Bull. Malays. Math. Sci. Soc. (2020) 43:783–807
https://doi.org/10.1007/s40840-018-00713-5

BULLETIN of the
MALAYSIAN MATHEMATICAL
SCIENCES SOCIETY
http://www.springer.com/mathematics/journal/40840

Well-Balanced Numerical Schemes for Shallow Water Equations with Horizontal Temperature Gradient

Mai Duc Thanh[1] · Nguyen Xuan Thanh[2]

Received: 8 September 2018 / Revised: 9 November 2018 / Published online: 17 December 2018

Abstract

A class of well-balanced numerical schemes for the one-dimensional shallow water equations with temperature gradient is constructed. The construction of the schemes is based on two steps: the first step to absorb the nonconservative term and the second step to deal with the evolution of the system. Algorithms for computing contact waves which absorb the nonconservative term are developed. Furthermore, to improve the accuracy, the underlying numerical fluxes can be formed as convex combinations of a pair of numerical fluxes of a low and stable scheme and a higher and fast scheme. The schemes are well balanced and can retain the positivity of the water height and the water temperature. Numerical tests show that the schemes are stable and have a good accuracy.

Keywords Shallow water equations · Ripa system · Conservation law · Well-balanced scheme · Convergence · Accuracy

Mathematics Subject Classification 35L65 · 65M08 · 76B15

Communicated by Yong Zhou.

✉ Mai Duc Thanh
mdthanh@hcmiu.edu.vn

Nguyen Xuan Thanh
mr.nxthanh@gmail.com

[1] Department of Mathematics, International University, Vietnam National University - Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

[2] Department of Mathematics and Computer Science, University of Science, Vietnam National University - Ho Chi Minh City, 227 Nguyen Van Cu str., District 5, Ho Chi Minh City, Vietnam

1 Introduction

In this paper, we will construct a class of well-balanced numerical schemes for the following one-dimensional shallow water equations with variable topography and temperature gradient (see Ripa [25,26])

$$\begin{aligned} &h_t + (hu)_x = 0,\\ &(hu)_t + \Big(hu^2 + \frac{g}{2}h^2\theta\Big)_x = -gh\theta a_x,\\ &(h\theta)_t + (uh\theta)_x = 0, \end{aligned} \tag{1.1}$$

where h is the height of the water from the bottom to surface, u is the velocity, g is the gravity constant, θ is the temperature and a is the height of the bottom from a given level.

System (1.1) is a hyperbolic system of balance laws with a nonconservative source term on the right-hand side. Often, nonconservative terms cause lots of inconvenience for standard numerical schemes. For example, errors may be increasing when the mesh sizes get smaller. Therefore, the study on numerical approximations for systems of balance laws containing nonconservative terms is interesting and attracts many authors.

It has been shown that, by supplementing with the trivial equation

$$a_t = 0, \tag{1.2}$$

one can rewrite system (1.1) in the form of nonconservative system of conservation laws

$$\mathbf{U}_t + \mathbf{A}(\mathbf{U})\mathbf{U}_x = 0. \tag{1.3}$$

Recall that solutions of (1.3) can be understood in the sense of nonconservative products; see [12]. Recently, the Riemann problem for the shallow water equations with horizontal temperature gradients (1.1)–(1.2) was investigated in [33].

To deal with the nonconservativeness of system (1.1), we use the stationary waves. These waves result in equilibrium states, which are independent of time. The equilibrium states are incorporated into a suitable numerical flux. To improve the efficiency, we form a class of numerical fluxes to be convex combinations of a pair of numerical fluxes, where the first one is of a first-order (stable) scheme and the second one is of a high-order (fast) scheme. For example, such a pair can be the numerical fluxes of the Lax–Friedrichs (first-order, stable) scheme and of the Lax–Wendroff (second-order, fast) scheme. Schemes of this kind are fast and well balanced in the sense that they can capture exactly stationary waves. Many numerical tests are conducted, which all show that the schemes can give a good accuracy. Furthermore, we will show that the scheme using the underlying numerical flux of the Lax–Friedrichs scheme possesses interesting properties: The positivity of the water height and water temperature is conserved.

We note that numerical schemes for the Ripa system were constructed in [7,16, 28,36]. Well-balanced schemes for shallow water equations with variable topography were considered in [13,14,22–24,27]. Positively conservative schemes were built in [11,35]. Godunov-type schemes for hyperbolic systems of balance laws in nonconservative forms are considered in [2,9,21,29]. Well-balanced numerical schemes for a single conservation law with source term were studied in [3,5,6,15]. Well-balanced schemes for the model of a fluid flow in a nozzle with variable cross section were constructed in [18,19]. Numerical schemes for two-phase flow models were presented in [1,4,8,10,30]. The Riemann problem for other hyperbolic systems in nonconservative form is considered in [17,20,31,32]. See also the references therein.

The organization of this paper is as follows. Section 2 is devoted to basic properties of system (1.1)–(1.2). Numerical schemes will be constructed in Sect. 3. Furthermore, properties of the schemes are also established. Numerical tests are conducted in Sect. 4, where the errors are computed for different mesh sizes. Finally, we will make several conclusions and discussions in Sect. 5.

2 Basic Properties and Stationary Waves

In this section, we recall basic properties and investigate the admissible stationary waves of system (1.1).

2.1 Basic Properties of System

To study basic properties of system (1.1), one often supplement the system with the trivial equation

$$a_t = 0.$$

Then, the system can be transformed to the following system

$$\begin{aligned} &h_t + uh_x + hu_x = 0,\\ &u_t + uu_x + g\theta h_x + \frac{gh}{2}\theta_x + g\theta a_x = 0,\\ &\theta_t + u\theta_x = 0,\\ &a_t = 0. \end{aligned} \tag{2.1}$$

Thus, if one formally sets the unknown function in the form $\mathbf{U} = (h, u, \theta, a)^{\mathrm{T}}$, one can rewrite system (2.1) in the nonconservative form

$$\mathbf{U}_t + \mathbf{A}(\mathbf{U})\mathbf{U}_x = 0, \tag{2.2}$$

where the matrix $\mathbf{A}(\mathbf{U})$ is given by

$$\mathbf{A}(\mathbf{U}) = \begin{pmatrix} u & h & 0 & 0 \\ g\theta & u & \frac{gh}{2} & g\theta \\ 0 & 0 & u & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.3}$$

The characteristic equation of the matrix $\mathbf{A}(\mathbf{U})$ is given by

$$|\mathbf{A}(\mathbf{U}) - \lambda I| = 0,$$

which gives us four eigenvalues

$$\lambda_1 = u - \sqrt{g\theta h}, \quad \lambda_2 = u, \quad \lambda_3 = u + \sqrt{g\theta h}, \quad \lambda_4 = 0. \tag{2.4}$$

The corresponding eigenvectors can be chosen as

$$\begin{aligned}
\mathbf{r}_1 &= \left(-\sqrt{gh\theta}, g\theta, 0, 0\right)^{\mathrm{T}}, \\
\mathbf{r}_2 &= (-h, 0, 2\theta, 0)^{\mathrm{T}}, \\
\mathbf{r}_3 &= \left(\sqrt{gh\theta}, g\theta, 0, 0\right)^{\mathrm{T}}, \\
\mathbf{r}_4 &= \left(gh\theta, -g\theta u, 0, u^2 - gh\theta\right)^{\mathrm{T}}.
\end{aligned} \tag{2.5}$$

From these formulas, one can see that the first and the fourth characteristic fields may coincide. Indeed, letting

$$(\lambda_1(\mathbf{U}), \mathbf{r}_1(\mathbf{U})) = (\lambda_4(\mathbf{U}), \mathbf{r}_4(\mathbf{U})),$$

we obtain a hypersurface of the space (h, u, θ, a) on which the first and the fourth characteristic fields coincide

$$\mathbf{C}_+ = \{(h, u, \theta, a) | u = \sqrt{g\theta h}\}. \tag{2.6}$$

Similarly, the third and the fourth characteristic fields may coincide:

$$(\lambda_3(\mathbf{U}), \mathbf{r}_3(\mathbf{U})) = (\lambda_4(\mathbf{U}), \mathbf{r}_4(\mathbf{U}))$$

on the hypersurface of the space (h, u, θ, a)

$$\mathbf{C}_- = \{(h, u, \theta, a) | u = -\sqrt{g\theta h}\}. \tag{2.7}$$

Furthermore, the second and the fourth eigenvalues may coincide when $u = 0$:

$$\mathbf{C}_0 = \{(h, u, \theta, a) | u = 0\}. \tag{2.8}$$

Therefore, system (2.2) may not be strictly hyperbolic in the entire domain.

On the other hand, it holds that

$$\begin{aligned}
D\lambda_2(\mathbf{U}) \cdot \mathbf{r}_2(\mathbf{U}) &= D\lambda_4(\mathbf{U}) \cdot \mathbf{r}_4(\mathbf{U}) = 0, \\
D\lambda_1(\mathbf{U}) \cdot \mathbf{r}_1(\mathbf{U}) &= D\lambda_3(\mathbf{U}) \cdot \mathbf{r}_3(\mathbf{U}) = \frac{3}{2}g\theta \neq 0, \quad h > 0.
\end{aligned}$$

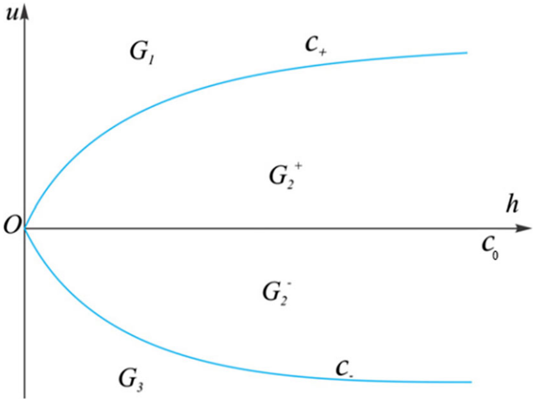

Fig. 1 Projection of strictly hyperbolic areas in the (h, u) plane

This means that the second and the fourth characteristic fields $(\lambda_2, \mathbf{r}_2)$, $(\lambda_4, \mathbf{r}_4)$ are linearly degenerate, and the first and the third characteristic fields $(\lambda_1, \mathbf{r}_1)$, $(\lambda_3, \mathbf{r}_3)$ are genuinely nonlinear in the open half-space $\{(h, u, \theta, a) | h > 0\}$.

