

# On well-posed boundary conditions and energy stable finite volume method for the linear shallow water wave equation

Rudi Prihandoko<sup>1</sup>      Kenneth Duru<sup>2</sup>      Stephen Roberts<sup>3</sup>  
Christopher Zoppou<sup>4</sup>

31 January 2023

## Abstract

We derive and analyse well-posed boundary conditions for the linear shallow water wave equation. The analysis is based on the energy method and it identifies the number, location and form of the boundary conditions so that the initial boundary value problem is well-posed. We propose a finite volume method encapsulated in the summation-by-parts framework and implement boundary conditions weakly. We prove stability by deriving discrete energy estimate analogous to the continuous estimate. The continuous and discrete analysis covers all flow regimes, and can be extended to the nonlinear problem. Numerical experiments are presented verifying the analysis.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Continuous analysis</b>	<b>3</b>
<b>3</b>	<b>Numerical scheme</b>	<b>6</b>
3.1	The finite volume method . . . . .	7
3.2	Numerical boundary treatment and stability . . . . .	8
<b>4</b>	<b>Numerical Experiments</b>	<b>10</b>
<b>5</b>	<b>Conclusion</b>	<b>12</b>
<b>A</b>	<b>Appendix</b>	<b>12</b>

# 1 Introduction

Numerical models that solve the shallow water wave equations (SWWE) have become a common tool for modeling environmental problems. This system of nonlinear hyperbolic partial differential equations (PDE) represent the conservation of mass and momentum of unsteady free surface flow subject to gravitational forces. The SWWE assume that the fluid is inviscid, incompressible and the wavelength of the wave is much greater than its height. Typically these waves are associated with flows caused for example by storm surges and floods in riverine systems. The shallow water wave equations are also a fundamental component for predicting a range of aquatic processes, including sediment transport and the transport of pollutants. All these processes can have a significant impact on the environment, vulnerable communities and infrastructure, making accurate predictions using the shallow water wave equations crucial for urban, rural and environmental planners.

For practical problems, the SWWE has been solved numerically using finite difference methods [7], finite volume methods [12], discontinuous Galerkin method [11] and the method of characteristics [2]. Although, the shallow water wave equations are in common use, a rigorous theoretical investigation of boundary conditions necessary for their solution is still an area of active research [3].

In this paper, we investigate well-posed boundary conditions for the linearized SWWE using the energy method [4, 5] and develop provably stable numerical method for the model. As in [3], our analysis identifies the type, location and number of boundary conditions that are required to yield a well-posed initial boundary value problem (IBVP). More importantly, we formulate the boundary conditions so that they can be readily implemented in a stable manner for numerical approximations that obey the summation-by-parts (SBP) principle [6]. We demonstrate this by deriving a stable finite volume method encapsulated in the SBP framework and impose the boundary conditions weakly using the Simultaneous Approximation Term (SAT) method [1]. This SBP-SAT approach enables us to prove that the numerical scheme satisfies the discrete counterparts of energy estimates required for well-posedness of the IBVP, resulting in a provably stable and conservative numerical scheme.

The continuous and discrete analysis covers all flow regimes, including sub-critical, critical and super-critical flows. We have included the proofs, of the theorems and lemmas presented in this paper, in the Appendix. Numerical experiments are performed to verify the theoretical analysis of the continuous and discrete models.

## 2 Continuous analysis

The one dimensional SWWE are

$$\frac{\partial h}{\partial t} + \frac{\partial(uh)}{\partial x} = 0, \quad \frac{\partial(uh)}{\partial t} + \frac{\partial(u^2h + \frac{1}{2}gh^2)}{\partial x} = 0, \quad (1)$$

where  $x \in \mathbb{R}$  is spatial variable,  $t \geq 0$  is time,  $h(x, t)$  and  $u(x, t)$  are the water height and the depth averaged fluid velocity respectively,  $g > 0$  is the gravitational acceleration.

To make our analysis tractable we linearise the SWWE by substituting  $h = H + \tilde{h}$  and  $u = U + \tilde{u}$  into (1), where  $\tilde{h}$  and  $\tilde{u}$  denote perturbations of the constant water depth  $H > 0$  and fluid velocity  $U$  respectively.

After simplifying, the linearised SWWEs are

$$\frac{\partial h}{\partial t} + U \frac{\partial h}{\partial x} + H \frac{\partial u}{\partial x} = 0, \quad \frac{\partial u}{\partial t} + g \frac{\partial h}{\partial x} + U \frac{\partial u}{\partial x} = 0, \quad (2)$$

where we have dropped the tilde on the perturbed variables.

Introducing the unknown vector field  $\mathbf{q} = [h, u]^\top$ , the linear equation (2) can be rewritten in a more compact form as

$$\frac{d\mathbf{q}}{dt} = D\mathbf{q}, \quad D = -M \frac{\partial}{\partial x}, \quad M = \begin{bmatrix} U & H \\ g & U \end{bmatrix}. \quad (3)$$

We will now consider (3) in a bounded domain and augment it with initial and boundary conditions. Let our domain be  $\Omega = [0, 1]$  and  $\Gamma = \{0, 1\}$  be the boundary points. We consider the IBVP

$$\frac{\partial \mathbf{q}}{\partial t} = D\mathbf{q}, \quad (4a)$$

$$\mathbf{q}(x, 0) = \mathbf{f}(x), \quad x \in \Omega, \quad (4b)$$

$$\mathcal{B}\mathbf{q} = 0, \quad x \in \Gamma. \quad (4c)$$

One objective of this study is to investigate linear boundary operators  $\mathcal{B}$  which ensure that the IBVP (4) is well-posed. We have considered zero boundary data, but the results can be extended to nontrivial boundary data.

Let  $\mathbf{p}$  and  $\mathbf{q}$  be real valued functions, and define the weighted scalar product and the norm

$$(\mathbf{p}, \mathbf{q})_W = \int_{\Omega} \mathbf{p}^\top W \mathbf{q} dx, \quad \|\mathbf{q}\|_W^2 = (\mathbf{q}, \mathbf{q})_W, \quad (5)$$

where  $W = W^\top$  and  $\mathbf{q}^\top W \mathbf{q} > 0$  for all non-zero  $\mathbf{q} \in \mathbb{R}^2$ . If  $W = I$  we get the standard  $L_2$  scalar product, and we omit the subscript  $W$ .

