



**HAL**  
open science

## A well-balanced scheme for the shallow-water equations with topography.

Victor Michel-Dansac, Christophe Berthon, Stéphane Clain, Françoise Foucher

### ► To cite this version:

Victor Michel-Dansac, Christophe Berthon, Stéphane Clain, Françoise Foucher. A well-balanced scheme for the shallow-water equations with topography.. 2015. hal-01201825v1

**HAL Id: hal-01201825**

**<https://hal.science/hal-01201825v1>**

Preprint submitted on 18 Sep 2015 (v1), last revised 10 Jan 2017 (v2)

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

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

# A well-balanced scheme for the shallow-water equations with topography.

Victor Michel-Dansac<sup>a,\*</sup>, Christophe Berthon<sup>a</sup>, Stéphane Clain<sup>b</sup>, Françoise Foucher<sup>a,c</sup>

<sup>a</sup>Laboratoire de Mathématiques Jean Leray, CNRS UMR 6629, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, France

<sup>b</sup>Centre of Mathematics, Minho University, Campus de Gualtar - 4710-057 Braga, Portugal

<sup>c</sup>École Centrale de Nantes, 1 rue de La Noë, BP 92101 44321 Nantes Cedex 3, France

---

## Abstract

A non-negativity preserving and well-balanced scheme that exactly preserves all the steady states of the shallow water system, including the moving ones, is proposed. In addition, the scheme must deal with vanishing water heights and transitions between wet and dry areas. A Godunov-type method is derived by using a relevant average of the source terms in order to enforce the required well-balance property. A second-order well-balanced MUSCL extension is also designed. Numerical experiments are carried out to check the properties of the scheme and assess the ability to exactly preserve all the steady states.

*Keywords:* shallow-water equations, Godunov-type schemes, well-balanced schemes, moving steady states

*2000 MSC:* 65M08, 65M12

---

## 1. Introduction

During the last two decades, numerous schemes have been derived to preserve exactly (or, at least, accurately) the lake at rest. For instance, we refer the reader to the previous work by Bermudez and Vasquez [3], and next to Greenberg and LeRoux [24]. These works introduced the definition and the relevance of the well-balanced procedure. Such approaches were extended by Gosse [21] for nonlinear systems, by involving a nonlinear equation to be solved. More recently, in [1], a simplification (by enforcing vanishing velocities) of the work by Gosse [21] was proposed, yielding the so-called hydrostatic reconstruction (see [8, 11, 15, 16, 27, 30, 31, 32] for related works).

---

\*Corresponding author

*Email addresses:* `victor.michel-dansac@univ-nantes.fr` (Victor Michel-Dansac), `christophe.berthon@univ-nantes.fr` (Christophe Berthon), `clain@math.uminho.pt` (Stéphane Clain), `francoise.foucher@ec-nantes.fr` (Françoise Foucher)

The steady states for the shallow-water equations with nonzero discharge are known to be more delicate to exactly capture than the usual lake at rest. The critical role played by these specific solutions was illustrated in [23], where several benchmarks were exhibited. Next, in [21], a pioneer fully well-balanced scheme was designed to deal with these sensitive steady states. This numerical technique is based on a suitable resolution of the Bernoulli equation. Next, several methods preserving the moving steady states were designed by involving high-order accurate techniques (see [10, 33, 34, 39, 40]). More recently, in [4, 5], the authors have proposed an extension of the work by Gosse [21] in order to deal with Godunov-type schemes. Such Godunov-type schemes (see [18, 19]) are based on approximate Riemann solvers, whose intermediate states are obtained by solving a Bernoulli-type equation. This process allows the authors to get a fully well-balanced scheme preserving the entropy stability. Since the resolution of the Bernoulli-type equation is costly, in the present paper, we adopt a linear formulation to deal with a general form of well-balanced states.

In the present paper, we propose a generic approach to provide a well-balanced scheme. As a consequence, the objectives of the paper are to derive a numerical scheme to approximate the solutions of the shallow-water equations, and that satisfies the following properties:

1. exact preservation of all the steady states for the system with topography;
2. non-negativity preservation for the water height;
3. ability to model transitions between wet and dry areas.

The paper is organized as follows. First, we devote Section 2 to the study of smooth steady states for the shallow-water equations with topography. Some comments are also given to establish steady states with a dry/wet transition. Afterwards, Section 3 is devoted to the construction of a Godunov-type scheme. We then show a general procedure to obtain a well-balanced scheme for the shallow-water equations with a source term on the discharge equation. At this level, the definition of the source terms is not specified and the resulting scheme depends on the source term discretization. To put it in other words, the well-balance property is obtained according to the PDE governing the steady states. Next, in Section 4, we fix the source term according to the topography, to get a suitable discretization of this source term in order to exactly preserve the steady states governed by the topography. An extension to apply the scheme to dry/wet transitions is also designed. A second-order extension with a MUSCL technique is then studied in Section 5, and we conclude the paper with numerical experiments in Section 6. The paper is achieved in Section 7 by a brief conclusion.

## 2. Steady states for the shallow-water equations with topography

### 2.1. The shallow-water model

This paper is devoted to designing a numerical scheme to approximate the solutions of the shallow-water equations with topography. The model of interest

is governed by the following system:

$$\begin{cases} \partial_t h + \partial_x q & = 0, \\ \partial_t q + \partial_x \left( \frac{q^2}{h} + \frac{1}{2}gh^2 \right) & = -gh\partial_x Z. \end{cases} \quad (2.1)$$

These equations describe the behavior of water in a one-dimensional channel with a non-flat bottom. The modeled quantities are the water height  $h(x, t) \geq 0$  and its depth-averaged discharge  $q(x, t)$ . The depth-averaged velocity  $u$  of the water is such that  $q = hu$ . The constant  $g > 0$  stands for the gravity, while the function  $x \mapsto Z(x)$  is the topography. We define the admissible states space by

$$\Omega = \{W = {}^t(h, q) \in \mathbb{R}^2 ; h \geq 0, q \in \mathbb{R}\}.$$

Let us note that the water height may vanish, which accounts for dry areas.

For the sake of simplicity in the notations, we rewrite (2.1) under the following condensed form:

$$\partial_t W + \partial_x F(W) = s(W), \quad W \in \Omega, \quad (2.2)$$

where

$$W = \begin{pmatrix} h \\ q \end{pmatrix}, \quad F(W) = \begin{pmatrix} q \\ \frac{q^2}{h} + \frac{1}{2}gh^2 \end{pmatrix}, \quad s(W) = \begin{pmatrix} 0 \\ -gh\partial_x Z \end{pmatrix}.$$

The homogeneous system deriving from canceling the source terms in (2.1) turns out to be hyperbolic with characteristic velocities given by  $u - c$  and  $u + c$  (see [9, 20, 28] for instance), where  $c$  is the sound speed, defined by

$$c = \sqrt{gh}. \quad (2.3)$$

## 2.2. Smooth steady states with positive water heights

In the present paper, we focus on the smooth steady state solutions of (2.2), which thus satisfy  $\partial_t W = 0$ . We assume a smooth topography, that is to say continuous and differentiable in space. Such steady states are governed by the following set of partial differential equations:

$$\begin{cases} \partial_x q & = 0, \\ \partial_x \left( \frac{q^2}{h} + \frac{1}{2}gh^2 \right) & = -gh\partial_x Z. \end{cases} \quad (2.4)$$

The first equation immediately imposes a constant discharge  $q = q_0$ . The second equation of (2.4) then becomes:

$$\partial_x \left( \frac{q_0^2}{h} + \frac{1}{2}gh^2 \right) = -gh\partial_x Z. \quad (2.5)$$

This equation is a nonlinear ordinary differential equation, with unknown  $h$ . From (2.5), we can extract solutions in specific cases. Some of them have been

extensively studied. For instance, by adopting  $q_0 = 0$  and assuming  $h > 0$ , we recover the notable *lake at rest* steady state (see for instance [2, 8, 9, 12, 17, 24]), defined by

$$\begin{cases} q &= 0, \\ h + Z &= \text{cst}. \end{cases} \quad (2.6)$$

We now turn to the study of the equation (2.5) for  $q_0 \neq 0$ . Since functions are smooth, we rewrite (2.5) under the following algebraic form:

$$\partial_x \left( \frac{q_0^2}{2h^2} + g(h + Z) \right) = 0. \quad (2.7)$$

Note that (2.7) is nothing but a statement of Bernoulli's principle. Now, let us consider a fixed  $x_0 \in \mathbb{R}$ . We introduce  $h(x_0) = h_0$  and  $Z(x_0) = Z_0$ . For all  $x \in \mathbb{R}$ , integration (2.7) over  $(x_0, x)$  provides

$$\frac{q_0^2}{2h^2} + g(h + Z) - \frac{q_0^2}{2h_0^2} - g(h_0 + Z_0) = 0. \quad (2.8)$$

To shorten the notations, we set

$$\xi(h; Z, h_0, q_0, Z_0) = \frac{q_0^2}{2h^2} + g(h + Z) - \frac{q_0^2}{2h_0^2} - g(h_0 + Z_0).$$

such that (2.8) rewrites

$$\xi(h; Z, h_0, q_0, Z_0) = 0. \quad (2.9)$$

Here,  $Z$  is effectively a parameter of the function  $\xi$  since the topography is not an unknown of the problem.

Now, we study  $h = h(x)$ , solution of (2.7), or equivalently (2.9), with initial condition  $h(x_0) = h_0$ . In order to exhibit the solutions of (2.9), we first differentiate  $\xi$  with respect to  $h$ :

$$\frac{\partial \xi}{\partial h}(h; Z, h_0, q_0, Z_0) = -\frac{q_0^2}{h^3} + g.$$

If  $h > h_c$ , the function  $h \mapsto \xi(h; Z, h_0, q_0, Z_0)$  is strictly increasing, while if  $h < h_c$ ,  $h \mapsto \xi(h; Z, h_0, q_0, Z_0)$  is a strictly decreasing function, with  $h_c$  such that  $\frac{\partial \xi}{\partial h}(h_c) = 0$ , given by

$$h_c = \left( \frac{q_0^2}{g} \right)^{1/3}. \quad (2.10)$$

To define solutions  $h \in \mathbb{R}_+^*$  of (2.9), we evaluate the sign of  $\xi$ . After straightforward computations, this function is proven to satisfy the following limits:

- $\lim_{h \rightarrow +\infty} \xi(h; Z, h_0, q_0, Z_0) = +\infty$ ,
- $\lim_{h \rightarrow 0^+} \xi(h; Z, h_0, q_0, Z_0) = +\infty$ .

Moreover, the function  $\xi$  is immediately shown to verify the following evaluation for  $h = h_c$ :

$$\xi(h_c; Z, h_0, q_0, Z_0) = \frac{q_0^2}{2} \left( \frac{3}{h_c^2} - \frac{1}{h_0^2} \right) + g(Z - Z_0 - h_0).$$

Equipped with these properties of  $\xi$ , we have the following statement.

**Lemma 1.** *Assume  $h > 0$  and  $q_0 \neq 0$ . Thus,  $h_c > 0$  according to (2.10), and the following properties hold:*

- (i) *If  $\xi(h_c; Z, h_0, q_0, Z_0) > 0$ , then there is no solution to the equation (2.9).*
- (ii) *If  $\xi(h_c; Z, h_0, q_0, Z_0) = 0$ , then the equation (2.9) admits a unique solution. This solution,  $h = h_c$ , is a double root of the function  $h \mapsto \xi(h; Z, h_0, q_0, Z_0)$ .*
- (iii) *If  $\xi(h_c; Z, h_0, q_0, Z_0) < 0$ , then the equation (2.9) admits two distinct solutions,  $h^{sup} \in (0, h_c)$  and  $h^{sub} \in (h_c, +\infty)$ .*

*Proof.* The proof of this lemma relies on using all the properties of  $\xi$  we have obtained above. We recall that  $\xi$  tends to infinity as  $h$  tends to 0 or infinity. Moreover, the function  $\xi$  admits a unique minimum, reached for  $h = h_c$ .