From (2.6), (2.7), (2.8), it is convenient to set

$$\mathbf{C} = \mathbf{C}_+ \cup \mathbf{C}_0 \cup \mathbf{C}_-$$

which is the hypersurface on which the system fails to be strictly hyperbolic. We have seen that the system lacks strict hyperbolicity only on the surface **C**. However, this surface divides the phase domain into three sub-domains which are disjoint regions, or areas, denoted by $\mathbf{G}_1, \mathbf{G}_2$ and $\mathbf{G}_3$, so that in each region the system is strictly hyperbolic. More precisely,

$$\begin{aligned}
\mathbf{G}_1 &= \{(h, u, \theta, a) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ | \lambda_4 < \lambda_1 < \lambda_2 < \lambda_3\},\\
\mathbf{G}_2^+ &= \{(h, u, \theta, a) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ | \lambda_1 < \lambda_4 < \lambda_2 < \lambda_3\},\\
\mathbf{G}_2^- &= \{(h, u, \theta, a) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ | \lambda_1 < \lambda_2 < \lambda_4 < \lambda_3\},\\
\mathbf{G}_2 &= \mathbf{G}_2^+ \cup \mathbf{G}_2^-,\\
\mathbf{G}_3 &= \{(h, u, \theta, a) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ | \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4\};
\end{aligned}$$

see Fig. 1.

Let us recall the concept of supercritical and subcritical regions in the water resource engineering. The *Froude number* is defined by

$$Fr(\mathbf{U}) = \frac{|u|}{\sqrt{gh\theta}}.$$

If a state **U** such that $Fr(\mathbf{U}) = 1$, then it is said to be a *critical* state. If $Fr(\mathbf{U}) > 1$, then **U** is said to be a *supercritical* state. If $Fr(\mathbf{U}) < 1$, then **U** is said to be a *subcritical* state.

2.2 The Curve of Stationary Waves

First, consider stationary smooth solutions of system (1.1), which are time-independent solutions. It is not difficult to check that such a stationary smooth solution satisfies the following ordinary differential equations

$$\begin{aligned}
&(hu)' = 0,\\
&\left(\frac{u^2}{2} + g\theta(h+a)\right)' = 0,\\
&\theta' = 0,
\end{aligned} \tag{2.9}$$

where $(.)' = \mathrm{d}/\mathrm{d}x$. Thus, the trajectory of the system of three differential equations (2.9) passing through a fixed point $(h_0, u_0, \theta_0, a_0)$ satisfy the following algebraic equations

$$\begin{aligned}
&hu = h_0u_0,\\
&\frac{u^2}{2} + g\theta(h+a) = \frac{u_0^2}{2} + g\theta(h_0+a_0),\\
&\theta = \theta_0.
\end{aligned} \tag{2.10}$$

Stationary contact discontinuities associated with the fourth characteristic field λ_4 can be obtained as the limit of stationary smooth solutions; see [33]. Precisely, one can take a sequence of stationary smooth solutions of (1.1), which can be parameterized in terms of the water height h in the form $u = u(h), \theta = \theta(h), a = a(h)$. Then, by letting the water height h tend to a jump $h_\pm$, that is,

$$h \longrightarrow h_\pm,$$

and setting

$$u_\pm = u(h_\pm), \quad \theta_\pm = \theta(h_\pm), \quad a_\pm = a(h_\pm),$$

we can see that the states $\mathbf{U}_\pm = (h_\pm, u_\pm, \theta_\pm, a_\pm)^{\mathrm{T}}$ satisfy the jump conditions

$$\begin{aligned}
&[hu] = 0,\\
&\left[\frac{u^2}{2} + g\theta(h+a)\right] = 0,\\
&[\theta] = 0.
\end{aligned} \tag{2.11}$$

This implies that the curve of stationary contact waves which consists of all right-hand states $\mathbf{U}$ that can be connected to a given left-hand state $\mathbf{U}_0$ by a four-contact wave can be parameterized in h:

$$
\begin{aligned}
\mathcal{W}_0(\mathbf{U}_0) : u &= u(h) = \frac{h_0 u_0}{h},\\
\theta &= \theta(h) = \theta_0,\\
a &= a(h) = \frac{u_0^2 - u^2}{2g\theta_0} + h_0 - h + a_0.
\end{aligned}
\tag{2.12}
$$

Substituting for u in the third equation of system (2.12), and re-arranging terms, we obtain

$$
\begin{aligned}
u &= u(h) = \frac{h_0 u_0}{h},\\
\theta &= \theta(h) = \theta_0,\\
a &= a(h) = \frac{u_0^2}{2g\theta_0}\left(1 - \frac{h_0^2}{h^2}\right) + h_0 - h + a_0.
\end{aligned}
\tag{2.13}
$$

Therefore, if the topography levels on both sides of a discontinuity are known, we can determine the water height from the nonlinear algebraic equation

$$
a - a_0 + \frac{u_0^2}{2g\theta_0}\left(\frac{h_0^2}{h^2} - 1\right) + h - h_0 = 0.
$$

Thus, given a state $\mathbf{U}_0$ on the one side of a four-contact discontinuity, we can determine the state $\mathbf{U}$ on the other side by first evaluating the water height to be a zero of the function

$$
\varphi(h) = a - a_0 + \frac{u_0^2}{2g\theta_0}\left(\frac{h_0^2}{h^2} - 1\right) + h - h_0.
\tag{2.14}
$$

The other quantities of that state will follow from (2.13).

Let us now discuss about the finding zeros of the function φ in (2.14). Set

$$
\begin{aligned}
h_{\min}(\mathbf{U}_0) &:= \left(\frac{u_0^2 h_0^2}{g\theta_0}\right)^{\frac{1}{3}},\\
a_{\max}(\mathbf{U}_0) &:= \frac{u_0^2}{2g\theta_0}\left(1 - \frac{h_0^2}{h_{\min}^2}\right) + h_0 - h_{\min} + a_0.
\end{aligned}
$$

It is easy to verify that

$$
a_{\max}(\mathbf{U}_0) = \frac{h_0}{2}\left(\frac{u_0^{\frac{2}{3}}}{g^{\frac{1}{3}}\theta_0^{\frac{1}{3}} h_0^{\frac{1}{3}}} - 1\right)^2 \left(2 + \frac{u_0^{\frac{2}{3}}}{g^{\frac{1}{3}}\theta_0^{\frac{1}{3}} h_0^{\frac{1}{3}}}\right) + a_0.
\tag{2.15}
$$

We can see from Eq. (2.15) that $a_{\max}(\mathbf{U}_0) \geq a_0$ and the equality happens only along the curve $u^2 = gh\theta$ on which the system is not strictly hyperbolic.

Interesting properties of the function φ defined by (2.14) are obtained in the following lemma.

Lemma 2.1 *Suppose $u_0 \neq 0$. The function $\varphi(h)$, $h > 0$ is smooth, is convex, is decreasing in the interval $(0, h_{\min})$ and is increasing in the interval $(h_{\min}, +\infty)$, and satisfies the limit conditions*

$$\lim_{h\to 0} \varphi(h) = \lim_{h\to+\infty} \varphi(h) = +\infty. \tag{2.16}$$

Consequently, if $a \leq a_{\max}$, the function φ has two zeros $h_*(\mathbf{U}_0, a)$, $h^*(\mathbf{U}_0, a)$ such that $h_*(\mathbf{U}_0, a) \leq h_{\min}(\mathbf{U}_0) \leq h^*(\mathbf{U}_0, a)$. The inequalities are strict whenever $a < a_{\max}(\mathbf{U}_0)$.

Proof The smoothness of the function φ and the limit conditions at infinity (2.16) are obvious. Moreover, we have

$$\varphi'(h) = 1 - \frac{u_0^2 h_0^2}{g\theta_0 h^3} \tag{2.17}$$

(for $u_0 \neq 0$). Thus, $\varphi'(h)$ is positive if

$$h > \left(\frac{u_0^2 h_0^2}{g\theta_0}\right)^{\frac{1}{3}} = h_{\min}(\mathbf{U}_0),$$

and $\varphi'(h) < 0$ if $0 < h < h_{\min}(\mathbf{U}_0)$. This establishes the monotonicity properties of φ. Furthermore, we have

$$\varphi''(h) = 3\frac{u_0^2 h_0^2}{g\theta_0 h^4} \geq 0$$

which establishes the convexity of φ. Consequently, φ attains its minimum value at $h_{\min}(\mathbf{U}_0)$. That is,

$$\min \varphi = \varphi(h_{\min}(\mathbf{U}_0)).$$

If $a < a_{\max}(\mathbf{U}_0)$, then the minimum value $\varphi(h_{\min}(\mathbf{U}_0))$ is negative. It is derived from the limit conditions (2.16) that the equation $\varphi(h) = 0$ has exactly two distinct roots. In addition, it is not difficult to see that the two roots of the equation $\varphi(h) = 0$ coincide when $a = a_{\max}(\mathbf{U}_0)$. □

It is arisen from the above lemma that there are two choices of a contact wave from a given state. To select a physical contact wave, we need to impose an additional admissibility criterion as follows:

> (MC) (Monotonicity criterion)—Along any stationary curve $\mathcal{W}_0(\mathbf{U}_0)$, the bottom level a is monotone as a function of h. The total variation in the bottom-level component of any Riemann solution must not exceed $|a_L - a_R|$, where a_L and a_R are left-hand and right-hand bottom levels.