**Definition 1.** The IBVP (4) is well-posed if a unique solution  $\mathbf{q}$  satisfies

$$\|\mathbf{q}(\cdot, t)\|_W \leq \kappa e^{\nu t} \|\mathbf{f}\|_W, \quad (6)$$

for some constants  $\kappa > 0$  and  $\nu$  independent of  $\mathbf{f}$ .

The well-posedness of the IBVP (4) can be related to the boundedness of the differential operator  $D$ . The following two definitions are useful.

**Definition 2.** The operator  $D$  is said to be **semi-bounded** if it satisfies

$$(\mathbf{q}, D\mathbf{q})_W \leq \nu \|\mathbf{q}\|_W^2, \quad \nu \in \mathbb{R}. \quad (7)$$

**Definition 3.** The differential operator  $D$  is **maximally semi-bounded** if it is semi-bounded in the function space

$$\mathbb{V} = \{\mathbf{p} \mid \mathbf{p}(x) \in \mathbb{R}^2, \quad \|\mathbf{p}\|_W < \infty, \quad 0 \leq x \leq 1, \quad \{\mathcal{B}\mathbf{p} = 0, \quad x \in \Gamma\}\}, \quad (8)$$

but not semi-bounded in any space with fewer boundary conditions.

It is well-known that the maximally semi-boundedness of differential operator  $D$  is a necessary and sufficient condition for the well-posedness of the IBVP (4) [5]. Thus to ensure that the IBVP (4) is well-posed, we need: 1) the differential operator  $D$  to be semi-bounded and; 2) minimal number of boundary conditions such that  $D$  is maximally semi-bounded.

To begin, we will show that the differential operator  $D$  is semi-bounded in a certain weighted  $L_2$  scalar product.

**Lemma 4.** Consider the weighted  $L_2$  scalar product defined in (5) with  $W = W^\top$  and  $\mathbf{q}^\top W \mathbf{q} \geq 0$  for all non-zero  $\mathbf{q} \in \mathbb{R}^2$ . If the matrix product  $\widetilde{M} = WM$  is symmetric,  $M = M^\top$ , then  $D$  is semi-bounded.

The next step will be to derive boundary operators  $\{\mathcal{B}\mathbf{p} = 0, \quad x \in \Gamma\}$  with minimal number of boundary conditions such that the boundary term is never negative,  $\left(\mathbf{q}^\top \widetilde{M} \mathbf{q}\right)\big|_0^1 \geq 0$ . We will now chose a weight matrix  $W$  such that the weighted  $L_2$ -norm is related to the mechanical energy in the medium. Note in particular, if

$$W = \begin{bmatrix} g & 0 \\ 0 & H \end{bmatrix}, \quad (9)$$

then the weighted  $L_2$ -norm is related to the mechanical energy  $E$ , that is

$$\frac{1}{2} \|\mathbf{q}\|_W^2 = E := \int_{\Omega} \frac{1}{2} (gh^2 + Hu^2) dx \geq 0. \quad (10)$$

We introduce the boundary term

$$BT := -\frac{1}{2gH} \left(\mathbf{q}^\top \widetilde{M} \mathbf{q}\right)\big|_0^1 = \frac{U}{H} \left(\frac{1}{2} h^2\big|_1^0\right) + \left(uh\big|_1^0\right) + \frac{U}{g} \left(\frac{1}{2} u^2\big|_1^0\right). \quad (11)$$

By using the eigen-decomposition of the symmetric matrix  $\widetilde{M}$  the boundary term can be re-written as

$$BT = \frac{1}{2} (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_{x=0} - \frac{1}{2} (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_{x=1}, \quad (12)$$

where

$$[w_1, w_2]^\top = S^\top \mathbf{q}, \quad S = \begin{bmatrix} \frac{1}{c} \left( \lambda_1 - \frac{U}{g} \right) & \frac{1}{d} \left( \lambda_2 - \frac{U}{g} \right) \\ \frac{1}{c} & \frac{1}{d} \end{bmatrix}, \quad (13)$$

and  $c = \sqrt{\left( \lambda_1 - \frac{U}{g} \right)^2 + 1}$ ,  $d = \sqrt{\left( \lambda_2 - \frac{U}{g} \right)^2 + 1}$ . Here,  $S$  is a matrix of orthonormal eigenvectors and so  $S^\top S = I$ . The eigenvalues,  $\lambda_1, \lambda_2$ , are real and given by

$$\lambda_1 = \frac{1}{2gH} \left( U(g+H) + \sqrt{U^2(g+H)^2 + 4gH(gH - U^2)} \right), \quad (14)$$

$$\lambda_2 = \frac{1}{2gH} \left( U(g+H) - \sqrt{U^2(g+H)^2 + 4gH(gH - U^2)} \right). \quad (15)$$

The number of boundary conditions will depend on the signs of the eigenvalues  $\lambda_1, \lambda_2$ . Depending on the magnitude of the flow  $U$  we will have three main cases: sub-critical, super-critical, and critical flow regimes. We can also discriminate positive  $U > 0$  and negative  $U < 0$ . When  $U > 0$ ,  $x = 0$  is an inflow boundary and  $x = 1$  is an outflow boundary. The situation reverses when  $U < 0$ , that is  $x = 0$  becomes an outflow boundary and  $x = 1$  is an inflow boundary.

**Sub-critical flow.** The flow is sub-critical when  $U^2 < gH$  which implies  $\lambda_1 > 0$  and  $\lambda_2 < 0$ . We need one boundary condition at  $x = 0$  and one boundary condition at  $x = 1$ . Thus at the sub-critical flow regime we will always need an inflow boundary condition and an outflow boundary condition for any  $U$ . We formulate the boundary conditions

$$\{\mathcal{B}\mathbf{p} = 0, x \in \Gamma\} \equiv \{w_1 = \gamma_0 w_2, x = 0; w_2 = \gamma_1 w_1, x = 1\}, \quad (16)$$

where  $\gamma_0, \gamma_1 \in \mathbb{R}$  are boundary reflection coefficients. The following Lemma constraints the parameters  $\gamma_0, \gamma_1$ .

