A second-order well-balanced scheme for the shallow-water equations with topography

Christophe Berthon, Raphaël Loubère, Victor Michel-Dansac

A second-order well-balanced scheme for the shallow-water equations with topography

Christophe Berthon, Raphaël Loubère and Victor Michel-Dansac

Abstract We consider the well-balanced numerical scheme for the shallow-water equations with topography introduced in [8] and its second-order well-balanced extension, which requires two heuristic parameters. The goal of the present contribution is to derive a parameter-free second-order well-balanced scheme. To that end, we consider a convex combination between the well-balanced scheme and a second-order scheme. We then prove that a relevant choice of the parameter of this convex combination ensures that the resulting scheme is both second-order accurate and well-balanced. Afterwards, we perform several numerical experiments, in order to illustrate both the second-order accuracy and the well-balance property of this numerical scheme. Finally, we outline some perspectives in a short conclusion.

1 Introduction

We consider the shallow-water system with topography, governed by the following set of equations:

$$\begin{cases} \partial_t h + \partial_x q = 0, \\ \partial_t q + \partial_x \left(\dfrac{q^2}{h} + \dfrac{1}{2} g h^2 \right) = -g h \partial_x Z, \end{cases} \tag{1}$$

Christophe Berthon
Laboratoire de Mathématiques Jean Leray, Université de Nantes, 2 rue de la Houssinière, BP 92208 F-44322 Nantes Cedex 3. e-mail: christophe.berthon@univ-nantes.fr

Raphaël Loubère
Institut de Mathématiques de Bordeaux and CNRS, Université de Bordeaux, 351 cours de la Libération, F-33405 Talence. e-mail: raphael.loubere@u-bordeaux.fr

Victor Michel-Dansac
Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse Cedex 9. e-mail: victor.michel-dansac@math.univ-toulouse.fr

where $h(t,x) > 0$ is the water height, $q(t,x)$ is the water discharge, $Z(x)$ is the smooth topography, and g is the gravity constant. To shorten the notations, we rewrite this system under the condensed form $\partial_t W + \partial_x F(W) = S(W)$, where we have set:

$$W = \begin{pmatrix} h \\ q \end{pmatrix} \quad ; \quad F(W) = \begin{pmatrix} q \\ \dfrac{q^2}{h} + \dfrac{1}{2} g h^2 \end{pmatrix} \quad ; \quad S(W) = \begin{pmatrix} 0 \\ -g h \partial_x Z \end{pmatrix}.$$

In particular, we focus on smooth steady solutions, free from time and governed by:

$$\begin{cases} q = \text{cst}, \\ \dfrac{q}{2h^2} + g(h + Z) = \text{cst}. \end{cases} \tag{2}$$

Well-balanced schemes, i.e. schemes that exactly preserve such steady solutions, have been derived in the last decade (see for instance [2, 4, 5, 8]).

Namely, in [8], the authors suggested a well-balanced Godunov-type scheme based on a two-state approximate Riemann solver. We briefly recall the general form of a numerical scheme that falls within this classification (see for instance [10]). In the finite volume framework, the space domain $\mathbb{R}$ is discretized into cells, assumed to be of constant width Δx. The center of the i-th cell is denoted by x_i and its bounds are labeled $x_{i-\frac{1}{2}}$ and $x_{i+\frac{1}{2}}$; this cell shall henceforth be referred to by its center x_i. The approximate solution is piecewise constant, and it is denoted by W_i^n within the cell x_i and at time t^n. In order to provide a time update of this approximate solution, we note that Riemann problems (i.e. Cauchy problems with discontinuous initial data) are present at each interface between cells. However, the exact solution to such Riemann problems is usually difficult or impossible to compute exactly. To address this issue, an approximate Riemann solver is introduced. More specifically, in [8], the authors develop an approximate Riemann solver satisfying several crucial properties: consistency, well-balance and preservation of the water height non-negativity. The time update of the approximate solution in the cell x_i reads:

$$W_i^{n+1,WB} = W_i^n - \frac{\Delta t}{\Delta x}\left[\lambda_{i+\frac{1}{2}}\left(W_{i+\frac{1}{2}}^{L,*} - h_i^n\right) + \lambda_{i-\frac{1}{2}}\left(W_{i-\frac{1}{2}}^{R,*} - h_i^n\right)\right], \tag{3}$$

where $W_{i+\frac{1}{2}}^{L,*}$ and $W_{i-\frac{1}{2}}^{R,*}$ are the intermediate states of the approximate Riemann solver, respectively approximations of the Riemann solutions at the interfaces $x_{i+\frac{1}{2}}$ and $x_{i-\frac{1}{2}}$. In addition, $\lambda_{i+\frac{1}{2}}$ and $\lambda_{i-\frac{1}{2}}$ are approximations of the wave velocities from the exact Riemann solution. The authors of [8] prove that the Godunov-type scheme (3) is consistent, well-balanced and non-negativity-preserving.