Note that a similar condition was used in [20,33]. Under the monotonicity criterion, a stationary contact wave always remain in the closure of each of the strictly hyperbolic domains. That is, if a state on the one side of an admissible four-contact discontinuity $\mathbf{U}_0 \in \bar{\mathbf{G}}_i$, where $\bar{\mathbf{G}}_i$ denotes the closure of the region $\mathbf{G}_i$, then the state on the other side of that contact $\mathbf{U}$ still belongs to $\bar{\mathbf{G}}_i$, $i = 1, 2, 3$; see [33]. Since the point $(h_{\min}(\mathbf{U}_0), h_0u_0/h_{\min}(\mathbf{U}_0))$ is critical, we deduce the following algorithm for the computation of the admissible contact wave with a given state on the one side and given left-hand and right-hand levels of topography.

(i) If $\mathbf{U}_0$ is a supercritical state, then the smaller root $h_*(\mathbf{U}_0, a)$ of φ defined by (2.14) is selected and can be computed by Newton's method with a starting point in the interval $(0, h_{\min}(\mathbf{U}_0))$;

(ii) If $\mathbf{U}_0$ is a subcritical state, then the larger root $h^*(\mathbf{U}_0, a)$ of φ defined by (2.14) is selected and can be computed by Newton's method with a starting point in the interval $(h_{\min}(\mathbf{U}_0), +\infty)$.

3 Numerical Schemes

In this section, we construct well-balanced numerical schemes for approximating solutions of system (1.1), relying on the arguments in the previous sections. Given a uniform time step $\triangle t$ and a special mesh size $\triangle x$, define $x_j = j \triangle x, j \in \mathbb{Z}$, $t_n = n \triangle t, n \in \mathbb{N}$, and denote by $\mathbf{U}_j^n$ the approximate value of the exact solution $\mathbf{U} = (h, hu, h\theta)^{\mathrm{T}}$ of system (1.1) at the time t^n in the interval $(x_{j-1/2}, x_{j+1/2})$. Set

$$\lambda = \frac{\triangle t}{\triangle x},$$

where λ is required to satisfy the CFL condition

$$\lambda < \frac{\text{CFL coefficient}}{\max_{h,u}\{|u| + \sqrt{gh\theta}\}}, \quad 0 < \text{CFL coefficient} \le 1.$$

3.1 Constructing the Well-Balanced Schemes

The method we use to construct the well-balanced scheme consists of two steps:

Step 1: First, the source term on the right-hand side of system (1.1) will be absorbed in stationary contact waves at each grid node: For each state $\mathbf{U}_j^n$ at the time t^n in the interval $(x_{j-1/2}, x_{j+1/2})$, there is a state $\mathbf{U}_{j-1,+}^n$ on the left and a state $\mathbf{U}_{j+1,-}^n$ on the right of that interval that can be connected to $\mathbf{U}_j^n$ by a contact wave.

Step 2: Second, the states on both sides of the contact waves in Step 1 will be incorporated in a standard numerical flux for conservation laws: $\mathbf{U}_{j-1,+}^n$ and $\mathbf{U}_{j+1,-}^n$ which will replace $\mathbf{U}_{j-1}^n$ and $\mathbf{U}_{j+1}^n$, respectively, in a standard numerical flux for conservation laws.

Precisely, the scheme is defined by

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \lambda\left(k(\mathbf{U}_j^n, \mathbf{U}_{j+1,-}^n) - k(\mathbf{U}_{j-1,+}^n, \mathbf{U}_j^n)\right), \tag{3.1}$$

where k, which will be referred to as the *underlying numerical flux*, can be any standard numerical flux for conservation laws. For example, one may take the underlying numerical flux to be the one of the Lax–Friedrichs schemes:

$$k_{LF}(\mathbf{U},\mathbf{V})=\frac{1}{2}(\mathbf{f}(\mathbf{U})+\mathbf{f}(\mathbf{V}))-\frac{1}{2\lambda}(\mathbf{V}-\mathbf{U}). \tag{3.2}$$

The computation of the states $\mathbf{U}^n_{j+1,-}$, $\mathbf{U}^n_{j-1,+}$ is now discussed. In scheme (3.1), the states

$$\mathbf{U}^n_{j+1,-}=[(h,hu,h\theta)^{\mathrm{T}}]^n_{j+1,-},\mathbf{U}^n_{j-1,+}=[(h,hu,h\theta)^{\mathrm{T}}]^n_{j-1,+}$$

are defined by observing that the entropy is constant across each stationary jump, and by computing $h^n_{j+1,-}$, $u^n_{j+1,-}$, $\theta^n_{j+1,-}$ from the system

$$\begin{aligned}
&h^n_{j+1}u^n_{j+1}=h^n_{j+1,-}u^n_{j+1,-},\\
&\frac{(u^n_{j+1})^2}{2}+g\theta^n_{j+1}(h^n_{j+1}+a_{j+1})=\frac{(u^n_{j+1,-})^2}{2}+g\theta^n_{j+1}(h^n_{j+1,-}+a_j),\\
&\theta^n_{j+1}=\theta^n_{j+1,-},
\end{aligned}\tag{3.3}$$

and computing $h^n_{j-1,+}$, $u^n_{j-1,+}$, $\theta^n_{j-1,+}$ from the system

$$\begin{aligned}
&h^n_{j-1}u^n_{j-1}=h^n_{j-1,+}u^n_{j-1,+},\\
&\frac{(u^n_{j-1})^2}{2}+g\theta^n_{j-1}(h^n_{j-1}+a_{j-1})=\frac{(u^n_{j-1,+})^2}{2}+g\theta^n_{j-1}(h^n_{j-1,+}+a_j),\\
&\theta^n_{j-1}=\theta^n_{j-1,+}.
\end{aligned}\tag{3.4}$$

Observe that from Eq. (2.15), we have

$$a_{\max}(\mathbf{U}^n_{j+1})=\frac{h^n_{j+1}}{2}\left(\frac{(u^n_{j+1})^{\frac{2}{3}}}{g^{\frac{1}{3}}(\theta^n_{j+1})^{\frac{1}{3}}(h^n_{j+1})^{\frac{1}{3}}}-1\right)^2\left(2+\frac{(u^n_{j+1})^{\frac{2}{3}}}{g^{\frac{1}{3}}(\theta^n_{j+1})^{\frac{1}{3}}(h^n_{j+1})^{\frac{1}{3}}}\right)+a^n_{j+1}. \tag{3.5}$$

Thus, $a_{\max}(\mathbf{U}^n_{j+1})\geq a_{j+1}$ and therefore system (3.3) have a solution provided

$$a_{\max}(\mathbf{U}^n_{j+1})\geq a_j.$$

To ensure that the scheme always works, we may modify the value a_j and re-assign it to a new value $a_{\max}(\mathbf{U}^n_{j+1})$, if necessary. A similar procedure is used for system (3.4). In particular, the bottom function a is expected not to have too large jumps near the critical surface.

Scheme (3.1) is well balanced in the sense that it can capture exactly stationary waves. Indeed, it holds for any stationary waves that

$$\begin{aligned}
&h_{j+1}^n u_{j+1}^n = h_j u_j,\\
&\frac{(u_{j+1}^n)^2}{2} + g\theta_{j+1}^n(h_{j+1}^n + a_{j+1}) = \frac{u_j^2}{2} + g\theta_{j+1}^n(h_j + a_j),\\
&\theta_{j+1}^n = \theta_j,
\end{aligned} \tag{3.6}$$

and

$$\begin{aligned}
&h_{j-1}^n u_{j-1}^n = h_j u_j,\\
&\frac{(u_{j-1}^n)^2}{2} + g\theta_{j-1}^n(h_{j-1}^n + a_{j-1}) = \frac{u_j^2}{2} + g\theta_{j-1}^n(h_j + a_j),\\
&\theta_{j-1}^n = \theta_j.
\end{aligned} \tag{3.7}$$

This yields

$$\begin{aligned}
&h_{j+1,-}^n = h_j, u_{j+1,-}^n = u_j, \theta_{j+1,-}^n = \theta_j,\\
&h_{j-1,+}^n = h_j, u_{j-1,+}^n = u_j, \theta_{j-1,+}^n = \theta_j.
\end{aligned}$$

Thus,

$$\mathbf{U}_{j+1,-}^n = \mathbf{U}_j^n, \quad \mathbf{U}_{j-1,+}^n = \mathbf{U}_j^n,$$

and therefore, the scheme gives

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n.$$

This means that the solution is stationary.

Now, it is interesting to observe that a particular choice of scheme (3.1), which takes the underlying numerical flux (3.2), can preserve the positivity of the height and temperature of water.

