Journal of Computational Physics 422 (2020) 109784

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

An efficient finite difference method for the shallow water equations

Lukas Lundgren *, Ken Mattsson

Uppsala University, Department of Information Technology, P O Box 337, S-751 05 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 15 March 2019
Received in revised form 6 August 2020
Accepted 12 August 2020
Available online 18 August 2020

Keywords:
Finite difference methods
Shallow water equations
High-order accuracy
Stability
Boundary treatment
Well-balanced

ABSTRACT

A high-order explicit finite difference scheme is derived solving the shallow water equations. The boundary closures are based on the diagonal-norm summation-by-parts (SBP) framework and the boundary conditions are imposed using a penalty (SAT) technique. Flux-splitting combined with upwind SBP operators is used to naturally introduce artificial dissipation. The scheme is tested against various benchmark problems where high-order convergence is verified for smooth solutions. A particular discretization of the source term is used leading to a well-balanced scheme. We also present an application: A simplified incident wave simulation with wave-channel interaction using a multi-block setup. Experiments suggest that a bathymetry consisting of many spikes could provide a dispersing effect on an incoming wave.

1. Introduction

The shallow water equations were first proposed in 1871 by Saint-Venant to model flow in an open channel [1]. They can be viewed as a special 2D case of the Navier-Stokes equations and are valid when the wavelength of the modeled phenomena is much larger compared to its amplitude. With wide use in ocean and hydraulic engineering the equations are used to model many flows such as: rivers [2,3], flooding [4], dam breaks [5], overland flow [6,7] and tsunami [8–10]. In this paper we focus on smooth and wet solutions far away from dry states. Under these conditions it is well-known that higher order methods (as compared to first and second order methods) capture wave dominated phenomena, such as gravity wave propagation, well since they allow a considerable reduction in the degrees of freedom for a given error tolerance. In particular, high-order finite difference methods are ideally suited for problems of this kind, see the pioneering paper by Kreiss and Oliger concerning hyperbolic problems [11].

The major difficulty with high-order finite difference methods is to obtain a stable boundary treatment. A robust and well-proven high-order finite difference methodology, for well-posed initial boundary value problems (IBVPs), is to combine summation-by-parts (SBP) operators [12,13] and the simultaneous approximation term (SAT) method [14] to impose boundary conditions (BCs). Recent examples of the SBP-SAT approach can be found in [15–20]. An added benefit of the SBP-SAT method is that it naturally extends to multi-block geometries while retaining the essential single-block properties: strict stability, accuracy, and conservation [21]. Thus, problems involving complex domains or non-smooth geometries are amenable to the approach. References [22–24] report applications of the SBP-SAT method to problems involving nontrivial geometries.

* Corresponding author.
E-mail addresses: lukas.lundgren@it.uu.se (L. Lundgren), ken.mattsson@it.uu.se (K. Mattsson).

https://doi.org/10.1016/j.jcp.2020.109784
0021-9991/

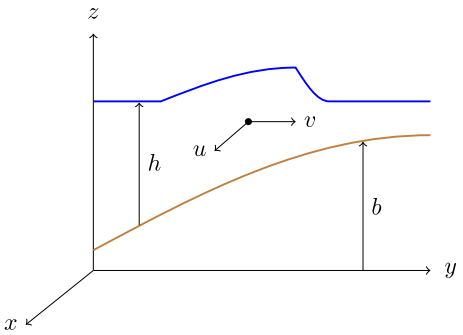


Fig. 1. Cross section of flow governed by the 2D shallow water equations.

SBP-SAT only guarantees linear stability. For nonlinear conservation laws, such as the shallow water equations, a purely central difference based finite difference scheme is incapable of handling non-smooth solutions without the addition of artificial dissipation (AD). Flux-splitting combined with upwind SBP operators naturally introduce AD to damp spurious oscillations without destroying linear stability [19]. Rydin [20] reports that the AD from the upwind operators adds robustness in the presence of point sources and when imposing characteristic non-reflecting BCs. Additionally, the numerical experiments presented in [19] report favorable convergence properties for first and second order hyperbolic systems. In this paper we introduce a high-order accurate explicit upwind SBP-SAT finite difference approximation for the shallow water equations.

There is now a known relationship between discontinuous Galerkin methods and SBP-SAT finite difference schemes [25]. Recently, this relationship was used to derive an entropy conserving discontinuous Galerkin method for the 1D shallow water equations [26]. This was later extended into 2D [27] and then further extended with a shock-capturing technique as well as positivity preservation capability [28]. This is far from the only example of this. Considerable attention has been paid on shock-capturing schemes applicable to the shallow water equations as well as schemes that capture wet/dry transitions correctly. See the review article [29] for high-order numerical methods for the shallow water equations and the references therein as well as recent examples [30–35,10,36] involving positivity-preserving schemes. The aim of this paper, however, is to obtain a high-order accurate finite difference approximation with stable boundary treatment. Although the focus is on smooth and wet solutions where nonlinear effects are small, the scheme is still robust when shocks are present in the solution. We also present an application: A simplified incident wave simulation with wave-channel interaction using a multi-block setup.

One main difficulty in solving the shallow water equations is preserving non-trivial steady state solutions, such as the "lake at rest" [37] problem. A numerical scheme that achieves this to machine precision is well-balanced. Methods that are not well-balanced can produce spurious waves proportional to the grid size (a so-called "numerical storm" [38,39]) which pollutes the solution quality. To achieve a well-balanced scheme we have used a particular discretization of the source term as well as a modified flux-splitting following the ideas introduced in [40].

This paper is organized as follows. In Section 2 the shallow water equations are introduced and rewritten into quasilinear form. In Section 3 well-posedness of the continuous 1D problem as well as semi-discrete stability are analyzed. In Section 4 the 2D problem is analyzed. In Sections 5 and 6 the numerical method is verified against some benchmark problems in 1D and 2D, respectively. In Section 7 an application is presented where some experiments are performed.

2. The shallow water equations

The shallow water equations can be derived from depth integrating the Navier-Stokes equations under the assumption that the wavelength of the modeled phenomena is much larger than its amplitude and under the hydrostatic assumption [41]. The variables solved for are the water height $h = h(x, y, t)$ and the vertically averaged velocity $u = u(x, y, t)$ in the x-direction and $v = v(x, y, t)$ in the y-direction. The height h is defined in relation to the bottom topography $b = b(x, y)$, see Fig. 1. Neglecting friction and viscosity the shallow water equations in conservative form are given by

$$\begin{aligned} &h_t + (hu)_x + (hv)_y = 0, \\ &(hu)_t + \left(hu^2 + \frac{1}{2}gh^2\right)_x + (huv)_y + ghb_x = 0, \\ &(hv)_t + (huv)_x + \left(hv^2 + \frac{1}{2}gh^2\right)_y + ghb_y = 0, \end{aligned} \tag{1}$$

where the first equation represents mass balance and the other two equations represent conservation of momentum in the x- and y-direction, respectively. The constant g denotes the gravitational acceleration. Since friction and viscosity have been neglected, (1) represents a special 2D case of the isentropic Euler equations with pressure given by a hydrostatic law. The external force acting upon the fluid in (1) is gravity, which means that the focus of this system is to model gravity waves.

2.1. The 1D problem

The 1D version of the shallow water equations (1) is given by

$$\begin{aligned} &h_t + (hu)_x = 0, \\ &(hu)_t + \left(hu^2 + \frac{1}{2} g h^2 \right)_x + ghb_x = 0, \end{aligned} \tag{2}$$

which can be reformulated as

$$\mathbf{q}_t + \mathbf{F}(\mathbf{q})_x + \mathbf{G}(\mathbf{q}) = 0, \tag{3}$$

where $\mathbf{q} = \begin{bmatrix} q_1 & q_2 \end{bmatrix}^T = \begin{bmatrix} h & hu \end{bmatrix}^T$, $\mathbf{F}(\mathbf{q}) = \begin{bmatrix} q_2 & \frac{q_2^2}{q_1} + \frac{1}{2} g q_1^2 \end{bmatrix}^T$ and $\mathbf{G}(\mathbf{q}) = \begin{bmatrix} 0 & g(q_1 + b)b_x - (\frac{1}{2} g b^2)_x \end{bmatrix}^T$. Here the source term $\mathbf{G}(\mathbf{q})$ has been written in a special form. When (3) is discretized, this will lead to a well-balanced scheme as was shown by Xing and Shu [40]. To analyze well-posedness for linearized problems it is useful to rewrite the problem into quasilinear form. Equation (3) rewritten into quasilinear form is given by

$$\mathbf{q}_t + \mathbf{A}(\mathbf{q})\mathbf{q}_x + \mathbf{C}\mathbf{q} = \mathbf{E}, \tag{4}$$

where

$$\mathbf{A} = \mathbf{A}(\mathbf{q}) = \mathbf{F}'(\mathbf{q}) = \begin{bmatrix} 0 & 1 \\ g q_1 - \frac{q_2^2}{q_1^2} & \frac{2 q_2}{q_1} \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} 0 & 0 \\ g b_x & 0 \end{bmatrix}, \qquad \mathbf{E} = \begin{bmatrix} 0 \\ (\frac{1}{2} g b^2)_x - g b b_x \end{bmatrix}. \tag{5}$$

The eigenvalues of $\mathbf{A}$ are given by

$$\Lambda_\pm = \frac{q_2}{q_1} \pm \sqrt{g q_1} = u \pm \sqrt{gh}. \tag{6}$$

These eigenvalues represent the wave speeds. In the subsequent sections we will use that $\mathbf{A}$ can be expressed in terms of its eigenvalues and eigenvectors

$$\mathbf{A} = \mathbf{W} \mathbf{\Lambda} \mathbf{W}^{-1}, \tag{7}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \Lambda_+ & 0 \\ 0 & \Lambda_- \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} W_+ & W_- \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \Lambda_+ & \Lambda_- \end{bmatrix}. \tag{8}$$

In Section 3 well-posedness of (2) is analyzed. The discrete version of (2) is also considered where a stable upwind SBP-SAT approximation is derived.

3. 1D analysis

In this section we first analyze well-posedness of (2) and then analyze stability of the corresponding semi-discrete approximation. The 1D shallow water equations with forcing function $\mathbf{E} = \mathbf{E}(x, t)$ and boundary operators $L_{l,r}$ are given by

$$\begin{aligned} &\mathbf{q}_t = -\mathbf{A}\mathbf{q}_x - \mathbf{C}\mathbf{q} + \mathbf{E}, && x_l \le x \le x_r, && t > 0, \\ &L_l \mathbf{q} = g_l(t), && x = x_l, && t \ge 0, \\ &L_r \mathbf{q} = g_r(t), && x = x_r, && t \ge 0, \\ &\mathbf{q} = q_0(x), && x_l \le x \le x_r, && t = 0. \end{aligned} \tag{9}$$

An IBVP is well-posed if a unique solution exists that depends continuously on the problem's data (initial data, boundary data and forcing) [42]. For nonlinear IBVPs existence and uniqueness are in general hard to prove and we will assume that the problem has a unique solution. In the present study we will only consider linear stability. Assuming that the solution is smooth, linear stability also implies nonlinear stability [43]. For more details about the relation between nonlinear and linear stability via localization principles, we refer to [44]. In this section we analyze the linearized (frozen coefficient) problem and derive well-posed BCs. The analysis is presented for the sake of completeness and more details on the procedure can be found here [42].

Definition 3.1. Given two real valued vector functions (with k unknowns) $\mathbf{f} = \begin{bmatrix} f^{(1)}(x,t) & f^{(2)}(x,t) & \dots & f^{(k)}(x,t) \end{bmatrix}^T$ and $\mathbf{g} = \begin{bmatrix} g^{(1)}(x,t) & g^{(2)}(x,t) & \dots & g^{(k)}(x,t) \end{bmatrix}^T$ and let $\mathbf{f}, \mathbf{g} \in L^2[x_l, x_r]$, then the inner product is defined as

$$(\mathbf{f}, \mathbf{g}) = \int_{x_l}^{x_r} \mathbf{f}^T \mathbf{g} \, dx,$$

with the following implied norm $||\mathbf{f}(\cdot, t)||^2 = (\mathbf{f}, \mathbf{f})$.

Definition 3.2. A linear IBVP is well-posed if for every $\mathbf{E} \in C^\infty$ there exists a smooth and unique solution such that

$$||\mathbf{q}(\cdot, t)|| \leq K e^{\alpha t} \left(||q_0|| + \max_{0 \leq \tau \leq t} ||\mathbf{E}(\cdot, \tau)|| \right),$$

where α and K are constants independent of the forcing function $\mathbf{E}$ and the initial data q_0, with a minimal number of BCs.

Lemma 3.1. *The linearized IBVP of (9) is well-posed according to Definition 3.2 if the BCs are chosen such that*

$$-\left[\boldsymbol{q}^T \boldsymbol{A} \boldsymbol{q} \right]_{x_l}^{x_r} \leq 0,$$

with a minimal number of BCs which are given by the number of positive and negative eigenvalues of $\boldsymbol{A}$.

Proof. We start from the frozen coefficient version of (9)

$$\bar{\mathbf{q}}_t = -\bar{\mathbf{A}} \bar{\mathbf{q}}_x - \mathbf{C} \bar{\mathbf{q}} + \mathbf{E}, \tag{10}$$

where $\bar{\mathbf{A}}$ is a constant coefficient matrix. The problem is rewritten in symmetric form. We know that

$$\exists \, \mathbf{T} \quad \text{s.t.} \quad \mathbf{T}^{-1} \bar{\mathbf{A}} \mathbf{T} = \tilde{\mathbf{A}}, \tag{11}$$

where $\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}$. The problem in symmetric form is given by

$$\begin{aligned}
&\tilde{\mathbf{q}}_t = -\tilde{\mathbf{A}} \tilde{\mathbf{q}}_x - \tilde{\mathbf{C}} \tilde{\mathbf{q}} + \tilde{\mathbf{E}}, && x_l \leq x \leq x_r, && t > 0, \\
&\tilde{L}_l \tilde{\mathbf{q}} = g_l(t), && x = x_l, && t \geq 0, \\
&\tilde{L}_r \tilde{\mathbf{q}} = g_r(t), && x = x_r, && t \geq 0, \\
&\tilde{\mathbf{q}} = \tilde{q}_0(x), && x_l \leq x \leq x_r, && t = 0,
\end{aligned} \tag{12}$$

where $\tilde{\mathbf{q}} = \mathbf{T}^{-1} \bar{\mathbf{q}}$, $\tilde{L}_l = L_l \mathbf{T}$, $\tilde{L}_r = L_r \mathbf{T}$, $\tilde{\mathbf{E}} = \mathbf{T}^{-1} \mathbf{E}$ and $\tilde{\mathbf{C}} = \mathbf{T}^{-1} \mathbf{C} \mathbf{T}$. To obtain an energy estimate, we multiply (12) by $\tilde{\mathbf{q}}^T$ from the left and integrate by parts to obtain

$$(\tilde{\mathbf{q}}, \tilde{\mathbf{q}}_t) = (\tilde{\mathbf{A}} \tilde{\mathbf{q}}_x, \tilde{\mathbf{q}}) - \left[\tilde{\mathbf{q}}^T \tilde{\mathbf{A}} \tilde{\mathbf{q}} \right]_{x_l}^{x_r} - (\tilde{\mathbf{q}}, \tilde{\mathbf{C}} \tilde{\mathbf{q}}) + (\tilde{\mathbf{q}}, \tilde{\mathbf{E}}). \tag{13}$$

The transpose is given by

$$(\tilde{\mathbf{q}}_t, \tilde{\mathbf{q}}) = -(\tilde{\mathbf{A}} \tilde{\mathbf{q}}_x, \tilde{\mathbf{q}}) - (\tilde{\mathbf{q}}, \tilde{\mathbf{C}}^T \tilde{\mathbf{q}}) + (\tilde{\mathbf{E}}, \tilde{\mathbf{q}}). \tag{14}$$

Combining (13) and (14) the following energy estimate is obtained

$$\frac{d}{dt} ||\tilde{\mathbf{q}}||^2 = (\tilde{\mathbf{q}}, (-\tilde{\mathbf{C}} - \tilde{\mathbf{C}}^T) \tilde{\mathbf{q}}) - \left[\tilde{\mathbf{q}}^T \tilde{\mathbf{A}} \tilde{\mathbf{q}} \right]_{x_l}^{x_r} + (\tilde{\mathbf{E}}, \tilde{\mathbf{q}}) + (\tilde{\mathbf{q}}, \tilde{\mathbf{E}}). \tag{15}$$