**Lemma 5.** *Consider the boundary term  $BT$  defined in (12) and the boundary condition (16) for sub-critical flows  $U^2 < gH$  with  $\lambda_1 > 0$  and  $\lambda_2 < 0$ . If  $\gamma_0^2 \leq -\lambda_2/\lambda_1$  and  $\gamma_1^2 \leq -\lambda_1/\lambda_2$  then the boundary term is never positive,  $BT \leq 0$ .*

**Super-critical flow.** When  $U^2 > gH$  the flow is super-critical, then  $\lambda_1$  and  $\lambda_2$  both take the sign of the average flow velocity  $U$ . That is if  $U > 0$  then  $\lambda_1 > 0$ ,  $\lambda_2 > 0$  and if  $U < 0$  then  $\lambda_1 < 0$ ,  $\lambda_2 < 0$ . Thus when  $U > 0$  we need two boundary conditions at  $x = 0$  and no boundary conditions at  $x = 1$ . Similarly, when  $U < 0$  we need two boundary conditions at  $x = 1$  and no boundary conditions at  $x = 0$ . Thus at the super-critical flow regime there are no outflow boundary conditions for any  $U$ . We formulate the boundary conditions

$$\{\mathcal{B}\mathbf{q} = 0, x \in \Gamma\} \equiv \{w_1 = 0, w_2 = 0, x = 0; \text{ if } U > 0\}, \quad (17a)$$

$$\{\mathcal{B}\mathbf{q} = 0, x \in \Gamma\} \equiv \{w_1 = 0, w_2 = 0, x = 1; \text{ if } U < 0\}. \quad (17b)$$

**Lemma 6.** Consider the boundary term  $BT$  defined in (12) and the boundary condition (17) for super-critical flows  $U^2 > gH$ , we have  $BT \leq 0$ .

**Critical flow.** The flow is critical when  $U^2 = gH$ . Note that this case is degenerate, since there is only one nonzero eigenvalue, that is  $U > 0$  implies  $\lambda_1 > 0$ ,  $\lambda_2 = 0$  and  $U < 0$  implies  $\lambda_1 = 0$ ,  $\lambda_2 < 0$ . However, it can also be treated by prescribing only one boundary condition for the system. The location of the boundary condition will be determined by the sign of  $U$ , similar to the super-critical flow regime. We prescribe the boundary conditions

$$\{\mathcal{B}\mathbf{q} = 0, x \in \Gamma\} \equiv \{w_1 = 0, x = 0; \text{ if } U > 0 \text{ and } \lambda_2 = 0\}, \quad (18a)$$

$$\{\mathcal{B}\mathbf{q} = 0, x \in \Gamma\} \equiv \{w_2 = 0, x = 1; \text{ if } U < 0 \text{ and } \lambda_1 = 0\}. \quad (18b)$$

**Lemma 7.** Consider the boundary term  $BT$  defined in (12) and the boundary condition (18) for critical flows  $U^2 = gH$ , we have  $BT \leq 0$ .

We will conclude this section with the theorem which proves the well-posedness of the IBVP (4).

**Theorem 8.** Consider the IBVP (4) where the boundary operator  $\mathcal{B}\mathbf{q} = 0$  is defined by (16) with  $\gamma_0^2 \leq -\lambda_2/\lambda_1$  and  $\gamma_1^2 \leq -\lambda_1/\lambda_2$  for sub-critical flows,  $U^2 < gH$ , by (17) for the super-critical flow regime,  $U^2 > gH$ , and by (18) for critical flows,  $U^2 = gH$ , we have the energy estimate

$$\frac{1}{2} \frac{d}{dt} \|\mathbf{q}\|_W^2 = gH \times BT \leq 0. \quad (19)$$

This energy estimate (19) is what a stable method should emulate.

### 3 Numerical scheme

We will now derive a stable finite volume method for the IBVP (4) encapsulated in the SBP framework. We will prove numerical stability by deriving discrete energy estimates analogous to Theorem 8.

### 3.1 The finite volume method

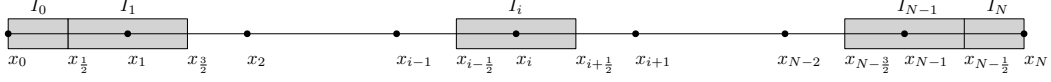


Figure 1: Finite volume nodes  $x_i$  and control cells  $I_i$ .

To begin, we split the domain,  $\Omega = [0, 1]$ , into  $N + 1$  computational nodes having  $x_i = x_{i-1} + \Delta x_i$ , for  $i = 1, 2, \dots, N$ , with  $x_0 = 0$ ,  $\Delta x_i > 0$  and  $\sum_{i=1}^N \Delta x_i = 1$ . We consider the control cell  $I_i = [x_{i-1/2}, x_{i+1/2}]$  for each interior node  $1 \leq i \leq N - 1$ , and for the boundary nodes  $\{x_0, x_N\}$  the control cells are  $I_0 = [x_0, x_{1/2}]$  and  $I_N = [x_{N-1/2}, x_N]$ , see Figure 1. Note that  $|I_i| = \Delta x_i/2 + \Delta x_{i+1}/2$  for the interior nodes  $1 \leq i \leq N - 1$ , and for the boundary nodes  $i \in \{0, N\}$  we have  $|I_0| = \Delta x_0/2$  and  $|I_N| = \Delta x_{N-1}/2$ . The control cells  $I_i$  are connected and do not overlap, and  $\sum_{i=0}^N |I_i| = \sum_{i=1}^N \Delta x_i = 1$ .