Theorem 3.1 *Scheme* (3.1)–(3.2) *is positively conservative for the water height. That is, if $h_j^0 \geq 0$ for all $j \in \mathbb{Z}$, then $h_j^n \geq 0$ for all $n \in \mathbb{N}$, $j \in \mathbb{Z}$.*

Proof We need only to verify that for an arbitrary fixed n, if $h_j^n \geq 0, \forall j$, then $h_j^{n+1} \geq 0, \forall j$. Indeed, it is not to difficult to check that $h_{j-1,+}^n$ and $h_{j+1,-}^n$ are also nonnegative for all j. Moreover, it follows from Eq. (3.1) that

$$\begin{aligned}
h_j^{n+1} &= \frac{h_{j-1,+}^n + h_{j+1,-}^n}{2} + \frac{\lambda}{2}\left(h_{j-1,+}^n u_{j-1,+}^n - h_{j+1,-}^n u_{j+1,-}^n\right)\\
&\geq \frac{h_{j-1,+}^n + h_{j+1,-}^n}{2} - \frac{\lambda}{2}\max\{|u_{j-1,+}^n|, |u_{j+1,-}^n|\}\left(h_{j-1,+}^n + h_{j+1,-}^n\right)\\
&\geq \frac{h_{j-1,+}^n + h_{j+1,-}^n}{2}\left(1 - \lambda\max\{|u_{j-1,+}^n|, |u_{j+1,-}^n|\}\right)\\
&\geq 0
\end{aligned}$$

due to the CFL condition. This completes the proof of Theorem 3.1.

Theorem 3.2 *Scheme (3.1)–(3.2) is positively conservative for the water temperature: if $\theta_j^0 \geq 0$ for all $j \in \mathbb{Z}$, then $\theta_j^n \geq 0$ for all $n \in \mathbb{N}$, $j \in \mathbb{Z}$.*

Proof By induction, we need only to show that for any fixed $n \geq 0$, if $\theta_j^n \geq 0$ for all integers j, then $\theta_j^{n+1} \geq 0$ for all integers j. Indeed, since the temperature θ remains constant across stationary waves, it holds that

$$\theta_{j-1,+}^n = \theta_{j-1}^n \geq 0, \quad \theta_{j+1,-}^n = \theta_{j+1}^n \geq 0,$$

for all integers j. It therefore follows from Eq. (3.1) that

$$\begin{aligned}
h_j^{n+1}\theta_j^{n+1} &= \frac{h_{j-1,+}^n\theta_{j-1,+}^n + h_{j+1,-}^n\theta_{j+1,-}^n}{2} \\
&\quad + \frac{\lambda}{2}\left(h_{j-1,+}^n\theta_{j-1,+}^n u_{j-1,+}^n - h_{j+1,-}^n\theta_{j+1,-}^n u_{j+1,-}^n\right) \\
&\geq \frac{h_{j-1,+}^n\theta_{j-1,+}^n + h_{j+1,-}^n\theta_{j+1,-}^n}{2} - \frac{\lambda}{2}\max\{|u_{j-1,+}^n|, |u_{j+1,-}^n|\} \\
&\quad \left(h_{j-1,+}^n\theta_{j-1,+}^n + h_{j+1,-}^n\theta_{j+1,-}^n\right) \\
&\geq \frac{h_{j-1,+}^n\theta_{j-1,+}^n + h_{j+1,-}^n\theta_{j+1,-}^n}{2}\left(1 - \lambda\max\{|u_{j-1,+}^n|, |u_{j+1,-}^n|\}\right) \\
&\geq 0
\end{aligned}$$

due to the CFL condition. Since $h_j^{n+1} \geq 0$, we get $\theta_j^{n+1} \geq 0$. □

3.2 Fast and Stable Schemes by Underlying Numerical Fluxes

In general, one may choose any standard numerical flux k in (3.1) as the underlying numerical flux. To make the scheme fast and stable, we form the underlying numerical flux to be any convex combination of the numerical fluxes of a first-order and stable scheme and a high-order one. For example, one can take the following convex combinations

$$k(\mathbf{U}, \mathbf{V}) = \theta k_{LF}(\mathbf{U}, \mathbf{V}) + (1-\theta)k_{LW}(\mathbf{U}, \mathbf{V}), \tag{3.8}$$

for $0 \leq \theta \leq 1$, where k_{LW} is the Lax–Wendroff numerical flux:

$$k_{LW}(\mathbf{U}, \mathbf{V}) = \frac{1}{2}(\mathbf{f}(\mathbf{U}) + \mathbf{f}(\mathbf{V})) - \frac{\lambda}{2}\mathbf{A}^2(\mathbf{U}, \mathbf{V})(\mathbf{V} - \mathbf{U}), \tag{3.9}$$

and

$$\begin{aligned}
\mathbf{U} &= (h, hu, h\theta)^{\mathrm{T}}, \\
\mathbf{f}(\mathbf{U}) &= \left(hu, hu^2 + \frac{gh^2\theta}{2}, uh\theta\right)^{\mathrm{T}}, \\
\mathbf{A}(\mathbf{U}, \mathbf{V}) &= \hat{\mathbf{A}}_{j-\frac{1}{2}} \text{ is a Roe matrix.}
\end{aligned}$$

In particular, taking $\theta = 1/2$ in (3.8) leads us to the one of the FORCE schemes (see [34]):

$$k_{FORCE}(\mathbf{U},\mathbf{V}) = \frac{1}{2}k_{LF}(\mathbf{U},\mathbf{V}) + \frac{1}{2}k_{LW}(\mathbf{U}.\mathbf{V}). \tag{3.10}$$

Let us construct a Roe matrix of system (1.1) as follows. The matrix $\hat{\mathbf{A}}_{j-\frac{1}{2}} = \mathbf{A}(\mathbf{U}_{j-1},\mathbf{U}_j)$ is chosen to be some approximation to $\mathbf{f}'(\mathbf{U})$ valid a neighborhood of the data $\mathbf{U}_{j-1}$ and $\mathbf{U}_j$. So $\hat{\mathbf{A}}_{j-\frac{1}{2}}$ satisfies condition:

$$\mathbf{f}(\mathbf{U}_j) - \mathbf{f}(\mathbf{U}_{j-1}) = \hat{\mathbf{A}}_{j-\frac{1}{2}}(\mathbf{U}_j - \mathbf{U}_{j-1}). \tag{3.11}$$

Parameter vector $\mathbf{Z}$: $\mathbf{U} = \mathbf{U}(\mathbf{Z})$ invertible $\mathbf{Z} = \mathbf{Z}(\mathbf{U})$. We will integrate along the path

$$\mathbf{Z}(\xi) = \mathbf{Z}_{j-1} + (\mathbf{Z}_j - \mathbf{Z}_{j-1})\xi, \quad 0 \le \xi \le 1, \tag{3.12}$$

where $\mathbf{Z}_i = \mathbf{Z}(\mathbf{U}_i)$ for $i = j-1, j$. Then $\mathbf{Z}'(\xi) = \mathbf{Z}_j - \mathbf{Z}_{j-1}$ is independent of ξ, and so

$$\mathbf{f}(\mathbf{U}_j) - \mathbf{f}(\mathbf{U}_{j-1}) = \int_0^1 \frac{d}{d\xi}\mathbf{f}(\mathbf{U}(\mathbf{Z}(\xi)))d\xi = \int_0^1 \mathbf{f}'(\mathbf{U}(\mathbf{Z}(\xi)))d\xi(\mathbf{Z}_j - \mathbf{Z}_{j-1}). \tag{3.13}$$

We also have

$$\mathbf{U}_j - \mathbf{U}_{j-1} = \int_0^1 \frac{d}{d\xi}\mathbf{U}(\mathbf{Z}(\xi))d\xi = \int_0^1 \mathbf{U}'(\mathbf{Z}(\xi))d\xi(\mathbf{Z}_j - \mathbf{Z}_{j-1}). \tag{3.14}$$

Setting

$$\hat{\mathbf{C}}_{j-\frac{1}{2}} = \int_0^1 \mathbf{f}'(\mathbf{U}(\mathbf{Z}(\xi)))d\xi, \tag{3.15}$$

$$\hat{\mathbf{B}}_{j-\frac{1}{2}} = \int_0^1 \mathbf{U}'(\mathbf{Z}(\xi))d\xi. \tag{3.16}$$

From (3.13), (3.14), we obtain

$$\mathbf{f}(\mathbf{U}_j) - \mathbf{f}(\mathbf{U}_{j-1}) = \hat{\mathbf{C}}_{j-\frac{1}{2}}(\mathbf{Z}_j - \mathbf{Z}_{j-1}), \tag{3.17}$$

$$\mathbf{U}_j - \mathbf{U}_{j-1} = \hat{\mathbf{B}}_{j-\frac{1}{2}}(\mathbf{Z}_j - \mathbf{Z}_{j-1}). \tag{3.18}$$

From (3.18), and from using

$$\hat{\mathbf{A}}_{j-\frac{1}{2}} = \hat{\mathbf{C}}_{j-\frac{1}{2}}\hat{\mathbf{B}}^{-1}_{j-\frac{1}{2}},$$

we obtain the relation (3.11). Now, set in system (1.1),

$$\mathbf{U} = \begin{bmatrix} h \\ hu \\ h\theta \end{bmatrix} = \begin{bmatrix} U^1 \\ U^2 \\ U^3 \end{bmatrix}. \tag{3.19}$$

It holds that

$$\mathbf{f}(\mathbf{U}) = \begin{bmatrix} hu \\ hu^2 + \dfrac{gh^2\theta}{2} \\ uh\theta \end{bmatrix} = \begin{bmatrix} U^2 \\ \dfrac{(U^2)^2}{U^1} + \dfrac{gU^1U^3}{2} \\ \dfrac{U^2U^3}{U^1} \end{bmatrix}. \tag{3.20}$$

From (3.19), (3.20), we obtain Jacobian matrix for the shallow equations with flat bottom (a is constant):

$$\frac{\partial \mathbf{f}(\mathbf{U})}{\partial \mathbf{U}} = \begin{bmatrix} 0 & 1 & 0 \\ -\left(\dfrac{U^2}{U^1}\right)^2 + \dfrac{gU^3}{2} & \dfrac{2U^2}{U^1} & \dfrac{gU^1}{2} \\ -\dfrac{U^2U^3}{(U^1)^2} & \dfrac{U^3}{U^1} & \dfrac{U^2}{U^1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -u^2 + \dfrac{gh\theta}{2} & 2u & \dfrac{gh}{2} \\ -u\theta & \theta & u \end{bmatrix}. \tag{3.21}$$