By choosing a minimal number of BCs such that $-\left[\tilde{\mathbf{q}}^T \tilde{\mathbf{A}} \tilde{\mathbf{q}} \right]_{x_l}^{x_r} \leq 0$ holds, we obtain

$$\frac{d}{dt} ||\tilde{\mathbf{q}}||^2 \leq 2\alpha \, ||\tilde{\mathbf{q}}||^2 + 2 ||\tilde{\mathbf{q}}|| \, ||\tilde{\mathbf{E}}||, \tag{16}$$

where $\alpha = \max_x(|-\mathbf{C} - \mathbf{C}^T|)$. Employing the chain rule and dividing (16) by $2||\tilde{\mathbf{q}}||$ the estimate

$$\frac{d}{dt} ||\tilde{\mathbf{q}}|| \leq \alpha ||\tilde{\mathbf{q}}|| + ||\tilde{\mathbf{E}}||, \tag{17}$$

is given. By introducing the variable $r = e^{-\alpha t} ||\tilde{\mathbf{q}}||$,

$$e^{\alpha t}\left(\frac{dr}{dt}+\alpha r\right)\leq \alpha e^{\alpha t} r+||\tilde{\mathbf{E}}|| \tag{18}$$

is obtained, which gives

$$\frac{dr}{dt}\leq e^{-\alpha t}||\tilde{\mathbf{E}}||. \tag{19}$$

By integrating over time,

$$r(t)\leq r(0)+\int_0^t e^{-\alpha\tau}||\tilde{\mathbf{E}}(\cdot,\tau)||d\tau \tag{20}$$

is obtained. Changing variables back to $\tilde{\mathbf{q}}$ yields

$$||\tilde{\mathbf{q}}||\leq e^{\alpha t}\left(||\tilde{q}_0||+\int_0^t e^{-\alpha\tau}||\tilde{\mathbf{E}}(\cdot,\tau)||d\tau\right)\leq K e^{\alpha t}\left(||\tilde{q}_0||+\max_{0\leq\tau\leq t}||\tilde{\mathbf{E}}(\cdot,\tau)||\right), \tag{21}$$

where K is a constant. This leads to

$$||\tilde{\mathbf{q}}||\leq \text{cond}(\mathbf{T}) K e^{\alpha t}\left(||\tilde{q}_0||+\max_{0\leq\tau\leq t}||\mathbf{E}(\cdot,\tau)||\right). \tag{22}$$

Hence, the problem is well-posed according to Definition 3.2 by choosing a minimal number of BCs such that $-\left[\tilde{\mathbf{q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{q}}\right]_{x_l}^{x_r}\leq 0$ holds. This implies that $-\left[\mathbf{q}^T\mathbf{A}\mathbf{q}\right]_{x_l}^{x_r}\leq 0$ holds. □

3.1. Boundary analysis

By Lemma 3.1, the 1D shallow water equations are linearly well-posed if

$$-\left[\mathbf{q}^T\mathbf{A}\mathbf{q}\right]_{x_l}^{x_r}\leq 0. \tag{23}$$

When choosing BCs, (23) must hold with a minimal number of BCs. One way to analyze this is to consider the eigenvalues of $\mathbf{A}$ which are $u\pm\sqrt{gh}$. Depending on the sign of these eigenvalues different numbers of BCs at each boundary need to be specified. To this end a dimensionless quantity called the Froude number is useful

$$\text{Fr}=\frac{|u|}{\sqrt{gh}}. \tag{24}$$

This quantity is analogous to the Mach number in gas dynamics. If $\text{Fr}<1$ the flow is said to be subcritical meaning that there is one positive and one negative eigenvalue. When this occurs, one BC at each boundary has to be specified. When $\text{Fr}>1$, the flow is supercritical which means that two BCs have to be specified at one of the boundaries depending on the sign of u. A transcritical regime can exist in the solution if parts of the flow are subcritical and other parts are supercritical.

3.1.1. Transformation of BCs

The aim of this section is to rewrite the BCs using a similar technique as in [45,20] which will be useful in the semi-discrete stability analysis. Let $\mathbf{X}=\begin{bmatrix}X_+ & X_-\end{bmatrix}$ be the matrix that diagonalizes $\tilde{\mathbf{A}}$ such that $\mathbf{X}^T\tilde{\mathbf{A}}\mathbf{X}=\mathbf{\Lambda}$. Note that this also means that $\mathbf{W}$ from (8) is given by $\mathbf{W}=\mathbf{T}\mathbf{X}=\begin{bmatrix}\mathbf{T}X_+ & \mathbf{T}X_-\end{bmatrix}$. Let

$$\mathbf{w}=\begin{bmatrix}w_+\\ w_-\end{bmatrix}=\begin{bmatrix}X_+^T\tilde{\mathbf{q}}\\ X_-^T\tilde{\mathbf{q}}\end{bmatrix}=\mathbf{X}^T\tilde{\mathbf{q}}, \tag{25}$$

denote the so-called characteristic variables. In the subcritical case, the BCs from (12) (homogeneous) can be expressed as

$$\begin{aligned} w_+&=-r_L w_-, \quad x=x_l,\\ w_-&=-r_R w_+, \quad x=x_r, \end{aligned} \tag{26}$$

where $r_L=\left(\tilde{L}_l X_+\right)^{-1}\tilde{L}_l X_-$ and $r_R=\left(\tilde{L}_r X_-\right)^{-1}\tilde{L}_r X_+$. The condition of well-posedness can be expressed as

$$
\begin{aligned}
-\left[\tilde{\mathbf{q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{q}}\right]_{x_l}^{x_r} &= \tilde{\mathbf{q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{q}}\Big|^{x=x_l} - \tilde{\mathbf{q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{q}}\Big|^{x=x_r} \\
&= \tilde{\mathbf{q}}^T\left(X_+\Lambda_+X_+^T + X_-\Lambda_-X_-^T\right)\tilde{\mathbf{q}}\Big|^{x=x_l} - \tilde{\mathbf{q}}^T\left(X_+\Lambda_+X_+^T + X_-\Lambda_-X_-^T\right)\tilde{\mathbf{q}}\Big|^{x=x_r} \\
&= \left(w_+^2\Lambda_+ + w_-^2\Lambda_-\right)\Big|^{x=x_l} - \left(w_+^2\Lambda_+ + w_-^2\Lambda_-\right)\Big|^{x=x_r}.
\end{aligned}
\tag{27}
$$

Inserting the BCs (26) into (27) yields

$$
-\left[\tilde{\mathbf{q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{q}}\right]_{x_l}^{x_r} = w_-^2\left(r_L^2\Lambda_+ + \Lambda_-\right)\Big|^{x=x_l} - w_+^2\left(\Lambda_+ + r_R^2\Lambda_-\right)\Big|^{x=x_r} \le 0, \tag{28}
$$

provided that r_L and r_R are sufficiently small. Similarly, the BCs (12) (homogeneous) for supercritical inflow at the left boundary, i.e. $\Lambda_\pm > 0$, can be expressed as

$$
\mathbf{w} = 0, \quad x = x_l, \tag{29}
$$

and the condition of well-posedness can then be expressed as

$$
\begin{aligned}
-\left[\tilde{\mathbf{q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{q}}\right]_{x_l}^{x_r} &= \tilde{\mathbf{q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{q}}\Big|^{x=x_l} - \tilde{\mathbf{q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{q}}\Big|^{x=x_r} = \tilde{\mathbf{q}}^T\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^T\right)\tilde{\mathbf{q}}\Big|^{x=x_l} - \tilde{\mathbf{q}}^T\left(\mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^T\right)\tilde{\mathbf{q}}\Big|^{x=x_r} \\
&= -\mathbf{w}^T\boldsymbol{\Lambda}\mathbf{w}\Big|^{x=x_r} \le 0.
\end{aligned}
\tag{30}
$$

3.2. Semi-discrete stability analysis

The well-posed IBVP (9) will be solved using a high-order finite difference SBP-SAT method. For more information concerning the SBP-SAT method, we refer to the two review papers [46,47].

3.2.1. Definitions

The following definitions and notations will be used in the semi-discrete analysis. The computational domain $x \in [x_l, x_r]$ is discretized using m grid points

$$
x_i = x_l + (i-1)\Delta x, \quad i = 1, ..., m, \quad \Delta x = \frac{x_r - x_l}{m-1}.
$$

Definition 3.3. Let $L = x_r - x_l$ denote the width of the domain. A symmetric positive definite matrix H defines a discrete norm if $\begin{bmatrix} 1 & \dots & 1 \end{bmatrix} H \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^T = L$, independent of the number of grid points m.

The discrete solution vector q and discrete bottom topography $\underline{b}$ are defined as the $2m \times 1$ and $m \times 1$ vectors

$$
q = \begin{bmatrix} q_1^{(1)} & q_2^{(1)} & \dots & q_m^{(1)} & q_1^{(2)} & \dots & q_m^{(2)} \end{bmatrix}^T \quad \text{and} \quad \underline{b} = \begin{bmatrix} b_1 & b_2 & \dots & b_m \end{bmatrix}^T,
$$

where $b_i = b(x_i)$. To mimic the continuous inner product, the following discrete weighted inner product is introduced $(u, v)_H = u^T H v$. The corresponding norm is $||u||_H^2 = (u, u)_H$. Let I_k denote the unit matrix of size $k \times k$. In the analysis we will be utilizing the Kronecker product which is defined as

$$
C \otimes D = \begin{bmatrix} c_{1,1}D & \dots & c_{1,q}D \\ \vdots & & \vdots \\ c_{p,1}D & \dots & c_{p,q}D \end{bmatrix},
$$

where C is a $p \times q$ matrix and D is a $m \times n$ matrix. The following $m \times 1$ vectors will be used frequently

$$
e_1 = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T, \quad e_m = \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix}^T, \quad \mathbf{0} = \begin{bmatrix} 0 & \dots & 0 & 0 \end{bmatrix}^T,
$$

as well as the operators

$$
e_L = I_2 \otimes e_1, \quad e_R = I_2 \otimes e_m, \quad B = e_m e_m^T - e_1 e_1^T, \quad \bar{H} = I_2 \otimes H.
$$

The definition of first derivative operators [48] is restated for the sake of completeness:

Definition 3.4. A difference operator $D_1 = H^{-1}\left(Q + \frac{B}{2}\right)$ approximating $\partial/\partial x$, using pth-order accurate interior stencils, is said to be a pth-order accurate first derivative SBP operator if H is a discrete norm and $Q + Q^T = 0$.

Recently, upwind SBP operators with non-central interior stencils were introduced [19] and are given by the following definition:

Definition 3.5. The difference operators $D_+ = H^{-1}\left(Q_+ + \frac{B}{2}\right)$ and $D_- = H^{-1}\left(Q_- + \frac{B}{2}\right)$ approximating $\partial/\partial x$, using pth-order accurate interior stencils, are said to be pth-order accurate upwind SBP operators if H is a discrete norm, $Q_+ + Q_-^T = 0$, and $\frac{Q_+ + Q_+^T}{2} = S$, where S is negative semi-definite.

A few useful relations involving first derivative SBP operators are given by

$$\frac{D_+ + D_-}{2} = H^{-1}\left(\frac{Q_+ + Q_-}{2} + \frac{B}{2}\right) = H^{-1}\left(Q + \frac{B}{2}\right) = D_1, \tag{31}$$

$$\frac{D_+ - D_-}{2} = H^{-1}\left(\frac{Q_+ - Q_-}{2}\right) = H^{-1}\left(\frac{Q_+ + Q_+^T}{2}\right) = H^{-1}S. \tag{32}$$

3.2.2. Flux-splitting for artificial dissipation

For nonlinear problems it is most often necessary to add some sort of AD to damp spurious oscillations (often resulting from shocks) in the solution. When done properly [49,50], the added AD guarantees that the solution converges to the physically correct solution. In this study upwind SBP operators will be used which have built in AD when combined with flux-splitting [19]. Note that to guarantee convergence in the presence of strongly nonlinear phenomena such as shocks, AD in the form of a second derivative is necessary. The upwind operators only introduce AD in the form of higher order derivatives (commonly referred to as hyperviscosity) which damp high frequency modes. Hence, additional techniques [51–54] are required to make the scheme truly shock-capturing and are the subject of ongoing work. To simplify notation, we drop the source term $\mathbf{G}(\mathbf{q})$ in the coming analysis, leading to

$$\mathbf{q}_t + \mathbf{F}(\mathbf{q})_x = 0, \tag{33}$$

which in quasilinear form is given by

$$\mathbf{q}_t + \mathbf{A}\mathbf{q}_x = 0. \tag{34}$$

Performing Lax-Friedrichs flux-splitting on $\mathbf{A}$ yields

$$2\mathbf{A}_\pm = \mathbf{W}(\mathbf{\Lambda} \pm \alpha(t) I_2)\mathbf{W}^{-1} = \mathbf{A} \pm \mathbf{R}(t) = \mathbf{A} \pm \alpha(t) I_2, \tag{35}$$

where $\alpha(t)$ is the globally largest (in magnitude) eigenvalue of $\mathbf{A}$ for a fixed time t. The matrices $\mathbf{A}_+$ and $\mathbf{A}_-$ are positive and negative definite, respectively. Additionally, $\mathbf{A} = \mathbf{A}_+ + \mathbf{A}_-$ holds which means that (34) can be rewritten as

$$\mathbf{q}_t + \mathbf{A}_+\mathbf{q}_x + \mathbf{A}_-\mathbf{q}_x = 0. \tag{36}$$

The symmetrized frozen coefficient version of (36) is given by

$$\tilde{\mathbf{q}}_t + \tilde{\mathbf{A}}_+\tilde{\mathbf{q}}_x + \tilde{\mathbf{A}}_-\tilde{\mathbf{q}}_x = 0. \tag{37}$$

Below the scheme is derived in terms of the symmetrized frozen coefficient problem. The semi-discrete approximation of (37) using upwind SBP operators is given by

$$\begin{aligned}
&\tilde{q}_t + \left(\tilde{A}_+ \otimes D_-\right)\tilde{q} + \left(\tilde{A}_- \otimes D_+\right)\tilde{q} = \tilde{q}_t + \left(\frac{\tilde{A} + R}{2} \otimes D_-\right)\tilde{q} + \left(\frac{\tilde{A} - R}{2} \otimes D_+\right)\tilde{q} = 0 \\
&\qquad\qquad \Downarrow \\
&\tilde{q}_t + \tilde{A} \otimes \frac{D_+ + D_-}{2}\tilde{q} - R \otimes \frac{D_+ - D_-}{2}\tilde{q} = 0. \\
&\qquad\qquad \Downarrow \\
&\qquad \tilde{q}_t + \tilde{A} \otimes D_1\tilde{q} - R \otimes H^{-1}S\,\tilde{q} = 0.
\end{aligned} \tag{38}$$

Hence, in the symmetrized frozen coefficient case, the combination of upwind operators and flux-splitting is equivalent to employing a central-difference first derivative SBP operator to the non-split form, with the addition of AD from the term $-R \otimes H^{-1}S\,q$. The scheme we propose for the nonlinear problem (33) is based on this argument leading to

$$q_t + (I_2 \otimes D_1)F(q) - R(t) \otimes H^{-1}S\,q = 0, \tag{39}$$

where $R(t)$ is based on (35) and is time-dependent. This means that for a fixed time t, $R(t)$ is constructed by computing the globally largest eigenvalue, $\alpha(t)$, using the solution vector q. In order for this AD not to violate the well-balanced property a slight modification is made leading to the following discretization

$$q_t + (I_2 \otimes D_1)\, F(q) - R(t) \otimes H^{-1} S \left(q + \bar{b}\right) = 0, \tag{40}$$

where $\bar{b} = \begin{bmatrix} \underline{b} \\ \mathbf{0} \end{bmatrix}$. This change represents a modification of the usual Lax-Friedrichs flux-splitting and was introduced by Xing and Shu [40] to good result. The change is justified by viewing the first equation in (2) as an evolution equation for $h + b$ instead of h since $b_t = 0$.