The accuracy of the first-order scheme (3) could be significantly improved by introducing a well-balanced second-order extension. The MUSCL framework is well suited to this extension. It consists in providing a piecewise linear approximation, instead of piecewise constant, of the solution in each cell. This is achieved using slope reconstructions, supplemented with slope limiters in order to recover the non-

negativity preservation. For more details on such procedures, the reader is referred for instance to [7]. The resulting MUSCL scheme reads as follows:

$$W_i^{n+1,MUSCL} = W_i^n - \frac{\Delta t}{\Delta x}\Delta\mathcal{F}_i + \Delta t\mathcal{S}_i, \tag{4}$$

where $\Delta\mathcal{F}_i$ and $\mathcal{S}_i$ are, respectively, second-order approximations of the physical flux F and the source term S in the cell x_i. Note that the intermediate states of the scheme (3) are used to define $\Delta\mathcal{F}_i$ and $\mathcal{S}_i$. In addition to the second-order in space time update (4), a specific treatment of the steady states is necessary, because this scheme is no longer naturally well-balanced due to the reconstruction procedure. For instance, in [4, 3], the authors suggest a reconstruction based on the steady states, which leads to a well-balanced second-order scheme. However, the downside of this approach is that, in each cell, the nonlinear steady relations (2) have to be solved, thus leading to extra computational cost.

In [8], the authors proposed a convex combination, in each cell c_i and at time t^{n+1}, between the well-balanced scheme and a MUSCL reconstruction to recover the well-balance property without having to solve nonlinear equations, as follows:

$$W_i^{n+1} = \theta_i^n W_i^{n+1,MUSCL} + (1-\theta_i^n)W_i^{n+1,WB}. \tag{5}$$

On the one hand, for a steady state, the well-balanced scheme is exact, and therefore is of order at least two. In this case, we wish to use the well-balanced scheme. On the other hand, for an unsteady state, the well-balanced scheme is of order one, and it should not be used. The MUSCL scheme is second-order accurate in both these cases. Therefore, we wish to use the MUSCL scheme when the approximate solution is unsteady and the well-balanced scheme when it is steady. As a consequence, the convex combination (5) becomes relevant when its parameter θ_i^n allows switching between the MUSCL scheme and the well-balanced scheme to ensure both a second-order accuracy and the well-balance property. To that end, θ_i^n must be equal to 1 in the unsteady case, and it must vanish for a steady state.

In [8], this convex combination relied on two heuristic parameters used to define θ_i^n with respect to some error to a steady state. The goal of the present manuscript is to propose a parameter-free formula for θ_i^n, that ensures both the well-balance property and the second-order accuracy of the scheme.

To that end, we first introduce such an expression of θ_i^n. We then prove that the required properties are satisfied by the resulting scheme. Finally, several numerical experiments confirm the second-order accuracy and the well-balance of the scheme. A short conclusion outlines several perspectives to this work.

2 A second-order accurate convex combination

The goal of this section is to introduce a parameter-free expression of θ_i^n such that the convex combination (5) yields a second-order accurate and well-balanced scheme.

To that end, let us first define the following potential Φ:

$$\Phi = \frac{q^2}{2h^2} + g(h + Z).$$

Note that, for a steady state, we have $\Phi = \text{cst}$ as well as $q = \text{cst}$, as per (2). Let us then define the spatial errors to a steady state, as follows:

$$\begin{aligned} \hat{e}_i^n &= \max\big(|q_i^n - q_{i-1}^n|, |q_{i+1}^n - q_i^n|\big), \\ \check{e}_i^n &= \max\big(|\Phi_i^n - \Phi_{i-1}^n|, |\Phi_{i+1}^n - \Phi_i^n|\big). \end{aligned}$$

In order to provide a relevant definition of θ_i^n, we make the following remarks:

- the well-balanced scheme is stationary (i.e. $W_i^{n+1,WB} = W_i^n$) if and only if the solution is stationary, i.e. $\hat{e}_i^n = 0$ and $\check{e}_i^n = 0$;
- the MUSCL scheme is not well-balanced; it can however become stationary (i.e. $W_i^{n+1,MUSCL} = W_i^n$), but, in this case, $\hat{e}_i^n \neq 0$ and $\check{e}_i^n \neq 0$.