As a parameter vector, one may choose

$$\mathbf{Z} = \frac{\mathbf{U}}{\sqrt{h}} = \begin{bmatrix} \sqrt{h} \\ \sqrt{h}u \\ \sqrt{h}\theta \end{bmatrix} = \begin{bmatrix} Z^1 \\ Z^2 \\ Z^3 \end{bmatrix}.$$

A straight calculation gives us

$$\frac{\partial \mathbf{U}(\mathbf{Z})}{\partial \mathbf{Z}} = \begin{bmatrix} 2Z^1 & 0 & 0 \\ Z^2 & Z^1 & 0 \\ Z^3 & 0 & Z^1 \end{bmatrix}, \tag{3.22}$$

and

$$\frac{\partial \mathbf{f}(\mathbf{U}(\mathbf{Z}))}{\partial \mathbf{Z}} = \begin{bmatrix} Z^2 & Z^1 & 0 \\ \dfrac{3g(Z^1)^2Z^3}{2} & 2Z^2 & \dfrac{g(Z^1)^3}{2} \\ 0 & Z^3 & Z^2 \end{bmatrix}. \tag{3.23}$$

From (3.12), it holds for each component $p = 1, 2, 3$ that

$$Z^p = Z^p_{j-1} + (Z^p_j - Z^p_{j-1})\xi, \quad p = \overline{1,3}, 0 \le \xi \le 1.$$

This yields

$$\begin{aligned}\int_0^1 Z^p(\xi)d\xi =&\frac{1}{2}(Z^p_{j-1}+Z^p_j)=\overline{Z}^p,\\ \int_0^1 (Z^1)^2Z^3d\xi =&\frac{1}{4}(Z^1_{j-1})^2Z^3_{j-1}+\frac{1}{12}(Z^1_{j-1})^2Z^3_j+\frac{1}{6}Z^1_{j-1}Z^1_jZ^3_{j-1}+\frac{1}{6}Z^1_{j-1}Z^1_jZ^3_j\\ &+\frac{1}{12}(Z^1_j)^2Z^3_{j-1}+\frac{1}{4}(Z^1_j)^2Z^3_j=\overline{\overline{Z}},\\ \int_0^1 (Z^1)^3d\xi =&\frac{1}{2}(Z^1_j+Z^1_{j-1}).\frac{1}{2}\left[(Z^1_j)^2+(Z^1_{j-1})^2\right]=\overline{Z}^1\overline{h},\end{aligned}$$

where

$$\begin{aligned}&\overline{h}=\frac{h_{j-1}+h_j}{2},\\ &\overline{\overline{Z}}=\frac{h^{\frac{3}{2}}_{j-1}\theta_{j-1}}{4}+\frac{h_{j-1}\sqrt{h_j}\theta_j}{12}+\frac{\sqrt{h_j}h_{j-1}\theta_{j-1}}{6}+\frac{\sqrt{h_{j-1}}h_j\theta_j}{6}+\frac{h_j\sqrt{h_{j-1}}\theta_{j-1}}{12}+\frac{h^{\frac{3}{2}}_j\theta_j}{4}.\end{aligned}$$

Substituting (3.22) into (3.16), we have

$$\hat{\mathbf{B}}_{j-\frac{1}{2}}=\begin{bmatrix}2\overline{Z}^1 & 0 & 0\\ \overline{Z}^2 & \overline{Z}^1 & 0\\ \overline{Z}^3 & 0 & \overline{Z}^1\end{bmatrix},$$

which yields the inverse matrix

$$\hat{\mathbf{B}}^{-1}_{j-\frac{1}{2}}=\begin{bmatrix}\frac{1}{2\overline{Z}^1} & 0 & 0\\ -\frac{\overline{Z}^2}{2(\overline{Z}^1)^2} & \frac{1}{\overline{Z}^1} & 0\\ -\frac{\overline{Z}^3}{2(\overline{Z}^1)^2} & 0 & \frac{1}{\overline{Z}^1}\end{bmatrix}.$$

Similarly, substituting (3.23) into (3.16), we have

$$\hat{\mathbf{C}}_{j-\frac{1}{2}}=\begin{bmatrix}\overline{Z}^2 & \overline{Z}^1 & 0\\ \frac{3g\overline{\overline{Z}}}{2} & 2\overline{Z}^2 & \frac{g\overline{Z}^1\overline{h}}{2}\\ 0 & \overline{Z}^3 & \overline{Z}^2\end{bmatrix}.$$

So

$$
\begin{aligned}
\hat{\mathbf{A}}_{j-\frac{1}{2}} &= \mathbf{A}(\mathbf{U}_{j-1}, \mathbf{U}_j) = \mathbf{A}(\mathbf{U}_j, \mathbf{U}_{j-1}) = \hat{\mathbf{C}}_{j-\frac{1}{2}} \hat{\mathbf{B}}^{-1}_{j-\frac{1}{2}} \\
&= \begin{bmatrix} 0 & 1 & 0 \\ \dfrac{3g\overline{\overline{Z}}}{4\overline{Z}^1} - \left(\dfrac{\overline{Z}^2}{\overline{Z}^1}\right)^2 - \dfrac{g\overline{h}}{4}\dfrac{\overline{Z}^3}{\overline{Z}^1} & \dfrac{2\overline{Z}^2}{\overline{Z}^1} & \dfrac{g\overline{h}}{2} \\ -\dfrac{\overline{Z}^2}{\overline{Z}^1}\dfrac{\overline{Z}^3}{\overline{Z}^1} & \dfrac{\overline{Z}^3}{\overline{Z}^1} & \dfrac{\overline{Z}^2}{\overline{Z}^1} \end{bmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\overline{Z}^2}{\overline{Z}^1} &= \frac{\sqrt{h_{j-1}}u_{j-1} + \sqrt{h_j}u_j}{\sqrt{h_{j-1}} + \sqrt{h_j}}, \\
\frac{\overline{Z}^3}{\overline{Z}^1} &= \frac{\sqrt{h_{j-1}}\theta_{j-1} + \sqrt{h_j}\theta_j}{\sqrt{h_{j-1}} + \sqrt{h_j}}, \\
\overline{Z}^1 &= \frac{\sqrt{h_{j-1}} + \sqrt{h_j}}{2}.
\end{aligned}
$$

Specially, if $\mathbf{U}_{j-1} = \mathbf{U}_j$ then $\hat{\mathbf{A}}_{j-\frac{1}{2}}$ is Jacobian matrix in (3.21) where $h = h_j, u = u_j, \theta = \theta_j$.

4 Numerical Tests

This section is devoted to numerical tests, where the exact solutions and the approximate solutions are computed and compared. The errors are computed for different mesh sizes, where we take the underlying numerical flux in our well-balanced scheme (3.1) to be the one of the Lax–Friedrichs schemes (3.2) and the FORCE scheme (3.10). The exact solution is denoted by $\mathbf{U} = (h, u, \theta)^{\mathrm{T}}$, and the approximate solution at the step size h is denoted by $\mathbf{U}_h^{\mathrm{LF}}, \mathbf{U}_h^{\mathrm{FORCE}}$ corresponding to the scheme using the numerical flux of the Lax–Friedrichs scheme and of the FORCE scheme, respectively.

Exact solutions and approximate solutions of the Riemann problem for (1.1) will be computed on the interval $x \in [-1, 1]$ at the time $t = 0.1$. The CFL constant is chosen to be $\lambda = 0.5$ in all of the tests.

4.1 Stationary Waves

In the last section, scheme (3.1) is shown to be well balanced in the sense that it can capture exactly stationary waves. It is interesting to see the numerical demonstration of this property. One can see that the errors are very small and almost stable, since they are caused by the errors from the input data.

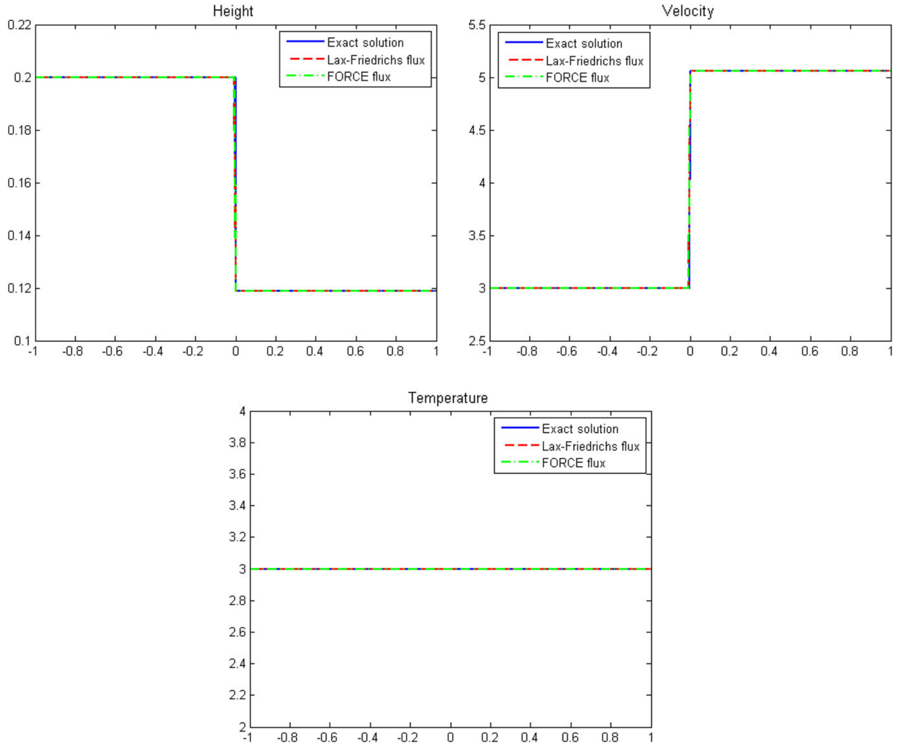

Fig. 2 Exact stationary wave (4.1) approximated by the well-balanced scheme with 250 mesh points