Consider the integral form of (4a) over the control cells  $I_i$

$$\frac{d}{dt} \int_{I_0} \mathbf{q}(x, t) dx + M\mathbf{q}(x_{1/2}, t) - M\mathbf{q}(x_0, t) = 0, \quad (20a)$$

$$\frac{d}{dt} \int_{I_i} \mathbf{q}(x, t) dx + M\mathbf{q}(x_{i+1/2}, t) - M\mathbf{q}(x_{i-1/2}, t) = 0, \quad 1 \leq i \leq N - 1, \quad (20b)$$

$$\frac{d}{dt} \int_{I_N} \mathbf{q}(x, t) dx + M\mathbf{q}(x_N, t) - M\mathbf{q}(x_{N-1/2}, t) = 0. \quad (20c)$$

Introduce the cell-average

$$\bar{\mathbf{q}}_i = \frac{1}{|I_i|} \int_{I_i} \mathbf{q}(x, t) dx, \quad (21)$$

and approximate the PDE flux  $M\mathbf{q}$  with the local Lax-Friedrich flux

$$M\mathbf{q}(x_{i+1/2}, t) \approx \frac{M\bar{\mathbf{q}}_{i+1} + M\bar{\mathbf{q}}_i}{2} - \frac{\alpha}{2} (\bar{\mathbf{q}}_{i+1} - \bar{\mathbf{q}}_i), \quad \alpha \geq 0, \quad (22)$$

and  $M\mathbf{q}(x_0, t) \approx M\bar{\mathbf{q}}_0$ ,  $M\mathbf{q}(x_N, t) \approx M\bar{\mathbf{q}}_N$ . The evolution of the cell-average is governed by the semi-discrete system

$$|I_0| \frac{d\bar{\mathbf{q}}_0}{dt} + M \frac{\bar{\mathbf{q}}_1 - \bar{\mathbf{q}}_0}{2} - \frac{\alpha}{2} (\bar{\mathbf{q}}_1 - \bar{\mathbf{q}}_0) = 0, \quad (23a)$$

$$|I_i| \frac{d\bar{\mathbf{q}}_i}{dt} + M \frac{\bar{\mathbf{q}}_{i+1} - \bar{\mathbf{q}}_{i-1}}{2} - \frac{\alpha}{2} (\bar{\mathbf{q}}_{i+1} - 2\bar{\mathbf{q}}_i + \bar{\mathbf{q}}_{i-1}) = 0, \quad 1 \leq i \leq N - 1, \quad (23b)$$

$$|I_N| \frac{d\bar{\mathbf{q}}_N}{dt} + M \frac{\bar{\mathbf{q}}_N - \bar{\mathbf{q}}_{N-1}}{2} - \frac{\alpha}{2} (\bar{\mathbf{q}}_{N-1} - \bar{\mathbf{q}}_N) = 0. \quad (23c)$$

Introducing the discrete solution vector  $\bar{\mathbf{q}} = [\bar{\mathbf{q}}_0, \bar{\mathbf{q}}_1, \dots, \bar{\mathbf{q}}_N]^\top$  and rewriting (23) in a more compact form, we have

$$(I \otimes P) \frac{d\bar{\mathbf{q}}}{dt} + (M \otimes Q) \bar{\mathbf{q}} - \frac{\alpha}{2} (I \otimes A) \bar{\mathbf{q}} = 0, \quad (24)$$

where  $\otimes$  denotes the Kronecker product and

$$Q = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad A = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix},$$

and  $P = \text{diag}([|I_0|, |I_1|, \dots, |I_N|])$ . The matrix  $Q$  is related to the spatial derivative operator,  $A$  is a numerical dissipation operator, and  $\alpha \geq 0$  controls the amount of numerical dissipation applied. The important stability property of the semi-discrete approximation (24) is that the associated discrete derivative operator satisfies the SBP property. To see this, we rewrite equation (24) as

$$\frac{d\bar{\mathbf{q}}}{dt} + (M \otimes D_x) \bar{\mathbf{q}} - \frac{\alpha}{2} (I \otimes P^{-1}A) \bar{\mathbf{q}} = 0, \quad (25)$$

where

$$D_x = P^{-1}Q, \quad Q + Q^\top = \text{diag}([-1, 0, \dots, 0, 1]). \quad (26)$$

The relation (26) is the so-called SBP property [6, 4] for the first derivative  $d/dx$ , which can be useful in proving numerical stability of the discrete approximation (24). Note that we have not enforced any boundary condition yet, the boundary condition (4c) will be implemented weakly using penalties.

### 3.2 Numerical boundary treatment and stability

We will now implement the boundary conditions and prove numerical stability. The boundary conditions are implemented using the SAT method, by appending the boundary operators (16)–(18) to the right hand-side of (24) with penalty weights, we have

$$(I \otimes \mathbf{P}) \frac{d\bar{\mathbf{q}}}{dt} + (M \otimes \mathbf{Q}) \bar{\mathbf{q}} - \frac{\alpha}{2} (I \otimes \mathbf{A}) \bar{\mathbf{q}} = \text{SAT}. \quad (27)$$



With  $\mathbf{e}_0 = [1, 0, \dots, 0]^T$  and  $\mathbf{e}_N = [0, 0, \dots, 1]^T$ , the SAT for sub-critical flow is

$$\text{SAT} = -\frac{1}{2} (W^{-1}SW \otimes \mathbf{I}) \begin{bmatrix} \tau_{01}H\mathbf{e}_0(\bar{w}_1 - \gamma_0\bar{w}_2) + \tau_{N1}H\mathbf{e}_N(\bar{w}_2 - \gamma_1\bar{w}_1) \\ \tau_{02}g\mathbf{e}_0(\bar{w}_1 - \gamma_0\bar{w}_2) + \tau_{N2}g\mathbf{e}_N(\bar{w}_2 - \gamma_1\bar{w}_1) \end{bmatrix}, \quad (28)$$

and for critical/super-critical flow regimes we have

$$\text{SAT} = -\frac{1}{2} (W^{-1}SW \otimes \mathbf{I}) \begin{bmatrix} \tau_{01}H\mathbf{e}_0\bar{w}_1 \\ \tau_{02}g\mathbf{e}_0\bar{w}_2 \end{bmatrix}, \quad U > 0, \quad (29a)$$

$$\text{SAT} = -\frac{1}{2} (W^{-1}SW \otimes \mathbf{I}) \begin{bmatrix} \tau_{N1}H\mathbf{e}_N\bar{w}_1 \\ \tau_{N2}g\mathbf{e}_N\bar{w}_2 \end{bmatrix}, \quad U < 0. \quad (29b)$$

Here  $S$  is the eigenvector matrix given in (13) and  $W$  is the weight matrix defined in (9). The coefficients  $\tau_{01}$ ,  $\tau_{02}$ ,  $\tau_{N1}$ ,  $\tau_{N2}$  are real penalty parameters to be determined by requiring stability. Note that (27) is a consistent semi-discrete approximation of the IBVP (4) for all nontrivial choices of the penalty parameters. The semi-discrete approximation (27), given that the discrete derivative operator satisfies the SBP property (26), is often referred to as the SBP-SAT scheme. We introduce the discrete weighted  $L_2$ -norm

$$\|\bar{\mathbf{q}}\|_{WP}^2 := \bar{\mathbf{q}}^T (W \otimes P) \bar{\mathbf{q}} \geq 0.$$

We say that the semi-discrete approximation (27) is stable if the discrete energy norm  $\|\bar{\mathbf{q}}\|_{WP}^2$  is non-increasing with time. We will now prove the stability of the semi-discrete approximation (27) for sub-critical flows.