Equipped with these properties, the proofs of (i), (ii) and (iii) are obvious. The proof is thus achieved.  $\square$

*Remark 1.* Defining the Froude number as  $Fr = u/c$ , we have for steady states  $Fr = q_0/\sqrt{gh^3}$ . Thus,  $h^{sub} > h_c$  corresponds to the subcritical case ( $Fr < 1$ ) while  $h^{sup} < h_c$  is to the supercritical solution ( $Fr > 1$ ).

*Remark 2.* Assume that, for all  $x \in \mathbb{R}$ ,  $h(x) \neq h_c$ . From (2.5), the derivative of  $h$  with respect to  $x$  writes

$$\partial_x h = \frac{h^3 \partial_x Z}{h_c^3 - h^3}.$$

Therefore, if the solution is subcritical, i.e.  $h(x) > h_c$ , then the sign of  $\partial_x h$  is the opposite of the sign of  $\partial_x Z$ , whereas, the sign of  $\partial_x h$  is that of  $\partial_x Z$  if the solution is supercritical. These results are in accordance with the subcritical and supercritical experiments presented in [23] for instance.

As a consequence, we have obtained the general form of steady states for the shallow-water equations (2.1) assuming  $q_0 \neq 0$  and  $h > 0$ .

### 2.3. Smooth steady states with dry areas

Now, we turn to the study of steady states with dry areas, that is to say areas where the water height is zero. Indeed, the condition  $h = 0$  does not imply immediately that  $q = 0$  since the velocity may be unbounded. The goal of this subsection is to obtain a characterization of a steady state where such an area is present.

We begin by defining the kinetic energy in a wet area where  $h > 0$ , as follows:

$$E = \frac{1}{2}qu.$$

In order to recover some essential physical properties, the kinetic energy has to be bounded, i.e.

$$\|E\|_\infty < +\infty. \tag{2.11}$$

To study a dry/ wet transition, we recall that  $h$ ,  $q$  and  $Z$  are continuous differentiable functions, and introduce two points in  $\mathbb{R}$ ,  $x_D$  and  $x_W$ . Since  $h$  is continuous, the dry zone is a closed or half-closed domain, and we denote by  $x_D$  a point on its boundary, such that  $h(x_D) = 0$ . Moreover, we define  $x_W$  such that  $h(x_W) > 0$ . The three cases corresponding to this situation are displayed on Figure 1. In all three cases, the point  $x_D$  corresponds to the transition between a wet area and a dry area.

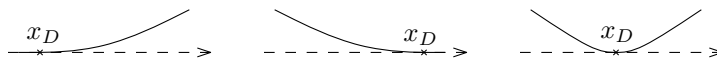


Figure 1: Graph of the height function in the three different configurations of dry/wet transitions.

For the sake of simplicity, we study the left case of Figure 1. Similar sequences of arguments can be applied to the other two cases. In this case, we have  $x_W > x_D$ , and  $h$  is an increasing function.

**Proposition 2.** *Assume the kinetic energy is bounded. Then, as soon as a dry area is involved, smooth steady states must be at rest.*

*Proof.* We begin by proving that  $q(x_D) = 0$ . Assume that  $q(x_D) \neq 0$ . For  $x_W > x_D$ , we have  $q(x_W) = h(x_W)u(x_W)$ . Therefore, going to the limit as  $x_W$  tends to  $x_D$ , we obtain  $|u(x_W)| \rightarrow +\infty$ . However, since the kinetic energy at point  $x_W$  is  $E(x_W) = \frac{1}{2}q(x_W)u(x_W)$ , the limit of the kinetic energy as  $x_W$  tends to  $x_D$  is  $+\infty$ . This result contradicts the boundedness assumption (2.11). Therefore, we deduce that  $q(x_D) = 0$ .

Now, for a steady state, the relation  $\partial_x q = 0$  holds. Since  $q(x_D) = 0$ , we obtain that  $q \equiv 0$ . Therefore, the only steady state with a dry/wet transition is a steady state at rest, where  $q_0 = 0$ . The proof is thus achieved.  $\square$

### 3. A generic well-balanced scheme

We propose a generic strategy to derive numerical schemes that able to exactly capture all the steady state solutions. To address such an issue, we first briefly recall some important steps of the construction of a Godunov-type scheme based on a two-state Riemann solver [25] (see also [9, 36]). Then, the scheme is extended to the shallow-water equations with a general source term in order to preserve the steady states.

### 3.1. Godunov-type schemes

First, we introduce the discretization of space and time. Let  $\Delta x$  be the space step, assumed to be constant, and  $\Delta t$  the time step. The space discretization consists in cells  $(x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$ , of volume  $\Delta x$  and centered at  $x_i = x_{i-\frac{1}{2}} + \frac{\Delta x}{2}$ , for all  $i \in \mathbb{Z}$ . Now, we assume known a piecewise constant approximation of  $W(x, t)$  at time  $t^n$ , denoted by  $W^\Delta(x, t^n)$  and defined, for all  $x \in (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$ , by  $W^\Delta(x, t^n) = W_i^n$ . In order to evolve this approximation in time, we suggest a Godunov-type finite volume method. The scheme is obtained by setting an approximate Riemann solver, denoted by  $\widetilde{W}(\frac{x}{t}; W_L, W_R)$ , at each interface  $x_{i+\frac{1}{2}}$ . This solver is designed in order to mimic the exact solution of a Riemann problem for the initial system (2.1).

From now on, let us consider an approximate Riemann solver made of four constant states, separated by three discontinuities. We define it as follows (see Figure 2):

$$\widetilde{W}\left(\frac{x}{t}; W_L, W_R\right) = \begin{cases} W_L & \text{if } x/t < \lambda_L, \\ W_L^* & \text{if } \lambda_L < x/t < 0, \\ W_R^* & \text{if } 0 < x/t < \lambda_R, \\ W_R & \text{if } x/t > \lambda_R, \end{cases} \quad (3.1)$$

where  $\lambda_L$  and  $\lambda_R$  denote some characteristic velocities, and  $W_L^*$  and  $W_R^*$  stand for the intermediate states, to be detailed later. To enforce  $\lambda_L < 0$  and  $\lambda_R > 0$ , we choose the following expressions of the characteristic velocities (see for instance [35] and references therein), where  $c$  denotes the sound speed, defined by (2.3):

$$\begin{aligned} \lambda_L &= \min(-|u_L| - c_L, -|u_R| - c_R, -\varepsilon_\lambda), \\ \lambda_R &= \max(|u_L| + c_L, |u_R| + c_R, \varepsilon_\lambda), \end{aligned} \quad (3.2)$$

with  $\varepsilon_\lambda$  to be fixed in the numerical applications.

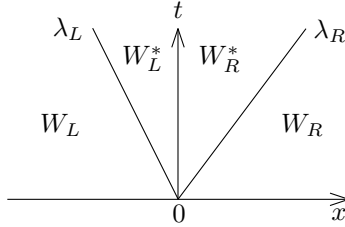


Figure 2: Structure of the chosen approximate Riemann solver.

*Remark 3.* Note that  $\lambda_L < 0$  and  $\lambda_R > 0$  even for the supercritical case, where both velocities should have the same sign. Because of this choice, both intermediate states  $W_L^*$  and  $W_R^*$  are always considered. This is a very important ingredient in the construction of the well-balanced approximate Riemann solver.

Equipped with  $\widetilde{W}$ , we determine the updated states  $W_i^{n+1}$ . Indeed, let us introduce the juxtaposition of the approximate Riemann solvers stated at each



interface (see Figure 3). Such a juxtaposition is denoted by  $W^\Delta(x, t^n + t)$ , for all  $t \in (0, \Delta t]$ . As a consequence, we have

$$\forall x \in [x_i, x_{i+1}), \forall t \in (0, \Delta t], W^\Delta(x, t^n + t) = \widetilde{W} \left( \frac{x - x_{i+\frac{1}{2}}}{t}; W_i^n, W_{i+1}^n \right). \quad (3.3)$$

Moreover, to prevent the waves from interacting, we impose the following CFL-like condition:

$$\Delta t \leq \frac{\Delta x}{2\Lambda}, \quad \text{where } \Lambda = \max_{i \in \mathbb{Z}} \left( \lambda_{i+\frac{1}{2}}^L, \lambda_{i+\frac{1}{2}}^R \right). \quad (3.4)$$

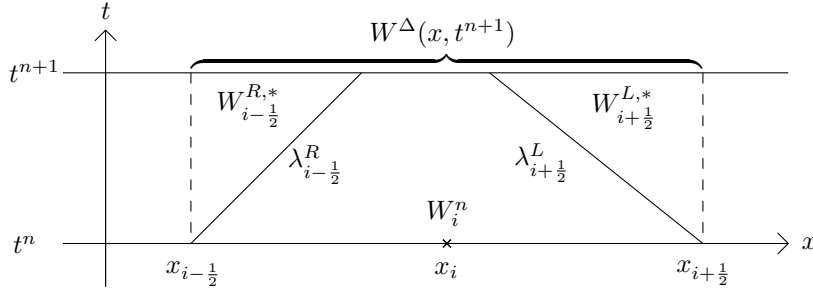


Figure 3: The full Godunov-type scheme using the prescribed approximate Riemann solver.

We now define  $W_i^{n+1}$  as the average of  $W^\Delta(x, t^n + \Delta t)$  over the cell  $(x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$ , as follows:

$$W_i^{n+1} = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} W^\Delta(x, t^n + \Delta t) dx.$$

Arguing the definition (3.3) of  $W^\Delta$ , we get:

$$\begin{aligned} W_i^{n+1} &= \frac{1}{\Delta x} \int_{-\Delta x/2}^0 \widetilde{W} \left( \frac{x}{\Delta t}; W_{i-1}^n, W_i^n \right) dx \\ &\quad + \frac{1}{\Delta x} \int_0^{\Delta x/2} \widetilde{W} \left( \frac{x}{\Delta t}; W_i^n, W_{i+1}^n \right) dx. \end{aligned} \quad (3.5)$$

Finally, from the definition (3.1) of the approximate Riemann solver  $\widetilde{W}$ , we have, after straightforward computations and adopting clear notations, the following expression of  $W_i^{n+1}$ :

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x} \left[ \lambda_{i+\frac{1}{2}}^L \left( W_{i+\frac{1}{2}}^{L,*} - W_i^n \right) - \lambda_{i-\frac{1}{2}}^R \left( W_{i-\frac{1}{2}}^{R,*} - W_i^n \right) \right]. \quad (3.6)$$

Consequently, the scheme (3.6) is fully defined by giving explicit values to the intermediate states  $W_{i+\frac{1}{2}}^{L,*}$  and  $W_{i-\frac{1}{2}}^{R,*}$  that ensure the scheme is consistent and preserves all the steady states.

### 3.2. A well-balanced approximate Riemann solver for general source terms

We are now interested in deriving generic well-balanced schemes, that is to say schemes that preserve the steady states. Concerning the source term, at this level, we do not specify the definition of  $s(W) = {}^t(0, S(W))$ . Thus,  $S(W)$  may represent the contribution of the topography and/or any other source term on the discharge equation. The well-balanced scheme is obtained as soon as relevant definitions of the intermediate states  $W_L^* = {}^t(h_L^*, q_L^*)$  and  $W_R^* = {}^t(h_R^*, q_R^*)$  are prescribed. We remark that, to find such  $W_L^*$  and  $W_R^*$ , we need four equations characterizing the four unknowns  $h_L^*$ ,  $h_R^*$ ,  $q_L^*$  and  $q_R^*$ . To address such an issue, we look for intermediate states that enforce both consistency and well-balancedness.

The generic non-conservative shallow-water system is given by

$$\begin{cases} \partial_t h + \partial_x q & = 0, \\ \partial_t q + \partial_x \left( \frac{q^2}{h} + \frac{1}{2} g h^2 \right) & = S(W). \end{cases} \quad (3.7)$$

Here, the steady states are governed by

$$\begin{cases} q = q_0, \\ \partial_x \left( \frac{q_0^2}{h} + \frac{1}{2} g h^2 \right) = S(W). \end{cases} \quad (3.8)$$

We note that the ordinary differential equation (3.8) cannot be rewritten under an algebraic form for any generic source term  $S(W)$ . However, in the specific case of the topography source term, we can rewrite (3.8) as an algebraic relation. For instance, the lake at rest is given by (2.6), and the moving steady states are given by (2.7). Such an algebraic relation will allow us to obtain a relevant numerical approximation of the source term.