Table 1 Errors of numerical approximations for different mesh sizes for test case of height bottom discontinuous

N	$\|\|\mathbf{U} - \mathbf{U}_h^{\mathrm{LF}}\|\|_{L^1}$	$\|\|\mathbf{U} - \mathbf{U}_h^{\mathrm{FORCE}}\|\|_{L^1}$
250	1.1786×10^{-5}	1.155×10^{-5}
500	1.1504×10^{-5}	1.1385×10^{-5}
1000	1.1369×10^{-5}	1.1311×10^{-5}
2000	1.1303×10^{-5}	1.1274×10^{-5}
4000	1.127×10^{-5}	1.1256×10^{-5}

4.1.1 Test 1: Stationary Contact Discontinuities

This test is devoted to a stationary discontinuity of system (1.1)

$$(h, u, \theta, a)(x, t) = \begin{cases} (h_L, u_L, \theta_L, a_L), & \text{if } x < 0, \\ (h_R, u_R, \theta_R, a_R), & \text{if } x > 0, \end{cases} \tag{4.1}$$

where the left-hand and the right-hand states of the wave are *approximately* given by

$$\begin{aligned} (h_L, u_L, \theta_L, a_L) &= (0.2, 3, 3, 1.2), \\ (h_R, u_R, \theta_R, a_R) &= (0.118727337921731, 5.053595999899697, 3, 1). \end{aligned}$$

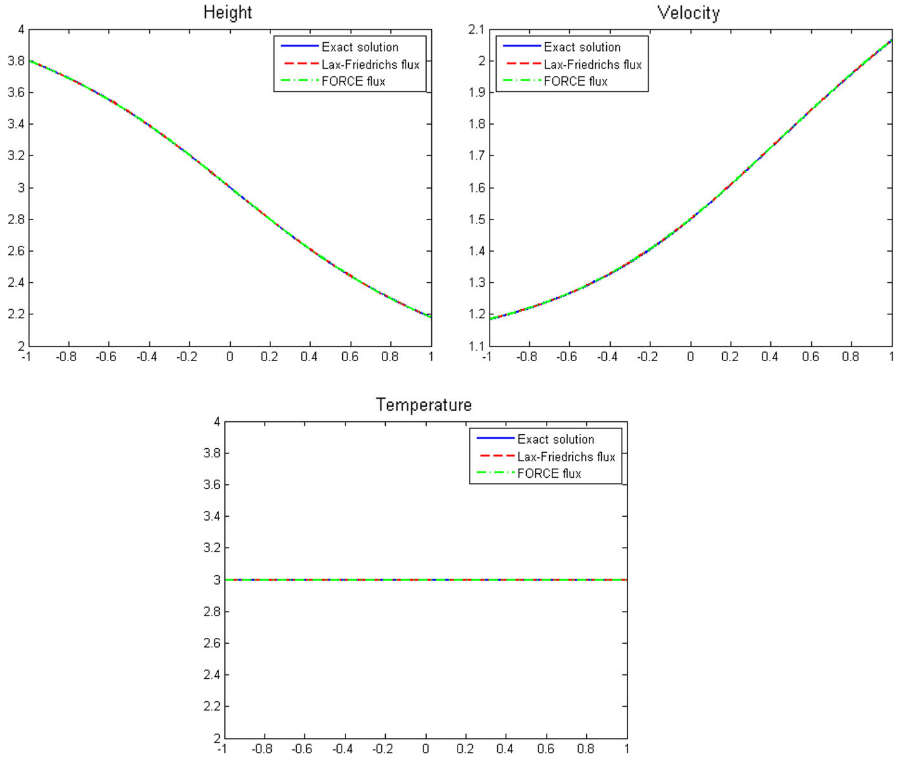

Fig. 3 Exact stationary wave with bottom height continuous approximated by the well-balanced scheme with 250 mesh points

Table 2 Errors of numerical approximations for different mesh sizes for test case of height bottom continuous

N	$\|\|\mathbf{U}-\mathbf{U}_h^{\mathrm{LF}}\|\|_{L^1}$	$\|\|\mathbf{U}-\mathbf{U}_h^{\mathrm{FORCE}}\|\|_{L^1}$
250	1.3074×10^{-4}	1.3237×10^{-4}
500	1.4068×10^{-4}	1.4147×10^{-4}
1000	1.367×10^{-4}	1.3699×10^{-4}
2000	1.3601×10^{-4}	1.364×10^{-4}
4000	1.3424×10^{-4}	1.3448×10^{-4}

Table 3 States for the exact solution near dry zone in Test 3

	h	u	θ	a
U_L	0.01	4	3	1.2
U_1	0.00757245265656	5.28230440178	3	1
U_2	0.0553068463584	3.0397850844	3	1
U_3	0.0478971339495	3.0397850844	4	1
U_R	0.02	2	4	1

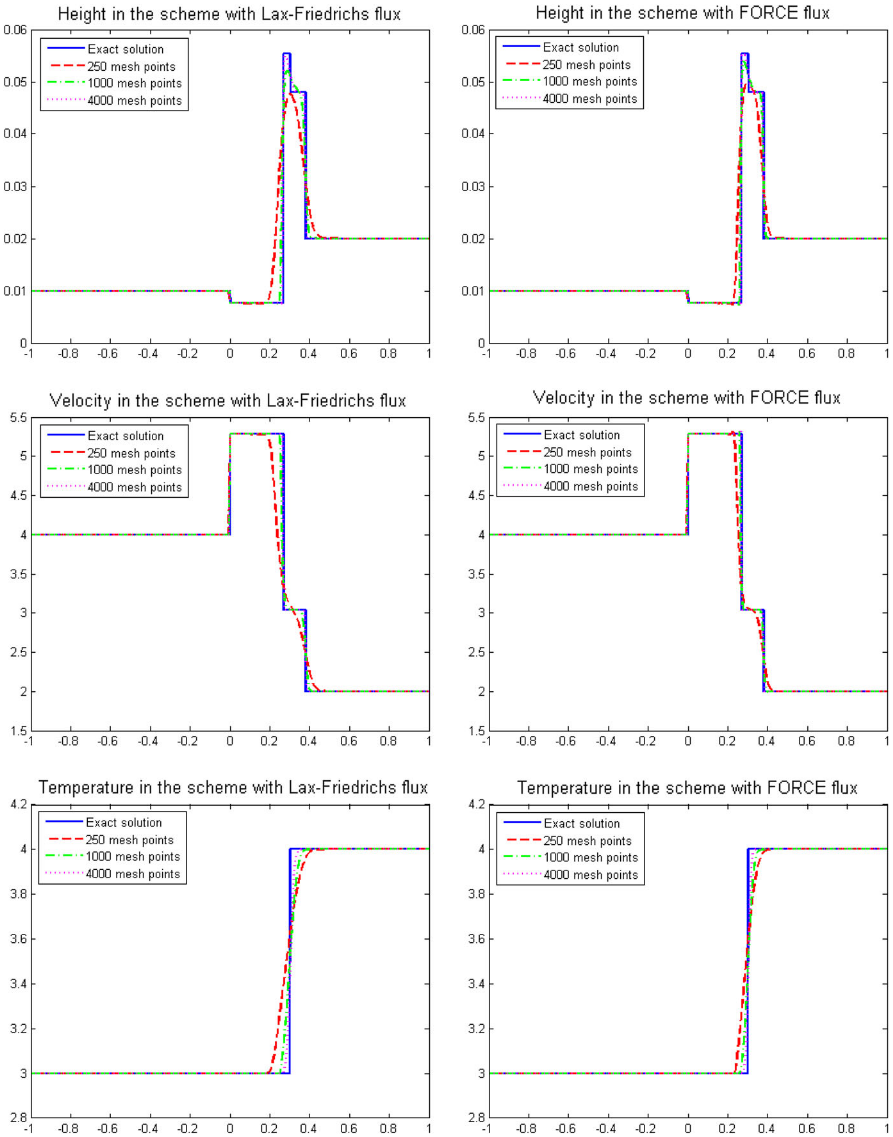

Fig. 4 Test 3: Exact Riemann solution with Riemann data (4.3) approximated by the well-balanced scheme with Lax–Friedrichs and FORCE flux

As mentioned above, the approximate solutions are computed by the well-balanced scheme with the numerical fluxes (3.2) and (3.10). The exact solution, which is a stationary contact discontinuity, and the approximate solutions are plotted in Fig. 2. The errors for different mesh sizes are given in Table 1.

Table 4 Errors of numerical approximations for different mesh sizes for Test 3

N	$\|\|\mathbf{U}-\mathbf{U}_h^{\text{LF}}\|\|_{L^1}$	$\|\|\mathbf{U}-\mathbf{U}_h^{\text{FORCE}}\|\|_{L^1}$
250	0.16423	0.10371
500	0.096432	0.061294
1000	0.055003	0.034385
2000	0.032295	0.020713
4000	0.019649	0.012904

Figure 2 and Table 1 show that the approximate solutions almost coincide with the exact solution. The very errors are caused by the errors from the input data and the tolerance of the code iterative algorithm.

4.1.2 Test 2: Smooth Stationary Waves

Let us take the exact solution to be a smooth stationary wave. The solution is thus independent of time. Precisely, let us take the smooth topography as

$$a(x) = 2 + \tan^{-1}(x), \tag{4.2}$$

and the initial water height, water velocity and water temperature as

$$(h_0, u_0, \theta_0) = (3, 1.5, 3).$$

Then, the exact solution is given by algebraic system (2.10). Note that the values of the exact solution at any mesh point x_j can be computed by the exact solution and the approximate solutions are displayed in Fig. 3. The corresponding errors are given in Table 2.