**Theorem 9.** *Consider the semi-discrete finite volume approximation (27) with the SAT (28) for sub-critical flow regimes where  $\lambda_1 > 0$ ,  $\lambda_2 < 0$  and  $\gamma_0^2 \leq -\lambda_2/\lambda_1$ ,  $\gamma_1^2 \leq -\lambda_1/\lambda_2$ . If the penalty parameters are chosen such that  $\tau_{01} = \lambda_1$ ,  $\tau_{02} = \gamma_0\lambda_1$ ;  $\tau_{N2} = -\lambda_2$ ,  $\tau_{N1} = -\gamma_1\lambda_2$ , then*

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 \leq 0, \quad \forall t \geq 0.$$

The next theorem will prove the stability of the semi-discrete approximation (27) for super-critical flows.

**Theorem 10.** *Consider the semi-discrete finite volume approximation (27) with the SAT (29) for super-critical flows. If the penalty parameters are chosen such that  $\tau_{01} \geq \lambda_1$ ,  $\tau_{02} \geq \lambda_2$ ;  $\tau_{N1} \geq -\lambda_1$ ,  $\tau_{N2} \geq -\lambda_2$ , then*

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 \leq 0, \quad \forall t \geq 0.$$

Finally, we will prove the stability of the semi-discrete approximation (27) for critical flows.

**Theorem 11.** *Consider the semi-discrete finite volume approximation (27) with the SAT (29) for critical flows. If the penalty parameters are chosen such that  $\tau_{01} \geq \lambda_1$ ,  $\tau_{02} = 0$ ;  $\tau_{N1} = 0$ ,  $\tau_{N2} \geq -\lambda_2$ , then*

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 \leq 0, \quad \forall t \geq 0.$$

## 4 Numerical Experiments

Here, we will perform some numerical experiments to verify the analysis performed in the previous sections. There are two numerical experiments conducted. In both cases we used  $U = 0.1$  m/s,  $H = 0.7$  m, and  $g = 9.8$  m/s. The flow is sub-critical, that is  $U < \sqrt{Hg}$ . We set transmissive boundary conditions both on the inflow and outflow boundaries,

$$\sqrt{\frac{g}{H}}h + u = 0, \quad \text{at } x = 0, \quad \sqrt{\frac{g}{H}}h - u = 0, \quad \text{at } x = 1.$$

The corresponding reflection coefficients  $\gamma_0$  and  $\gamma_1$  satisfy the conditions for well-posedness.

The semi-discrete system (27) is integrated in time using the classical fourth order explicit Runge-Kutta method with the explicit time step

$$\Delta t = \text{CFL} \frac{\Delta x}{U + \sqrt{gH}}, \quad \text{CFL} = 0.25$$

where  $\Delta x = 1/(N - 1)$  and  $N$  is the number of nodes.

**Gaussian profile.** This test aims at verifying the numerical implementation of the transmissive boundary conditions. The initial Gaussian profile is considered for the water height with the zero initial condition for the velocity,

$$h(x, 0) = e^{-\frac{(x-0.5)^2}{0.01}}, \quad u(x, 0) = 0. \quad (30)$$

We set the final time  $T = 2\pi$  so that waves leave the interval of interest. The snapshots of the solutions are displayed in Figure 2, showing the evolution of the wave with time. At  $t = 0.74$ s the right-going solution has left the domain and  $t = 2.96$ s the left-going wave leaves the domain, without reflections.

**Convergence test.** Here, we verify the convergence properties of the numerical method. We will use the method of manufactured solution [9]. That is we force the system to have the exact smooth solution

$$h(x, t) = \cos(2\pi t) \sin(6\pi x), \quad u(x, t) = \sin(2\pi t) \cos(4\pi x). \quad (31)$$