3.2.3. Semi-discrete analysis for BCs

The SBP-SAT approximation of (9) using flux-splitting combined with upwind SBP operators is given by

$$q_t + (I_2 \otimes D_1)\, F(q) - R(t) \otimes H^{-1} S \left(q + \bar{b}\right) + G(q) = \text{SAT}, \tag{41}$$

where SAT imposes the continuous BCs weakly and

$$G(q) = \begin{bmatrix} \mathbf{0} \\ g\left(\underline{\underline{q}}^{(1)} + \underline{\underline{b}}\right) D_1 \underline{b} - D_1 \left(\frac{1}{2} g \underline{\underline{b}}\, \underline{b}\right) \end{bmatrix}, \tag{42}$$

with $\underline{\underline{q}}^{(1)} = diag\left(q^{(1)}\right)$ and $\underline{\underline{b}} = diag(\underline{b})$. The source term G has been discretized exactly as in [40] where it was shown to lead to a well-balanced scheme. For subcritical flow one BC at each boundary has to be specified, leading to

$$\text{SAT} = -\bar{H}^{-1} e_L W_+ \Lambda_+ \left(L_l W_+\right)^{-1} \left(L_l e_L^T q - g_l(t)\right) + \bar{H}^{-1} e_R W_- \Lambda_- \left(L_r W_-\right)^{-1} \left(L_r e_R^T q - g_r(t)\right), \tag{43}$$

where $L_{l,r}$ are 1×2 vectors and $g_{l,r}(t)$ are scalars. For supercritical flow two BCs needs to be specified at one boundary. As an example, supercritical inflow at the left boundary leads to

$$\text{SAT} = -\bar{H}^{-1} e_L W \Lambda \left(L_l W\right)^{-1} \left(L_l e_L^T q - g_l(t)\right), \tag{44}$$

where L_l is a 2×2 matrix and g_l are 2×1 vector. The symmetrized frozen coefficient version of (41) is given by

$$\tilde{q}_t + \tilde{A} \otimes D_1 \tilde{q} - R \otimes H^{-1} S\, \tilde{q} + \tilde{C} \tilde{q} = \widetilde{\text{SAT}} + \tilde{E}, \tag{45}$$

where $\widetilde{\text{SAT}} = \left(T^{-1} \otimes I_m\right) \text{SAT}$,

$$\tilde{C} = \left(T^{-1} \otimes I_m\right) \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ g D_1 \underline{b} & \mathbf{0} \end{bmatrix}, \quad \text{and} \quad \tilde{E} = \left(T^{-1} \otimes I_m\right) \begin{bmatrix} \alpha H^{-1} S \underline{b} \\ D_1 \left(\frac{1}{2} g \underline{\underline{b}}\, \underline{b}\right) - g \underline{\underline{b}} D_1 \underline{b} \end{bmatrix}. \tag{46}$$

The scheme (45) approximates (12) and stability is given by the following lemma:

Lemma 3.2. *Let $b(x)$ be a smooth function. If the continuous problem (12) is well-posed the scheme (45) is stable.*

Proof. Since lower order terms don't affect stability, they are neglected. In the proof we assume subcritical flow which means that $\Lambda_+ > 0$ and $\Lambda_- < 0$. The proof is similar for supercritical flow. To make the proof more compact, we only consider the left BC. The scheme (45), without lower order terms and only considering the left BC, is given by

$$\tilde{q}_t + \tilde{A} \otimes D_1 \tilde{q} = R \otimes H^{-1} S\, \tilde{q} - \bar{H}^{-1} e_L X_+ \Lambda_+ \left(\tilde{L}_l X_+\right)^{-1} \left(\tilde{L}_l e_L^T \tilde{q} - g_l(t)\right). \tag{47}$$

By setting $g_l = 0$, multiplying (47) by $\tilde{q}^T \bar{H}$ from the left and adding the transpose the following energy estimate is obtained

$$\begin{aligned} \frac{d}{dt} ||\tilde{q}||_{\bar{H}}^2 &= \tilde{q}_l^T \left(\tilde{A} - 2 X_+ \Lambda_+ \left(\tilde{L}_l X_+\right)^{-1} \tilde{L}_l\right) \tilde{q}_l + 2\, \tilde{q}^T R \otimes S\, \tilde{q} \\ &\leq \tilde{q}_l^T \left(\tilde{A} - 2 X_+ \Lambda_+ \left(\tilde{L}_l X_+\right)^{-1} \tilde{L}_l\right) \tilde{q}_l, \end{aligned} \tag{48}$$

where $\tilde{q}_l = e_L^T \tilde{q}$. The term $2\, \tilde{q}^T R \otimes S\, \tilde{q}$ is negative semi-definite introducing AD. Note that the boundary term from the right boundary in (48) has been neglected. By setting $\tilde{q}_l = X X^T \tilde{q}_l = X \begin{bmatrix} w_{l+} \\ w_{l-} \end{bmatrix} = X w_l$ in (48), $\tilde{A}$ is diagonalized leading to

$$\begin{aligned}\frac{d}{dt}||\tilde{q}||_{\tilde{H}}^2 &\leq w_l^T \left(\Lambda - 2\begin{bmatrix}1 & 0\end{bmatrix}^T \Lambda_+ \left(\tilde{L}_l X_+\right)^{-1} \tilde{L}_l X\right) w_l \\ &= w_l^T \begin{bmatrix} -\Lambda_+ & -2\Lambda_+ \left(\tilde{L}_l X_+\right)^{-1} \tilde{L}_l X_- \\ 0 & \Lambda_- \end{bmatrix} w_l \\ &= -\Lambda_+ w_{l+}^2 + \Lambda_- w_{l-}^2 - 2\Lambda_+ \left(\tilde{L}_l X_+\right)^{-1} \tilde{L}_l X_- w_{l+} w_{l-} \\ &= -\Lambda_+ w_{l+}^2 + \Lambda_- w_{l-}^2 - 2\Lambda_+ r_L w_{l+} w_{l-}.\end{aligned} \tag{49}$$

Inserting the homogeneous BCs (26) into (49) yields

$$\begin{aligned}\frac{d}{dt}||\tilde{q}||_{\tilde{H}}^2 &\leq -\Lambda_+ \left(-r_L w_{l-}\right)^2 + \Lambda_- w_{l-}^2 + 2\Lambda_+ r_L^2 w_{l-} w_{l-} \\ &= -\Lambda_+ r_L^2 w_{l-}^2 + \Lambda_- w_{l-}^2 + 2\Lambda_+ r_L^2 w_{l-}^2 \\ &= w_{l-}^2 \left(r_L^2 \Lambda_+ + \Lambda_-\right) \leq 0.\end{aligned} \tag{50}$$

Here, (50) mimics the continuous estimate (28) at the left boundary and is ≤ 0 provided that the BCs are well-posed. Hence, (45) is stable for subcritical flow when considering the left BC. The analysis is similar for the right boundary as well as for supercritical flow. □

In [55] it was proven that the in the case of variable coefficients, i.e. $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}(x,t)$, discrete stability of the frozen coefficient case is equivalent to stability of the variable coefficient case assuming smoothness of $\tilde{\mathbf{A}}(x,t)$. See Strang [43] as well as Kreiss and Lorenz [44] for additional details regarding nonlinear stability. Thus, assuming smoothness of the solution, the linear energy estimates are still useful in guiding the discrete boundary treatment for the nonlinear scheme (41). It is worth mentioning that the frozen coefficient approach has been successfully used for the compressible Euler and Navier-Stokes equations [56–59].

Remark. The $\widetilde{\text{SAT}}$ we propose for the symmetrized frozen coefficient version is in fact exactly the same as proposed by Eriksson [45] for general constant coefficient hyperbolic systems. In [45] it was proven (in the constant coefficient case) that $\widetilde{\text{SAT}}$ leads to a dual consistent scheme. For more information on dual-consistent discretizations in the SBP-SAT finite difference method setting, see e.g. [45,15,60,61].

4. 2D analysis

In this section we will analyze well-posedness for the continuous 2D problem. A stable SBP-SAT approximation for different BCs will also be derived.

4.1. The 2D problem

The shallow water equations in 2D are given by (1) and are reformulated as

$$\mathbf{q}_t + \mathbf{F}_1(\mathbf{q})_x + \mathbf{F}_2(\mathbf{q})_y + \mathbf{G}(\mathbf{q}) = 0, \tag{51}$$

where $\mathbf{q} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix}^T = \begin{bmatrix} h & hu & hv \end{bmatrix}^T$, $\mathbf{F}_1(\mathbf{q}) = \begin{bmatrix} q_2 & \frac{q_2^2}{q_1} + \frac{1}{2} g q_1^2 & \frac{q_2 q_3}{q_1} \end{bmatrix}^T$, $\mathbf{F}_2(\mathbf{q}) = \begin{bmatrix} q_3 & \frac{q_2 q_3}{q_1} & \frac{q_3^2}{q_1} + \frac{1}{2} g q_1^2 \end{bmatrix}^T$ and $\mathbf{G}(\mathbf{q}) = \begin{bmatrix} 0 & g(q_1 + b) b_x - \left(\frac{1}{2} g b^2\right)_x & g(q_1 + b) b_y - \left(\frac{1}{2} g b^2\right)_y \end{bmatrix}^T$. Here the source term $\mathbf{G}(\mathbf{q})$ has been rewritten in a similar way as in Section 2.1 to make the discretization well-balanced. This can further be rewritten into quasilinear form as

$$\mathbf{q}_t + \mathbf{A}_1(\mathbf{q})\mathbf{q}_x + \mathbf{A}_2(\mathbf{q})\mathbf{q}_y + \mathbf{C}\mathbf{q} = \mathbf{E}, \tag{52}$$

where

$$\begin{aligned}&\mathbf{A}_1 = \mathbf{A}_1(\mathbf{q}) = \begin{bmatrix} 0 & 1 & 0 \\ g q_1 - \frac{q_2^2}{q_1^2} & \frac{2 q_2}{q_1} & 0 \\ -\frac{q_2 q_3}{q_1^2} & \frac{q_3}{q_1} & \frac{q_2}{q_1} \end{bmatrix}, \quad \mathbf{A}_2 = \mathbf{A}_2(\mathbf{q}) = \begin{bmatrix} 0 & 0 & 1 \\ -\frac{q_2 q_3}{{q_1}^2} & \frac{q_3}{q_1} & \frac{q_2}{q_1} \\ g\, q_1 - \frac{{q_3}^2}{{q_1}^2} & 0 & \frac{2 q_3}{q_1} \end{bmatrix}, \\ &\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 \\ g b_x & 0 & 0 \\ g b_y & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{E} = \begin{bmatrix} 0 \\ \left(\frac{1}{2} g b^2\right)_x - g b b_x \\ \left(\frac{1}{2} g b^2\right)_y - g b b_y \end{bmatrix}.\end{aligned} \tag{53}$$

The eigenvalues of $\mathbf{A}_1$ are given by

$$\Lambda_{x\pm} = \frac{q_2}{q_1} \pm \sqrt{gq_1} = u \pm \sqrt{gh}, \quad \Lambda_{x0} = \frac{q_2}{q_1} = u, \tag{54}$$

and the eigenvalues of $\mathbf{A}_2$ are given by

$$\Lambda_{y\pm} = \frac{q_3}{q_1} \pm \sqrt{gq_1} = v \pm \sqrt{gh}, \quad \Lambda_{y0} = \frac{q_3}{q_1} = v. \tag{55}$$

Analogous to the 1D case, these eigenvalues represent the different wave speeds in the x- and y-direction, respectively. Additionally, $\mathbf{A}_1$ and $\mathbf{A}_2$ can be written in terms of its eigenvalues and eigenvectors

$$\mathbf{A}_1 = \mathbf{W}_x \mathbf{\Lambda}_x \mathbf{W}_x^{-1}, \qquad \mathbf{A}_2 = \mathbf{W}_y \mathbf{\Lambda}_y \mathbf{W}_y^{-1}, \tag{56}$$

where

$$\begin{aligned} &\mathbf{\Lambda}_{x,y} = \begin{bmatrix} \Lambda_{x,y+} & 0 & 0 \\ 0 & \Lambda_{x,y-} & 0 \\ 0 & 0 & \Lambda_{x,y0} \end{bmatrix}, \quad \mathbf{W}_x = \begin{bmatrix} W_{x+} & W_{x-} & W_{x_0} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ \Lambda_{x+} & \Lambda_{x-} & 0 \\ \frac{q_3}{q_1} & \frac{q_3}{q_1} & 1 \end{bmatrix} \quad \text{and} \\ &\mathbf{W}_y = \begin{bmatrix} W_{y+} & W_{y-} & W_{y_0} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ \frac{q_2}{q_1} & \frac{q_2}{q_1} & 1 \\ \Lambda_{y+} & \Lambda_{y-} & 0 \end{bmatrix}. \end{aligned} \tag{57}$$

4.2. Boundary analysis

The energy method will be used to analyze well-posedness of the boundary and interface conditions used in the computations in Sections 6 and 7. First wall- and characteristic BCs on a single block will be analyzed in Section 4.2.1. Then an interface condition across two blocks will be analyzed in Section 4.2.2. In the analysis the following inner product is used:

Definition 4.1. Denote the 2D bounding box $x_l \le x \le x_r$, $y_l \le y \le y_r$ by $(x, y) \in \Omega_{x_l,y_l}^{x_r,y_r}$. Given two real valued vector functions with k components $\mathbf{f} = \begin{bmatrix} f^{(1)}(x,y,t) & f^{(2)}(x,y,t) & \dots & f^{(k)}(x,y,t) \end{bmatrix}^T$ and $\mathbf{g} = \begin{bmatrix} g^{(1)}(x,y,t) & g^{(2)}(x,y,t) & \dots & g^{(k)}(x,y,t) \end{bmatrix}^T$ and let $\mathbf{f}, \mathbf{g} \in L^2\left[\Omega_{x_l,y_l}^{x_r,y_r}\right]$, then the inner product is defined as

$$(\mathbf{f}, \mathbf{g}) = \int_{y_l}^{y_r} \int_{x_l}^{x_r} \mathbf{f}^T \mathbf{g} \, dx \, dy,$$

with the following implied norm $||\mathbf{f}||^2 = (\mathbf{f}, \mathbf{f})$.

If the analysis is done analogous to Section 3, it can be shown that the 2D shallow water equations are linearly well-posed if

$$BT_W + BT_E + BT_S + BT_N \le 0, \tag{58}$$

for homogeneous BCs (zero boundary data) where

$$BT_W = \int_{y_l}^{y_r} \mathbf{q}^T \mathbf{A}_1 \mathbf{q} \Big|^{x=x_l} dy, \qquad BT_E = -\int_{y_l}^{y_r} \mathbf{q}^T \mathbf{A}_1 \mathbf{q} \Big|^{x=x_r} dy, \tag{59}$$

$$BT_S = \int_{x_l}^{x_r} \mathbf{q}^T \mathbf{A}_2 \mathbf{q} \Big|^{y=y_l} dx, \qquad BT_N = -\int_{x_l}^{x_r} \mathbf{q}^T \mathbf{A}_2 \mathbf{q} \Big|^{y=y_r} dx. \tag{60}$$

This holds assuming that $\mathbf{A}_1$ and $\mathbf{A}_2$ can be symmetrized simultaneously, i.e.

$$\exists \, \mathbf{T} \quad \text{s.t.} \quad \mathbf{T}^{-1} \mathbf{A}_1 \mathbf{T} + \mathbf{T}^{-1} \mathbf{A}_2 \mathbf{T} = \tilde{\mathbf{A}}_1 + \tilde{\mathbf{A}}_2,$$

where $\tilde{\mathbf{A}}_1^T = \tilde{\mathbf{A}}_1$ and $\tilde{\mathbf{A}}_2^T = \tilde{\mathbf{A}}_2$. As in Section 3, an equivalent condition of linear well-posedness is

$$\widetilde{BT}_W + \widetilde{BT}_E + \widetilde{BT}_S + \widetilde{BT}_N \le 0, \tag{61}$$

where

$$\widetilde{BT}_W = \int_{y_l}^{y_r} \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_1 \tilde{\mathbf{q}} \Big|^{x=x_l} dy, \qquad \widetilde{BT}_E = -\int_{y_l}^{y_r} \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_1 \tilde{\mathbf{q}} \Big|^{x=x_r} dy, \tag{62}$$

$$\widetilde{BT}_S = \int_{x_l}^{x_r} \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_2 \tilde{\mathbf{q}} \Big|^{y=y_l} dx, \qquad \widetilde{BT}_N = -\int_{x_l}^{x_r} \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_2 \tilde{\mathbf{q}} \Big|^{y=y_r} dx. \tag{63}$$

For a more rigorous analysis of well-posed BCs for the 2D shallow water equations (although with an added Coriolis term) as well as general non-reflecting BCs, see [62].

4.2.1. Single block

The BCs that will be analyzed are the following: wall BCs at the north and east boundaries and characteristic non-reflecting BCs at the west and south boundaries. The BCs are given by

$$\begin{cases} q_3 = hv = 0, & y = y_r, \\ q_2 = hu = 0, & x = x_r, \\ \mathbf{A}_{1+}\mathbf{q} = \mathbf{A}_{1+} g_C, & x = x_l, \\ \mathbf{A}_{2+}\mathbf{q} = \mathbf{A}_{2+} g_C, & y = y_l, \end{cases} \tag{64}$$

where g_C is boundary data. For these BCs to be well-posed (58) (or equivalently (61)) needs to hold. We first consider the west boundary where a characteristic non-reflecting BC is imposed. An alternative formulation of this BC is obtained by multiplying by $\mathbf{T}^{-1}$ from the left leading to

$$\tilde{\mathbf{A}}_{1+}\tilde{\mathbf{q}} = \tilde{\mathbf{A}}_{1+}\tilde{g}_C, \quad x = x_l, \tag{65}$$

where $\tilde{g}_C = \mathbf{T}^{-1} g_C$. The integrand in $\widetilde{BT}_W$ is then given by

$$\begin{aligned} \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_1 \tilde{\mathbf{q}} \Big|^{x=x_l} &= \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_{1+} \tilde{\mathbf{q}} \Big|^{x=x_l} + \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_{1-} \tilde{\mathbf{q}} \Big|^{x=x_l} \\ &= \tilde{g}_C^T \tilde{\mathbf{A}}_{1+} \tilde{g}_C \Big|^{x=x_l} + \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_{1-} \tilde{\mathbf{q}} \Big|^{x=x_l} \le 0 \quad \text{if } \tilde{g}_C = 0 \Rightarrow \widetilde{BT}_W \le 0. \end{aligned}$$

We repeat the process for the integrand in $\widetilde{BT}_S$ where we also have a characteristic non-reflecting BC

$$\begin{aligned} \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_2 \tilde{\mathbf{q}} \Big|^{y=y_l} &= \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_{2+} \tilde{\mathbf{q}} \Big|^{y=y_l} + \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_{2-} \tilde{\mathbf{q}} \Big|^{y=y_l} \\ &= \tilde{g}_C^T \tilde{\mathbf{A}}_{2+} \tilde{g}_C \Big|^{y=y_l} + \tilde{\mathbf{q}}^T \tilde{\mathbf{A}}_{2-} \tilde{\mathbf{q}} \Big|^{y=y_l} \le 0 \quad \text{if } \tilde{g}_C = 0 \Rightarrow \widetilde{BT}_S \le 0. \end{aligned}$$

Then we consider the integrand in BT_N where we have a wall BC

$$-\mathbf{q}^T \mathbf{A}_2 \mathbf{q} \Big|^{y=y_r} = -q_3 \left. \frac{g q_1^3 + q_1^2 + q_2^2 + q_3^2}{q_1} \right|^{y=y_r} = 0 \Rightarrow BT_N = 0.$$

Lastly, we consider the integrand in BT_E where we also have a wall BC

$$-\mathbf{q}^T \mathbf{A}_1 \mathbf{q} \Big|^{x=x_r} = -q_2 \left. \frac{g q_1^3 + q_1^2 + q_2^2 + q_3^2}{q_1} \right|^{x=x_r} = 0 \Rightarrow BT_E = 0.$$

We have now shown that the boundary terms from each boundary is negative semi-definite. As a consequence (58) (or equivalently (61)) is fulfilled which means that the BCs given by (64) leads to a linearly well-posed IBVP.