*Remark 4.* If no algebraic relation is available, we can still derive a suitable approximation of the generic source term  $S(W)$  if it can be split into a sum of source terms for which (3.8) rewrites under an algebraic form.

By the scheme (3.6), we observe that the solution is obviously stationary, i.e.  $W_i^{n+1} = W_i^n$  for all  $i \in \mathbb{Z}$ , if we have  $W_L^* = W_L$  and  $W_R^* = W_R$ . Therefore, we will look for  $W_L^*$  and  $W_R^*$  such that  $W_L^* = W_L$  and  $W_R^* = W_R$  as soon as  $W_L$  and  $W_R$  define a steady state. Here, the pair  $(W_L, W_R)$  is said to define a steady state if the identity (3.8) is satisfied in a sense to be determined later. Such intermediate states will enforce the well-balancedness of our scheme. To that end, we state the following well-balance principle:

**Principle (WB).** The intermediate states  $W_L^*$  and  $W_R^*$  are such that  $W_L^* = W_L$  and  $W_R^* = W_R$  as soon as  $W_L$  and  $W_R$  define a steady state.

### 3.2.1. Consistency

We introduce a necessary condition for the scheme to be consistent with (3.7). We denote by  $W_{\mathcal{R}}\left(\frac{x}{\Delta t}; W_L, W_R\right)$  the exact solution of the Riemann problem for (3.7). From [25], the consistency condition states that the average of the approximate Riemann solver  $\widetilde{W}$ , defined by (3.1), over a cell has to be equal to the average of the exact solution  $W_{\mathcal{R}}$  over the same cell. As a consequence, we have to impose the following equality:

$$\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}; W_L, W_R\right) dx = \frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} W_{\mathcal{R}}\left(\frac{x}{\Delta t}; W_L, W_R\right) dx. \quad (3.9)$$

On the one hand, we rewrite the right-hand side of (3.9) as:

$$\begin{aligned} \frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} W_{\mathcal{R}}\left(\frac{x}{\Delta t}; W_L, W_R\right) dx &= \frac{1}{2} (W_L + W_R) - \frac{\Delta t}{\Delta x} (F(W_R) - F(W_L)) \\ &\quad + \frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \int_0^{\Delta t} s\left(W_{\mathcal{R}}\left(\frac{x}{t}; W_L, W_R\right)\right) dt dx. \end{aligned} \quad (3.10)$$

On the other hand, a direct computation of the integral in the left-hand side of (3.9) yields:

$$\begin{aligned} \frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \widetilde{W}\left(\frac{x}{\Delta t}; W_L, W_R\right) dx &= \frac{1}{2} (W_L + W_R) \\ &\quad + \lambda_L \frac{\Delta t}{\Delta x} (W_L - W_L^*) - \lambda_R \frac{\Delta t}{\Delta x} (W_R - W_R^*). \end{aligned} \quad (3.11)$$

Recall that the first component of  $s(W)$  is 0. Therefore, combining (3.10) with (3.11), within (3.9), leads us to the following relations:

$$\lambda_R h_R^* - \lambda_L h_L^* = \lambda_R h_R - \lambda_L h_L - [q], \quad (3.12a)$$

$$\begin{aligned} \lambda_R q_R^* - \lambda_L q_L^* &= \lambda_R q_R - \lambda_L q_L - \left[ \frac{q^2}{h} + \frac{1}{2} g h^2 \right] \\ &\quad + \frac{1}{\Delta t} \int_{-\Delta x/2}^{\Delta x/2} \int_0^{\Delta t} S\left(W_{\mathcal{R}}\left(\frac{x}{t}; W_L, W_R\right)\right) dt dx, \end{aligned} \quad (3.12b)$$

where  $[X] = X_R - X_L$ . For the sake of simplicity, let us now introduce the following notations:

$$(\lambda_R - \lambda_L) h_{HLL} = \lambda_R h_R - \lambda_L h_L - [q], \quad (3.13a)$$

$$(\lambda_R - \lambda_L) q_{HLL} = \lambda_R q_R - \lambda_L q_L - \left[ \frac{q^2}{h} + \frac{1}{2} g h^2 \right]. \quad (3.13b)$$

Let us emphasize that  $h_{HLL}$  and  $q_{HLL}$  coincide with the well-known intermediate state introduced by Harten, Lax and van Leer in [25]. We remark that

$h_{HLL} > 0$  for large enough  $-\lambda_L$  and  $\lambda_R$ . Thanks to these notations, (3.12) can be rewritten as follows:

$$\lambda_R h_R^* - \lambda_L h_L^* = (\lambda_R - \lambda_L) h_{HLL}, \quad (3.14a)$$

$$\begin{aligned} \lambda_R q_R^* - \lambda_L q_L^* &= (\lambda_R - \lambda_L) q_{HLL} \\ &+ \frac{1}{\Delta t} \int_{-\Delta x/2}^{\Delta x/2} \int_0^{\Delta t} S\left(W_{\mathcal{R}}\left(\frac{x}{t}; W_L, W_R\right)\right) dt dx. \end{aligned} \quad (3.14b)$$

After [4, 5], we now choose  $q_L^* = q_R^*$ , and we denote this value by  $q^*$ . We will see later on that this choice allows the recovery of the required well-balance property. Using  $q^*$ , we rewrite (3.14b) under the form

$$q^* = q_{HLL} + \frac{1}{\lambda_R - \lambda_L} \frac{1}{\Delta t} \int_{-\Delta x/2}^{\Delta x/2} \int_0^{\Delta t} S\left(W_{\mathcal{R}}\left(\frac{x}{t}; W_L, W_R\right)\right) dt dx. \quad (3.15)$$

### 3.2.2. Well-balancedness parametrization

By adopting a non-explicit source term, given by  $S(W)$ , we have to introduce two parameters, denoted  $\bar{S}$  and  $\tilde{q}$ . These parameters are such that the solution characterized by  $h_L^*$ ,  $h_R^*$  and  $q^*$  corresponds to a well-balanced one according to (3.8), and a consistent one according to (3.14a) and (3.15). In the next section, the source term will be chosen as the topography, and we will impose the additional suitable algebraic relation to govern the steady states and thus to find both additional parameters.

We begin by introducing a first parameter  $\bar{S}$ . It is a consistent approximation of the mean value of the source term  $S(W)$ , and it implicitly depends on  $W_L$  and  $W_R$ . The approximation of the source term average is thus given by:

$$\frac{1}{\Delta x} \frac{1}{\Delta t} \int_{-\Delta x/2}^{\Delta x/2} \int_0^{\Delta t} S(W) dt dx \simeq \bar{S}. \quad (3.16)$$

We now plug  $\bar{S}$  into (3.15). As a consequence, we impose that the three unknowns  $h_L^*$ ,  $h_R^*$  and  $q^*$  satisfy the following two relations:

$$\lambda_R h_R^* - \lambda_L h_L^* = (\lambda_R - \lambda_L) h_{HLL}, \quad (3.17a)$$

$$q^* = q_{HLL} + \frac{\bar{S} \Delta x}{\lambda_R - \lambda_L}. \quad (3.17b)$$

A relevant definition of  $\bar{S}$  is required in order to fully determine  $q^*$ . Suitable expressions will be given in the next section for the specific case of varying topography.

Before completing the system (3.17) to fully determine the intermediate water heights  $h_L^*$  and  $h_R^*$ , let us assume that  $h_L \neq 0$  and  $h_R \neq 0$ . We need to specify discrete steady states associated to (3.8). Indeed, the left and right states,  $W_L$  and  $W_R$ , define a steady state if the following relations hold:

$$\begin{cases} q_L = q_R = q_0, \end{cases} \quad (3.18a)$$

$$\begin{cases} q_0^2 \left[ \frac{1}{h} \right] + \frac{g}{2} [h^2] = \bar{S} \Delta x. \end{cases} \quad (3.18b)$$

Such equations are nothing but a discretization of the steady relation (3.8). We rewrite (3.18b) as follows:

$$\alpha(h_R - h_L) = \bar{S} \Delta x, \quad \text{where} \quad \alpha = \frac{-q_0^2}{h_L h_R} + \frac{g}{2}(h_L + h_R). \quad (3.19)$$

Now, to complete the determination of the intermediate water heights, we adopt an extension of the above relation, given by:

$$\alpha(h_R^* - h_L^*) = \bar{S} \Delta x, \quad \text{where} \quad \alpha = \frac{-\tilde{q}^2}{h_L h_R} + \frac{g}{2}(h_L + h_R), \quad (3.20)$$

where we have introduced a parameter  $\tilde{q}(W_L, W_R)$  which must be such that  $\tilde{q} = q_0$  as soon as  $W_L$  and  $W_R$  define a steady state, i.e. as soon as the relations (3.18) are verified.

Therefore, (3.17a) and (3.20) define the following linear system:

$$\begin{cases} \lambda_R h_R^* - \lambda_L h_L^* = (\lambda_R - \lambda_L) h_{HLL}, \\ \alpha(h_R^* - h_L^*) = \bar{S} \Delta x, \end{cases}$$

from which we deduce

$$h_L^* = h_{HLL} - \frac{\lambda_R \bar{S} \Delta x}{\alpha(\lambda_R - \lambda_L)}, \quad (3.21a)$$

$$h_R^* = h_{HLL} - \frac{\lambda_L \bar{S} \Delta x}{\alpha(\lambda_R - \lambda_L)}. \quad (3.21b)$$

### 3.3. Properties of the approximate Riemann solver

The goal of this subsection is to assert the properties verified by the approximate Riemann solver. The intermediate states have been chosen to yield a consistent and well-balanced approximate Riemann solver. In addition, it is possible to preserve the positivity of the water height while retaining the consistency and well-balancedness of the scheme. Indeed, let us emphasize that the definitions (3.21) do not ensure positive intermediate water heights. To address such an issue, we follow the procedure proposed in [2] (see also [6]) and presented in Figure 4. It consists in enforcing the positivity of  $h_L^*$  and  $h_R^*$ , while still ensuring that they satisfy the consistency relation (3.17a). To that end, we introduce the parameter  $\varepsilon$ , given by

$$\varepsilon = \min(h_L, h_R, h_{HLL}). \quad (3.22)$$

Note that  $h_L > 0$ ,  $h_R > 0$  and  $h_{HLL} > 0$ : hence  $\varepsilon > 0$ . If  $h_L^* < \varepsilon$ , we take  $h_L^* = \varepsilon$ , and  $h_R^*$  is chosen according to (3.17a), which guarantees  $h_R^* > 0$  (see Figure 4). A similar procedure is applied if  $h_R^* < \varepsilon$ . After the correction procedure, we have  $h_L^* \geq \varepsilon$  and  $h_R^* \geq \varepsilon$ .