These remarks lead us to consider switching between the well-balanced and the MUSCL scheme when the time update of the MUCSL scheme becomes very small. Indeed, we will show that, in this case, the MUSCL scheme approximates a steady solution, and switching to the well-balanced scheme ensures its preservation. The following result states how to define θ_i^n in order to make sure that the scheme (5) is both well-balanced and second-order accurate.

Theorem 1. *We first introduce the following two conditions:*

$$\begin{aligned} &(C_1) \quad \hat{e}_i^n < \varepsilon_m \text{ and } \check{e}_i^n < \varepsilon_m, \\ &(C_2) \quad |h_i^{n+1,MUSCL} - h_i^n| \leq (e_h)_i^n \text{ and } |q_i^{n+1,MUSCL} - q_i^n| \leq (e_q)_i^n, \end{aligned}$$

where the errors $(e_h)_i^n$ and $(e_q)_i^n$ are defined by:

$$\begin{aligned} (e_h)_i^n &= \hat{e}_i^n \Delta t \Delta x \frac{\Delta t}{\Delta x} \frac{q_i^n}{(h_i^n)^3} + \check{e}_i^n \Delta t \Delta x \frac{\Delta t^2}{\Delta x^2} \frac{q_i^n}{(h_i^n)^2} + \frac{\Delta x^3}{(h_i^n)^2}, \\ (e_q)_i^n &= \hat{e}_i^n \Delta t \Delta x \frac{q_i^n}{(h_i^n)^3} + \check{e}_i^n \Delta t \Delta x \frac{\Delta t}{\Delta x} \frac{q_i^n}{(h_i^n)^2} + \Delta x^3 \frac{q_i^n}{(h_i^n)^3}, \end{aligned} \tag{6}$$

and where ε_m is a measure of the machine precision, usually taken equal to 10^{-12} in the numerical simulations. Then, let us define θ_i^n as follows:

$$\theta_i^n = \begin{cases} 0 & \text{if } (C_1) \text{ or } (C_2) \text{ holds,} \\ 1 & \text{otherwise.} \end{cases} \tag{7}$$

The above definition of θ_i^n ensures the scheme (5) is well-balanced and second-order accurate.

Remark 2. Note that the respective units of $(e_h)_i^n$ and $(e_q)_i^n$, as defined by (6), are those of the height and the discharge. In addition, note that $\hat{e}_i^n = O(\Delta x)$

and $\check{e}_i^n = O(\Delta x)$, and that $\Delta x = O(\Delta t)$ because of the CFL condition. As a consequence, we remark that $(e_h)_i^n = O(\Delta x^2 \Delta t)$ and $(e_q)_i^n = O(\Delta x^2 \Delta t)$.

The remainder of this section is dedicated to a proof of Theorem 1. First, we prove a preliminary result related to the time update of the well-balanced scheme. Then, this result is used to complete the proof of Theorem 1.

2.1 Time update of the well-balanced scheme with respect to the steady state deviation

The goal of this section is to prove a result that will be used to complete the proof of Theorem 1. It is an estimation of the time update of the well-balanced scheme with respect to the error to a steady state. Indeed, we know that, if this error vanishes, then so does the time update of the well-balanced scheme. The following statement provides us with such an estimation.

Lemma 3. *Let us consider the following (almost steady) configuration:*

$$\begin{cases} q_{i-1}^n = q_i^n + \hat{\varepsilon}_-, \\ q_{i+1}^n = q_i^n + \hat{\varepsilon}_+, \end{cases} \qquad \begin{cases} \Phi_{i-1}^n = \Phi_i^n + \check{\varepsilon}_-, \\ \Phi_{i+1}^n = \Phi_i^n + \check{\varepsilon}_+. \end{cases} \tag{8}$$

Then the time update of the well-balanced scheme (3) *satisfies:*

$$W_i^{n+1,WB} = W_i^n + O(\hat{\varepsilon}_+) + O(\check{\varepsilon}_+) + O(\hat{\varepsilon}_-) + O(\check{\varepsilon}_-). \tag{9}$$

Proof. In order to prove Lemma 3, we first provide an estimation of the intermediate states of the approximate Riemann solver involved in the scheme (3) and derived in [8]. This estimation will then act as a stepping stone towards proving Lemma 3, by being used at each interface of the cell x_i and injected within the time update (3).