Remark. For characteristic BCs the matrices $\mathbf{A}_{1\pm}$ and $\mathbf{A}_{2\pm}$ are constructed using Steger-Warming flux-splitting

$$\mathbf{A}_{1,2\pm} = \frac{1}{2} \mathbf{W}_{x,y} \left(\mathbf{\Lambda}_{x,y} \pm |\mathbf{\Lambda}_{x,y}| \right) \mathbf{W}_{x,y}^{-1}. \tag{66}$$

Characteristic BCs are first order accurate non-reflecting BCs that are well-proven for the compressible Euler and Navier–Stokes equations [58,63]. An alternative approach to impose non-reflecting BCs is the Perfectly Matched Layer (PML). For a comparison between characteristic BCs and the PML approach, to impose non-reflecting BCs for the shallow water equations, see [64,65]. Another approach is the Highdon condition which was implemented for the shallow water equations in [66].

Remark. To impose non-reflecting BCs with respect to the surface level $h_0 + b(x, y)$, we set $g_C = \begin{bmatrix} h_0 - b(x,y) & 0 & 0 \end{bmatrix}^T$.

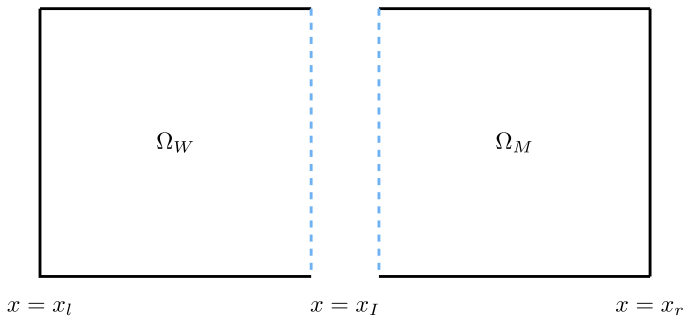


Fig. 2. Multi-block setup with an interface condition between the blocks Ω_W and Ω_M at $x = x_I$.

4.2.2. Multi-block

In this section the well-posedness of an interface condition across two blocks Ω_W and Ω_M, see Fig. 2, will be analyzed. The model for the two blocks is given by

$$\begin{aligned}(\mathbf{q}_W)_t + \mathbf{A}_1(\mathbf{q}_W)_x + \mathbf{A}_2(\mathbf{q}_W)_y + \mathbf{C}\mathbf{q}_W = \mathbf{E}, \quad (x,y) \in \Omega_W, \\ (\mathbf{q}_M)_t + \mathbf{A}_1(\mathbf{q}_M)_x + \mathbf{A}_2(\mathbf{q}_M)_y + \mathbf{C}\mathbf{q}_M = \mathbf{E}, \quad (x,y) \in \Omega_M,\end{aligned} \qquad t > 0, \tag{67}$$

together with initial conditions and well-posed BCs. The symmetrized frozen coefficient version of (67) is given by

$$\begin{aligned}(\tilde{\mathbf{q}}_W)_t + \tilde{\mathbf{A}}_1(\tilde{\mathbf{q}}_W)_x + \tilde{\mathbf{A}}_2(\tilde{\mathbf{q}}_W)_y + \tilde{\mathbf{C}}\tilde{\mathbf{q}}_W = \tilde{\mathbf{E}}, \quad (x,y) \in \Omega_W, \\ (\tilde{\mathbf{q}}_M)_t + \tilde{\mathbf{A}}_1(\tilde{\mathbf{q}}_M)_x + \tilde{\mathbf{A}}_2(\tilde{\mathbf{q}}_M)_y + \tilde{\mathbf{C}}\tilde{\mathbf{q}}_M = \tilde{\mathbf{E}}, \quad (x,y) \in \Omega_M,\end{aligned} \qquad t > 0. \tag{68}$$

Assuming that the solution is smooth, there is continuity at the interface between the two blocks, i.e. $\mathbf{q}_W = \mathbf{q}_M$ at $x = x_I$. In the frozen coefficient case, this continuity is guaranteed with the following interface condition

$$\begin{cases}\bar{\mathbf{A}}_{1-}\mathbf{q}_W = \bar{\mathbf{A}}_{1-}\bar{\mathbf{q}}_M, \\ \bar{\mathbf{A}}_{1+}\bar{\mathbf{q}}_M = \bar{\mathbf{A}}_{1+}\bar{\mathbf{q}}_W,\end{cases} \qquad x = x_I, \tag{69}$$

where $\bar{\mathbf{A}}_{1,2}$ is a constant coefficient matrix. After multiplication by $\mathbf{T}^{-1}$ from the left, (69) is equivalent to

$$\begin{cases}\tilde{\mathbf{A}}_{1-}\tilde{\mathbf{q}}_W = \tilde{\mathbf{A}}_{1-}\tilde{\mathbf{q}}_M, \\ \tilde{\mathbf{A}}_{1+}\tilde{\mathbf{q}}_M = \tilde{\mathbf{A}}_{1+}\tilde{\mathbf{q}}_W,\end{cases} \qquad x = x_I. \tag{70}$$

The integrand from the boundary terms at the interface $\widetilde{BT}_I$ (when considering the energy method) from both domains are given by

$$\begin{aligned}&-\tilde{\mathbf{q}}_W^T\tilde{\mathbf{A}}_1\tilde{\mathbf{q}}_W\Big|^{x=x_I} + \tilde{\mathbf{q}}_M^T\tilde{\mathbf{A}}_1\tilde{\mathbf{q}}_M\Big|^{x=x_I} \\ =\ &-\tilde{\mathbf{q}}_W^T\tilde{\mathbf{A}}_{1+}\tilde{\mathbf{q}}_W\Big|^{x=x_I} - \tilde{\mathbf{q}}_W^T\tilde{\mathbf{A}}_{1-}\tilde{\mathbf{q}}_W\Big|^{x=x_I} + \tilde{\mathbf{q}}_M^T\tilde{\mathbf{A}}_{1+}\tilde{\mathbf{q}}_M\Big|^{x=x_I} + \tilde{\mathbf{q}}_M^T\tilde{\mathbf{A}}_{1-}\tilde{\mathbf{q}}_M\Big|^{x=x_I}.\end{aligned} \tag{71}$$

Inserting the interface conditions (70) into (71) yields

$$-\tilde{\mathbf{q}}_W^T\tilde{\mathbf{A}}_{1+}\tilde{\mathbf{q}}_W\Big|^{x=x_I} - \tilde{\mathbf{q}}_M^T\tilde{\mathbf{A}}_{1-}\tilde{\mathbf{q}}_M\Big|^{x=x_I} + \tilde{\mathbf{q}}_W^T\tilde{\mathbf{A}}_{1+}\tilde{\mathbf{q}}_W\Big|^{x=x_I} + \tilde{\mathbf{q}}_M^T\tilde{\mathbf{A}}_{1-}\tilde{\mathbf{q}}_M\Big|^{x=x_I} = 0 \Rightarrow \widetilde{BT}_I = 0.$$

4.3. Semi-discrete stability analysis

4.3.1. Definitions

In this section the 2D SBP operators used in the subsequent sections are defined. Each sub-domain to the 2D continuous domain is discretized with $m \times m$ grid points as follows:

$$(x_i, y_j) = (x_l + (i-1)\Delta x)\,,\ y_l + (j-1)\Delta x), \quad i = 1, ..., m, \quad j = 1, ..., m, \quad \Delta x = \frac{x_r - x_l}{m-1} = \frac{y_r - y_l}{m-1}.$$

In this study we only consider square domains for the sake of simplicity which means that $x_r - x_l = y_r - y_l$. Additionally, the grid spacing has been set to Δx in both the x- and y-direction. The discrete solution vector is defined as a vector of vectors, at each time step

$$q = \begin{bmatrix} q_{11}^{(1)} & q_{12}^{(1)} & \dots & q_{1m}^{(1)} & q_{21}^{(1)} & \dots & q_{mm}^{(1)} & q_{11}^{(2)} & \dots & q_{mm}^{(3)} \end{bmatrix}^T$$

and the discrete bottom topography $\underline{b}$ is similarly defined as

$$\underline{b} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} & b_{21} & \dots & b_{mm} \end{bmatrix}^T,$$

where $b_{ij} = b(x_i, y_j)$. The column vector $\mathbf{0}$ is redefined as a vector with length m^2 filled with zeros. Lastly, the operators previously defined are extended to two dimensions as follows:

$$\begin{aligned}
e_E &= I_3 \otimes e_m \otimes I_m, & H_x &= H \otimes I_m, & D_x &= D_1 \otimes I_m, \\
e_W &= I_3 \otimes e_1 \otimes I_m, & H_y &= I_m \otimes H, & D_y &= I_m \otimes D_1, \\
e_S &= I_3 \otimes I_m \otimes e_1, & \bar{H}_x &= I_3 \otimes H_x, & S_x &= S \otimes I_m, \\
e_N &= I_3 \otimes I_m \otimes e_m, & \bar{H}_y &= I_3 \otimes H_y, & S_y &= I_m \otimes S, \qquad \bar{H} = \bar{H}_x \bar{H}_y.
\end{aligned}$$

4.3.2. Semi-discrete analysis for BCs

The SBP-SAT approximation of (51) using flux-splitting combined with upwind operators is given by

$$q_t + (I_3 \otimes D_x)F_1 + (I_3 \otimes D_y)F_2 - \left(R_x(t) \otimes H_x^{-1} S_x + R_y(t) \otimes H_y^{-1} S_y\right)\left(q + \bar{b}\right) + G = \text{SAT}, \tag{72}$$

where

$$\bar{b} = \begin{bmatrix} \underline{b} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad R_{x,y}(t) = \alpha_{x,y}(t) I_3 \quad \text{and} \quad G = \begin{bmatrix} \mathbf{0} \\ g\left(\underline{\underline{q}}^{(1)} + \underline{\underline{b}}\right) D_x \underline{b} - D_x\left(\tfrac{1}{2}\underline{\underline{b}}\,\underline{b}\right) \\ g\left(\underline{\underline{q}}^{(1)} + \underline{\underline{b}}\right) D_y \underline{b} - D_y\left(\tfrac{1}{2}\underline{\underline{b}}\,\underline{b}\right) \end{bmatrix}, \tag{73}$$

with $\underline{\underline{q}}^{(1)} = diag(q^{(1)})$ and $\underline{\underline{b}} = diag(\underline{b})$. Again, the source term G has been discretized exactly as in [40] where it was shown to lead to a well-balanced scheme. For a fixed time t, $\alpha_x(t)$ is the globally largest eigenvalue (in magnitude) of $A_1(q)$ and $\alpha_y(t)$ is the globally largest eigenvalue (in magnitude) of $A_2(q)$. The SAT imposes the continuous BCs weakly. As an example, for the well-posed BCs analyzed in Section 4.2.1 the SAT is given by

$$\begin{aligned}
\text{SAT} = \bar{H}_y^{-1} e_N W_{y-} \Lambda_{y-} \left(L_N W_{y-}\right)^{-1} L_N e_N^T q \\
- A_{2+} \otimes H_y^{-1} e_S \left(e_S^T q - e_S^T g_C\right) \\
- A_{1+} \otimes H_x^{-1} e_W \left(e_W^T q - e_W^T g_C\right) \\
+ \bar{H}_x^{-1} e_E W_{x-} \Lambda_{x-} \left(L_E W_{x-}\right)^{-1} L_E e_E^T q,
\end{aligned} \tag{74}$$

where $L_{N,E}$ are boundary operators given by $L_N = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ and $L_E = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$. The first and last terms in (74) impose the wall BCs at the north and east boundary, respectively. The remaining terms impose the characteristic non-reflecting BCs. The symmetrized frozen coefficient version of (72) is given by

$$\tilde{q}_t + \tilde{A}_1 \otimes D_x \tilde{q} + \tilde{A}_2 \otimes D_y \tilde{q} + \tilde{C}\tilde{q} - \left(R_x \otimes H_x^{-1} S_x + R_y \otimes H_y^{-1} S_y\right)\tilde{q} = \tilde{E} + \widetilde{\text{SAT}}, \tag{75}$$

where $\widetilde{\text{SAT}} = \left(T^{-1} \otimes I_{m^2}\right) \text{SAT}$,

$$\tilde{C} = \left(T^{-1} \otimes I_{m^2}\right) \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ g D_x \underline{b} & \mathbf{0} & \mathbf{0} \\ g D_y \underline{b} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \text{and} \quad \tilde{E} = \left(T^{-1} \otimes I_{m^2}\right) \begin{bmatrix} \left(\alpha_x H_x^{-1} S_x + \alpha_y H_y^{-1} S_y\right) \underline{b} \\ D_x\left(\tfrac{1}{2} g \underline{\underline{b}}\,\underline{b}\right) - g \underline{\underline{b}} D_x \underline{b} \\ D_y\left(\tfrac{1}{2} g \underline{\underline{b}}\,\underline{b}\right) - g \underline{\underline{b}} D_y \underline{b} \end{bmatrix}. \tag{76}$$

Stability is given by the following lemma:

Lemma 4.1. *Let $b(x, y)$ be a smooth function of x and y. If the continuous problem is well-posed the scheme (75) is stable.*

Proof. See Appendix B. □

4.3.3. Semi-discrete analysis for interface

In this section we analyze the multi-block setup presented in Section 4.2.2. The upwind SBP-SAT approximation of the setup is given by

$$\begin{aligned}
&(q_W)_t + (I_3 \otimes D_x)F_1 + (I_3 \otimes D_y)F_2 \\
&- \left(R_x(t) \otimes H_x^{-1} S_x + R_y(t) \otimes H_y^{-1} S_y\right)\left(q_W + \bar{b}\right) + G = \mathrm{SAT}_W + \mathrm{SAT}_{IW}, \\
&(q_M)_t + (I_3 \otimes D_x)F_1 + (I_3 \otimes D_y)F_2 \\
&- \left(R_x(t) \otimes H_x^{-1} S_x + R_y(t) \otimes H_y^{-1} S_y\right)\left(q_M + \bar{b}\right) + G = \mathrm{SAT}_M + \mathrm{SAT}_{IM},
\end{aligned} \tag{77}$$

where SAT_W and SAT_M impose well-posed BCs for Ω_W and Ω_M, respectively. See Section 4.3.2 for more details. SAT_{IW} and SAT_{IM} impose the interface condition (69) and are given by

$$\begin{aligned}
\mathrm{SAT}_{IW} &= A_{1-} \otimes H_x^{-1} e_E \left(e_E^T q_W - e_W^T q_M\right), \\
\mathrm{SAT}_{IM} &= -A_{1+} \otimes H_x^{-1} e_W \left(e_W^T q_M - e_E^T q_W\right).
\end{aligned} \tag{78}$$

Lemma 4.2. *Let $b(x, y)$ be a smooth function. If the continuous problem presented in Section 4.2.2 is well-posed, the symmetrized frozen coefficient version of the scheme (77) is stable.*

Proof. The semi-discrete approximation of the symmetric frozen coefficient version of (77) without lower order terms is given by

$$\begin{aligned}
\tilde{q}_{Wt} + \tilde{A}_1 \otimes D_x \tilde{q}_W + \tilde{A}_2 \otimes D_y \tilde{q}_W &= \widetilde{\mathrm{SAT}}_W + \tilde{A}_{1-} \otimes H_x^{-1} e_E \left(e_E^T \tilde{q}_W - e_W^T \tilde{q}_M\right) \\
&\quad + \left(R_x \otimes H_x^{-1} S_x + R_y \otimes H_y^{-1} S_y\right) \tilde{q}_W, \\
\tilde{q}_{Mt} + \tilde{A}_1 \otimes D_x \tilde{q}_M + \tilde{A}_2 \otimes D_y \tilde{q}_M &= \widetilde{\mathrm{SAT}}_M - \tilde{A}_{1+} \otimes H_x^{-1} e_W \left(e_W^T \tilde{q}_M - e_E^T \tilde{q}_W\right) \\
&\quad + \left(R_x \otimes H_x^{-1} S_x + R_y \otimes H_y^{-1} S_y\right) \tilde{q}_M,
\end{aligned} \tag{79}$$

where $\widetilde{\mathrm{SAT}}_{W,M} = \left(T^{-1} \otimes I_{m^2}\right) \mathrm{SAT}_{W,M}$. The lower order terms have been neglected since they don't concern stability. Multiplying the first and second equations in (79) by $\tilde{q}_W \bar{H}$ and $\tilde{q}_M \bar{H}$, respectively, and adding the transpose leads to

$$\begin{aligned}
\frac{d}{dt}\left(||\tilde{q}_W||_{\bar{H}}^2 + ||\tilde{q}_M||_{\bar{H}}^2\right) &= BT + AD \\
&\quad - \left(e_W^T \tilde{q}_M - e_E^T \tilde{q}_W\right)^T \left(\left(\tilde{A}_{1+} - \tilde{A}_{1-}\right) \otimes H\right)\left(e_W^T \tilde{q}_M - e_E^T \tilde{q}_W\right).
\end{aligned} \tag{80}$$

Here the interface contribution $-\left(e_W^T \tilde{q}_M - e_E^T \tilde{q}_W\right)^T \left(\left(\tilde{A}_{1+} - \tilde{A}_{1-}\right) \otimes H\right)\left(e_W^T \tilde{q}_M - e_E^T \tilde{q}_W\right)$ is a negative term which doesn't lead to growth in the semi-discrete energy. In Appendix B it was proven that $BT + AD \leq 0$ for well-posed BCs. □

5. Computations 1D

Several benchmark solutions for the shallow water equations have been proposed to verify the accuracy of numerical methods. The community SWASHES [67] compiles a set of analytic solutions for the different forms of the shallow water equations. We consider the following three cases: the "lake at rest" problem, a shock problem and a problem focusing on a Gaussian pulse with wall BCs. The convergence rate c is calculated as

$$c = \log_{10}\left(\frac{||\mathbf{q} - q^{(m_1)}||_H}{||\mathbf{q} - q^{(m_2)}||_H}\right) / \log_{10}\left(\frac{m_2}{m_1}\right), \tag{81}$$

where $\mathbf{q}$ is the exact solution and $m_{1,2}$ are the number of grid points for two discretizations.