We now state the expressions of the intermediate states, for given  $W_L$ ,  $W_R$ ,  $\bar{S}$  and  $\tilde{q}$ , with the positivity correction. The parameters  $\bar{S}$  and  $\tilde{q}$  will be defined

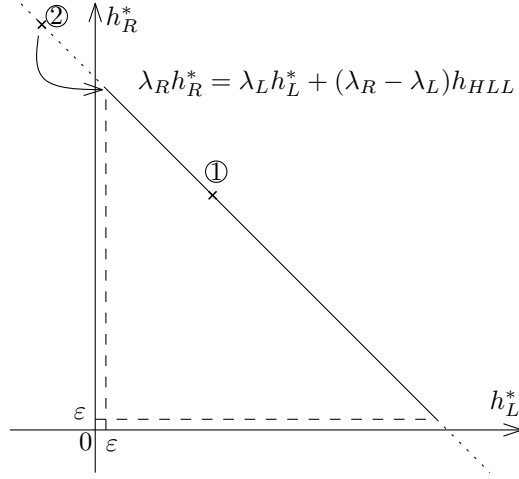


Figure 4: Procedure to ensure positive and consistent intermediate water heights, showing the line given by the consistency equation (3.17a). If the point  $(h_L^*, h_R^*)$  is at position 1, then  $h_L^*$  and  $h_R^*$  are not modified. However, if it is at position 2, we replace  $(h_L^*, h_R^*)$  with  $(\epsilon, (1 - \frac{\lambda_L}{\lambda_R})h_{HLL} + \frac{\lambda_L}{\lambda_R}\epsilon)$  according to (3.17a).

in the next section for a specific source term. The intermediate states of the approximate Riemann solver are then given by  $W_L^* = {}^t(h_L^*, q_L^*)$  and  $W_R^* = {}^t(h_R^*, q_R^*)$ , with

$$\alpha = \frac{-\tilde{q}^2}{h_L h_R} + \frac{g}{2}(h_L + h_R), \quad (3.23a)$$

$$q_L^* = q_R^* = q^* = q_{HLL} + \frac{\bar{S} \Delta x}{\lambda_R - \lambda_L}, \quad (3.23b)$$

$$h_L^* = \min \left( \max \left( h_{HLL} - \frac{\lambda_R \bar{S} \Delta x}{\alpha(\lambda_R - \lambda_L)}, \epsilon \right), \left( 1 - \frac{\lambda_R}{\lambda_L} \right) h_{HLL} + \frac{\lambda_R}{\lambda_L} \epsilon \right), \quad (3.23c)$$

$$h_R^* = \min \left( \max \left( h_{HLL} - \frac{\lambda_L \bar{S} \Delta x}{\alpha(\lambda_R - \lambda_L)}, \epsilon \right), \left( 1 - \frac{\lambda_L}{\lambda_R} \right) h_{HLL} + \frac{\lambda_L}{\lambda_R} \epsilon \right), \quad (3.23d)$$

where  $q^*$  and  $\alpha$  have respectively been defined in (3.17b) and (3.20), and the quantities  $h_{HLL}$  and  $q_{HLL}$  are defined by (3.13).

**Lemma 3.** *Assume  $h_L$  and  $h_R$  to be positive. Then, the intermediate states  $W_L^* = {}^t(h_L^*, q_L^*)$  and  $W_R^* = {}^t(h_R^*, q_R^*)$  given by (3.23) satisfy the following properties:*

- (i) *consistency: the quantities  $h_L^*$ ,  $h_R^*$ ,  $q_L^*$  and  $q_R^*$  satisfy the equations (3.17);*
- (ii) *positivity preservation:  $h_L^* \geq \epsilon$  and  $h_R^* \geq \epsilon$ ;*

(iii) *well-balancedness*:  $W_L^*$  and  $W_R^*$  satisfy the property (WB).

*Proof.* Since  $\varepsilon$  is defined by (3.22), we obviously get the required property (ii). Indeed, after (3.23),  $h_L^*$  and  $h_R^*$  stand for the minimum of quantities that are greater than or equal to  $\varepsilon$ .

Next, let us set

$$\widetilde{h}_L^* = h_{HLL} - \frac{\lambda_R \bar{S} \Delta x}{\alpha(\lambda_R - \lambda_L)} \quad \text{and} \quad \widetilde{h}_R^* = h_{HLL} - \frac{\lambda_L \bar{S} \Delta x}{\alpha(\lambda_R - \lambda_L)}.$$

We immediately get the following identity:

$$\lambda_R \widetilde{h}_R^* - \lambda_L \widetilde{h}_L^* = (\lambda_R - \lambda_L) h_{HLL},$$

which means that the heights  $\widetilde{h}_L^*$  and  $\widetilde{h}_R^*$  satisfy the consistency relation (3.17a). Since (3.17b) is obviously verified by  $q^*$ , the property (i) is established as soon as  $h_L^*$  and  $h_R^*$  are proven to satisfy (3.17a). Assume that  $|\lambda_L|$  and  $|\lambda_R|$  are chosen large enough to ensure  $h_{HLL} > 0$  with  $\lambda_L < 0 < \lambda_R$ . We have the following three configurations.

- If  $\widetilde{h}_L^* \geq \varepsilon$  and  $\widetilde{h}_R^* \geq \varepsilon$ , then the relations (3.23) give  $h_L^* = \widetilde{h}_L^*$  and  $h_R^* = \widetilde{h}_R^*$ .
- If  $\widetilde{h}_L^* < \varepsilon$ , then from (3.23) we get  $h_L^* = \varepsilon$  and  $h_R^* = \left(1 - \frac{\lambda_L}{\lambda_R}\right) h_{HLL} + \frac{\lambda_L}{\lambda_R} \varepsilon$ .
- Similarly, if  $\widetilde{h}_R^* < \varepsilon$ , then we have  $h_R^* = \varepsilon$  and  $h_L^* = \left(1 - \frac{\lambda_R}{\lambda_L}\right) h_{HLL} + \frac{\lambda_R}{\lambda_L} \varepsilon$ .

We note that the consistency relation (3.17a) systematically holds and property (i) is proven.

We now need to ensure that the positivity procedure, involving relations (3.23c) and (3.23d), does not interfere with the well-balance property. In order to prove the well-balancedness, we assume that  $W_L$  and  $W_R$  define a steady state, given by (3.18), and we show that, in this case,  $W_L^* = W_L$  and  $W_R^* = W_R$ . If  $h_L^* = \widetilde{h}_L^*$  and  $h_R^* = \widetilde{h}_R^*$ , the property (iii) holds by construction. Now, we complete the proof as the intermediate states involve the water heights  $h_L^*$  and  $h_R^*$  given by (3.23). From the definition (3.23b) of  $q^*$  and the steady relations (3.18) satisfied by  $W_L$  and  $W_R$ , we deduce

$$\begin{aligned} q^* &= \frac{\lambda_R q_0 - \lambda_L q_0}{\lambda_R - \lambda_L} - \frac{1}{\lambda_R - \lambda_L} \left[ \frac{q_0^2}{h} + \frac{1}{2} g h^2 \right] + \frac{\bar{S} \Delta x}{\lambda_R - \lambda_L} \\ &= q_0 - \frac{1}{\lambda_R - \lambda_L} \left( q_0^2 \left[ \frac{1}{h} \right] + \frac{g}{2} [h^2] - q_0^2 \left[ \frac{1}{h} \right] - \frac{g}{2} [h^2] \right) \\ &= q_0. \end{aligned}$$

We now have to prove that  $h_L^* = h_L$  and  $h_R^* = h_R$ . First, let us compute  $\frac{\bar{S} \Delta x}{\alpha}$  at the equilibrium using (3.18) and (3.20), to get:

$$\frac{\bar{S} \Delta x}{\alpha} = \frac{q_0^2 \left[ \frac{1}{h} \right] + \frac{g}{2} [h^2]}{\frac{-q_0^2}{h_L h_R} + \frac{g}{2} (h_L + h_R)} = [h].$$

We can then compute  $\widetilde{h}_L^*$  at the equilibrium, according to (3.21a):

$$\begin{aligned}\widetilde{h}_L^* &= \frac{\lambda_R h_R - \lambda_L h_L}{\lambda_R - \lambda_L} - \frac{[q]}{\lambda_R - \lambda_L} - \frac{\lambda_R \bar{S} \Delta x}{\alpha(\lambda_R - \lambda_L)} \\ &= \frac{\lambda_R h_R - \lambda_L h_L - \lambda_R h_R + \lambda_R h_L}{\lambda_R - \lambda_L} \\ &= h_L.\end{aligned}$$

Similar computations with (3.21b) lead to  $\widetilde{h}_R^* = h_R$ . Moreover, from the definition (3.22) of  $\varepsilon$ , we have  $h_L \geq \varepsilon$  and  $h_R \geq \varepsilon$ . Therefore  $\widetilde{h}_L^* \geq \varepsilon$  and  $\widetilde{h}_R^* \geq \varepsilon$ , and we have  $h_L^* = \widetilde{h}_L^* = h_L$  and  $h_R^* = \widetilde{h}_R^* = h_R$ . This proves the (WB) property, that is to say  $W_L^* = W_L$  and  $W_R^* = W_R$  as soon as  $W_L$  and  $W_R$  define a steady state. This concludes the proof of the well-balancedness, and therefore Lemma 3 is established.  $\square$

Equipped with the suitable intermediate states, we can assert the following result concerning the full scheme (3.6).

**Theorem 4.** *Consider  $W_i^n \in \Omega^*$  for all  $i \in \mathbb{Z}$ , where  $\Omega^*$  is a restricted admissible states space defined as follows:*

$$\Omega^* = \{W = {}^t(h, q) \in \mathbb{R}^2 ; h > 0, q \in \mathbb{R}\}.$$

*Assume that the intermediate states  $W_{i+\frac{1}{2}}^{L,*}$  and  $W_{i+\frac{1}{2}}^{R,*}$  are given, for all  $i \in \mathbb{Z}$ , by*

$$W_{i+\frac{1}{2}}^{L,*} = \begin{pmatrix} h_L^*(W_i^n, W_{i+1}^n) \\ q^*(W_i^n, W_{i+1}^n) \end{pmatrix} \quad \text{and} \quad W_{i+\frac{1}{2}}^{R,*} = \begin{pmatrix} h_R^*(W_i^n, W_{i+1}^n) \\ q^*(W_i^n, W_{i+1}^n) \end{pmatrix},$$

*where  $q^*$ ,  $h_L^*$  and  $h_R^*$  are given by (3.23b), (3.23c) and (3.23d), respectively. Also, assume that the source term approximation  $\bar{S}$  is consistent with the source term  $S$  according to (3.16). Finally, assume that, as soon as  $(W_i^n)_{i \in \mathbb{Z}}$  defines a steady state,  $\bar{S}$  verifies (3.18b) and  $\tilde{q} = q_0$ . Then the Godunov-type scheme, given by (3.6) under the CFL restriction (3.4), satisfies the following properties:*

1. *consistency with the shallow-water system (3.7);*
2. *positivity preservation:  $\forall i \in \mathbb{Z}, W_i^{n+1} \in \Omega^*$ ;*
3. *well-balancedness: if  $(W_i^n)_{i \in \mathbb{Z}}$  defines a steady state, then  $\forall i \in \mathbb{Z}, W_i^{n+1} = W_i^n$ .*

*Proof.* The consistency is a direct consequence of Lemma 3.

We turn to proving that  $\forall i \in \mathbb{Z}, h_i^{n+1} > 0$  as soon as  $h_i^n > 0$ . From Lemma 3, we have  $h_{i+\frac{1}{2}}^{L,*} \geq \varepsilon_i^n$  and  $h_{i+\frac{1}{2}}^{R,*} \geq \varepsilon_i^n$ , where  $\varepsilon_i^n = \min(h_i^n, h_{i+1}^n, h_{i+\frac{1}{2}}^{HLL})$  and  $h_{i+\frac{1}{2}}^{HLL}$  is given by evaluating (3.13a) between states  $W_i^n$  and  $W_{i+1}^n$ . Since the scheme under consideration is given by (3.5),  $h_i^{n+1}$  turns out to be the sum of positive quantities.



We finally need to prove the well-balancedness of the scheme. Once again, this property comes from Lemma 3. Indeed, let us consider that  $(W_i^n)_{i \in \mathbb{Z}}$  defines a steady state. Therefore,  $\forall i \in \mathbb{Z}$ ,  $W_i^n$  and  $W_{i+1}^n$  define a steady state, and Lemma 3 gives  $W_{i+\frac{1}{2}}^{L,*} = W_i^n$  and  $W_{i+\frac{1}{2}}^{R,*} = W_{i+1}^n$ , which in turn yields  $W_i^{n+1} = W_i^n$  for all  $i \in \mathbb{Z}$ . This concludes the proof of Theorem 4.  $\square$

*Remark 5.* Because of the arbitrary small parameter  $\varepsilon$ , introduced in (3.23) to enforce the positivity of the intermediate water heights, the updated water height never vanishes. In a specific section, devoted to dry/wet transitions, we will present an extension of the scheme to deal with dry areas. At this level, we reject vanishing water heights because of the definition of  $\bar{S}$  as well as the term  $\bar{S}/\alpha$  involved within the definitions (3.23) of  $h_L^*$  and  $h_R^*$ . As soon as the full characterization of  $\bar{S}$  is established, the scheme will be extended to allow  $\varepsilon = 0$  in the definition (3.23).

