235–269. https://doi.org/10.1016/0021-9991(83)90066-9
[43] C.-W. Shu, S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, J. Comput. Phys. 77 (2) (1988) 439–471.
[44] Y. Cao, A. Kurganov, Y. Liu, V. Zeitlin, Flux globalization based well-balanced path-conservative central-upwind scheme for two-layer thermal rotating shallow water equations, J. Comput. Phys. 474 (2023) 111790. https://doi.org/10.1016/j.jcp.2022.111790
[45] C. Parés, M.J. Castro, On the well-balance property of Roe's method for nonconservative hyperbolic systems. Applications to shallow-water systems, ESAIM: Modélisation mathématique et analyse numérique 38 (5) (2004) 821–852. https://doi.org/10.1051/m2an:2004041
[46] M.J. Castro, E.D. Fernández-Nieto, A.M. Ferreiro, J.A. García-Rodríguez, C. Parés, High order extensions of roe schemes for two-dimensional nonconservative hyperbolic systems, J. Sci. Comput. 39 (1) (2009) 67–114.
[47] J. Schijf, J.C. Schönfled, Theoretical considerations on the motion of salt and fresh water, Proceedings Minnesota International Hydraulic Convention (1953) 321–333. https://api.semanticscholar.org/CorpusID:14860277.
[48] N. Krvavica, M. Tuhtan, G. Jelenić, Analytical implementation of Roe solver for two-layer shallow water equations with accurate treatment for loss of hyperbolicity, Adv. Water Resour. 122 (2018) 187–205.
[49] E.D. Fernández-Nieto, M.J. Castro, C. Parés, On an intermediate field capturing Riemann solver based on a parabolic viscosity matrix for the two-layer shallow water system, J. Sci. Comput. 48 (2011) 117–140.
[50] A. Kurganov, G. Petrova, Central-upwind schemes for two-layer shallow water equations, SIAM J. Sci. Comput. 31 (3) (2009) 1742–1773. https://doi.org/10.1137/080719091
[51] I. Gómez-Bueno, M.J. Castro, C. Parés, G. Russo, Collocation methods for high-order well-balanced methods for systems of balance laws, Mathematics 9 (15) (2021). https://doi.org/10.3390/math9151799
[52] D.S. Balsara, S. Garain, C.-W. Shu, An efficient class of WENO schemes with adaptive order, J. Comput. Phys. 326 (2016) 780–804. https://doi.org/10.1016/j.jcp.2016.09.009
[53] D. Levy, G. Puppo, G. Russo, On the behavior of the total variation in CWENO methods for conservation laws, Appl. Numer. Math. 33 (1) (2000) 407–414. https://doi.org/10.1016/S0168-9274(99)00107-5