The operators used in this study are the odd pth order upwind SBP operators from [19], i.e. 3rd, 5th, 7th and 9th order accurate SBP operators. These operators have an interior accuracy of p and a boundary accuracy of $(p-1)/2$ [19]. The expected convergence rate is then $(p+1)/2$ [68]. In the simulations the time step is set to $\Delta t = 0.1\Delta x$, and the classical fourth order Runge-Kutta method is employed. In the third case the sixth order Runge-Kutta method from [69] is employed to make sure that the time-discretization error in the solution is negligible, see Section 5.3. In all computations the gravitational acceleration is set to $g = 9.81$.

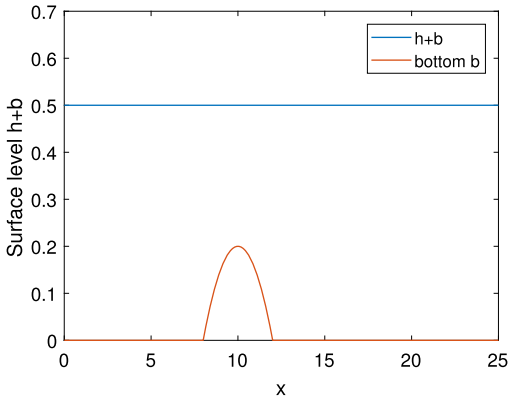


Fig. 3. Analytic solution for the "lake at rest" problem.

Table 1
L_2 errors for the "Lake at rest" problem.

m	$\log_{10} L^2_{(3rd)}$	$\log_{10} L^2_{(5th)}$	$\log_{10} L^2_{(7th)}$	$\log_{10} L^2_{(9th)}$
50	-14.625	-14.881	-14.829	-14.809
100	-14.141	-14.218	-14.350	-14.582
200	-14.283	-14.111	-14.520	-13.926
400	-13.721	-14.001	-13.963	-14.141

5.1. Lake at rest

A well-known (1D) benchmark for verification of shallow water codes is the "lake at rest" problem [37], see Fig. 3 where the analytic solution is presented. A numerical scheme that solves this problem to machine precision is well-balanced. Here the surface level of the solution, $h+b$, is initialized as a constant and the discharge hu is set to zero. The numerical scheme should preserve this. This was done for the following bottom topography

$$b(x) = \begin{cases} 0.2 - 0.05\,(x-10)^2\,, & \text{if } 8 < x < 12, \\ 0, & \text{else}, \end{cases} \tag{82}$$

with domain length $L = 25$ and $h + b = 0.5$. Note that the bottom topography is non-smooth at $x = 8$ and $x = 12$. This means that the solution h will be non-smooth at these points which may impact the numerical accuracy unless special care is taken. Here this is handled by utilizing well-known techniques from the literature [40], see Sections 3.2.2 and 3.2.3 for more details. Periodic BCs were used and the end time was set to $t = 10$. In Table 1 computed L_2 errors are presented. The error is close to machine precision independent of the number of grid points used in the discretization verifying that the scheme is well-balanced.

5.2. Dam break on a wet domain

The dam break problem was first introduced in [70] and can be used to test if the numerical method catches the transitory behavior and shock behavior of the solution properly. The initial conditions are given by

$$h(x,0) = \begin{cases} h_l & 0 \le x \le x_0, \\ h_r & x_0 < x \le L, \end{cases}$$
$$(hu)(x,0) = 0.$$

Here the parameters were chosen as: $h_l = 1$, $h_r = 0.5$, $L = 1$ and $x_0 = 0.5$. The initial condition is plotted with a yellow dashed line in Fig. 4a. The solution is given by a left going rarefaction wave and a right going shock, see [67] for more details. The analytic solution is plotted at $t = 0.1$ in the blue line in Fig. 4a. In Figs. 4b, 4c and 4d the numerical solutions are presented using 3rd order, 5th order and 9th order SBP operators, respectively. 501 grid points were used. The method is able to capture the location of shocks and rarefactions correctly but with oscillations and an overshoot at the shock location as a side effect. These side effects are amplified as the order of the operators is increased, see Fig. 4. Additionally, the overshoot does not vanish with grid refinement meaning that the maximum error of the solution does not converge. Note that to guarantee convergence in the presence of strongly nonlinear phenomena such as shocks, AD in the form of a second derivative is necessary. The upwind operators only introduce AD in the form of higher order derivatives (commonly

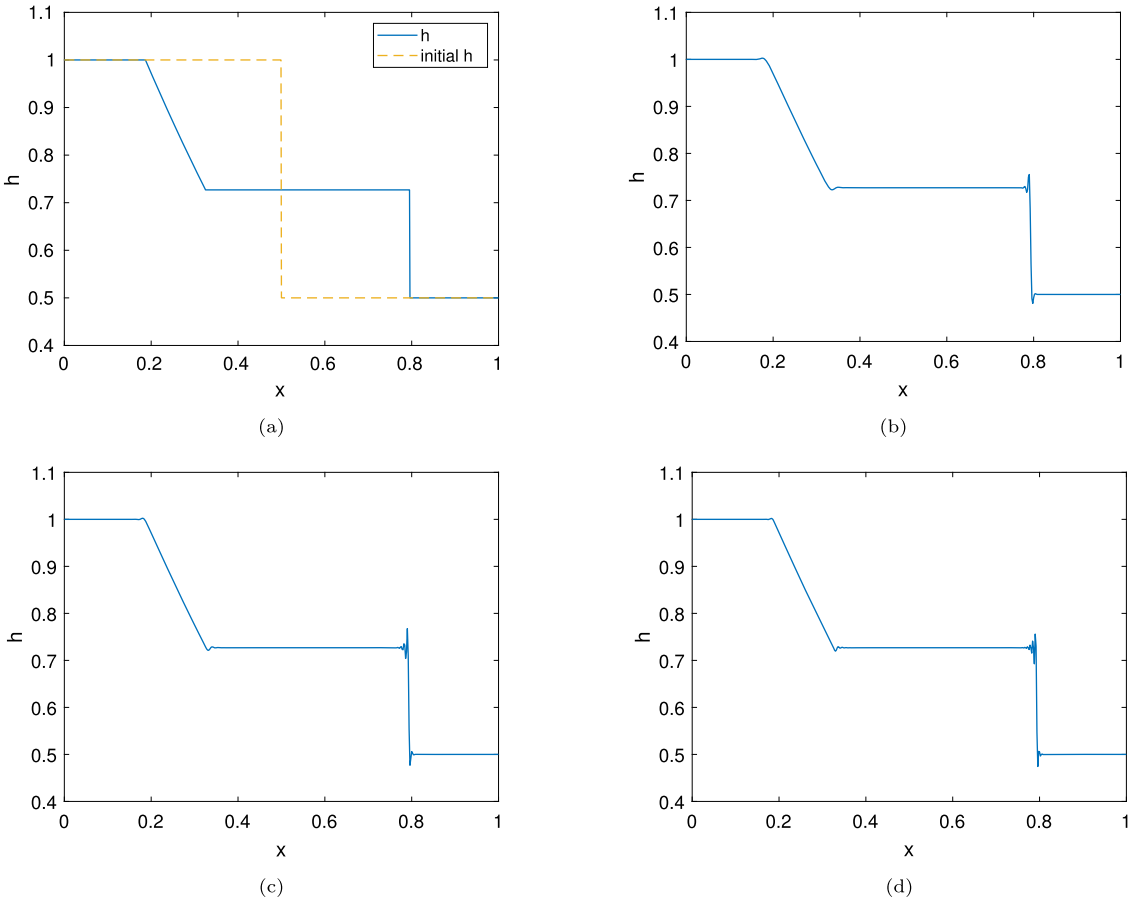


Fig. 4. (a) Analytic solution to the "Dam break on a wet domain" problem. The dashed yellow line is the initial condition and the blue line is the analytic solution at $t = 0.1$. The three remaining figures show the numerical solution computed using 501 grid points and (b) 3rd order operators, (c) 5th order operators and (d) 9th order operators. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

referred to as hyperviscosity). Hence, additional techniques are required to make the scheme truly shock-capturing [51–54] and are the subject of ongoing work.

5.3. Gaussian pulse

One area where the SBP-SAT method is advantageous is for smooth solutions including BCs. To this end we solved the shallow water equations for a Gaussian pulse with wall BCs at both boundaries, i.e. $(hu)(x_{l,r}, t) = 0$. The following initial Gaussian was used

$$\begin{cases} h(x,0) = h_0 + G_{Amp} \exp\left(-\left(\dfrac{x - x_0}{r_0}\right)^2\right), \\ (hu)(x,0) = 0, \end{cases}$$

where G_{Amp} is the amplitude, x_0 is the initial position, r_0 is the radius of the Gaussian. h_0 is the unperturbed initial height. The following parameters were used in the computations: $G_{Amp} = 0.1$, $h_0 = 1$, $x_0 = 0.5$ and $r_0 = 0.1$. These parameters were chosen to ensure that the solution remained smooth at $t = 0.22$.

The error of the numerical solution was measured against a reference solution computed on a fine mesh using 1601 grid points and using the 9th order operators. The solution was time integrated using the sixth order Runge-Kutta method from [69] to make sure the error in time was negligible to be able to see grid convergence rates more clearly. The time evolution for this solution is plotted in Fig. 5 where the solution at the top ($t = 0.22$) was used as reference solution. The convergence results are presented in Table 2. The convergence rates are somewhat higher than the expected convergence rate $(p + 1)/2$.

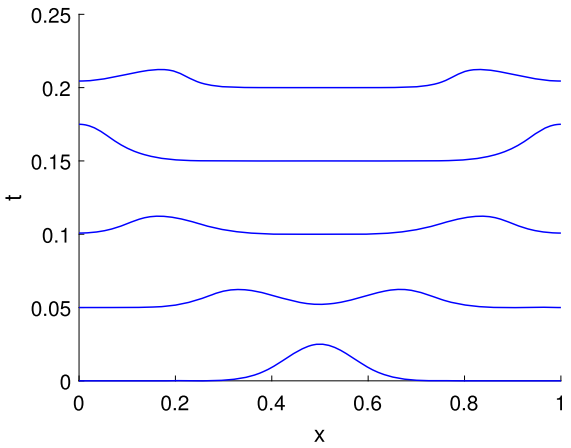


Fig. 5. Time series evolution of two Gaussian pulses moving in opposite directions with a wall BC. The solution at the top, plotted at $t = 0.22$, was used as reference solution.

Table 2
L_2 errors and convergence rates for a Gaussian pulse with a wall BC.

m	$\log_{10} L^2_{(3rd)}$	c	$\log_{10} L^2_{(5th)}$	c	$\log_{10} L^2_{(7th)}$	c	$\log_{10} L^2_{(9th)}$	c
51	-2.320	-	-2.795	-	-2.917	-	-2.985	-
101	-2.964	2.169	-3.779	3.315	-4.138	4.113	-4.319	4.498
201	-3.736	2.585	-5.048	4.246	-5.824	5.642	-6.248	6.452
401	-4.597	2.869	-6.500	4.840	-7.837	6.711	-8.679	8.104
801	-5.491	2.974	-7.979	4.922	-9.927	6.953	-10.789	7.022

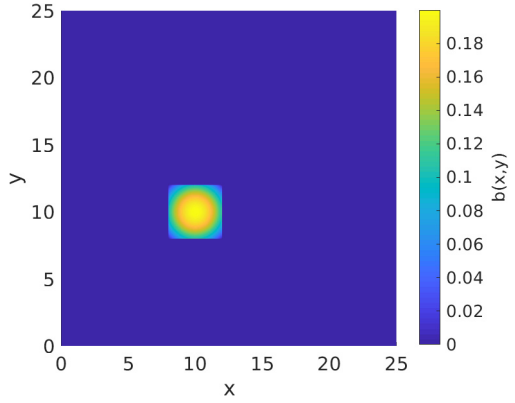


Fig. 6. The bottom topography used for the "lake at rest" problem in 2D.

6. Computations 2D

The time step is set to $\Delta t = 0.1\Delta x$ and the simulations are time integrated by employing the classical fourth order Runge-Kutta method. In all computations the gravitational acceleration is set to $g = 9.81$.

6.1. Lake at rest 2D

This problem is a 2D extension to the 1D problem presented in Section 5.1. The bottom topography used is given by

$$b(x,y) = \begin{cases} 0.2 - 0.025\left((x-10)^2 + (y-10)^2\right), & \text{if } 8 < x < 12 \quad \text{and} \quad 8 < y < 12, \\ 0, & \text{else}, \end{cases} \tag{83}$$

and is plotted in Fig. 6. The computed L_2 errors are presented in Table 3 and are very similar to the 1D case.

Table 3
L_2 errors for the "Lake at rest problem" in 2D.

m	$\log_{10} L^2_{(3rd)}$	$\log_{10} L^2_{(5th)}$	$\log_{10} L^2_{(7th)}$	$\log_{10} L^2_{(9th)}$
50	-13.730	-13.998	-14.088	-13.846
100	-13.286	-13.364	-13.551	-13.033
200	-13.416	-13.270	-13.168	-12.580
400	-13.003	-13.214	-13.003	-13.067

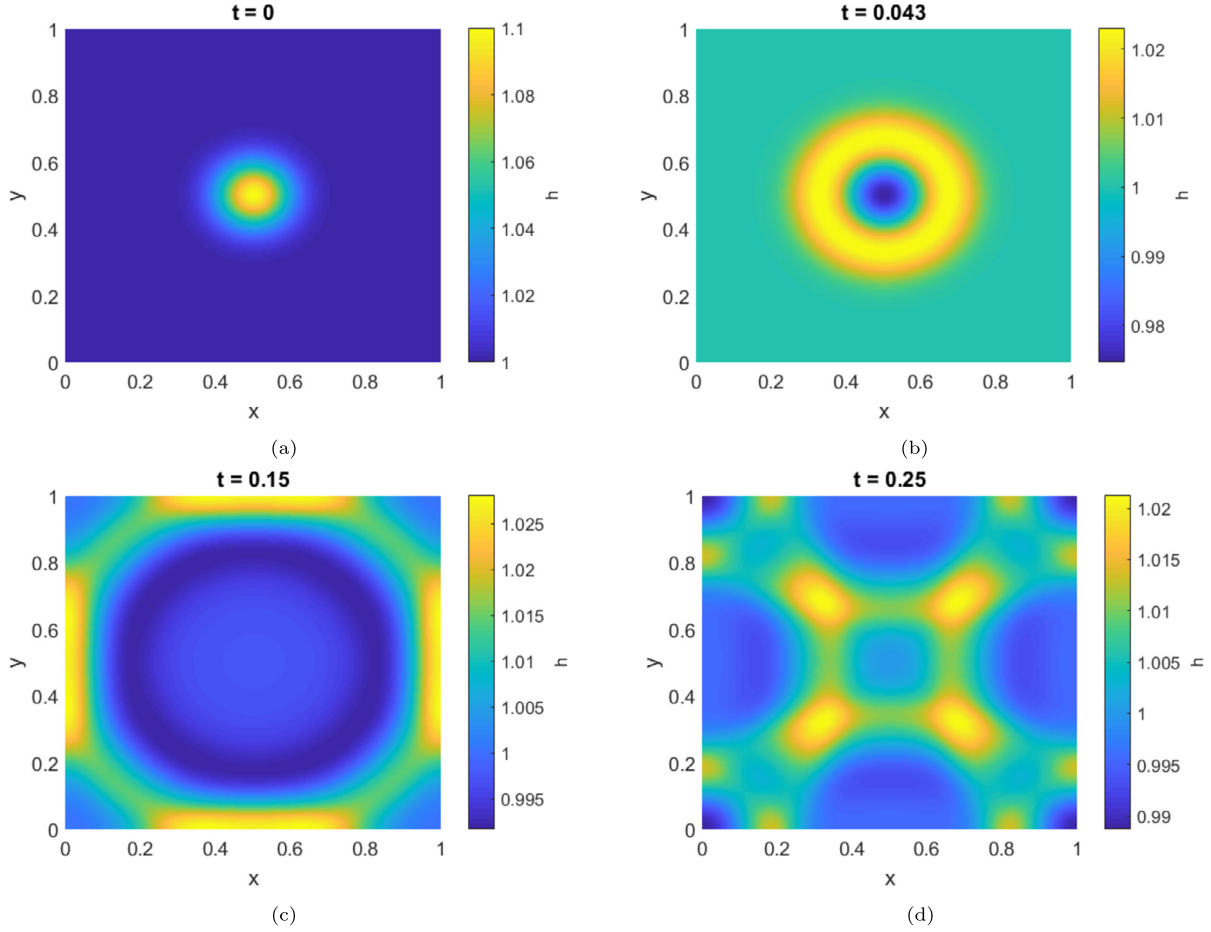


Fig. 7. Time evolution of a Gaussian pulse in 2D. (d) was used as reference solution for the convergence results.