#### 4. Application to the topography source term

In this section, our concern is the derivation of both parameters  $\bar{S}$  and  $\tilde{q}$  in order to complete the scheme (3.6). Here, we consider the topography source term, i.e.

$$S = S^t = -gh\partial_x Z.$$

In this specific case, we shall denote these parameters  $\bar{S}^t$  and  $\tilde{q}_t$ .

##### 4.1. Determination of the parameters

We begin by determining  $\bar{S}^t$ . Let us recall that the steady states are governed by the following PDE:

$$q_0^2 \partial_x \frac{1}{h} + \frac{g}{2} \partial_x h^2 = S^t. \quad (4.1)$$

Since we are only considering the topography source term and smooth steady states, (4.1) can be rewritten under the following algebraic form:

$$\frac{q_0^2}{2} \left[ \frac{1}{h^2} \right] + g[h + Z] = 0. \quad (4.2)$$

Consider  $W_L$  and  $W_R$  defining a steady state: the states  $W_L$  and  $W_R$  are thus given by (3.18), and they must satisfy the algebraic relation (4.2). As a consequence,  $W_L$  and  $W_R$  verify the following two identities:

$$q_0^2 \left[ \frac{1}{h} \right] + \frac{g}{2} [h^2] = \bar{S}^t \Delta x, \quad (4.3a)$$

$$\frac{q_0^2}{2} \left[ \frac{1}{h^2} \right] + g[h + Z] = 0. \quad (4.3b)$$

The first identity (4.3a) corresponds to the generic steady relation applied to the topography, while the second one (4.3b) is nothing but the algebraic relation

specific to the topography source term. We exhibit the expression of  $\bar{S}^t$  from the above relations as follows. First, from (4.3b), we extract the expression of  $q_0^2$ :

$$q_0^2 = 2g[h + Z] \frac{h_L^2 h_R^2}{h_R^2 - h_L^2},$$

which is plugged into (4.3a). We then get the following definition of  $\bar{S}^t$ :

$$\bar{S}^t \Delta x = \frac{g}{2} [h^2] - 2g[h + Z] \frac{h_L h_R}{h_L + h_R},$$

which can be rewritten as follows:

$$\bar{S}^t \Delta x = -2g[Z] \frac{h_L h_R}{h_L + h_R} + \frac{g}{2} \frac{[h]^3}{h_L + h_R}.$$

Let us emphasize that such a definition of the approximation of the topography source term can be found in the literature. For instance, the reader is referred to [4, 5] (see also [33] for related expressions).

Recall that an important ingredient in the consistency of the scheme is that the source term approximation  $\bar{S}^t$  be consistent with the actual source term  $-gh\partial_x Z$ , assuming positive water heights. However, as underlined in [4, 5],  $\bar{S}^t$  is no longer consistent with zero when the topography is flat, i.e.  $[Z] = 0$ . In order to recover the required consistency, we adopt the strategy proposed in [4, 5]. We modify  $\bar{S}^t$  as follows:

$$\bar{S}^t \Delta x := \bar{S}^t(h_L, h_R, Z_L, Z_R) \Delta x = -2g[Z] \frac{h_L h_R}{h_L + h_R} + \frac{g}{2} \frac{[h]_c^3}{h_L + h_R}, \quad (4.4)$$

where  $[h]_c$  is a cutoff of  $[h] = h_R - h_L$ , defined as follows, with  $C$  a positive constant that does not depend on  $\Delta x$ :

$$[h]_c = \begin{cases} h_R - h_L & \text{if } |h_R - h_L| \leq C \Delta x, \\ \text{sgn}(h_R - h_L) C \Delta x & \text{otherwise.} \end{cases}$$

This new expression of  $\bar{S}^t$  is consistent with the topography source term  $S^t$  and ensures that the scheme is well-balanced according to smooth steady states. Note that the expression of  $\bar{S}^t$  given by (4.4) has been obtained by considering  $W_L$  and  $W_R$  defining a steady state. Since this expression only depends on the left and right states, it is relevant to extend it to the case where these states do not define a steady state, and actually use this expression for all  $W_L$  and  $W_R$ .

To achieve the characterization of the scheme (3.6), we also need to define the parameter  $\tilde{q}$  introduced in (3.20). In fact, we can choose any expression for  $\tilde{q}$ , as long as it ensures  $\tilde{q} = q_0$  when  $W_L$  and  $W_R$  define a steady state. We decide to use the following expression for  $\tilde{q}_t$ :

$$\tilde{q}_t = q^*, \quad (4.5)$$

which yields, from (3.23a),

$$\alpha^t = \frac{-(q^*)^2}{h_L h_R} + \frac{g}{2} (h_L + h_R). \quad (4.6)$$

#### 4.2. Extension to dry/wet transitions

We finally study how the approximate average  $\bar{S}^t$ , as well as the term  $\bar{S}^t \Delta x / \alpha^t$ , behave when dealing with vanishing water heights. First, we make the following assumption.

**Assumption.** When the height vanishes, so does the velocity.

This assumption allows us to state the following result, concerning the approximate topography source term.

**Lemma 5.** *When  $h_L$  or  $h_R$  vanishes, the quantities  $\bar{S}^t$  and  $\bar{S}^t \Delta x / \alpha^t$  satisfy:*

$$\bar{S}^t \Delta x = -g(Z_R - Z_L) \frac{h_R + h_L}{2} \quad \text{and} \quad \frac{\bar{S}^t \Delta x}{\alpha^t} = -(Z_R - Z_L).$$

*Proof.* Assume that  $h_L$  or  $h_R$  vanishes. Therefore, from Proposition 2, we have  $q_0 = 0$  as soon as  $W_L$  and  $W_R$  define a steady state. Thus, this steady state is a lake at rest steady state, i.e.  $[h + Z] = 0$ .

Now, to determine  $\bar{S}^t$ , we assume that  $W_L$  and  $W_R$  define a steady state, with  $q_0 = 0$ . The above assumption ensures that  $u_L = u_R = 0$ . From (4.3a), we obtain  $\bar{S}^t \Delta x = \frac{g}{2}[h^2]$ . Plugging  $[h] = -[Z]$  into this equality, we get the above expression of  $\bar{S}^t \Delta x$ .

Now, recall the definition (4.6) of  $\alpha^t$ . In the context of this proof, we have  $q^* = q_0 = 0$ , and then  $\alpha^t = \frac{g}{2}(h_L + h_R)$  since  $u_L = u_R = 0$ . Therefore, we immediately obtain  $\bar{S}^t \Delta x / \alpha^t = -(Z_R - Z_L)$ , which concludes the proof.  $\square$

To handle the case where both  $h_L$  and  $h_R$  vanish (and thus  $q_L = q_R = 0$ ), we have to make sure that, in this case,  $q^* = 0$  and  $h_L^* = h_R^* = 0$ . This requirement is met by taking  $\bar{S}^t \Delta x = 0$  and  $\bar{S}^t \Delta x / \alpha = 0$  as soon as both  $h_L$  and  $h_R$  are zero.

We can now state the following result, concerning the approximate Riemann solver with non-negative water heights.

**Lemma 6.** *The intermediate states (3.23), using  $\bar{S}^t$  and  $\tilde{q}_t$  respectively defined by (4.4) and (4.5), satisfy the following properties:*

- (i) consistency with the shallow-water equations with topography (2.1);
- (ii) the well-balancedness principle (WB).

Moreover, we have:

- (iii) with  $\varepsilon > 0$  given by (3.22), the positivity is preserved: if  $h_L > 0$ ,  $h_R > 0$  and  $h_{HLL} > 0$ , then  $h_L^* \geq \varepsilon$  and  $h_R^* \geq \varepsilon$ ;
- (iv) with  $\varepsilon = 0$ , the non-negativity is preserved: if  $h_L \geq 0$ ,  $h_R \geq 0$  and  $h_{HLL} \geq 0$ , then  $h_L^* \geq 0$  and  $h_R^* \geq 0$ .

*Proof.* For  $\varepsilon > 0$ , the proofs of (i), (ii) and (iii) are immediate. They come from Lemma 3 as well as the results obtained in the current section. Now, assume  $\varepsilon = 0$ . The intermediate states then rewrite:

$$q^* = q_{HLL} + \frac{\bar{S}^t \Delta x}{\lambda_R - \lambda_L}, \quad (4.7a)$$

$$h_L^* = \min \left( \left( h_{HLL} - \frac{\lambda_R \bar{S}^t \Delta x}{\alpha^t (\lambda_R - \lambda_L)} \right)_+, \left( 1 - \frac{\lambda_R}{\lambda_L} \right) h_{HLL} \right), \quad (4.7b)$$

$$h_R^* = \min \left( \left( h_{HLL} - \frac{\lambda_L \bar{S}^t \Delta x}{\alpha^t (\lambda_R - \lambda_L)} \right)_+, \left( 1 - \frac{\lambda_L}{\lambda_R} \right) h_{HLL} \right), \quad (4.7c)$$

with  $\alpha^t$  given by (4.6).

From Lemma 5, the above expressions are well-defined for all  $h_L \geq 0$  and  $h_R \geq 0$ . Moreover, these new intermediate states can be easily shown (see the proof of Lemma 3) to satisfy properties (i), (ii) and (iv). Indeed, the proof of the consistency stated Lemma 3 is preserved, and thus (i) holds. Then, (iv) is a direct consequence of the fact that  $h_L^*$  and  $h_R^*$  are defined as the minimum of non-negative quantities.

Finally, concerning (ii) with  $\varepsilon = 0$ , we assume that  $W_L$  and  $W_R$  define a steady state according to (4.3). If  $h_L > 0$ ,  $h_R > 0$  and  $h_{HLL} > 0$ , then we know from Lemma 3 that  $W_L^* = W_L$  and  $W_R^* = W_R$ . Now, we assume that  $h_L = 0$ . Using Lemma 5, the definition (4.7a) immediately yields  $q^* = q_0$ . Moreover, from (4.7b) and (4.7c), we get

$$h_L^* = \frac{\lambda_R h_R}{\lambda_R - \lambda_L} + \frac{\lambda_R (Z_R - Z_L)}{\lambda_R - \lambda_L}, \quad (4.8a)$$

$$h_R^* = \min \left( \frac{\lambda_R h_R}{\lambda_R - \lambda_L} + \frac{\lambda_L (Z_R - Z_L)}{\lambda_R - \lambda_L}, \left( 1 - \frac{\lambda_L}{\lambda_R} \right) \frac{\lambda_R h_R}{\lambda_R - \lambda_L} \right). \quad (4.8b)$$

Recall from the proof of Lemma 5 that, when  $h_L$  or  $h_R$  vanishes and  $W_L$  and  $W_R$  define a steady state, we have  $[h+Z] = 0$ . Then, the relations (4.8) immediately yield  $h_L^* = 0 = h_L$  and  $h_R^* = h_R$ . A similar sequence of arguments is used to prove the well-balancedness when  $h_R$  vanishes, or when both  $h_L$  and  $h_R$  are zero.

Thus, the proof is achieved.  $\square$

This lemma allows us to state the following result, that concerns the full scheme (3.6).

**Theorem 7.** *Consider  $W_i^n \in \Omega$  for all  $i \in \mathbb{Z}$ . Assume that the intermediate states  $W_{i+\frac{1}{2}}^{L,*}$  and  $W_{i+\frac{1}{2}}^{R,*}$  are given, for all  $i \in \mathbb{Z}$ , by*

$$W_{i+\frac{1}{2}}^{L,*} = \left( h_L^*(W_i^n, W_{i+1}^n) \right) \quad \text{and} \quad W_{i+\frac{1}{2}}^{R,*} = \left( h_R^*(W_i^n, W_{i+1}^n) \right), \quad (4.9)$$

where  $q^*$ ,  $h_L^*$  and  $h_R^*$  are given by (4.7). Then the Godunov-type scheme, given by (3.6) under the CFL restriction (3.4), satisfies the following properties:

1. consistency with the shallow-water system (2.1);
2. positivity preservation:  $\forall i \in \mathbb{Z}, W_i^{n+1} \in \Omega$ ;
3. well-balancedness: if  $(W_i^n)_{i \in \mathbb{Z}}$  defines a steady state according to (4.3), then  $\forall i \in \mathbb{Z}, W_i^{n+1} = W_i^n$ .