We begin by considering two states W_L and W_R almost defining a steady state, i.e. we assume that there exist small $\hat{\varepsilon}$ and $\check{\varepsilon}$ such that $q_R = q_L + \hat{\varepsilon}$ and $\Phi_R = \Phi_L + \check{\varepsilon}$. Let us also define the following quantities:

- $[X] = X_R - X_L$ denotes the jump of a quantity X,
- $X^a = (X_L + X_R)/2$ its arithmetic mean, and
- $X^h = 2X_L X_R/(X_L + X_R)$ its harmonic mean.

In addition, we introduce

$$\beta = -\frac{q_L^2}{h_L h_R} + g h^a.$$

We now consider the intermediate states h_L^*, h_R^* and q^* of the Godunov-type scheme introduced in [8]. Equipped with these assumptions and definitions, these intermediate states are proven to satisfy, after straightforward but tedious computations:

$$
\begin{aligned}
h_L^* &= h_L - \frac{\hat{\varepsilon}}{2}\left(\frac{1}{\lambda} + \frac{q_L}{\beta h^a}\left(\frac{h_L}{h_R} + \frac{[h]}{h^h} - \frac{q_L}{\lambda}\frac{[h]}{h_L h_R}\right)\right) \\
&\quad + \frac{\check{\varepsilon}}{2\beta h^a}\left(h_L h_R + \frac{q_L}{\lambda}[h]\right) + O(\hat{\varepsilon}^2) + O(\check{\varepsilon}^2),
\end{aligned} \tag{10a}
$$

$$
\begin{aligned}
h_R^* &= h_R - \frac{\hat{\varepsilon}}{2}\left(\frac{1}{\lambda} - \frac{q_L}{\beta h^a}\left(\frac{h_L}{h_R} + \frac{[h]}{h^h} - \frac{q_L}{\lambda}\frac{[h]}{h_L h_R}\right)\right) \\
&\quad - \frac{\check{\varepsilon}}{2\beta h^a}\left(h_L h_R + \frac{q_L}{\lambda}[h]\right) + O(\hat{\varepsilon}^2) + O(\check{\varepsilon}^2),
\end{aligned} \tag{10b}
$$

$$
q^* = q_L + \frac{\hat{\varepsilon}}{2}\left(1 - \frac{1}{\lambda}\frac{q_L}{h^a}\right) - \frac{\check{\varepsilon}}{2}\frac{h^h}{\lambda} + O(\hat{\varepsilon}^2) + O(\check{\varepsilon}^2), \tag{10c}
$$

where $\lambda = -\lambda_L = \lambda_R$, as prescribed in [8]. Note that such expressions come from the fact that the scheme is well-balanced. Indeed, for a true steady state, we have $\hat{\varepsilon} = 0$ and $\check{\varepsilon} = 0$, which correctly yields $h_L^* = h_L$, $h_R^* = h_R$ and $q^* = q_L = q_R$.

Now, recall that the time update of the well-balanced scheme from [8] reads as follows:

$$
h_i^{n+1,WB} = h_i^n + \frac{\Delta t}{\Delta x}\left[\lambda_+\left(h_+^{L,*} - h_i^n\right) + \lambda_-\left(h_-^{R,*} - h_i^n\right)\right], \tag{11a}
$$

$$
q_i^{n+1,WB} = q_i^n + \frac{\Delta t}{\Delta x}\left[\lambda_+\left(q_+^* - q_i^n\right) + \lambda_-\left(q_-^* - q_i^n\right)\right], \tag{11b}
$$

where the subscript $\pm$ is a shorter notation for $i \pm 1/2$. Let us assume that the almost steady configuration (8) is satisfied for the cells x_{i-1}, x_i and x_{i+1}. As a consequence, we can use the formulas (10) to rewrite the update (11) as follows:

$$
\begin{aligned}
h_i^{n+1,WB} = h_i^n + \frac{\Delta t}{2\Delta x}\Bigg[& -\hat{\varepsilon}_+\left(1 + \frac{q_i^n}{\beta_+ h_+^a}\left(\lambda_+\left(\frac{h_i}{h_{i+1}} + \frac{[h]_+}{h_+^h}\right) - \frac{q_i^n [h]_+}{h_i h_{i+1}}\right)\right) \\
& + \frac{\check{\varepsilon}_+}{\beta_+ h_+^a}(\lambda_+ h_i h_{i+1} + q_i [h]_+) \\
& + \hat{\varepsilon}_-\left(1 - \frac{q_i^n}{\beta_- h_-^a}\left(\lambda_-\left(\frac{h_{i-1}}{h_i} + \frac{[h]_-}{h_-^h}\right) - \frac{q_i^n [h]_-}{h_{i-1} h_i}\right)\right) \\
& + \frac{\check{\varepsilon}_-}{\beta_- h_-^a}(\lambda_- h_{i-1} h_i + q_i [h]_-)\Bigg] \\
& + O(\hat{\varepsilon}_+^2) + O(\check{\varepsilon}_+^2) + O(\hat{\varepsilon}_-^2) + O(\check{\varepsilon}_-^2),
\end{aligned}
$$