6.2. Gaussian pulse 2D

This problem is a 2D extension to the 1D problem presented in Section 5.3. The following initial Gaussian was used

$$\begin{cases} h(x,y,0) = h_0 + G_{Amp} \exp\left(-\left(\dfrac{x-x_0}{r_0}\right)^2 - \left(\dfrac{y-y_0}{r_0}\right)^2\right), \\ (hu)(x,y,0) = 0, \\ (hv)(x,y,0) = 0, \end{cases}$$

with $r_0 = 0.1$, $G_{Amp} = 0.1$ and $(x_0, y_0) = (0.5, 0.5)$. As in the 1D case, these parameters were chosen to obtain a smooth solution. Wall BCs were set at each of the four boundaries and are given by

$$\begin{cases} \mathbf{q}^{(2)} = (hu)(x,y,t) = 0, \quad x = 0 \text{ and } x = 1, \\ \mathbf{q}^{(3)} = (hv)(x,y,t) = 0, \quad y = 0 \text{ and } y = 1. \end{cases} \tag{84}$$

The end time of the simulation was chosen to $t = 0.25$ and a time evolution of the solution is plotted in Fig. 7. To compute convergence the error was measured against a reference solution, computed on a fine mesh using the 9th order

Table 4
L_2 errors and convergence rates for the Gaussian pulse with a wall BC in 2D problem.

m	$\log_{10} L^2_{(3rd)}$	c	$\log_{10} L^2_{(5th)}$	c	$\log_{10} L^2_{(7th)}$	c	$\log_{10} L^2_{(9th)}$	c
51	-2.758	-	-3.421	-	-3.487	-	-3.430	-
101	-3.508	2.527	-4.599	3.970	-4.921	4.831	-5.020	5.357
201	-4.352	2.824	-5.693	3.660	-6.263	4.492	-6.679	5.552
401	-5.224	2.908	-6.641	3.160	-7.494	4.102	-8.225	5.155
801	-6.047	2.739	-7.578	3.118	-8.705	4.031	-10.226	6.658

Table 5
Time-step restrictions for the Gaussian pulse with a wall BC in 2D problem.

Order	3rd	5th	7th	9th
Δt	0.2966	0.3242	0.3281	0.3113

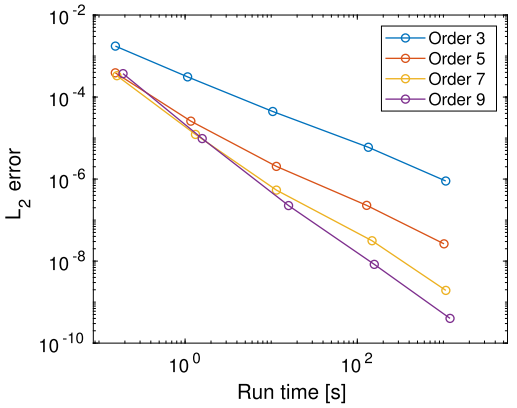


Fig. 8. L_2 error as a function of run time for the Gaussian pulse in 2D problem.

operators, which is plotted in Fig. 7d. To make sure the time error in the reference solution was negligible it was time integrated employing the sixth order Runge-Kutta method from [69]. 1601^2 grid points were used for the reference solution. The solution itself was time integrated using the classical fourth order Runge-Kutta method. The convergence results are presented in Table 4. The convergence rates are close to the expected boundary accuracy $(p+1)/2$.

6.2.1. Efficiency

In Table 5 the time-step restrictions (when considering this particular benchmark problem) for the odd order upwind SBP operators are presented. The restrictions were found numerically by searching for the lowest time-step which produced a complex valued solution for $t \in (0, 1]$ using $m = 101$. The results from the table are similar to [19, Fig. 1a] where, similarly, 3rd order upwind SBP operators were shown to have the largest spectral radius.

The time-step values from Table 5 (divided by two to make sure that the solution wasn't at the edge of stability) were used to measure L_2 errors and run time. In this way, the underlying stiffness of each order SBP operator was taken into account when measuring run time. In Fig. 8 the L_2 errors are plotted as a function of run time for the different operators. It is favorable with as low run time as possible for a given error. This means that the plot shows the efficiency of the different operators. While higher order operators use wider interior stencils, they are potentially much more efficient when we expect high-order convergence, which is confirmed in Fig. 8.

Remark. In order to make sure that the computed solution was obtained at exactly $t = 0.25$, the final time-step was chosen as $0.25 - \lfloor 0.25/\Delta t \rfloor * \Delta t$.

7. Application

7.1. A simplified incident wave simulation

In this section we solve the shallow water equations in a more realistic multi-block setting with wave-channel interaction and varying bathymetry. In Fig. 9 the basic set up of the problem is presented. The computational domain is divided into four blocks: Ω_W, Ω_M, Ω_E and Ω_N. The black lines in Fig. 9 represent wall BCs, the blue lines represent characteristic non-reflecting BCs and the dashed blue lines represent interface conditions. The idea behind this setup is that a wave is initialized as a Gaussian profile in Ω_M. This represents an incident wave from the ocean. This wave will travel and hit the

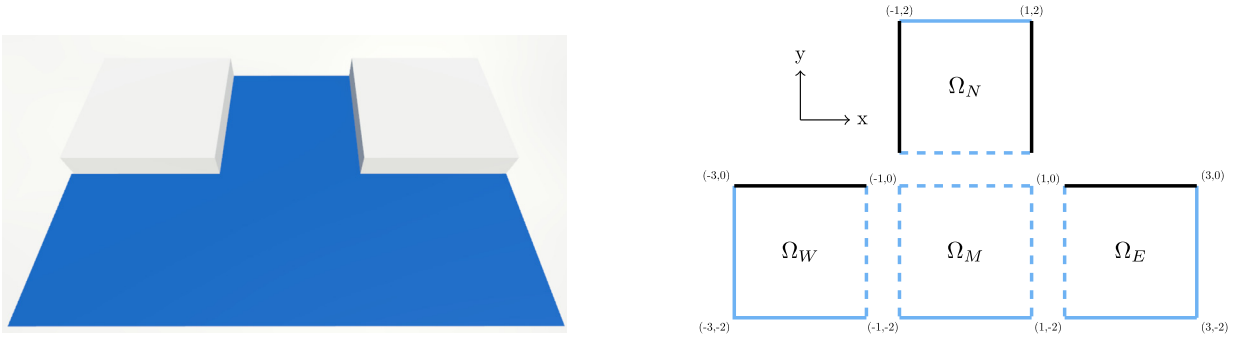


Fig. 9. The multi-block setup. The domain has been divided into four blocks. The black lines in (b) represent wall BCs, the blue lines represent characteristic non-reflecting BCs and the dashed blue lines represent interface conditions.

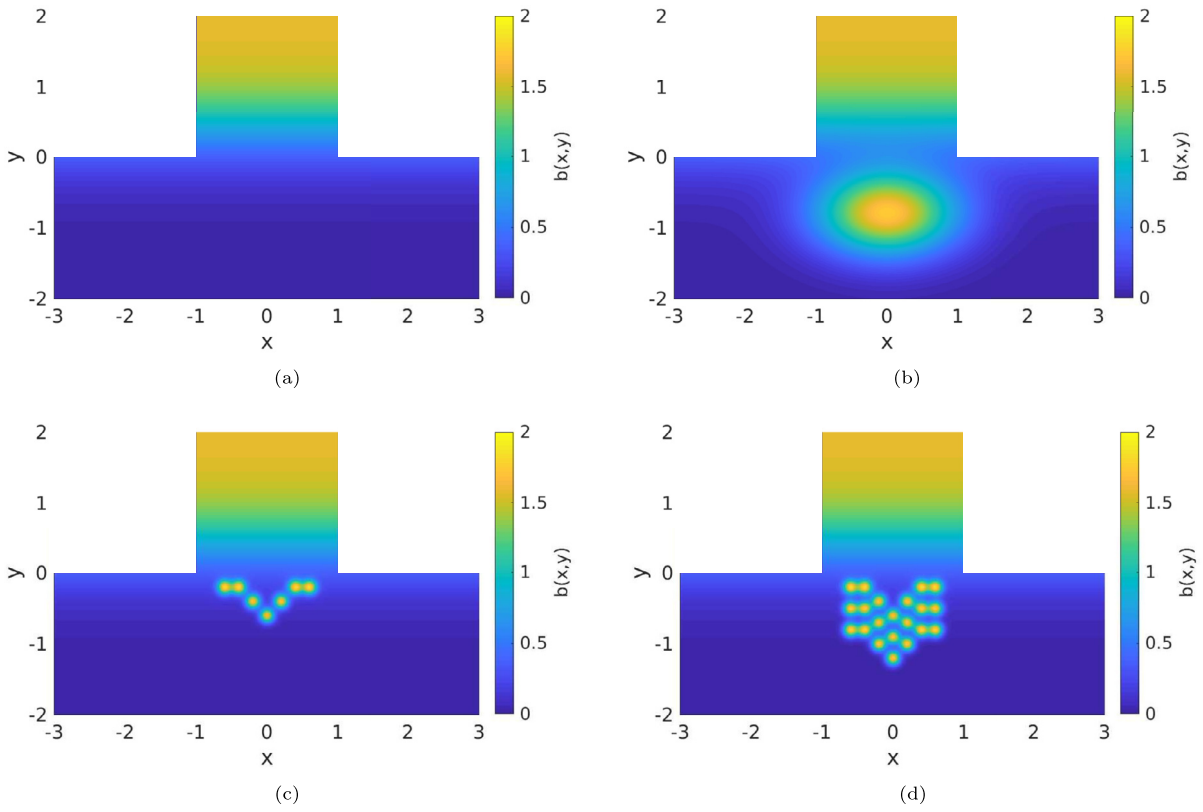


Fig. 10. The four bathymetries used in the simulations.

walls in western and eastern domain and also enter the channel in the northern domain. In Section 7.2 we simulate this setup with the four bathymetries in Fig. 10. The bottom topography is given by

$$b(x,y) = \sum_{i=1}^{n} b^{(i)}_{Gaussian}(x,y) + b_{Sigmoid}(x,y), \tag{85}$$

where $b_{Sigmoid}(x,y)$ is a Sigmoid function representing the increasing bottom topography from the ocean towards the shoreline and $b^{(i)}_{Gauss}(x,y)$ represents one of multiple Gaussian humps. The Sigmoid bottom topography is given by

$$b_{Sigmoid}(x,y) = S_{Amp} \tanh\left(1.5\,(y-0.4)\right) + S_{Amp}, \tag{86}$$

where S_{Amp} represents the height of the shoreline compared to the ocean. The Gaussian bottom topography is given by a Gaussian profile. The four bathymetries used in the simulations are presented in Fig. 10.

7.1.1. Continuous model

The full IBVP for the multi-block setup is given by

$$
\begin{aligned}
&(\mathbf{q}_{W,M,E,N})_t + \mathbf{A}_1(\mathbf{q}_{W,M,E,N})_x \\
&+\mathbf{A}_2(\mathbf{q}_{W,M,E,N})_y + \mathbf{C}\mathbf{q}_{W,M,E,N} = \mathbf{E}, && (x,y) \in \Omega_{W,M,E,N}, && t > 0, \\
&\mathbf{A}_{1+}\mathbf{q}_W = \mathbf{A}_{1+}g_C, && x = -3, \\
&\mathbf{A}_{2+}\mathbf{q}_W = \mathbf{A}_{2+}g_C, && y = -2, \\
&\mathbf{A}_{1-}\mathbf{q}_E = \mathbf{A}_{1-}g_C, && x = 3, \\
&\mathbf{A}_{2+}\mathbf{q}_E = \mathbf{A}_{2+}g_C, && y = -2, \\
&\mathbf{A}_{2+}\mathbf{q}_M = \mathbf{A}_{2+}g_C, && y = -2, && t \geq 0, \\
&\mathbf{A}_{2-}q_N = \mathbf{A}_{2-}g_C, && y = 2, \\
&\mathbf{q}_W^{(3)} = (hv)_W = 0, && y = 0, \\
&\mathbf{q}_E^{(3)} = (hv)_E = 0, && y = 0, \\
&\mathbf{q}_N^{(2)} = (hu)_N = 0, && x = -1 \text{ and } x = 1, \\
&\mathbf{q}_{W,M,E,N}^{(1)} = h_{initial}(x,y), \\
&\mathbf{q}_{W,M,E,N}^{(2)} = \mathbf{q}_{W,M,E,N}^{(3)} = 0, && (x,y) \in \Omega_{W,M,E,N} && t = 0,
\end{aligned}
\tag{87}
$$

where $g_C = \begin{bmatrix} h_0 - b(x,y) & 0 & 0 \end{bmatrix}^T$ and $h_{initial}(x,y) = h_0 - b(x,y) + G_{Amp} \exp\left(-\left(\frac{x-x_0}{r_0}\right)^2 - \left(\frac{y-y_0}{r_0}\right)^2 \right)$. The bottom topography $b(x,y)$ is given by (85). These BCs are well-posed which is shown for the western block in Section 4.2. If the analysis is repeated, it can be shown that the BCs are well-posed for the remaining blocks as well. Well-posedness at the interfaces is analyzed in Section 4.2.2.

7.1.2. Semi-discrete approximation

The semi-discrete approximation of the well-posed IBVP (87) is given by

$$
\begin{aligned}
&(q_W)_t + (I_3 \otimes D_x)F_1 + (I_3 \otimes D_y)F_2 - (R_x \otimes H_x^{-1}S_x + R_y \otimes H_y^{-1}S_y)(q_W + \bar{b}) + G = \mathrm{SAT}_W, \\
&(q_E)_t + (I_3 \otimes D_x)F_1 + (I_3 \otimes D_y)F_2 - (R_x \otimes H_x^{-1}S_x + R_y \otimes H_y^{-1}S_y)(q_E + \bar{b}) + G = \mathrm{SAT}_E, \\
&(q_M)_t + (I_3 \otimes D_x)F_1 + (I_3 \otimes D_y)F_2 - (R_x \otimes H_x^{-1}S_x + R_y \otimes H_y^{-1}S_y)(q_M + \bar{b}) + G = \mathrm{SAT}_M, \\
&(q_N)_t + (I_3 \otimes D_x)F_1 + (I_3 \otimes D_y)F_2 - (R_x \otimes H_x^{-1}S_x + R_y \otimes H_y^{-1}S_y)(q_N + \bar{b}) + G = \mathrm{SAT}_N,
\end{aligned}
\tag{88}
$$

where the SATs in (88) impose the continuous boundary and interface conditions in (87) weakly and are given by

$$
\begin{aligned}
\mathrm{SAT}_W = {} & \bar{H}_y^{-1} e_N W_{y-} \Lambda_{y-} \left(L_v W_{y-}\right)^{-1} L_v e_N^T q_W \\
& - A_{2+} \otimes H_y^{-1} e_S \left(e_S^T q_W - e_S^T g_C \right) \\
& - A_{1+} \otimes H_x^{-1} e_W \left(e_W^T q_W - e_W^T g_C \right) \\
& + A_{1-} \otimes H_x^{-1} e_E \left(e_E^T q_W - e_W^T q_M \right),
\end{aligned}
\tag{89}
$$

$$
\begin{aligned}
\mathrm{SAT}_M = {} & A_{2-} \otimes H_y^{-1} e_N \left(e_N^T q_M - e_S^T q_N \right) \\
& - A_{2+} \otimes H_y^{-1} e_S \left(e_S^T q_M - e_S^T g_C \right) \\
& - A_{1+} \otimes H_x^{-1} e_W \left(e_W^T q_M - e_E^T q_W \right) \\
& + A_{1-} \otimes H_x^{-1} e_E \left(e_E^T q_M - e_W^T q_E \right),
\end{aligned}
\tag{90}
$$

$$
\begin{aligned}
\mathrm{SAT}_E = {} & \bar{H}_y^{-1} e_N W_{y-} \Lambda_{y-} \left(L_v W_{y-}\right)^{-1} L_v e_N^T q_E - A_{2+} \otimes H_y^{-1} e_S \left(e_S^T q_E - e_S^T g_C \right) \\
& - A_{1+} \otimes H_x^{-1} e_W \left(e_W^T q_E - e_E^T q_M \right) + A_{1-} \otimes H_x^{-1} e_E \left(e_E^T q_E - e_E^T g_C \right),
\end{aligned}
\tag{91}
$$