*Proof.* The same arguments as used in the proof of [Theorem 4](#), while using the results of [Lemma 6](#), yield the proof.  $\square$

## 5. Second-order MUSCL extension

We devote this section to a second-order extension based on a MUSCL technique (for instance, see [[37](#), [38](#), [28](#), [29](#), [35](#)]). The purpose of this extension is to increase the space and time accuracy of the scheme. It consists in using a piecewise linear reconstruction improved by a time correction, instead of piecewise constant, in the Godunov-type scheme. In order to derive this second-order scheme, we first need to rewrite the scheme (3.6) in order to exhibit the numerical flux function and the source term contribution

### 5.1. Rewriting the scheme

The goal of this subsection is to rewrite the scheme (3.6), with intermediate states given by (4.7), under the form proposed above. After straightforward computations (see for instance [[25](#)]), the scheme (3.6) can be rewritten as

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x} \left( f_{i+\frac{1}{2}}^n - f_{i-\frac{1}{2}}^n \right) + \frac{\Delta t}{2} \left( s_{i+\frac{1}{2}}^n + s_{i-\frac{1}{2}}^n \right).$$

The quantities  $f_{i+\frac{1}{2}}^n$  and  $s_{i+\frac{1}{2}}^n$  denote numerical approximations of the flux and source term, respectively, at the interface  $x_{i+\frac{1}{2}}$ . They are defined by

$$f_{i+\frac{1}{2}}^n = \begin{pmatrix} (f^h)_{i+\frac{1}{2}}^n \\ (f^q)_{i+\frac{1}{2}}^n \end{pmatrix} \quad \text{and} \quad s_{i+\frac{1}{2}}^n = \begin{pmatrix} 0 \\ (S^t)_{i+\frac{1}{2}}^n \end{pmatrix}. \quad (5.1)$$

Here, adopting clear notations,  $(S^t)_{i+\frac{1}{2}}^n$  is given by:

$$(S^t)_{i+\frac{1}{2}}^n = \bar{S}^t (h_i^n, h_{i+1}^n, Z_i, Z_{i+1}), \quad (5.2)$$

where  $\bar{S}^t$  is defined by (4.4). The scheme then reads as follows:

$$\begin{aligned} h_i^{n+1} &= h_i^n - \frac{\Delta t}{\Delta x} \left( (f^h)_{i+\frac{1}{2}}^n - (f^h)_{i-\frac{1}{2}}^n \right), \\ q_i^{n+1} &= q_i^n - \frac{\Delta t}{\Delta x} \left( (f^q)_{i+\frac{1}{2}}^n - (f^q)_{i-\frac{1}{2}}^n \right) + \frac{\Delta t}{2} \left( (S^t)_{i+\frac{1}{2}}^n + (S^t)_{i-\frac{1}{2}}^n \right), \end{aligned}$$

where the approximate fluxes are defined by (5.1) and

$$\begin{aligned} f_{i+\frac{1}{2}}^n &= f(W_i^n, W_{i+1}^n) = \frac{1}{2} \left( F(W_i^n) + F(W_{i+1}^n) \right) \\ &\quad + \frac{\lambda_{i+\frac{1}{2}}^L}{2} \left( W_{i+\frac{1}{2}}^{L,*} - W_i^n \right) + \frac{\lambda_{i+\frac{1}{2}}^R}{2} \left( W_{i+\frac{1}{2}}^{R,*} - W_{i+1}^n \right) \end{aligned}$$

### 5.2. The MUSCL procedure

Equipped with the numerical flux function and the source term contribution, we can state the MUSCL procedure. Consider  $w \in \{h, q, h + Z\}$ . The reconstruction is carried out by replacing the constant state  $w_i^n$  with a linear approximation, given in each cell  $(x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$  by

$$w_i^n(x) = w_i^n + (x - x_i) \sigma_i^n,$$

where  $\sigma_i^n$  is the slope of the linear reconstruction, defined by

$$\sigma_i^n = \text{minmod} \left( \frac{w_{i+1}^n - w_i^n}{\Delta x}, \frac{w_i^n - w_{i-1}^n}{\Delta x} \right).$$

We have applied a limiter to this slope to improve the stability of the scheme. Here, we have chosen the minmod limiter (the reader is referred for instance to [29] for more details regarding the use of slope limiters and a wider range of limiters). The reconstruction of  $w_i^n$  at the interfaces within the cell  $(x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$  is then given by

$$w_i^- := w_i^n \left( x_i - \frac{\Delta x}{2} \right) = w_i^n - \frac{\Delta x}{2} \sigma_i^n, \quad (5.3a)$$

$$w_i^+ := w_i^n \left( x_i + \frac{\Delta x}{2} \right) = w_i^n + \frac{\Delta x}{2} \sigma_i^n. \quad (5.3b)$$

The reconstructed value of  $Z$  at the interfaces is finally obtained from the values of  $h + Z$  and  $h$ .

Therefore, the updated states are given by the following scheme:

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x} (f(W_i^+, W_{i+1}^-) - f(W_{i-1}^+, W_i^-)) + \frac{\Delta t}{2} (s_{i+\frac{1}{2}}^n + s_{i-\frac{1}{2}}^n), \quad (5.4)$$

where the source term contributions  $s_{i+\frac{1}{2}}^n$  and  $s_{i-\frac{1}{2}}^n$  use the following approximation of the source term instead of (5.2):

$$(S^t)_{i+\frac{1}{2}}^n = \bar{S}^t (h_i^+, h_{i+1}^-, Z_i^+, Z_{i+1}^-).$$

Finally, the scheme's time accuracy is improved by the use of the classical Heun's method (see [22]).

### 5.3. The well-balance property

Because of the reconstruction step (5.3), it is worth noting that the designed MUSCL scheme is not well-balanced. We therefore use a MOOD-like technique (see [13] for an overview of such techniques, and [7] for more recent applications) to restore this essential property. Consider that  $W_i^n$  and  $W_{i+1}^n$  define a steady state. At this level, the MUSCL scheme provides a second-order approximation of this steady state, while the first-order well-balanced scheme previously derived exactly captures such a state.

Therefore, we suggest to introduce a convex combination between the reconstructed state and the non-reconstructed one (see [26] for related work). As a consequence, we adopt the following reconstruction:

$$w_i^- = (1 - \theta_i^n) w_i^n + \theta_i^n \left( w_i^n - \frac{\Delta x}{2} \sigma_i^n \right) = w_i^n - \frac{\Delta x}{2} \sigma_i^n \theta_i^n, \quad (5.5a)$$

$$w_i^+ = (1 - \theta_i^n) w_i^n + \theta_i^n \left( w_i^n + \frac{\Delta x}{2} \sigma_i^n \right) = w_i^n + \frac{\Delta x}{2} \sigma_i^n \theta_i^n, \quad (5.5b)$$

where  $0 \leq \theta_i^n \leq 1$  is the parameter of the convex combination. Note that the states are not reconstructed if  $\theta_i^n = 0$ , while the full MUSCL scheme is obtained by taking  $\theta_i^n = 1$ .

The objective is now to propose a suitable process to define the parameter  $\theta_i^n$ . In order to have a relevant definition of  $\theta_i^n$ , we first define

$$\Delta \psi_{i+\frac{1}{2}}^n = \frac{(q_{i+1}^n)^2}{h_{i+1}^n} - \frac{(q_i^n)^2}{h_i^n} + \frac{g}{2} ((h_{i+1}^n)^2 - (h_i^n)^2) - \Delta x \bar{S}_{i+\frac{1}{2}}^t,$$

where  $\bar{S}_{i+\frac{1}{2}}^t = \bar{S}^t(h_i^n, h_{i+1}^n, Z_i, Z_{i+1})$ , with  $\bar{S}^t$  defined by (4.4). Note that this quantity turns out to be a residue that governs steady states, according to (4.3). As a consequence,  $\Delta \psi_{i+\frac{1}{2}}^n$  vanishes when  $W_i^n$  and  $W_{i+1}^n$  define a steady state. Next, we define a function to evaluate the deviation with respect to the equilibrium as follows:

$$\varphi_i^n = \left\| \begin{pmatrix} q_i^n - q_{i-1}^n \\ \Delta \psi_{i-\frac{1}{2}}^n \end{pmatrix} \right\|_2 + \left\| \begin{pmatrix} q_{i+1}^n - q_i^n \\ \Delta \psi_{i+\frac{1}{2}}^n \end{pmatrix} \right\|_2,$$

which vanishes when  $W_{i-1}^n$ ,  $W_i^n$  and  $W_{i+1}^n$  define a steady state. Equipped with  $\varphi_i^n$ , the parameter of the convex combination  $\theta_i^n$  is defined for some  $M > m > 0$  as follows (see Figure 5):

$$\theta_i = \begin{cases} 0 & \text{if } \varphi_i^n < m \Delta x \\ \frac{\varphi_i^n - m \Delta x}{M \Delta x - m \Delta x} & \text{if } m \Delta x \leq \varphi_i^n \leq M \Delta x \\ 1 & \text{if } \varphi_i^n > M \Delta x. \end{cases} \quad (5.6)$$

Such a definition ensures that the first-order well-balanced scheme is used if the equilibrium error  $\varphi_i^n$  is small enough. Moreover, the MUSCL scheme is used if the states are far from defining a steady state, i.e.  $\varphi_i^n$  is large enough. In addition, the closer the states are to the equilibrium, the more the convex combination will favor the first-order well-balanced scheme.

## 6. Numerical experiments

Numerical simulations are now carried out to test the scheme devised in the previous sections. We first check that the scheme preserves lake at rest steady

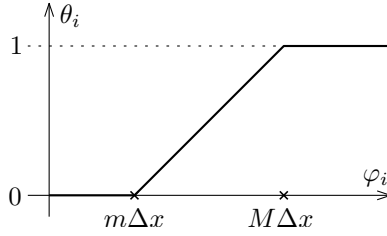


Figure 5: Graph of  $\theta_i$  with respect to  $\varphi_i$ , according to (5.6).

states. Then, we assess the well-balancedness of the scheme by studying moving steady states. Finally, we present a 2D experiment to provide an error analysis of the scheme. One purpose of these numerical tests is to compare the proposed scheme (3.6) - (4.9) (denoted *WBt* from now on) with two classical schemes: the *HLL* scheme (see [25]) and the hydrostatic reconstruction (*HR*) scheme (see [1]).

Assuming  $N$  discretization cells, we compute the  $L^1$ ,  $L^2$  and  $L^\infty$  errors for a bounded function  $w$  using the following expressions:

$$L^1 : \frac{1}{N} \sum_{i=1}^N |w_i - w_i^{ex}| ; \quad L^2 : \sqrt{\frac{1}{N} \sum_{i=1}^N (w_i - w_i^{ex})^2} ; \quad L^\infty : \max_{1 \leq i \leq N} |w_i - w_i^{ex}|,$$

where  $w_i$  and  $w_i^{ex}$  are respectively the approximate and the exact solution at cell  $i$  and at the final physical time  $t_{end}$ . We recall that the time step  $\Delta t$  is given by the CFL condition (3.4):

$$\Delta t \leq \frac{\Delta x}{2\Lambda}, \quad \text{where } \Lambda = \max_{i \in \mathcal{Z}} \left( \lambda_{i+\frac{1}{2}}^L, \lambda_{i+\frac{1}{2}}^R \right).$$

The constants will be chosen according to Table 1.

Constant	Equation	Value
$g$	(2.1)	$g = 9.81$
$\varepsilon_\lambda$	(3.2)	$\varepsilon_\lambda = 10^{-10}$
$\varepsilon$	(3.23)	$\varepsilon = 0$

Table 1: Values of the constants within the numerical experiments.

### 6.1. Steady states at rest

The first numerical experiments we showcase are steady states at rest, i.e. lake at rest steady states governed by (2.6). The first two experiments concern wet steady states at rest, with a continuous or discontinuous topography. The third experiment is a steady state at rest with a transition between wet and dry areas, with a discontinuous topography. The topography  $Z$  is different for each experiment, and will be provided for each test case. The initial data are



$h(x) = \max(1 - Z(x), 0)$  and  $q(x) = 0$ , and the space domain is  $[0, 1]$ . In all these lake at rest experiments, we set  $C = +\infty$ . The errors are presented at the physical time  $t_{end} = 1$ s. We chose to carry out these experiments with 200 discretization cells.