$$
\begin{aligned}
q_i^{n+1,WB} = q_i^n + \frac{\Delta t}{2\Delta x}\left[\hat{\varepsilon}_+\left(\lambda_+ - \frac{q_i^n}{h_+^a}\right) - \check{\varepsilon}_+ h_+^h + \hat{\varepsilon}_-\left(\lambda_- + \frac{q_i^n}{h_-^a}\right) + \check{\varepsilon}_- h_-^h\right] \\
+ O(\hat{\varepsilon}_+^2) + O(\check{\varepsilon}_+^2) + O(\hat{\varepsilon}_-^2) + O(\check{\varepsilon}_-^2).
\end{aligned}
$$

As a consequence, the estimation (9) holds, and the proof is achieved. □

2.2 Proof of Theorem 1

Proof (Theorem 1). The goal of this proof is to show that, with θ_i^n defined by (7), the scheme defined by the convex combination (5) is second-order accurate and well-balanced. More precisely, let $W^{ex}(t,x)$ be a smooth exact solution of the system (1) equipped with suitable initial and boundary conditions. We introduce the notation $(W^{ex})_i^n := W^{ex}(t^n, x_i)$. The expected result is established as soon as we have shown that $|h_i^{n+1} - (h^{ex})_i^{n+1}| = O(\Delta x^2)$ and $|q_i^{n+1} - (q^{ex})_i^{n+1}| = O(\Delta x^2)$ and that, if the states $(W^{ex})_{i-1}^n$, $(W^{ex})_i^n$ and $(W^{ex})_{i+1}^n$ define a steady solution, then $W_i^{n+1} = W_i^n$.

To that end, we consider the three possible cases: $\theta_i^n = 1$, $\theta_i^n = 0$ because (C_1) holds, and $\theta_i^n = 0$ because (C_2) holds.

- First, if $\theta_i^n = 1$, then the scheme is second-order accurate. Indeed, the contribution of the well-balanced scheme is multiplied to $1 - \theta_i^n$, and the convex combination is therefore reduced to contribution of the second-order MUSCL scheme. In addition, neither (C_1) nor (C_2) holds, and therefore the exact solution is unsteady. Thus, the well-balance property is irrelevant in this case.
- Second, if (C_1) holds, then $\theta_i^n = 0$, and the convex combination is reduced to the sole well-balanced scheme. Note that (C_1) is equivalent to the approximate solution being steady (up to the machine precision). Since only the well-balanced scheme is used in the update (5), the resulting scheme exactly preserves this steady solution, and it is, consequently, at least second-order accurate.
- Third, let us assume that (C_2) holds. As a consequence, $\theta_i^n = 0$ and the well-balanced scheme is used. However, contrary to the case where (C_1) held, the approximate solution is unsteady. For this third case, we need to prove that the well-balanced scheme is actually second-order accurate. To that end, we prove that (C_2) necessarily implies that the approximate solution is close to a steady state, up to Δx^2. Arguing Lemma 3 will then allow us to conclude that the well-balanced scheme is actually second-order accurate.
  Using Remark 2, we get that

$$|W_i^{n+1,MUSCL} - W_i^n| = O(\Delta x^2 \Delta t).$$

Arguing the expression (4) of the time update $W_i^{n+1,MUSCL}$ immediately yields:

$$\left| \frac{\Delta \mathcal{F}_i}{\Delta x} - \mathcal{S}_i \right| = O(\Delta x^2). \tag{12}$$

The above equation is a discretization of the steady relation $\partial_x F(W) = S(W)$. Therefore, since the MUSCL scheme is consistent, the states W_{i-1}^n, W_i^n and W_{i-1}^n are close to a steady state. Thus, there exist small $\hat{\varepsilon}_-$, $\hat{\varepsilon}_+$, $\check{\varepsilon}_-$ and $\check{\varepsilon}_+$ such that

$$\begin{cases} q_{i-1}^n = q_i^n + \hat{\varepsilon}_-, \\ q_{i+1}^n = q_i^n + \hat{\varepsilon}_+, \end{cases} \qquad \begin{cases} \Phi_{i-1}^n = \Phi_i^n + \check{\varepsilon}_-, \\ \Phi_{i+1}^n = \Phi_i^n + \check{\varepsilon}_+. \end{cases} \tag{13}$$