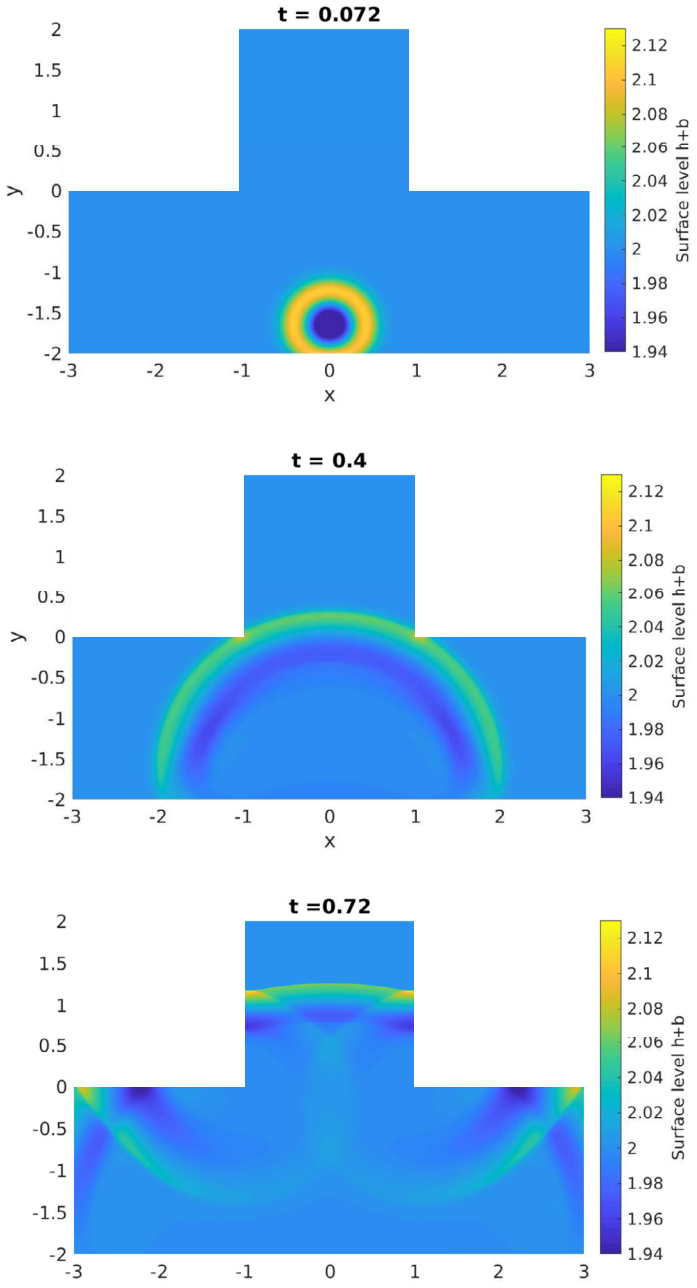


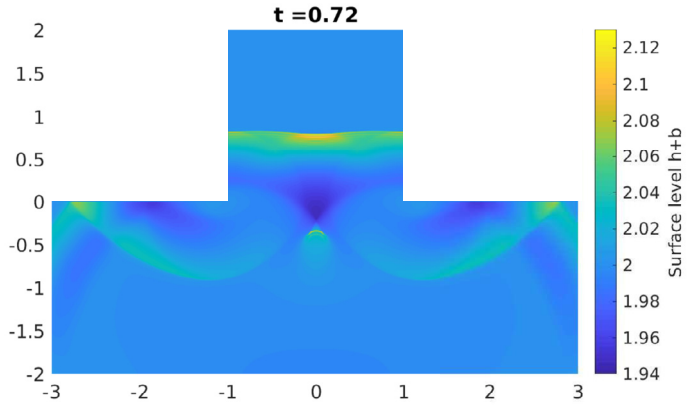
Fig. 11. Time evolution of the simulation with bathymetry given by Fig. 10a.

$$\begin{aligned}\mathrm{SAT}_N = \bar{H}_x^{-1} e_E W_{x-} \Lambda_{x-} \left(L_u W_{x-}\right)^{-1} L_u e_E^T q_N - \bar{H}_x^{-1} e_W W_{y+} \Lambda_{x+} \left(L_u W_{x+}\right)^{-1} L_u e_W^T q_N \\ + A_{2-} \otimes H_y^{-1} e_N \left(e_N^T q_N - e_N^T g_C\right) - A_{2+} \otimes H_y^{-1} e_S \left(e_S^T q_N - e_N^T q_M\right),\end{aligned} \tag{92}$$

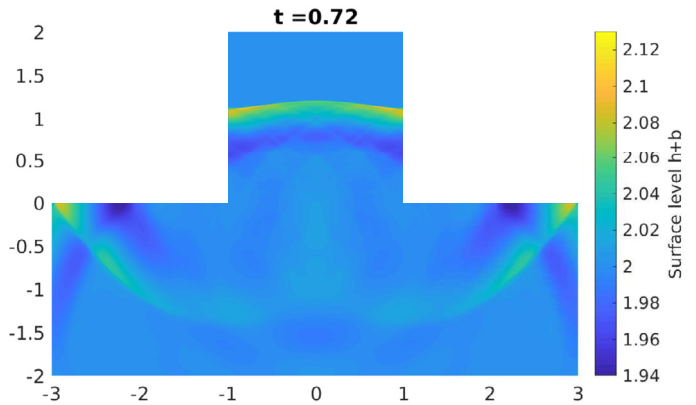
where $L_u = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$ and $L_v = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ are boundary operators. As an example, SAT_W is the SAT which imposes the boundary and interface conditions in Ω_W. The first term in (89) imposes the wall BC, the next two terms impose the characteristic non-reflecting BCs and the last term imposes the interface condition.

Lemma 7.1. *Let $b(x, y)$ be a smooth function of x and y. If the continuous problem (87) is well-posed, the symmetrized frozen coefficient version of the scheme (88) is stable.*

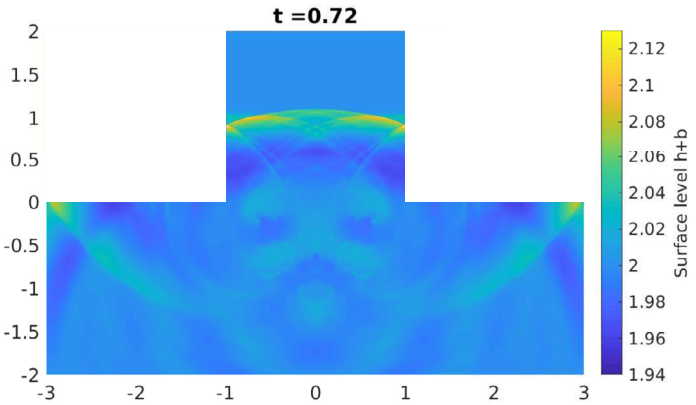
Proof. See Sections 4.3.2 and 4.3.3. □



(a)



(b)



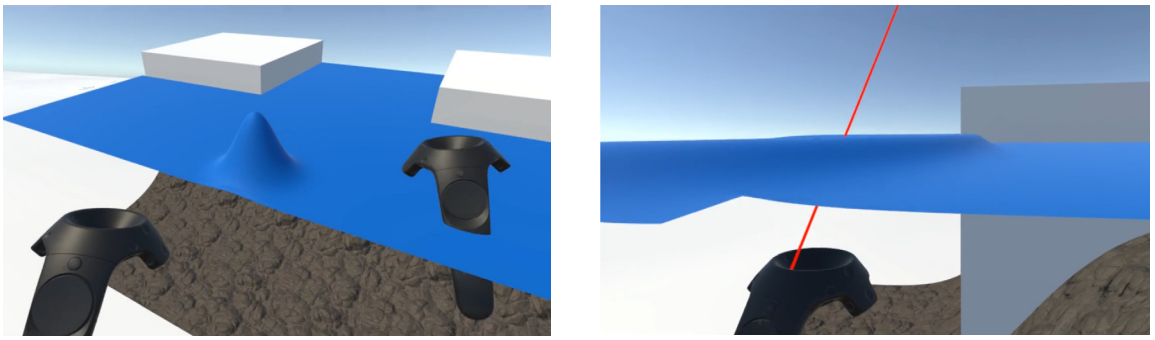
(c)

Fig. 12. Surface level of the simulations at $t = 0.72$. The bathymetry in Fig. 10b was used for (a), Fig. 10c was used for (b) and Fig. 10d was used for (c).

7.2. Experiments

In this section four simulations have been performed based on the four bathymetries in Fig. 10. Each simulation uses 250^2 grid points in each block and utilizes 9th order accurate operators. The bottom topography for the Sigmoid, given by equation (86) using $S_{Amp} = 0.8$, is the same for all simulations, see Fig. 10a. The following parameters were used as initial conditions: $h_0 = 2$, $G_{Amp} = 0.5$, $x_0 = 0$, $y_0 = -1.65$ and $r_0 = 0.2$. The gravitational acceleration is set to $g = 9.81$.

In the first simulation no other additions are added. The time evolution of this simulation is plotted in Fig. 11. The second simulation has a Gaussian hump added to the bathymetry with the following properties: $G_{Amp} = 1.7$, $x_0 = 0$, $y_0 = -0.8$, $r_0 = 1$ in the x-direction and $r_0 = 0.6$ in the y-direction, see Fig. 10b. The surface level is plotted at $t = 0.72$ in Fig. 12a. In



(a) (b)

Fig. 13. Interacting with the simulations using immersive virtual reality.

the third and fourth simulations multiple humps are added, all with $r_0 = 0.1$ and G_{Amp} between 1.6 and 1.75. The difference between the third and fourth simulations is that the fourth has 21 bumps placed instead of 7, see Figs. 10c and 10d. The surface level of the third and fourth simulation is plotted at $t = 0.72$ in Figs. 12b and 12c, respectively.

7.2.1. Visualization using immersive virtual reality

The experiments performed in Section 7.2 have also been visualized in immersive virtual reality to gain additional insight into the simulation results. The visualization was done using the game engine Unity [71]. A video showcasing this is available via the following link: https://youtu.be/oXPINFmHdEo. Two screenshots from the video are presented in Fig. 13.

8. Conclusion and future work

A stable high-order explicit upwind finite difference scheme solving the shallow water equations has been derived using the SBP-SAT method. The method was tested against various benchmark solutions and shows high-order convergence provided that the solution is smooth, which was also the main focus of the present study. To achieve a well-balanced scheme we used a particular discretization of the source term as well as a modified flux-splitting in order for the added AD not to violate the well-balanced property following the ideas presented in [40]. An application was also presented: A simplified incident wave simulation using a multi-block setup. Simulation results show that a bathymetry consisting of spikes could have a dispersing effect on an incoming wave.

Although the focus of this work is on smooth solutions, the scheme is still robust against shocks. We do not claim that the solution is overshoot or oscillation free however. Additional work is needed to make the scheme truly shock-capturing as well extending the scheme to be positivity-preserving, to be able to handle wet/dry transitions correctly.

CRediT authorship contribution statement

Lukas Lundgren: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Ken Mattsson:** Conceptualization, Formal analysis, Methodology, Supervision, Writing - review & editing.

Declaration of competing interest

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

Acknowledgements

We sincerely thank the anonymous reviewers for their helpful comments which significantly improved the quality of this manuscript. The research of the first author is funded by Uppsala University and Ingrid Cloud (https://www.ingridcloud.com/).

Appendix A. Transformation of BCs in 2D

This section is a 2D extension of Section 3.1.1. Let $\mathbf{X}_{x,y} = \begin{bmatrix} X_{x,y+} & X_{x,y-} & X_{x,y0} \end{bmatrix}$ be the matrix that diagonalizes $\tilde{\mathbf{A}}_{1,2}$ such that $\mathbf{X}_{x,y}^T \tilde{\mathbf{A}}_{1,2} \mathbf{X}_{x,y} = \mathbf{\Lambda}_{x,y}$. Note that this also means that $\mathbf{W}_{x,y}$ from (57) is given by $\mathbf{W}_{x,y} = \mathbf{T}\mathbf{X}_{x,y} = \begin{bmatrix} \mathbf{T}X_{x,y+} & \mathbf{T}X_{x,y-} & \mathbf{T}X_{x,y0} \end{bmatrix}$. One important observation is that the third component of the so-called characteristic variables

$$\mathbf{w}_{x,y} = \mathbf{W}_{x,y}^{-1}\mathbf{q} = \mathbf{X}_{x,y}^{T}\tilde{\mathbf{q}} = \begin{bmatrix} w_{x,y+} \\ w_{x,y-} \\ w_{x,y,0} \end{bmatrix} = \begin{bmatrix} \frac{q_1}{2} \\ \frac{q_1}{2} \\ 0 \end{bmatrix}, \tag{A.1}$$

is zero. As a consequence of $w_{x,y,0} = 0$, the BCs (homogeneous) in the subcritical case can be expressed as

$$\begin{aligned} w_{x+} &= -r_W w_{x-}, \quad x = x_l, \\ w_{x-} &= -r_E w_{x+}, \quad x = x_r, \\ w_{y+} &= -r_S w_{y-}, \quad y = y_l, \\ w_{y-} &= -r_N w_{y+}, \quad y = y_r, \end{aligned} \tag{A.2}$$

where $r_W = \left(\tilde{L}_W X_{x+}\right)^{-1} \tilde{L}_W X_{x-}$, $r_E = \left(\tilde{L}_E X_{x-}\right)^{-1} \tilde{L}_E X_{x+}$, $r_S = \left(\tilde{L}_S X_{y+}\right)^{-1} \tilde{L}_S X_{y-}$ and $r_N = \left(\tilde{L}_N X_{y-}\right)^{-1} \tilde{L}_N X_{y+}$. Performing similar steps as in Section 3.1.1, the boundary terms (62) can be expressed as

$$\widetilde{BT}_W = \int_{y_l}^{y_r} w_{x-}^2 \left(r_W^2 \Lambda_{x+} + \Lambda_{x-}\right)\Big|^{x=x_l} dy, \qquad \widetilde{BT}_E = -\int_{y_l}^{y_r} w_{x+}^2 \left(\Lambda_{x+} + r_E^2 \Lambda_{x-}\right)\Big|^{x=x_r} dy, \tag{A.3}$$

$$\widetilde{BT}_S = \int_{x_l}^{x_r} w_{y-}^2 \left(r_S^2 \Lambda_{y+} + \Lambda_{y-}\right)\Big|^{y=y_l} dx, \qquad \widetilde{BT}_N = -\int_{x_l}^{x_r} w_{y+}^2 \left(\Lambda_{y+} + r_N^2 \Lambda_{y-}\right)\Big|^{y=y_r} dx. \tag{A.4}$$

Here the 2D shallow water equations are linearly well-posed if $r_{W,E,S,N}$ are sufficiently small such that (61) holds.