### 6.1.1. Wet lake at rest experiments

We here propose two different wet configurations. The first one involves a continuous topography, given by  $Z_c(x) = \max(0, 0.5 - 2|x - 0.5|)$ , while we consider a discontinuous topography  $Z_d(x) = \mathbb{1}_{[\frac{1}{2}, 1]}(x)$  for the second one.

	$h$			$q$		
	$L^1$	$L^2$	$L^\infty$	$L^1$	$L^2$	$L^\infty$
HLL	5.09e-05	4.36e-04	5.04e-03	6.72e-03	9.56e-03	1.56e-02
HR	0	0	0	0	0	0
WBt	1.11e-18	1.11e-17	1.11e-16	0	0	0

Table 2: Height and discharge errors for the wet steady state at rest with continuous topography  $Z_c$ .

	$h$			$q$		
	$L^1$	$L^2$	$L^\infty$	$L^1$	$L^2$	$L^\infty$
HLL	2.47e-03	1.30e-02	1.33e-01	6.46e-03	3.48e-02	3.70e-01
HR	0	0	0	0	0	0
WBt	0	0	0	0	0	0

Table 3: Height and discharge errors for the wet steady state at rest with discontinuous topography  $Z_d$ .

Both these experiments show that both the HR and WBt schemes exactly capture these wet configurations of a lake at rest, while the HLL scheme is only first-order accurate.

### 6.1.2. Lake at rest with a dry/wet transition

Here, we focus on a lake at rest with a transition between a wet area and a dry area. The topography function is now  $Z(x) = \max(0, 2x - 0.5)\mathbb{1}_{[0.5, 1]}(x)$ .

As in the previous examples, the WBt scheme and the HR scheme allow the preservation of the lake at rest, even with dry/wet interfaces. On the contrary, the HLL scheme provides a poor first-order approximation.

## 6.2. Moving steady states

We now assess the scheme's ability to preserve steady states with a nonzero discharge.

Two experiments, deriving from Goutal and Maurel's test cases, are carried out. These experiments, the *subcritical flow* and the *transcritical flow without*

	$h$			$q$		
	$L^1$	$L^2$	$L^\infty$	$L^1$	$L^2$	$L^\infty$
HLL	2.35e-03	1.38e-02	1.32e-01	5.82e-03	3.55e-02	3.74e-01
HR	0	0	0	0	0	0
WBt	0	0	0	0	0	0

Table 4: Steady state at rest with discontinuous topography and a dry/wet transition.

*shock*, are presented in [23], and will be respectively named GM1 and GM2. The space domain is  $0 < x < 25$  and the topography is given by  $Z(x) = (0.2 - 0.05(x - 10)^2)_+$ . The boundary conditions for these test cases are given hereafter, with the values of  $q_0$  and  $h_0$  depending on the experiment studied:

- on the left boundary, the water height follows a homogeneous Neumann boundary condition and the discharge is set to some  $q_0$ ;
- on the right boundary, the water height is set to a prescribed value  $h_0$  when the flow is subcritical (and a homogeneous Neumann boundary condition is prescribed otherwise) and the discharge follows a homogeneous Neumann boundary condition.

In addition, the initial conditions are  $h + Z = h_0$  and  $q = 0$  throughout the domain. The values of  $q_0$  and  $h_0$  are:

- GM1:  $q_0 = 4.42\text{m}^3/\text{s}$  and  $h_0 = 2\text{m}$ ;
- GM2:  $q_0 = 1.53\text{m}^3/\text{s}$  and  $h_0 = 0.66\text{m}$ .

Both of these test cases present a transient state followed by a steady state, with uniform discharge  $q_0$ . The steady state should be exactly obtained by the fully well-balanced scheme WBt. For GM1, the final physical time is  $t_{end} = 500\text{s}$  and  $C = +\infty$ . For GM2, the final physical time is  $t_{end} = 125\text{s}$  and  $C = 2.5$ . We have elected to use 200 discretization cells.

For both these experiments, note that  $q = q_0$  and that the steady state equation (2.7) is verified. This equation is nothing but a statement of Bernoulli's principle, and can be rewritten:

$$\frac{q_0^2}{2h^2} + g(h + Z) = H,$$

where  $H$  is uniform throughout the domain, and is usually called the total head (see [23] for instance). As a consequence, to evaluate the well-balancedness of the scheme, we compute the error to the uniform discharge  $q_0$ , as well as the error to the total head.

### 6.2.1. Comparison with other schemes

To assess the relevance of the WBt scheme, we compare it with the HR and the HLL schemes on these moving steady states experiments. The results of the

WBt scheme are displayed in Figures 6 - 7 and the comparison with the HR and HLL schemes are presented in Tables 5 - 6.

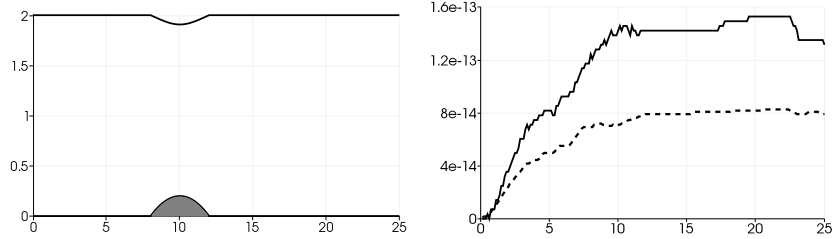


Figure 6: Left panel: free surface and topography for the subcritical flow test case. Right panel: errors for the subcritical flow; the solid line is the total head error and the dashed line is the discharge error.

	$H$			$q$		
	$L^1$	$L^2$	$L^\infty$	$L^1$	$L^2$	$L^\infty$
HLL	8.24e-03	1.19e-02	7.41e-02	4.31e-03	1.22e-02	5.19e-02
HR	1.32e-02	1.97e-02	7.48e-02	2.37e-03	6.74e-03	2.74e-02
WBt	1.18e-13	1.25e-13	1.53e-13	6.65e-14	6.99e-14	8.26e-14

Table 5: Total head and discharge errors for the subcritical flow experiment.

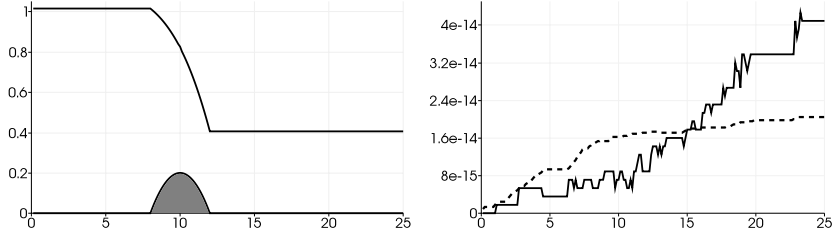


Figure 7: Left panel: free surface and topography for the transcritical flow test case. Right panel: errors for the transcritical flow; the solid line is the total head error and the dashed line is the discharge error.

Both HLL and HR schemes are shown to provide a first-order approximation of these moving steady states. However, the proposed WBt scheme allows an exact preservation (i.e. up to the machine precision) of such moving steady states. Moreover, the WBt scheme recovers these steady states after a transient state, even if the steady state is not prescribed as initial condition.

### 6.2.2. Comparison with the MUSCL scheme

Now, we compare the first-order WBt scheme with the second-order one, with or without the well-balance correction. The results obtained with 200

	$H$			$q$		
	$L^1$	$L^2$	$L^\infty$	$L^1$	$L^2$	$L^\infty$
HLL	2.72e-02	3.50e-02	7.45e-02	1.54e-03	6.16e-03	3.70e-02
HR	4.79e-02	6.07e-02	8.12e-02	8.28e-04	3.30e-03	1.82e-02
WBt	1.67e-14	2.13e-14	4.26e-14	1.47e-14	1.58e-14	2.04e-14

Table 6: Total head and discharge errors for the transcritical flow experiment.

discretization cells are shown in Tables 7 - 8. The *MUSCL* scheme refers to the scheme (5.4) - (5.5) without the well-balancedness correction, i.e.  $m = 0$  and  $M = 0$ . The hybrid version, denoted  $\theta$ -*MUSCL*, is given by (5.4) - (5.5) with the well-balancedness correction. In the numerical experiments, we choose  $m = 10^{-10}$  and  $M = 5.10^{-1}$ .

	$H$			$q$		
	$L^1$	$L^2$	$L^\infty$	$L^1$	$L^2$	$L^\infty$
WBt	1.18e-13	1.25e-13	1.53e-13	6.65e-14	6.99e-14	8.26e-14
MUSCL	1.32e-03	4.07e-03	3.38e-02	2.36e-04	9.92e-04	8.27e-03
$\theta$ -MUSCL	9.32e-14	1.08e-13	1.56e-13	5.51e-14	5.75e-14	8.88e-14

Table 7: Total head and discharge errors for the subcritical flow experiment.

	$H$			$q$		
	$L^1$	$L^2$	$L^\infty$	$L^1$	$L^2$	$L^\infty$
WBt	1.67e-14	2.13e-14	4.26e-14	1.47e-14	1.58e-14	2.04e-14
MUSCL	2.44e-03	5.43e-03	4.47e-02	2.55e-04	8.21e-04	5.75e-03
$\theta$ -MUSCL	4.94e-14	5.19e-14	6.93e-14	4.22e-14	4.50e-14	5.44e-14

Table 8: Total head and discharge errors for the transcritical flow experiment.

Tables 7 and 8 show that the correction of the MUSCL scheme indeed recovers the well-balance property of the first-order scheme. Without this correction, the MUSCL scheme is not exact for these moving steady states, and the accuracy is reduced to a second-order one.

### 6.3. 2D steady vortex

The proposed WBt scheme is exact for 1D stationary solutions. Therefore, the order of accuracy of the scheme would have to be tested on a smooth unstationary exact solution of the shallow-water equations with topography. To the extent of our knowledge, there is no such solution referenced. We turn to a two-dimensional experiment, by considering the steady vortex (see [14]), to

assess the order of accuracy of the scheme. Indeed, the scheme is exact for all the steady states in the  $x$  and  $y$  directions, but the question of the preservation of a fully 2D steady state arises.

Let us state the shallow-water equations with topography in two dimensions:

$$\begin{cases} \partial_t h + \partial_x hu + \partial_y hv & = 0, \\ \partial_t hu + \partial_x \left( hu^2 + \frac{1}{2}gh^2 \right) + \partial_y (huv) & = -gh\partial_x Z, \\ \partial_t hv + \partial_x (huv) + \partial_y \left( hv^2 + \frac{1}{2}gh^2 \right) & = -gh\partial_y Z, \end{cases}$$

where  $h(x, y, t)$  is the water height,  $u(x, y, t)$  and  $v(x, y, t)$  respectively represent the water velocities in the  $x$  and  $y$  directions, and  $Z(x, y)$  is the smooth topography.

For the steady vortex experiment, the space domain is  $[-3, 3] \times [-3, 3]$ . The topography is defined by  $Z(x, y) = 0.2e^{0.5(1-r^2)}$ , where  $r^2 = x^2 + y^2$ . The exact solution is given by

$$h(x, y) = 1 - \frac{1}{4g}e^{2(1-r^2)} - Z(x, y); \quad u(x, y) = ye^{1-r^2}; \quad v(x, y) = -xe^{1-r^2}.$$

The exact solution is chosen as the initial condition, and is prescribed as the boundary conditions. This exact solution is displayed on Figure 8.