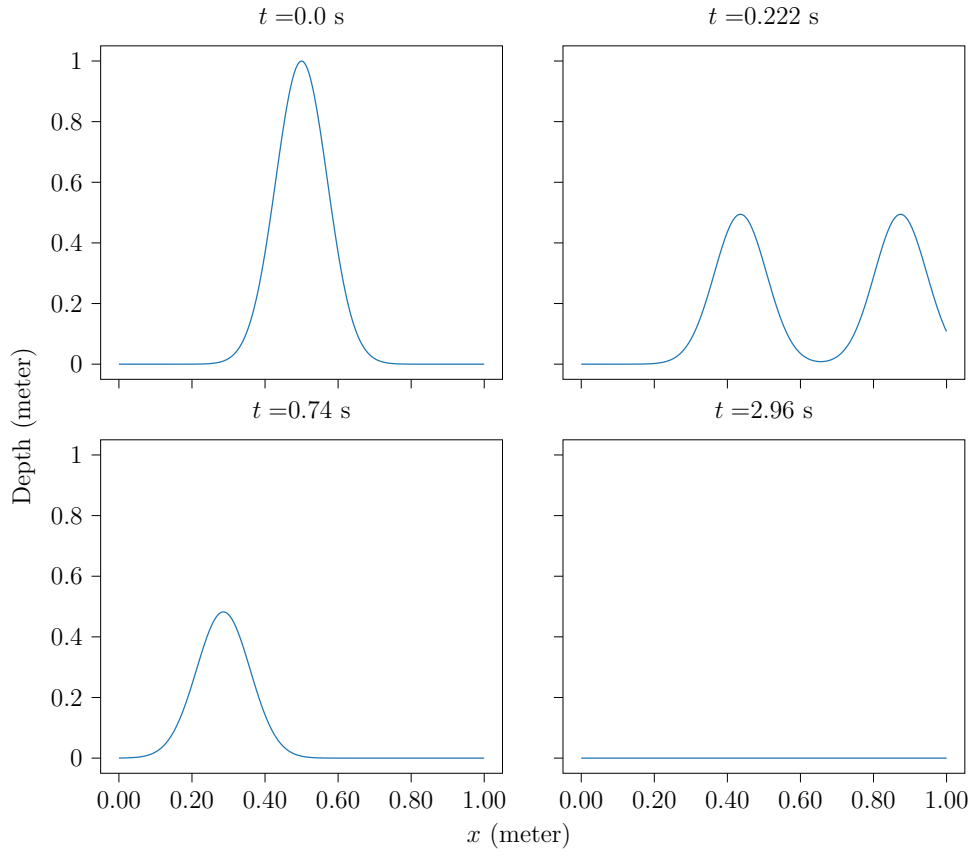


Figure 2: Subcritical flow with transmissive boundary conditions.

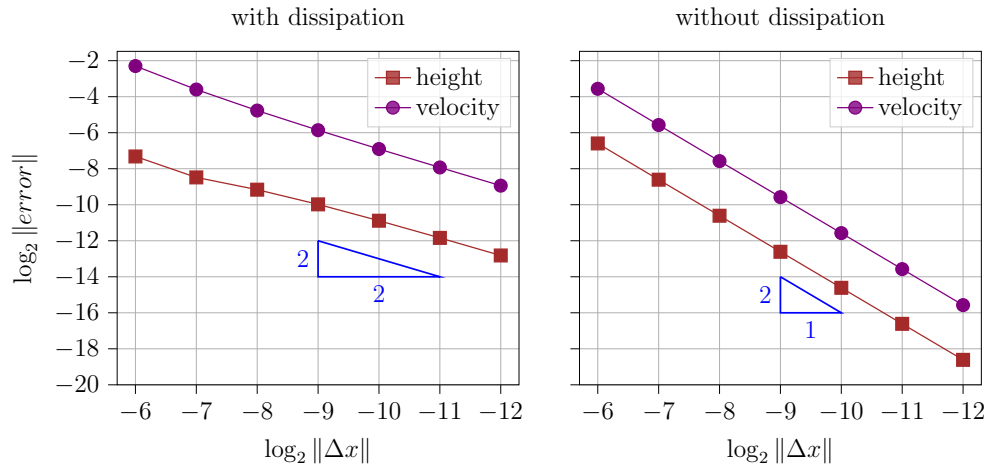


Figure 3: Error at different resolutions and convergence rate of the error.

We compute the numerical solution on a sequence of increasing number of nodes,  $N = 65, 129, 257, 513, 1025, 2049, 4097$ . The  $L_2$ -error and convergence rates of the error are shown in Figure 3. We have performed numerical experiments with no dissipation  $\alpha = 0$  and with numerical dissipation turned on  $\alpha > 0$ . From Figure 3 the method is first order accurate  $O(\Delta x)$  when  $\alpha > 0$  and second order accurate  $O(\Delta x^2)$  when  $\alpha = 0$ . These are in agreement with the theory.

## 5 Conclusion

We derived and analysed well-posed boundary conditions for the linear SWWE. Our analysis is based on the energy method and prescribes the number, location and form of the boundary conditions so that the IBVP is well-posed. We propose a finite volume method formulated in SBP framework and implement boundary conditions weakly using SAT. We derive stable penalty parameters and prove numerical stability by deriving discrete energy estimate analogous to the continuous estimate. Our continuous and numerical analysis covers all flow regimes, and can be extended to the nonlinear problem. Numerical experiments are performed to verify the analysis. The next step in our study will extend the 1D theory and results to 2D, and implement our scheme in open source 2D model [10, 8] for efficient and accurate simulations of the nonlinear shallow water equations.


**Acknowledgements** This research is conducted as part of doctoral study funded by Indonesian Endowment Fund for Education (LPDP).

## A Appendix

**Proof:** Lemma 4:

We consider  $(\mathbf{q}, \mathbf{Dq})_W$  and use integration-by-parts, we have

$$(\mathbf{q}, \mathbf{Dq})_W = - \int_{\Omega} \mathbf{q}^\top \widetilde{M} \frac{\partial \mathbf{q}}{\partial x} dx = - \frac{1}{2} \int_{\Omega} \frac{\partial}{\partial x} \left( \mathbf{q}^\top \widetilde{M} \mathbf{q} \right) dx = - \frac{1}{2} \left( \mathbf{q}^\top \widetilde{M} \mathbf{q} \right) \Big|_0^1.$$

Ignoring the boundary term  $\left( \mathbf{q}^\top \widetilde{M} \mathbf{q} \right) \Big|_0^1$  gives  $(\mathbf{q}, \mathbf{Dq})_W = 0$ , which satisfies Definition 2 with  $\nu = 0$ . 

**Proof:** Lemma 5:

Let  $w_1 = \gamma_0 w_2$  at  $x = 0$  and  $w_2 = \gamma_1 w_1$  at  $x = 1$ , and consider

$$(\lambda_1 w_1^2 + \lambda_2 w_2^2)|_0 - (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_1 = w_2^2 (\lambda_1 \gamma_0^2 + \lambda_2)|_0 - w_1^2 (\lambda_1 + \lambda_2 \gamma_1^2)|_1.$$

Thus if  $\gamma_0^2 \leq -\lambda_2/\lambda_1$  and  $\gamma_1^2 \leq -\lambda_1/\lambda_2$  then  $(\lambda_1 \gamma_0^2 + \lambda_2) \leq 0$  and  $(\lambda_1 + \lambda_2 \gamma_1^2) \geq 0$ , and we have

$$BT = \frac{1}{2} (w_2^2 (\lambda_1 \gamma_0^2 + \lambda_2)|_0 - w_1^2 (\lambda_1 + \lambda_2 \gamma_1^2)|_1) \leq 0.$$

♠

**Proof:** Lemma 6:

Let  $U > 0$  with  $\lambda_1 > 0$ ,  $\lambda_2 > 0$  if  $w_1 = 0$ ,  $w_2 = 0$ , at  $x = 0$ , then

$$BT = \frac{1}{2} ((\lambda_1 w_1^2 + \lambda_2 w_2^2)|_0 - (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_1) = -\frac{1}{2} (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_1 \leq 0.$$