Consequently, the above identities are a direct consequence of the condition (C_2), and they hold as soon as it is true. We now set out to prove that $\hat{\varepsilon}_-$, $\hat{\varepsilon}_+$, $\check{\varepsilon}_-$ and $\check{\varepsilon}_+$ are of order $O(\Delta x^2)$. Once this fact is established, applying Lemma 3 will conclude the proof in this third case.
To that end, let us introduce a truly steady state W^{steady}, such that

$$\begin{cases} q_{i-1}^{steady} = q_i^{steady}, \\ q_{i+1}^{steady} = q_i^{steady}, \end{cases} \qquad \begin{cases} \Phi_{i-1}^{steady} = \Phi_i^{steady}, \\ \Phi_{i+1}^{steady} = \Phi_i^{steady}. \end{cases}$$

Since the MUSCL scheme is consistent and second-order accurate in space, the following estimation holds:

$$\frac{|W_i^{steady,MUSCL} - W_i^{steady}|}{\Delta t} = O(\Delta x^2),$$

where $W_i^{steady,MUSCL}$ denotes the time update provided by the MUSCL scheme when considering W^{steady} as initial condition. Arguing the expression (4) of this time update, we get:

$$\left| \left\{ \frac{\Delta \mathcal{F}_i}{\Delta x} - \mathcal{S}_i \right\} (W^{steady}) \right| = O(\Delta x^2). \tag{14}$$

Moreover, we show after tedious computations that

$$\left| \frac{\Delta \mathcal{F}_i}{\Delta x} - \mathcal{S}_i \right| - \left| \left\{ \frac{\Delta \mathcal{F}_i}{\Delta x} - \mathcal{S}_i \right\} (W^{steady}) \right| = O(\hat{\varepsilon}_+) + O(\check{\varepsilon}_+) + O(\hat{\varepsilon}_-) + O(\check{\varepsilon}_-), \tag{15}$$

where the first term of the left-hand side corresponds to the MUSCL scheme applied to the current configuration (13). Plugging the estimation (14) into (15) yields:

$$\left| \frac{\Delta \mathcal{F}_i}{\Delta x} - \mathcal{S}_i \right| = O(\Delta x^2) + O(\hat{\varepsilon}_+) + O(\check{\varepsilon}_+) + O(\hat{\varepsilon}_-) + O(\check{\varepsilon}_-). \tag{16}$$

Using both (12) and (16), we obtain the result we had set out to prove:

$$\hat{\varepsilon}_- = O(\Delta x^2) \quad ; \quad \hat{\varepsilon}_+ = O(\Delta x^2) \quad ; \quad \check{\varepsilon}_- = O(\Delta x^2) \quad ; \quad \check{\varepsilon}_+ = O(\Delta x^2). \tag{17}$$

Finally, note that since the approximate solution satisfies (13), which is identical to (8). Therefore, we apply Lemma 3, using (17), to conclude that the time update of the well-balanced scheme satisfies the following estimation:

$$W_i^{n+1,WB} = W_i^n + O(\Delta x^2). \tag{18}$$

In addition, thanks to (17), the configuration (13) becomes

$$\begin{cases} q_{i-1}^n = q_i^n + O(\Delta x^2), \\ q_{i+1}^n = q_i^n + O(\Delta x^2), \end{cases} \qquad \begin{cases} \Phi_{i-1}^n = \Phi_i^n + O(\Delta x^2), \\ \Phi_{i+1}^n = \Phi_i^n + O(\Delta x^2), \end{cases}$$

and the sequence W_{i-1}^n, W_i^n, W_{i+1}^n corresponds to a steady state, up to Δx^2. Therefore, (18) yields

$$|W_i^{n+1,WB} - (W^{ex})_i^{n+1}| = O(\Delta x^2),$$

i.e. the well-balanced scheme is second-order accurate. Since the convex combination scheme (5) is reduced to the contribution of the well-balanced scheme, it is second-order accurate.

Therefore, in all three cases under consideration, the scheme (5) is at least second-order accurate. In addition, if a steady state is considered, then this scheme is exact. As a consequence, the convex combination (5) yields a well-balanced and second-order accurate scheme, which concludes the proof of Theorem 1. □

3 Numerical experiments

In this section, we propose three numerical experiments. The goal of these experiments is to check that the scheme (5) satisfies the required properties. To this end, we first present an experiment dedicated to the computation of the order of accuracy. Then, we check the well-balance property of the scheme by considering an unsteady state, which, in finite time and after a transient state, converges to a steady state. The last experiment consists in a "dam-break" problem over a non-flat topography.