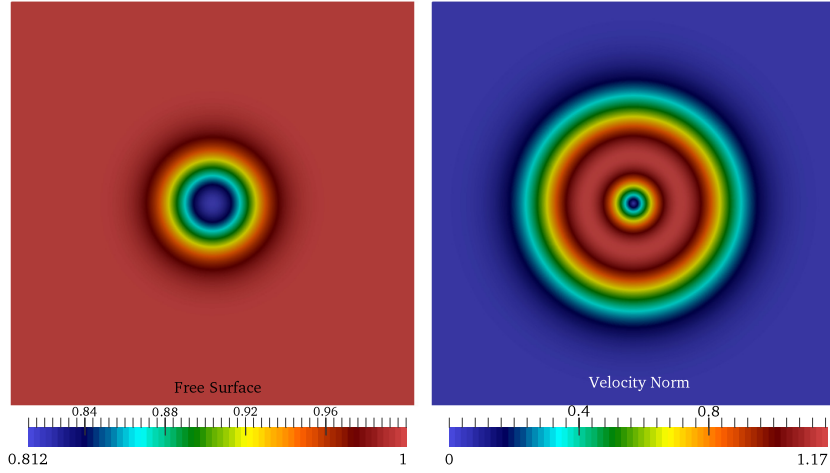


Figure 8: Exact solution for the steady vortex experiment. Left panel: free surface. Right panel: velocity norm (the flow is clockwise).

We use a uniform Cartesian mesh of  $N$  square cells, whose edges are of length  $\Delta x$ . We take the final time  $t_{end} = 1$ s. The time step  $\Delta t$  is given by the following CFL-like condition:

$$\Delta t \leq \frac{\Delta x}{4\Lambda}, \quad \text{where } \Lambda = \max_{i \in Z} \left( \lambda_{i+\frac{1}{2}}^L, \lambda_{i+\frac{1}{2}}^R \right).$$

We take  $C = +\infty$ . For the  $\theta$ -MUSCL scheme, we take  $m = 0$  and  $M = 0$ . The accuracy assessment of both schemes is presented in Tables 9 - 10 - 11 - 12. Note that only the  $x$ -discharge error is presented. Indeed, the  $y$ -discharge error is very similar to the  $x$ -discharge error, and the orders of accuracy are the same.

N	$L^1$		$L^2$		$L^\infty$	
1024	6.37e-03	—	1.61e-02	—	1.19e-01	—
4096	4.31e-03	0.74	1.08e-02	0.74	7.88e-02	0.75
16384	2.58e-03	0.83	6.43e-03	0.84	4.58e-02	0.86
65536	1.43e-03	0.91	3.54e-03	0.91	2.45e-02	0.93
262144	7.52e-04	0.95	1.86e-03	0.95	1.27e-02	0.97

Table 9: First-order scheme, height error.

N	$L^1$		$L^2$		$L^\infty$	
1024	3.90e-02	—	6.72e-02	—	3.00e-01	—
4096	2.42e-02	0.81	4.14e-02	0.81	1.88e-01	0.80
16384	1.37e-02	0.89	2.35e-02	0.88	1.10e-01	0.86
65536	7.29e-03	0.94	1.26e-02	0.93	6.05e-02	0.91
262144	3.77e-03	0.97	6.54e-03	0.96	3.20e-02	0.95

Table 10: First-order scheme,  $x$ -discharge error.

Tables 9 and 10 show the proposed Wbt scheme does not exactly preserve the steady vortex, but delivers a first-order approximation of the exact solution.

N	$L^1$		$L^2$		$L^\infty$	
1024	2.41e-03	—	7.55e-03	—	5.52e-02	—
4096	6.65e-04	1.81	2.14e-03	1.77	1.68e-02	1.64
16384	1.71e-04	1.95	5.44e-04	1.96	5.00e-03	1.68
65536	5.25e-05	1.63	1.38e-04	1.97	1.40e-03	1.79
262144	1.99e-05	1.32	3.71e-05	1.86	3.77e-04	1.85

Table 11: Second-order scheme, height error.

Similarly, Tables 11 and 12 show that the  $\theta$ -MUSCL scheme preserves the steady vortex up to a second-order approximation.

## 7. Conclusion

In this work, we have proposed ways to obtain smooth steady states for the shallow-water equations with topography. We have derived a scheme for this system that is well-balanced, positivity-preserving and allows dry/wet transitions. We have also obtained a well-balanced second-order MUSCL extension

N	$L^1$		$L^2$		$L^\infty$	
1024	1.33e-02	—	2.27e-02	—	1.04e-01	—
4096	3.55e-03	1.87	6.08e-03	1.87	3.26e-02	1.60
16384	8.93e-04	1.99	1.60e-03	1.91	1.09e-02	1.49
65536	2.25e-04	1.99	4.13e-04	1.93	3.25e-03	1.68
262144	5.93e-05	1.89	1.06e-04	1.95	9.86e-04	1.65

Table 12: Second-order scheme,  $x$ -discharge error.

for this scheme. Several numerical experiments have been performed in order to exhibit the relevant properties of this scheme, which is shown to exactly recover the steady states, both at rest and moving.

The question of the stability of the scheme is not raised in the present paper. The choice (3.2) of the characteristic velocities has been made to ensure that the scheme be viscous, which allows us to postpone the study of the stability.

### Acknowledgments

C. Berthon, F. Foucher and V. Michel-Dansac would like to thank the ANR-12-IS01-0004-01 GEONUM for financial support. S. Clain would like to thank the FCT-ANR/MAT-NAN/0122/2012 grant for financial support.

### References

- [1] E. Audusse, F. Bouchut, M.-O. Bristeau, R. Klein, and B. Perthame. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM J. Sci. Comput.*, 25(6):2050–2065, 2004.
- [2] E. Audusse, C. Chalons, and P. Ung. A simple well-balanced and positive numerical scheme for the shallow-water system. *Commun. Math. Sci.*, 13(5):1317–1332, 2015.
- [3] A. Bermudez and M. E. Vazquez. Upwind methods for hyperbolic conservation laws with source terms. *Comput. & Fluids*, 23(8):1049–1071, 1994.
- [4] C. Berthon and C. Chalons. A fully well-balanced, positive and entropy-satisfying Godunov-type method for the shallow-water equations. 2013.
- [5] C. Berthon, C. Chalons, S. Cornet, and G. Sperone. Fully well-balanced, positive and simple approximate Riemann solver for shallow-water equations. *Proceedings of HYP2014 international conference*, 2014.
- [6] C. Berthon, A. Crestetto, and F. Foucher. A well-balanced finite volume scheme for a mixed hyperbolic/parabolic system to model chemotaxis. 2015.

- [7] C. Berthon and V. Desveaux. An entropy preserving MOOD scheme for the Euler equations. *Int. J. Finite Vol.*, 11:39, 2014.
- [8] C. Berthon and F. Foucher. Efficient well-balanced hydrostatic upwind schemes for shallow-water equations. *J. Comput. Phys.*, 231(15):4993–5015, 2012.
- [9] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2004.
- [10] M. J. Castro Díaz, J. A. López-García, and C. Parés. High order exactly well-balanced numerical methods for shallow water systems. *J. Comput. Phys.*, 246:242–264, 2013.
- [11] C. Chalons, F. Coquel, E. Godlewski, P.-A. Raviart, and N. Seguin. Godunov-type schemes for hyperbolic systems with parameter-dependent source. The case of Euler system with friction. *Math. Models Methods Appl. Sci.*, 20(11):2109–2166, 2010.
- [12] A. Chertock, S. Cui, A. Kurganov, and T. Wu. Well-balanced positivity preserving central-upwind scheme for the shallow water system with friction terms. *Internat. J. Numer. Methods Fluids*, 78(6):355–383, 2015.
- [13] S. Clain, S. Diot, and R. Loubère. A high-order finite volume method for systems of conservation laws—Multi-dimensional Optimal Order Detection (MOOD). *J. Comput. Phys.*, 230(10):4028–4050, 2011.
- [14] S. Clain and J. Figueiredo. The MOOD method for the non-conservative shallow-water system. Oct 2014.
- [15] E. D. Fernández-Nieto, D. Bresch, and J. Monnier. A consistent intermediate wave speed for a well-balanced HLLC solver. *C. R. Math. Acad. Sci. Paris*, 346(13-14):795–800, 2008.
- [16] U. S. Fjordholm, S. Mishra, and E. Tadmor. Well-balanced and energy stable schemes for the shallow water equations with discontinuous topography. *J. Comput. Phys.*, 230(14):5587–5609, 2011.
- [17] J. M. Gallardo, M. Castro, C. Parés, and J. M. González-Vida. On a well-balanced high-order finite volume scheme for the shallow water equations with bottom topography and dry areas. In *Hyperbolic problems: theory, numerics, applications*, pages 259–270. Springer, Berlin, 2008.
- [18] G. Gallice. Solveurs simples positifs et entropiques pour les systèmes hyperboliques avec terme source. *C. R. Math. Acad. Sci. Paris*, 334(8):713–716, 2002.
- [19] G. Gallice. Positive and entropy stable Godunov-type schemes for gas dynamics and MHD equations in Lagrangian or Eulerian coordinates. *Numer. Math.*, 94(4):673–713, 2003.



- [20] E. Godlewski and P.-A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*, volume 118 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.
- [21] L. Gosse. A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms. *Comput. Math. Appl.*, 39(9-10):135–159, 2000.
- [22] Sigal Gottlieb, Chi-Wang Shu, and Eitan Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Rev.*, 43(1):89–112, 2001.
- [23] N. Goutal and F. Maurel. Proceedings of the 2nd Workshop on Dam-Break Wave Simulation. Technical report, Groupe Hydraulique Fluviale, Département Laboratoire National d’Hydraulique, Electricité de France, 1997.
- [24] J. M. Greenberg and A. Leroux. A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. *SIAM J. Numer. Anal.*, 33(1):1–16, 1996.
- [25] A. Harten, P. D. Lax, and B. van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Rev.*, 25(1):35–61, 1983.
- [26] X. Y. Hu, N. A. Adams, and C.-W. Shu. Positivity-preserving method for high-order conservative schemes solving compressible Euler equations. *J. Comput. Phys.*, 242:169–180, 2013.
- [27] S. Jin. A steady-state capturing method for hyperbolic systems with geometrical source terms. *M2AN Math. Model. Numer. Anal.*, 35(4):631–645, 2001.
- [28] R. J. LeVeque. *Numerical methods for conservation laws*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 1992.
- [29] R. J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2002.
- [30] Q. Liang and F. Marche. Numerical resolution of well-balanced shallow water equations with complex source terms. *Advances in Water Resources*, 32(6):873 – 884, 2009.
- [31] M. Lukáčová-Medvidová, S. Noelle, and M. Kraft. Well-balanced finite volume evolution Galerkin methods for the shallow water equations. *J. Comput. Phys.*, 221(1):122–147, 2007.
- [32] R. Natalini, M. Ribot, and M. Twarogowska. A well-balanced numerical scheme for a one dimensional quasilinear hyperbolic model of chemotaxis. *Commun. Math. Sci.*, 12(1):13–39, 2014.

- [33] S. Noelle, Y. Xing, and C.-W. Shu. High-order well-balanced finite volume WENO schemes for shallow water equation with moving water. *J. Comput. Phys.*, 226(1):29–58, 2007.
- [34] G. Russo and A. Khe. High order well balanced schemes for systems of balance laws. In *Hyperbolic problems: theory, numerics and applications*, volume 67 of *Proc. Sympos. Appl. Math.*, pages 919–928. Amer. Math. Soc., Providence, RI, 2009.
- [35] E. F. Toro. *Riemann solvers and numerical methods for fluid dynamics*. Springer-Verlag, Berlin, third edition, 2009. A practical introduction.
- [36] E. F. Toro, M. Spruce, and W. Speares. Restoration of the contact surface in the HLL-Riemann solver. *Shock Waves*, 4(1):25–34, 1994.
- [37] B. van Leer. Towards the ultimate conservative difference scheme, V. A second order sequel to Godunov’s method. *J. Com. Phys.*, 32:101–136, 1979.
- [38] B. van Leer. On the relation between the upwind-differencing schemes of Godunov, Engquist-Osher and Roe. *SIAM J. Sci. Statist. Comput.*, 5(1):1–20, 1984.
- [39] Y. Xing. Exactly well-balanced discontinuous Galerkin methods for the shallow water equations with moving water equilibrium. *J. Comput. Phys.*, 257(part A):536–553, 2014.
- [40] Y. Xing, C.-W. Shu, and S. Noelle. On the advantage of well-balanced schemes for moving-water equilibria of the shallow water equations. *J. Sci. Comput.*, 48(1-3):339–349, 2011.