If  $U < 0$  with  $\lambda_1 < 0$ ,  $\lambda_2 < 0$  and  $w_1 = 0$ ,  $w_2 = 0$ , at  $x = 1$ , then we have

$$BT = \frac{1}{2} ((\lambda_1 w_1^2 + \lambda_2 w_2^2)|_0 - (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_1) = \frac{1}{2} (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_0 \leq 0.$$

♠

**Proof:** Lemma 7:

Let  $U > 0$  with  $\lambda_1 > 0$ ,  $\lambda_2 = 0$  if  $w_1 = 0$ , at  $x = 0$ ,

$$BT = \frac{1}{2} ((\lambda_1 w_1^2 + \lambda_2 w_2^2)|_0 - (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_1) = -\frac{1}{2} \lambda_1 w_1^2|_1 \leq 0.$$

If  $U < 0$  with  $\lambda_1 = 0$ ,  $\lambda_2 < 0$  and  $w_2 = 0$ , at  $x = 1$  we also have

$$BT = \frac{1}{2} ((\lambda_1 w_1^2 + \lambda_2 w_2^2)|_0 - (\lambda_1 w_1^2 + \lambda_2 w_2^2)|_1) = \frac{1}{2} \lambda_2 w_2^2|_0 \leq 0.$$

♠

**Proof:** Theorem 8:

We use the energy method, that is, from the left we multiply (4a) with  $\mathbf{q}^\top W$  and integrate over the domain. As above integration-by-parts gives

$$\frac{1}{2} \frac{d}{dt} \|\mathbf{q}\|_W^2 = \left( \mathbf{q}, \frac{\partial \mathbf{q}}{\partial t} \right)_W = (\mathbf{q}, \mathbf{D}\mathbf{q})_W = gH \times BT.$$

Using Lemmas 5–7 for each flow regime gives  $BT \leq 0$ , which completes the proof.

♠

**Proof:** Theorem 9:

We use the energy method, that is from the left, we multiply (27) with  $\bar{\mathbf{q}}^T (W \otimes \mathbf{I})$  and add the transpose of the product, we have

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 + \frac{1}{2} \bar{\mathbf{q}}^T \left( \widetilde{M} \otimes (Q + Q^T) \right) \bar{\mathbf{q}} - \frac{\alpha}{2} \bar{\mathbf{q}}^T (W \otimes A) \bar{\mathbf{q}} = \bar{\mathbf{q}}^T (W \otimes \mathbf{I}) \text{SAT}.$$

Using the SBP property (26) and the eigen-decomposition of  $\widetilde{M}$  we have

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 - \frac{\alpha}{2} \bar{\mathbf{q}}^T (W \otimes A) \bar{\mathbf{q}} = \frac{1}{2} gH \times \text{BT}_{num},$$

where

$$\begin{aligned} \text{BT}_{num} = & \left( \lambda_1 \bar{w}_1^2 + \lambda_2 \bar{w}_2^2 - (\tau_{01} \bar{w}_1 (\bar{w}_1 - \gamma_0 \bar{w}_2) + \tau_{02} \bar{w}_2 (\bar{w}_1 - \gamma_0 \bar{w}_2)) \right) \Big|_{i=0} \\ & - \left( \lambda_1 \bar{w}_1^2 + \lambda_2 \bar{w}_2^2 + (\tau_{N1} \bar{w}_1 (\bar{w}_2 - \gamma_1 \bar{w}_1) + \tau_{N2} \bar{w}_2 (\bar{w}_2 - \gamma_1 \bar{w}_1)) \right) \Big|_{i=N}. \end{aligned}$$

Thus, if  $\tau_{01} = \lambda_1$ ,  $\tau_{02} = \gamma_0 \lambda_1$ ;  $\tau_{N2} = -\lambda_2$ ,  $\tau_{N1} = -\gamma_1 \lambda_2$ , then we have

$$\text{BT}_{num} = (\lambda_2 + \lambda_1 \gamma_0^2) \bar{w}_2^2 \Big|_{i=0} - (\lambda_1 + \lambda_2 \gamma_1^2) \bar{w}_1^2 \Big|_{i=N}.$$

Since  $\lambda_1 > 0$ ,  $\lambda_2 < 0$  and

$$(\lambda_2 + \lambda_1 \gamma_0^2) \leq 0 \iff \gamma_0^2 \leq -\lambda_2 / \lambda_1; \quad (\lambda_1 + \lambda_2 \gamma_1^2) \geq 0 \iff \gamma_1^2 \leq -\lambda_1 / \lambda_2,$$

then we must have  $\text{BT}_{num} \leq 0$ , and

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 = \frac{\alpha}{2} \bar{\mathbf{q}}^T (W \otimes A) \bar{\mathbf{q}} + \frac{1}{2} gH \times \text{BT}_{num} \leq 0.$$

♠

**Proof:** Theorem 10:

As above the energy method with the SBP property (26) and the eigen-decomposition of  $\widetilde{M}$  yields

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 - \frac{\alpha}{2} \bar{\mathbf{q}}^T (W \otimes A) \bar{\mathbf{q}} = \frac{1}{2} gH \times \text{BT}_{num},$$

where

$$\begin{aligned} \text{BT}_{num} &= ((\lambda_1 - \tau_{01}) \bar{w}_1^2 + (\lambda_2 - \tau_{02}) \bar{w}_2^2) \Big|_{i=0} - (\lambda_1 \bar{w}_1^2 + \lambda_2 \bar{w}_2^2) \Big|_{i=N}, \quad U > 0, \\ \text{BT}_{num} &= (\lambda_1 \bar{w}_1^2 + \lambda_2 \bar{w}_2^2) \Big|_{i=0} - ((\lambda_1 + \tau_{N1}) \bar{w}_1^2 + (\lambda_2 + \tau_{N2}) \bar{w}_2^2) \Big|_{i=N}, \quad U < 0. \end{aligned}$$