From this test, one can see that the approximate solutions almost coincide with the exact solution. The errors are stable and caused by the input data and the tolerance of the code for iterative algorithms.

4.2 Nonstationary Solutions

4.2.1 Test 3: Solutions Near Dry Zone and in the Supercritical Region

This test is aimed to show the convergence in the supercritical region and the positive conservation for the water height of the scheme, as demonstrated by Theorem 3.1. For this purpose, we take an exact nonstationary solution of the Riemann problem for (1.1) near the dry zone, i.e.,

$$h \approx 0,$$

in an interval. Then, we will check that the scheme still gives the approximate solutions whose water height remains well above zero. Precisely, the Riemann data are given by

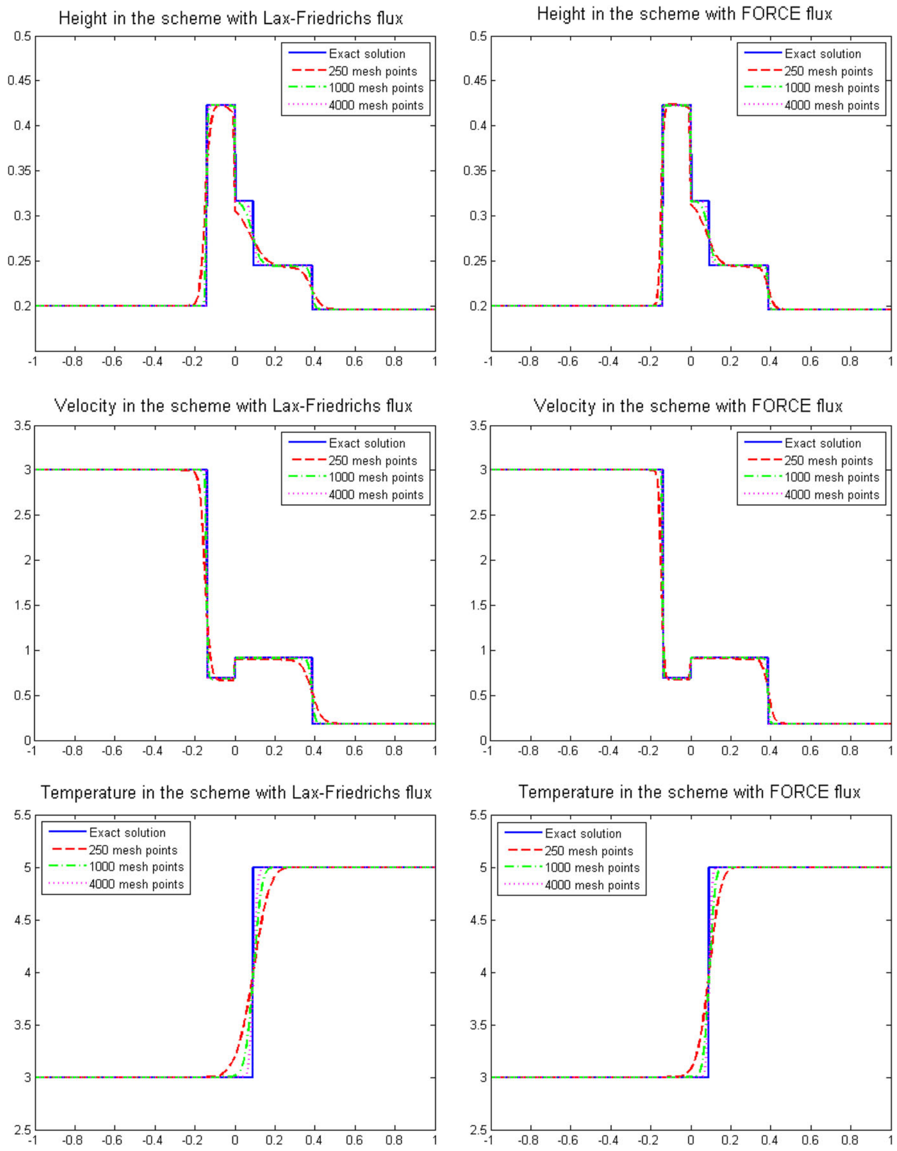

Fig. 5 Test 4: Exact Riemann solution with Riemann data (4.4) approximated by the well-balanced scheme with Lax–Friedrichs and FORCE flux

$$(h, u, \theta, a)(x, 0) = \begin{cases} (0.01, 4, 3, 1.2), & \text{if } x < 0, \\ (0.02, 2, 4, 1), & \text{if } x > 0. \end{cases} \tag{4.3}$$

The exact solution begins with a stationary contact wave from U_L to U_1, followed by a 1-shock wave from U_1 to U_2, a 2-contact from U_2 to U_3, and finally a 3-shock

Table 5 Errors of numerical approximations for different mesh sizes for Test 4

N	$\|\|\mathbf{U}-\mathbf{U}_h^{\text{LF}}\|\|_{L^1}$	$\|\|\mathbf{U}-\mathbf{U}_h^{\text{FORCE}}\|\|_{L^1}$
250	0.19767	0.13497
500	0.1282	0.08828
1000	0.081036	0.053296
2000	0.051389	0.034103
4000	0.03438	0.023685

wave from U_3 to U_R; see Table 3. It is not difficult to check that both left-hand and right-hand states U_L, U_R belong to the supercritical region.

The exact solution of the Riemann problem for (1.1) and the approximate solutions by our scheme are plotted in Fig. 4. The corresponding errors are given in Table 4. The errors become smaller as the mesh sizes are smaller.

The picture on the left upper corner of Fig. 4 shows approximate values of the water height by the scheme for different mesh sizes, which all remain well positive.

4.2.2 Test 4: Solutions in Both Supercritical and Subcritical Regions

In this test, we consider the approximation for the Riemann problem when the left-hand state is supercritical and the right-hand state is subcritical. Precisely, the Riemann data are given by

$$(h,u,\theta,a)(x,0)=\begin{cases}(0.2,3,3,1), & \text{if } x<0,\\ (0.195816152469433, 0.182801122801997, 5, 1.1), & \text{if } x>0.\end{cases} \tag{4.4}$$

See Fig. 5 and Table 5 for the comparison of the errors of the well-balanced scheme with Lax–Friedrichs and FORCE flux.

In this test, one can see that the errors are small and decrease significantly when the mesh sizes tends to zero.

4.2.3 Test 5: Approximation in Subsonic Region

In this test, we consider the approximation of the exact solution of the Riemann problem when the left-hand and right-hand states are subcritical. Precisely, the Riemann data are given by

$$(h,u,\theta,a)(x,0)=\begin{cases}(1,3,2,1), & \text{if } x<0,\\ (0.912012600264880, -0.176048159061341, 3, 1.1), & \text{if } x>0.\end{cases} \tag{4.5}$$

The exact solution and the approximate solutions are plotted in Fig. 6. The errors and orders of accuracy are given in Table 6.

Again, in this test, the scheme still possesses a very good accuracy. When using the underlying numerical flux of the FORCE scheme, the scheme has a much better accuracy than using the one of the Lax–Friedrichs schemes.

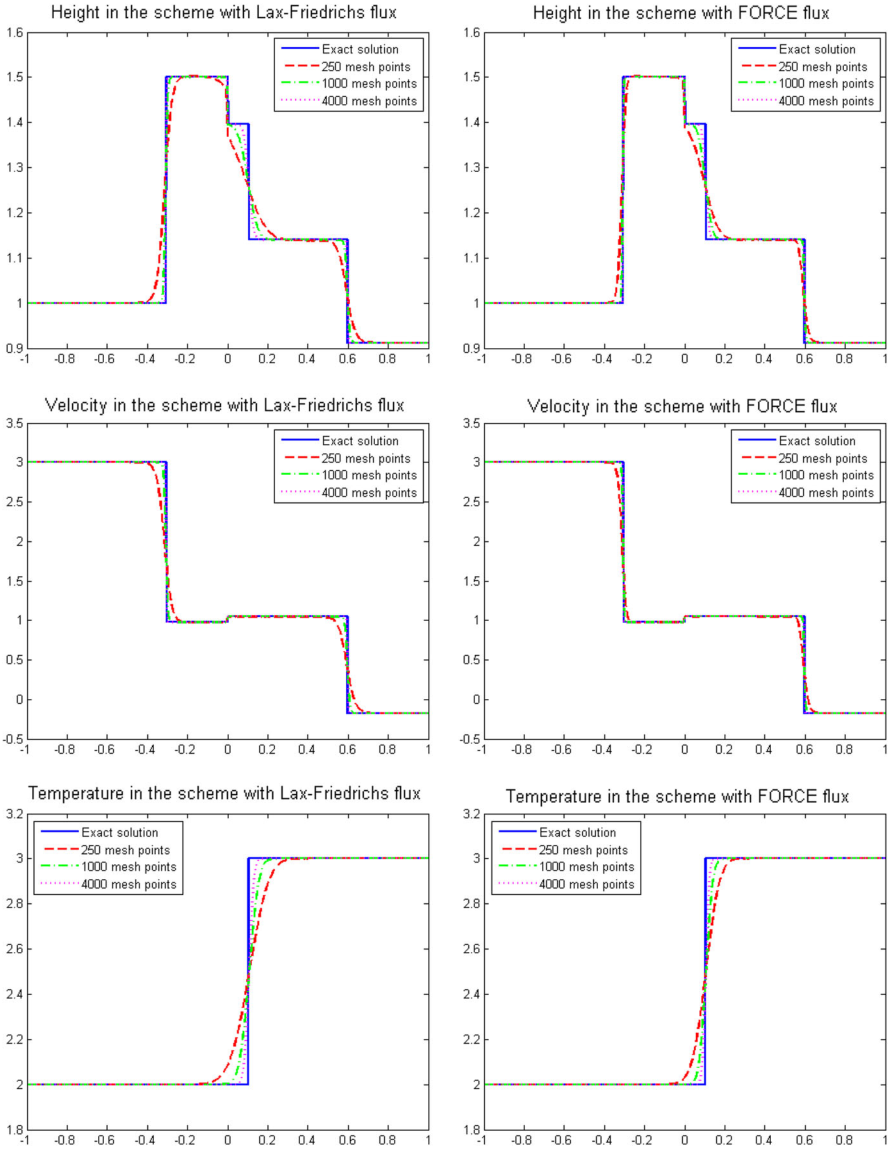

Fig. 6 Test 5: Exact Riemann solution with Riemann data (4.5) approximated by the well-balanced scheme with Lax–Friedrichs and FORCE flux