The numerical schemes tested in these experiments are labeled as follows: the first-order well-balanced scheme is called *WB*, the second-order scheme is labeled *MUSCL*, and the convex combination (5) is called θ-*WB*. In addition, the time accuracy of both second-order schemes is improved thanks to Heun's method.

3.1 Order of accuracy verification

This first experiment consists in the approximation of a smooth solution. This smooth solution is defined on the space domain $[0.9, 1.1]$ by

$$h(x) = 1 + \omega\left(\frac{2}{0.05}(x-1)\right), \quad q(x) = 0, \quad Z(x) = \frac{1}{4} + \frac{3}{4}\cos\left(\pi(x+0.05) + \frac{\pi}{4}\right)^2,$$

where we have set

$$\omega(y) = \begin{cases} \left(\dfrac{2-|y|}{2}\right)^4 (1+2|y|) & \text{if } |y| < 2, \\ 0 & \text{otherwise.} \end{cases}$$

The numerical simulation is carried out until the final physical time $t_{end} = 0.005$s.

In Table 1, we present the errors on Φ in the L^2-norm, as well as the corresponding orders of accuracy. These errors have been computed using a reference solution, provided by the hydrostatic reconstruction scheme from [1] with 25600 discretization cells. Note that similar results are obtained by considering the discharge or other norms. These results show that the θ-WB scheme is more accurate than both the WB and the MUSCL scheme, and that is is second-order accurate, as expected.

N	WB		MUSCL		θ-WB	
25	5.46e-01	—	2.89e-01	—	2.94e-01	—
50	2.84e-01	0.94	2.84e-02	3.34	2.41e-02	3.61
100	1.55e-01	0.87	7.36e-03	1.95	5.99e-03	2.01
200	8.11e-02	0.94	1.90e-03	1.95	1.51e-03	1.99
400	4.10e-02	0.98	5.15e-04	1.88	4.41e-04	1.78

Table 1 L^2-error on Φ for the approximation of a smooth solution.

3.2 Well-balance of the scheme: capture of a steady state

We now consider the capture of a steady state obtained after a transient state. Such steady states are exactly captured by the WB scheme, and we require the θ-WB scheme to exactly capture them as well.

Namely, we focus on the subcritical steady flow presented in [6]. We consider, over the space domain $[0, 25]$, the topography function $Z(x) = \left(0.2 - 0.05(x - 10)^2\right)_+$. We take initial conditions at rest, given by $q(0, x) = 0$ and $h(0, x) = h_0 - Z(x)$, where $h_0 = 2$. The boundary conditions, $q(t, 0) = q_0$ and $h(t, 25) = h_0$ (with $q_0 = 4.42$), ensure that the solution is a transient state followed by a smooth steady state with nonzero velocity. This steady state is governed by (2).

The numerical experiment is performed using 100 discretization cells and until the final physical time $t_{end} = 500$s. The results are presented in Figure 1, where we note that the steady state is exactly captured, even after the transient state, by the second-order θ-WB scheme. Indeed, the errors between the numerical discharge (resp. potential) and the steady state discharge (resp. potential) are of the order of the machine precision. Therefore, this scheme is well-balanced in the sense that it is able to exactly capture steady states, even after a transient state.

3.3 Dam-break experiment

This last experiment consists in a "dam-break" problem, on the space domain $[0, 1]$, whose initial data is:

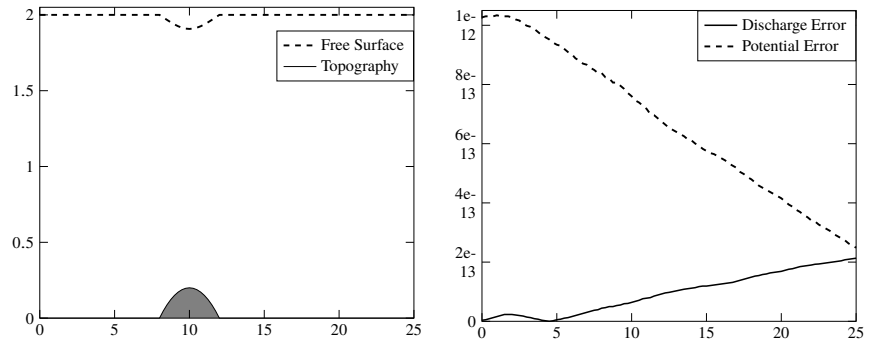

Fig. 1 Subcritical flow experiment. Left panel: free surface and topography for the steady state. Right panel: errors on the discharge q and the potential Φ.

$$\begin{cases} q_L = 5, \\ \Phi_L = 60, \end{cases} \qquad \begin{cases} q_R = 5, \\ \Phi_R = 30. \end{cases}$$

Note that the left and right states of this dam-break are moving steady states, which satisfy the equation (2). As a consequence, they will be exactly preserved by the first-order well-balanced scheme: the goal of this experiment is to display the accuracy gained by the use of the θ-WB scheme. To that end, we take the exact steady solution as boundary conditions, and we display the approximate solution obtained with 100 cells at the final time $t_{end} = 0.02$s, as well as a reference solution, in Figure 2.