Therefore if  $\tau_{01} \geq \lambda_1$ ,  $\tau_{02} \geq \lambda_2$ ;  $\tau_{N1} \geq -\lambda_1$ ,  $\tau_{N2} \geq -\lambda_2$ , then we have  $\text{BT}_{num} \leq 0$  and

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 = \frac{\alpha}{2} \bar{\mathbf{q}}^T (W \otimes A) \bar{\mathbf{q}} + \frac{1}{2} gH \times \text{BT}_{num} \leq 0.$$

♠

**Proof:** Theorem 11:

The zero penalties ensure consistency of the SAT, that is  $\tau_{02} = 0$  and  $\tau_{N1} = 0$  give

$$\begin{aligned} \text{SAT} &= -\frac{1}{2} (W^{-1}SW \otimes \mathbf{I}) \begin{bmatrix} \tau_{01} H \mathbf{e}_0 \bar{w}_1 \\ 0 \end{bmatrix}, & U > 0, \\ \text{SAT} &= -\frac{1}{2} (W^{-1}SW \otimes \mathbf{I}) \begin{bmatrix} 0 \\ \tau_{N2} g \mathbf{e}_N \bar{w}_2 \end{bmatrix}, & U < 0. \end{aligned}$$

Again the energy method with the SBP property (26) and the eigen-decomposition of  $\widetilde{M}$  yield

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 - \frac{\alpha}{2} \bar{\mathbf{q}}^T (W \otimes A) \bar{\mathbf{q}} = \frac{1}{2} gH \times \text{BT}_{num},$$

where

$$\begin{aligned} \text{BT}_{num} &= (\lambda_1 - \tau_{01}) \bar{w}_1^2|_{i=0} - \lambda_1 \bar{w}_1^2|_{i=N}, & U > 0, \lambda_1 > 0, \lambda_2 = 0 \\ \text{BT}_{num} &= \lambda_2 \bar{w}_2^2|_{i=0} - (\lambda_2 + \tau_{N2}) \bar{w}_2^2|_{i=N}, & U < 0, \lambda_1 = 0, \lambda_2 < 0. \end{aligned}$$

Therefore if  $\tau_{01} \geq \lambda_1$ ,  $\tau_{N2} \geq -\lambda_2$ , then we have  $\text{BT}_{num} \leq 0$  and

$$\frac{1}{2} \frac{d}{dt} \|\bar{\mathbf{q}}\|_{WP}^2 = \frac{\alpha}{2} \bar{\mathbf{q}}^T (W \otimes A) \bar{\mathbf{q}} + \frac{1}{2} gH \times \text{BT}_{num} \leq 0.$$



## References

- [1] Mark H Carpenter, David Gottlieb, and Saul Abarbanel. “Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: methodology and application to high-order compact schemes”. In: *Journal of Computational Physics* 111.2 (1994), pp. 220–236 (cit. on p. 2).
- [2] JA Cunge, FM Holly, and A Verwey. “Practical Aspects of Computational River Hydraulics, Pitman Adv”. In: *Pub. Program* (1980) (cit. on p. 2).
- [3] Sarmad Ghader and Jan Nordström. “Revisiting well-posed boundary conditions for the shallow water equations”. In: *Dynamics of Atmospheres and Oceans* 66 (2014), pp. 1–9. ISSN: 0377-0265. DOI: <https://doi.org/10.1016/j.dynatmoce.2014.01.002>. (Cit. on p. 2).

- [4] Bertil Gustafsson. *High order difference methods for time dependent PDE*. Vol. 38. Springer Science & Business Media, 2007 (cit. on pp. 2, 8).
- [5] Bertil Gustafsson, Heinz-Otto Kreiss, and Joseph Oliger. *Time dependent problems and difference methods*. Vol. 24. John Wiley & Sons, 1995 (cit. on pp. 2, 4).
- [6] H-O Kreiss and G Scherer. “Finite element and finite difference methods for hyperbolic partial differential equations”. In: *Mathematical aspects of finite elements in partial differential equations*. Elsevier, 1974, pp. 195–212 (cit. on pp. 2, 8).
- [7] Khalid Mahmood, Vujica M Yevjevich, and William Albert Miller. *Unsteady flow in open channels*. Vol. 2. Water Resources Publications, 1975 (cit. on p. 2).
- [8] O Nielsen et al. “Hydrodynamic modelling of coastal inundation”. In: *MODSIM 2005 International Congress on Modelling and Simulation* (2005), pp. 518–523 (cit. on p. 12).
- [9] Patrick J. Roache. “Code Verification by the Method of Manufactured Solutions”. In: *Journal of Fluids Engineering* 124.1 (Nov. 12, 2001), pp. 4–10. ISSN: 0098-2202. DOI: 10.1115/1.1436090. (Cit. on p. 10).
- [10] Stephen Roberts, Gareth Davies, and Ole Nielsen. *ANUGA Github Repository*. Version 3.1.9. June 2022. URL: [https://github.com/anuga-community/anuga\\_core](https://github.com/anuga-community/anuga_core) (cit. on p. 12).
- [11] Andrew R. Winters and Gregor J. Gassner. “A comparison of two entropy stable discontinuous Galerkin spectral element approximations for the shallow water equations with non-constant topography”. In: *Journal of Computational Physics* 301 (2015), pp. 357–376. ISSN: 0021-9991. DOI: <https://doi.org/10.1016/j.jcp.2015.08.034>. (Cit. on p. 2).
- [12] Christopher Zoppou and Stephen Roberts. “Explicit schemes for dam-break simulations”. In: *Journal of Hydraulic Engineering* 129.1 (2003), pp. 11–34 (cit. on p. 2).

## Author addresses

1. **Rudi Prihandoko**, Mathematical Science Institute, Australian National University, Australian Capital Territory 2600, AUSTRALIA.  
<mailto:u7107271@anu.edu.au>  
[orcid: '0000-0001-6376-7952'](https://orcid.org/0000-0001-6376-7952)



2. **Kenneth Duru**, Mathematical Science Institute, Australian National University, Australian Capital Territory 2600, AUSTRALIA.
3. **Stephen Roberts**, Mathematical Science Institute, Australian National University, Australian Capital Territory 2600, AUSTRALIA.
4. **Christopher Zoppou**, Mathematical Science Institute, Australian National University, Australian Capital Territory 2600, AUSTRALIA.