5 Conclusions

Numerical scheme (3.1) for the shallow water equations with variable topography and horizontal temperature gradient is constructed and tested. It is well balanced in the sense that it can capture stationary waves. In addition, the scheme with a particular choice of the underlying numerical flux to be the one of the Lax–Friedrichs schemes

Table 6 Errors and orders of accuracy of well-balanced scheme using Lax–Friedrichs and FORCE flux in Test 5

N	$\|\|\mathbf{U}-\mathbf{U}_h^{\mathrm{LF}}\|\|_{L^1}$	LF's order	$\|\|\mathbf{U}-\mathbf{U}_h^{\mathrm{FORCE}}\|\|_{L^1}$	FORCE's order
250	0.18749		0.11594	
500	0.11534	0.7009	0.072534	0.6766
1000	0.068111	0.7599	0.042699	0.7645
2000	0.042547	0.6788	0.027594	0.6298
4000	0.027285	0.6409	0.018083	0.6097

possesses interesting properties: it preserves the positivity of the water height and the positivity of the temperature. This well-balanced scheme provides us with very reasonable accuracy. It is interesting to note that we can improve the accuracy of the scheme by using the underlying numerical flux obtained as a convex combination of the numerical fluxes of a first-order and stable scheme and a high-order one, such as the FORCE scheme. Tests show that the scheme is convergent for all the circumstances when the exact solution belongs to either the supercritical region or the subcritical region, or expands in both supercritical and subcritical regions.

Acknowledgements The authors would like to thank the reviewers for their very valuable comments and suggestions. This research is funded by the Vietnam National University HoChiMinh City (VNU-HCM) under Grant No. B2018-28-01.

References

1. Ambroso, A., Chalons, C., Coquel, F., Galié, T.: Relaxation and numerical approximation of a two-fluid two-pressure diphasic model. Math. Mod. Numer. Anal. **43**, 1063–1097 (2009)
2. Ambroso, A., Chalons, C., Raviart, P.-A.: A Godunov-type method for the seven-equation model of compressible two-phase flow. Comput. Fluids **54**, 67–91 (2012)
3. Audusse, E., Bouchut, F., Bristeau, M.-O., Klein, R., Perthame, B.: A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows. SIAM J. Sci. Comput. **25**, 2050–2065 (2004)
4. Baudin, M., Coquel, F., Tran, Q.-H.: A semi-implicit relaxation scheme for modeling two-phase flow in a pipeline. SIAM J. Sci. Comput. **27**, 914–936 (2005)
5. Botchorishvili, R., Perthame, B., Vasseur, A.: Equilibrium schemes for scalar conservation laws with stiff sources. Math. Comput. **72**, 131–157 (2003)
6. Botchorishvili, R., Pironneau, O.: Finite volume schemes with equilibrium type discretization of source terms for scalar conservation laws. J. Comput. Phys. **187**, 391–427 (2003)
7. Chertock, A., Kurganov, A., Liu, Y.: Central-upwind schemes for the system of shallow water equations with horizontal temperature gradients. Numer. Math. **127**, 595–639 (2014)
8. Coquel, F., Godlewski, E., Perthame, B., In, Rascle, P.: Some new Godunov and relaxation methods for two-phase flow problems. In: Toro, E. F. (ed.) Godunov Methods (Oxford, 1999), pp. 179–188. Kluwer/Plenum, New York (2001)
9. Cuong, D.H., Thanh, M.D.: A Godunov-type scheme for the isentropic model of a fluid flow in a nozzle with variable cross-section. Appl. Math. Comput. **256**, 602–629 (2015)
10. Coquel, F., Hérard, J.-M., Saleh, K., Seguin, N.: Two properties of two-velocity two-pressure models for two-phase flows. Commun. Math. Sci. **12**, 593–600 (2014)
11. Dubroca, B.: Positively conservative Roe's matrixfor Euler equations. In: 16th International Conference on Numerical Methods in Fluid Dynamics (Arcachon, 1998), Lecture Notes in Phys., vol. 515, pp. 272–277. Springer, Berlin. https://doi.org/10.1007/BFb0106594

12. Dal Maso, G., LeFloch, P.G., Murat, F.: Definition and weak stability of nonconservative products. J. Math. Pures Appl. **74**, 483–548 (1995)
13. Gallardo, J.M., Parés, C., Castro, M.: On a well-balanced high-order finite volume scheme for shallow water equations with topography and dry areas. J. Comput. Phys. **227**, 574–601 (2007)
14. Gallouet, T., Herard, J.-M., Seguin, N.: Some approximate Godunov schemes to compute shallow-water equations with topography. Comput. Fluids **32**, 479–513 (2003)
15. Greenberg, J.M., Leroux, A.Y.: A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. SIAM J. Numer. Anal. **33**, 1–16 (1996)
16. Han, X., Li, G.: Well-balanced finite difference WENO schemes for the Ripa model. Comput. Fluids **134–135**, 1–10 (2016)
17. Isaacson, E., Temple, B.: Convergence of the 2 × 2 Godunov method for a general resonant nonlinear balance law. SIAM J. Appl. Math. **55**, 625–640 (1995)
18. Kröner, D., Thanh, M.D.: Numerical solutions to compressible flows in a nozzle with variable cross-section. SIAM J. Numer. Anal. **43**, 796–824 (2005)
19. Kröner, D., LeFloch, P.G., Thanh, M.D.: The minimum entropy principle for fluid flows in a nozzle with discntinuous crosssection. Math. Mod. Numer. Anal. **42**, 425–442 (2008)
20. LeFloch, P.G., Thanh, M.D.: The Riemann problem for fluid flows in a nozzle with discontinuous cross-section. Commun. Math. Sci. **1**, 763–797 (2003)
21. LeFloch, P.G., Thanh, M.D.: A Godunov-type method for the shallow water equations with variable topography in the resonant regime. J. Comput. Phys. **230**, 7631–7660 (2011)
22. Li, G., Caleffi, V., Qi, Z.K.: A well-balanced finite difference WENO scheme for shallow water flow model. Appl. Math. Comput. **265**, 1–16 (2015)
23. Li, G., Song, L.N., Gao, J.M.: High order well-balanced discontinuous Galerkin methods based on hydrostatic reconstruction for shallow water equations. J. Comput. Appl. Math. **340**, 546–560 (2018)
24. Qian, S.G., Shao, F.J., Li, G.: High order well-balanced discontinuous Galerkin methods for shallow water flow under temperature fields. Comput. Appl. Math. (2018). https://doi.org/10.1007/s40314-018-0662-y
25. Ripa, P.: Conservation laws for primitive equations models with inhomogeneous layers. Geophys. Astrophys. Fluid Dyn. **70**, 85–111 (1993)
26. Ripa, P.: On improving a one-layer ocean model with thermodynamics. J. Fluid Mech. **303**, 169–201 (1995)
27. Rosatti, G., Begnudelli, L.: The Riemann Problem for the one-dimensional, free-surface shallow water equations with a bed step: theoretical analysis and numerical simulations. J. Comput. Phys. **229**, 760–787 (2010)
28. Sanchez-Linares, C., Morales de Luna, T., Castro Diaz, M.J.: A HLLC scheme for Ripa model. Appl. Math. Comput. **72**, 369–384 (2016)
29. Saurel, R., Abgrall, R.: A multi-phase Godunov method for compressible multifluid and multiphase flows. J. Comput. Phys. **150**, 425–467 (1999)
30. Tian, B., Toro, E.F., Castro, C.E.: A path-conservative method for a five-equation model of two-phase flow with an HLLC-type Riemann solver. Comput. Fluids **46**, 122–132 (2011)
31. Thanh, M.D.: A phase decomposition approach and the Riemann problem for a model of two-phase flows. J. Math. Anal. Appl. **418**, 569–594 (2014)
32. Thanh, M.D.: The Riemann problem for a non-isentropic fluid in a nozzle with discontinuous cross-sectional area. SIAM J. Appl. Math. **69**, 1501–1519 (2009)
33. Thanh, M.D.: The Riemann problem for the shallow water equations with horizontal temperature gradients. Appl. Math. Comput. **325**, 159–178 (2018)
34. Thanh, M.D.: Building fast well-balanced two-stage numerical schemes for a model of two-phase flows. Commun. Nonlinear Sci. Num. Simul. **19**, 1836–1858 (2014)
35. Thanh, M.D., Kröner, D., Chalons, C.: A robust numerical method for approximating solutions of a model of two-phase flows and its properties. Appl. Math. Comput. **219**, 320–344 (2012)
36. Touma, R., Klingenberg, C.: Well-balanced central finite volume methods for the Ripa system. Appl. Numer. Math. **97**, 42–68 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.