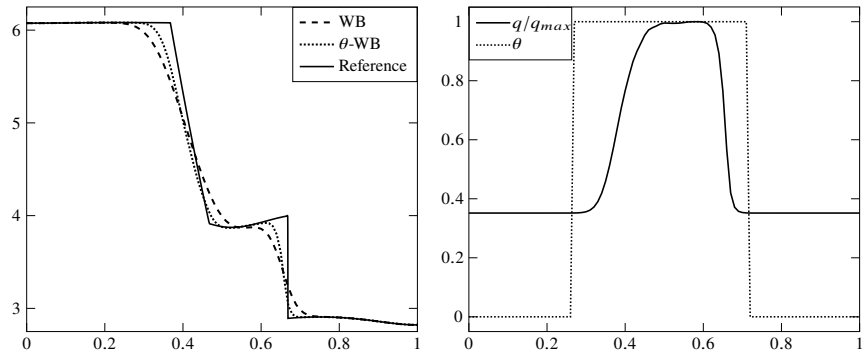

Fig. 2 Riemann problem at time $t = 0.02$s. Left panel: free surface. Right panel: value of θ_i^n for the θ-WB scheme.

The left panel of Figure 2 shows a comparison between the WB scheme, the θ-WB scheme and a reference solution. We observe that the θ-WB scheme is more accurate than the WB scheme, and that the steady areas are exactly preserved. In the right panel, we display the convex combination parameter θ_i^n on the space domain. We have added a plot of $q/max_i(q_i)$ to emphasize the steady areas. As expected, we

have $\theta_i^n = 0$ away from the waves, and $\theta_i^n = 1$ within and close to the waves. This means that the well-balanced scheme is used in the steady areas, while the MUSCL scheme is used within the dam-break itself, as expected.

4 Conclusion

We have developed a parameter-free, second-order and well-balanced extension of the scheme presented in [8]. This new scheme is a significant improvement over the second-order scheme suggested in [8], which relied on a heuristic parameter choice.

Several perspectives of this work naturally arise. Namely, in [9], the authors propose a well-balanced scheme for the nonlinear Manning friction source term. Due to the nonlinearity, providing a parameter-free second-order extension of this scheme would be an interesting challenge. Another challenge lies in a two-dimensional extension of this scheme. Indeed, the definitions of the conditions (C_1) and (C_2) in Theorem 1 strongly depend on the expression of the one-dimensional scheme.

Acknowledgements C. Berthon and V. Michel-Dansac extend their thanks to the ANR-12-IS01-0004-01 GEONUM for financial support. R. Loubère and V. Michel-Dansac acknowledge the financial support of the ANR-14-CE23-0007 MOONRISE.

References

1. C. Berthon and C. Chalons. A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. *SIAM J. Sci. Comput.*, 25(6):2050–2065, 2004.
2. C. Berthon and C. Chalons. A fully well-balanced, positive and entropy-satisfying Godunov-type method for the shallow-water equations. *Math. Comp.*, 85(299):1281–1307, 2016.
3. 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.
4. M. J. Castro, A. Pardo Milanés, and C. Parés. Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique. *Math. Models Methods Appl. Sci.*, 17(12):2055–2113, 2007.
5. 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.
6. 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.
7. R. J. LeVeque. *Finite volume methods for hyperbolic problems.* Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2002
8. V. Michel-Dansac, C. Berthon, S. Clain, and F. Foucher. A well-balanced scheme for the shallow-water equations with topography. *Comput. Math. Appl.*, 72(3):568–593, 2016.
9. V. Michel-Dansac, C. Berthon, S. Clain, and F. Foucher. A well-balanced scheme for the shallow-water equations with topography or Manning friction. *J. Comput. Phys.*, 335:115–154, 2017.
10. E. F. Toro. *Riemann solvers and numerical methods for fluid dynamics. A practical introduction.* Springer-Verlag, Berlin, third edition, 2009.