Appendix B. Stability proof for the 2D scheme

Since lower order terms don't affect stability, they are neglected. We focus on the well-posed BCs analyzed in Section 4.2.1, which are imposed by the SAT given by (74). The scheme (75) without lower order terms is then given by

$$\begin{aligned} \tilde{q}_t + \tilde{A}_1 \otimes D_x \tilde{q} + \tilde{A}_2 \otimes D_y \tilde{q} = {} & \bar{H}_y^{-1} e_N X_{y-} \Lambda_{y-} \left(\tilde{L}_N X_{y-}\right)^{-1} L_N e_N^T \tilde{q} \\ & - \tilde{A}_{2+} \otimes H_y^{-1} e_S \left(e_S^T \tilde{q} - e_S^T \tilde{g}_C\right) \\ & - \tilde{A}_{1+} \otimes H_x^{-1} e_W \left(e_W^T \tilde{q} - e_W^T \tilde{g}_C\right) \\ & + \bar{H}_x^{-1} e_E X_{x-} \Lambda_{x-} \left(\tilde{L}_E X_{x-}\right)^{-1} e_E^T \tilde{L}_E \tilde{q} \\ & + \left(R_x \otimes H_x^{-1} S_x + R_y \otimes H_y^{-1} S_y\right) \tilde{q}. \end{aligned} \tag{B.1}$$

Since $hu = 0$ is imposed at the east boundary, this leads to $\Lambda_{x+} > 0$ and $\Lambda_{x-} < 0$. Similarly, at the north boundary $hv = 0$ is imposed, which leads to $\Lambda_{y+} > 0$ and $\Lambda_{y-} < 0$. To continue with the proof we introduce the variables

$$\begin{aligned} w_{yN} &= \begin{bmatrix} w_{yN+} \\ w_{yN-} \\ w_{yN0} \end{bmatrix} = e_N^T \left(X_y^T \otimes I_{m^2}\right) \tilde{q}, \\ w_{xE} &= \begin{bmatrix} w_{xE+} \\ w_{xE-} \\ w_{xE0} \end{bmatrix} = e_E^T \left(X_x^T \otimes I_{m^2}\right) \tilde{q}. \end{aligned} \tag{B.2}$$

Setting $\tilde{g}_C = 0$, multiplying (B.1) by $\tilde{q}^T \bar{H}$ from the left and adding the transpose leads to

$$\frac{d}{dt} ||\tilde{q}||_{\bar{H}}^2 = BT_W + BT_E + BT_S + BT_N + AD, \tag{B.3}$$

where

$$AD = 2\tilde{q}^T \left(R_x \otimes H_y S_x + R_y \otimes H_x S_y\right) \tilde{q}, \tag{B.4}$$

is the added AD from the upwind operators which is a negative contribution to the semi-discrete energy. The boundary terms are analyzed one by one. The term BT_N is given by

$$\begin{aligned} BT_N = &- \Lambda_{y+}||w_{yN+}||_H^2 + \Lambda_{y-}||w_{yN-}||_H^2 - \Lambda_{y0}||w_{yN0}||_H^2 \\ &+ 2\Lambda_{y-}\left(\tilde{L}_N X_{y-}\right)^{-1}\tilde{L}_N X_{y+}(w_{yN+}, w_{yN-})_H + 2\Lambda_{y-}\left(\tilde{L}_N X_{y-}\right)^{-1}\tilde{L}_N X_{y0}(w_{yN0}, w_{yN-})_H. \end{aligned} \tag{B.5}$$

To arrive at (B.5) similar steps were performed as in the proof in Section 3.2.3. Since $w_{y0} = 0$ from (A.1), we obtain

$$\begin{aligned} BT_N &= -\Lambda_{y+}||w_{yN+}||_H^2 + \Lambda_{y-}||w_{yN-}||_H^2 + 2\Lambda_{y-}(\tilde{L}_N X_{y-})^{-1}\tilde{L}_N X_{y+}(w_{yN+}, w_{yN-})_H \\ &= -\Lambda_{y+}||w_{yN+}||_H^2 + \Lambda_{y-}||w_{yN-}||_H^2 + 2\Lambda_{y-} r_N (w_{yN+}, w_{yN-})_H. \end{aligned} \tag{B.6}$$

Inserting the BCs (A.2) into (B.6) yields

$$BT_N = -||w_{yN+}||_H^2 \left(\Lambda_{y+} + r_N^2 \Lambda_{y-}\right), \tag{B.7}$$

which mimics (A.3). This means that (B.7) is ≤ 0 provided that the problem is well-posed. Similarly, BT_E is given by

$$\begin{aligned} BT_E = &- \Lambda_{x+}||w_{xE+}||_H^2 + \Lambda_{x-}||w_{xE-}||_H^2 - \Lambda_{x0}||w_{xE0}||_H^2 + 2\Lambda_{-}\left(\tilde{L}_E X_{x-}\right)^{-1}\tilde{L}_E X_{x+}(w_{xE+}, w_{xE-})_H \\ &+ 2\Lambda_{-}\left(\tilde{L}_E X_{x-}\right)^{-1}\tilde{L}_E X_{x0}(w_{xE0}, w_{xE-})_H \\ = &- ||w_{xE+}||_H^2 \left(\Lambda_{x+} + r_E^2 \Lambda_{x-}\right), \end{aligned} \tag{B.8}$$

which is ≤ 0 provided that the continuous problem is well-posed. Lastly, BT_W and BT_S are given by

$$BT_W = \left(e_W^T \tilde{q}\right)^T \left(\left(\tilde{A}_{1-} - \tilde{A}_{1+}\right) \otimes H\right) e_W^T \tilde{q} \leq 0, \tag{B.9}$$

$$BT_S = \left(e_S^T \tilde{q}\right)^T \left(\left(\tilde{A}_{2-} - \tilde{A}_{2+}\right) \otimes H\right) e_S^T \tilde{q} \leq 0. \tag{B.10}$$

We have now shown that $\frac{d}{dt}||\tilde{q}||_H^2 \leq 0$, which means that the scheme (75) is stable. The proof is similar for other well-posed BCs.

References

[1] A.B. de Saint-Venant, Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et a l'introduction de marées dans leurs lits, C. R. Acad. Sci. 73 (1871) 147–154, pp. 237–240.
[2] N. Goutal, F. Maurel, A finite volume solver for 1D shallow-water equations applied to an actual river, J. Hydrol. 38 (1) (2002) 1–19.
[3] M. Heniche, Y. Secretan, P. Boudreau, M. Leclerc, A two-dimensional finite element drying-wetting shallow water model for rivers and estuaries, Adv. Water Resour. 23 (4) (2000) 359–372.
[4] E. Mignot, A. Paquier, S. Haider, Modeling floods in a dense urban area using 2D shallow water equations, J. Hydrol. 327 (1) (2006) 186–199.
[5] A. Valiani, V. Caleffi, A. Zanni, Case study: Malpasset dam-break simulation using a two-dimensional finite volume method, J. Hydraul. Eng. 128 (5) (2002) 460–472.
[6] M. Esteves, X. Faucher, S. Galle, M. Vauclin, Overland flow and infiltration modelling for small plots during unsteady rain: numerical results versus observed values, J. Hydrol. 228 (3) (2000) 265–282.
[7] G. Kirstetter, J. Hu, O. Delestre, F. Darboux, P.-Y. Lagrée, S. Popinet, J. Fullana, C. Josserand, Modeling rain-driven overland flow: empirical versus analytical friction terms in the shallow water approximation, J. Hydrol. 536 (2016) 1–9.
[8] D.L. George, Finite volume methods and adaptive refinement for tsunami propagation and inundation, PhD thesis, University of Washington, 2006.
[9] F. Løvholt, G.K. Pedersen, S. Glimsdal, Coupling of dispersive tsunami propagation and shallow water coastal response, Open Oceanogr. J. 4 (2010) 71–82.
[10] B. Bonev, J.S. Hesthaven, F.X. Giraldo, M.A. Kopera, Discontinuous Galerkin scheme for the spherical shallow water equations with applications to tsunami modeling and prediction, J. Comput. Phys. 362 (2018) 425–448.
[11] H.-O. Kreiss, J. Oliger, Comparison of accurate methods for the integration of hyperbolic equations, Tellus 24 (1972) 199–215.
[12] B. Strand, Summation by parts for finite difference approximations for d/dx, J. Comput. Phys. 110 (1) (1994) 47–67.
[13] K. Mattsson, J. Nordström, Summation by parts operators for finite difference approximations of second derivatives, J. Comput. Phys. 199 (2) (2004) 503–540.
[14] M.H. Carpenter, D. Gottlieb, S. Abarbanel, Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: methodology and application to high-order compact schemes, J. Comput. Phys. 111 (2) (1994) 220–236.
[15] J.E. Hicken, Output error estimation for summation-by-parts finite-difference schemes, J. Comput. Phys. 231 (9) (2012) 3828–3848.
[16] K. Mattsson, M. Almquist, M.H. Carpenter, Optimal diagonal-norm SBP operators, J. Comput. Phys. 264 (2014) 91–111.
[17] M. Almquist, K. Mattsson, T. Edvinsson, High-fidelity numerical solution of the time-dependent Dirac equation, J. Comput. Phys. 262 (2014) 86–103.
[18] K. Mattsson, J. Werpers, High-fidelity numerical simulation of solitons in the nerve axon, J. Comput. Phys. 305 (2016) 793–816.
[19] K. Mattsson, Diagonal-norm upwind SBP operators, J. Comput. Phys. 335 (2017) 283–310.
[20] Y. Rydin, K. Mattsson, J. Werpers, High-fidelity sound propagation in a varying 3D atmosphere, J. Sci. Comput. 77 (2) (2018) 1278–1302.
[21] M.H. Carpenter, J. Nordström, D. Gottlieb, A stable and conservative interface treatment of arbitrary spatial accuracy, J. Comput. Phys. 148 (2) (1999) 341–365.
[22] K. Mattsson, F. Ham, G. Iaccarino, Stable and accurate wave-propagation in discontinuous media, J. Comput. Phys. 227 (19) (2008) 8753–8767.
[23] K. Mattsson, M. Svärd, M. Carpenter, J. Nordström, High-order accurate computations for unsteady aerodynamics, Comput. Fluids 36 (3) (2007) 636–649.
[24] J.E. Hicken, D.W. Zingg, Aerodynamic optimization algorithm with integrated geometry parameterization and mesh movement, AIAA J. 48 (2) (2010) 400–413.

[25] G.J. Gassner, A skew-symmetric discontinuous Galerkin spectral element discretization and its relation to SBP-SAT finite difference methods, SIAM J. Sci. Comput. 35 (3) (2013) A1233–A1253.
[26] G.J. Gassner, A.R. Winters, D.A. Kopriva, A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations, Appl. Math. Comput. 272 (P2) (2016) 291–308.
[27] N. Wintermeyer, A.R. Winters, G.J. Gassner, D.A. Kopriva, An entropy stable nodal discontinuous Galerkin method for the two dimensional shallow water equations on unstructured curvilinear meshes with discontinuous bathymetry, J. Comput. Phys. 340 (2017) 200–242.
[28] N. Wintermeyer, A.R. Winters, G.J. Gassner, T. Warburton, An entropy stable discontinuous Galerkin method for the shallow water equations on curvilinear meshes with wet/dry fronts accelerated by GPUs, J. Comput. Phys. 375 (2018) 447–480.
[29] Y. Xing, C.-W. Shu, A survey of high order schemes for the shallow water equations, J. Math. Study 47 (3) (2014) 221–249.
[30] A. Chertock, S. Cui, A. Kurganov, T. Wu, Well-balanced positivity preserving central-upwind scheme for the shallow water system with friction terms, Int. J. Numer. Methods Fluids 78 (6) (2015) 355–383.
[31] A. Beljadid, A. Mohammadian, A. Kurganov, Well-balanced positivity preserving cell-vertex central-upwind scheme for shallow water flows, Comput. Fluids 136 (2016) 193–206.
[32] H. Shirkhani, A. Mohammadian, O. Seidou, A. Kurganov, A well-balanced positivity-preserving central-upwind scheme for shallow water equations on unstructured quadrilateral grids, Comput. Fluids 126 (2016) 25–40.
[33] M. Ricchiuto, An explicit residual based approach for shallow water flows, J. Comput. Phys. 280 (2015) 306–344.
[34] H. Ranocha, Shallow water equations: split-form, entropy stable, well-balanced, and positivity preserving numerical methods, GEM Int. J. Geomath. 8 (1) (2017) 85–133.
[35] S. Marras, M.A. Kopera, E.M. Constantinescu, J. Suckale, F.X. Giraldo, A residual-based shock capturing scheme for the continuous/discontinuous spectral element solution of the 2D shallow water equations, Adv. Water Resour. 114 (2018) 45–63.
[36] E. Audusse, C. Chalons, P. Ung, A simple well-balanced and positive numerical scheme for the shallow-water system, Commun. Math. Sci. 13 (5) (2015) 1317–1332.
[37] O. Delestre, Simulation du ruissellement d'eau de pluie sur des surfaces agricoles, PhD thesis, Université d'Orléans, 2010.
[38] A. Chertock, M. Dudzinski, A. Kurganov, M. Lukáčová-Medvid'ová, Well-balanced schemes for the shallow water equations with Coriolis forces, Numer. Math. 138 (4) (2018) 939–973.
[39] S. Noelle, N. Pankratz, G. Puppo, J.R. Natvig, Well-balanced finite volume schemes of arbitrary order of accuracy for shallow water flows, J. Comput. Phys. 213 (2) (2006) 474–499.
[40] Y. Xing, C.-W. Shu, High order finite difference WENO schemes with the exact conservation property for the shallow water equations, J. Comput. Phys. 208 (1) (2005) 206–227.
[41] J.-F. Gerbeau, B. Perthame, Derivation of viscous Saint-Venant system for laminar shallow water; numerical validation, Discrete Contin. Dyn. Syst., Ser. B 1 (1) (2001) 89–102.
[42] B. Gustafsson, H.-O. Kreiss, J. Oliger, Time-Dependent Problems and Difference Methods, second ed., Pure and Applied Mathematics (Hoboken), John Wiley & Sons, Inc., Hoboken, NJ, 2013.
[43] G. Strang, Accurate partial difference methods. II. Non-linear problems, Numer. Math. 6 (1964) 37–46.
[44] H.-O. Kreiss, J. Lorenz, Initial-Boundary Value Problems and the Navier-Stokes Equations, Pure and Applied Mathematics, vol. 136, Academic Press, Inc., Boston, MA, 1989.
[45] S. Eriksson, A dual consistent finite difference method with narrow stencil second derivative operators, J. Sci. Comput. 75 (2) (2018) 906–940.
[46] M. Svärd, J. Nordström, Review of summation-by-parts schemes for initial-boundary-value problems, J. Comput. Phys. 268 (2014) 17–38.
[47] D.C. Del Rey Fernández, J.E. Hicken, D.W. Zingg, Review of summation-by-parts operators with simultaneous approximation terms for the numerical solution of partial differential equations, Comput. Fluids 95 (2014) 171–196.
[48] H.-O. Kreiss, G. Scherer, Finite element and finite difference methods for hyperbolic partial differential equations, in: Mathematical Aspects of Finite Elements in Partial Differential Equations, Academic Press, 1974, pp. 195–212.
[49] M. Svärd, M.H. Carpenter, M. Parsani, Entropy stability and the no-slip wall boundary condition, SIAM J. Numer. Anal. 56 (1) (2018) 256–273.
[50] M. Svärd, S. Mishra, Shock capturing artificial dissipation for high-order finite difference schemes, J. Sci. Comput. 39 (3) (2009) 454–484.
[51] M. Nazarov, J. Hoffman, Residual-based artificial viscosity for simulation of turbulent compressible flow using adaptive finite element methods, Int. J. Numer. Methods Fluids 71 (3) (2013) 339–357.
[52] J.-L. Guermond, M. Nazarov, B. Popov, I. Tomas, Second-order invariant domain preserving approximation of the Euler equations using convex limiting, SIAM J. Sci. Comput. 40 (5) (2018) A3211–A3239.
[53] J.-L. Guermond, M. Nazarov, A maximum-principle preserving C^0 finite element method for scalar conservation equations, Comput. Methods Appl. Mech. Eng. 272 (2014) 198–213.
[54] M. Nazarov, Convergence of a residual based artificial viscosity finite element method, Comput. Math. Appl. 65 (4) (2013) 616–626.
[55] S. Mishra, M. Svärd, On stability of numerical schemes via frozen coefficients and the magnetic induction equations, BIT Numer. Math. 50 (1) (2010) 85–108.
[56] J. Nordström, M.H. Carpenter, Boundary and interface conditions for high-order finite-difference methods applied to the Euler and Navier-Stokes equations, J. Comput. Phys. 148 (2) (1999) 621–645.
[57] J. Nordström, M.H. Carpenter, High-order finite difference methods, multidimensional linear problems, and curvilinear coordinates, J. Comput. Phys. 173 (1) (2001) 149–174.
[58] M. Svärd, M.H. Carpenter, J. Nordström, A stable high-order finite difference scheme for the compressible Navier-Stokes equations, far-field boundary conditions, J. Comput. Phys. 225 (1) (2007) 1020–1038.
[59] M. Svärd, J. Nordström, A stable high-order finite difference scheme for the compressible Navier-Stokes equations. No-slip wall boundary conditions, J. Comput. Phys. 227 (10) (2008) 4805–4824.
[60] J.E. Hicken, D.W. Zingg, Superconvergent functional estimates from summation-by-parts finite-difference discretizations, SIAM J. Sci. Comput. 33 (2) (2011) 893–922.
[61] J. Berg, J. Nordström, Superconvergent functional output for time-dependent problems using finite differences on summation-by-parts form, J. Comput. Phys. 231 (20) (2012) 6846–6860.
[62] S. Ghader, J. Nordström, Revisiting well-posed boundary conditions for the shallow water equations, Dyn. Atmos. Ocean. 66 (2014) 1–9.
[63] M. Svärd, H. Özcan, Entropy-stable schemes for the Euler equations with far-field and wall boundary conditions, J. Sci. Comput. 58 (1) (2014) 61–89.
[64] S. Johansson, A modal PML for the 2D shallow water equations in conservative variables, Master's thesis, KTH, Numerical Analysis, NA, 2015.
[65] I. Navon, B. Neta, M. Hussaini, A perfectly matched layer formulation for the nonlinear shallow water equations models: the split equation approach, 2001.
[66] D. Givoli, B. Neta, High-order nonreflecting boundary conditions for the dispersive shallow water equations, in: Selected Papers from the Conference on Computational and Mathematical Methods for Science and Engineering, Alicante, 2002, vol. 158, 2003, pp. 49–60.
[67] O. Delestre, C. Lucas, P.-A. Ksinant, F. Darboux, C. Laguerre, T.-N.-T. Vo, F. James, S. Cordier, SWASHES: a compilation of shallow water analytic solutions for hydraulic and environmental studies, Int. J. Numer. Methods Fluids 72 (3) (2013) 269–300.

[68] M. Svärd, J. Nordström, On the order of accuracy for difference approximations of initial-boundary value problems, J. Comput. Phys. 218 (1) (2006) 333–352.
[69] E.A. Alshina, E.M. Zaks, N.N. Kalitkin, Optimal first- to sixth-order accurate Runge-Kutta schemes, Comput. Math. Math. Phys. 48 (3) (2008) 395–405.
[70] J.J. Stoker, Water Waves: The Mathematical Theory with Applications, Pure and Applied Mathematics, vol. IV, Interscience Publishers, Inc./Interscience Publishers Ltd., New York/London, 1957.
[71] Unity Technologies, Unity.