Combining a central scheme with the subtraction method for the shallow water equations

RONY TOUMA*
Computer Science and Mathematics, Lebanese American University, Byblos, Lebanon
rony.touma@lau.edu.lb

ELISSA MALAEB
Computer Science and Mathematics, Lebanese American University, Byblos, Lebanon
elissa.malaeb@lau.edu.lb

CHRISTIAN KLINGENBERG
Mathematics, University of Wuezburg, Wuezburg, Germany
klingen@mathematik.uni-wuerzburg.de

We present a new numerical scheme that is a well-balanced and second-order accurate for systems of shallow water equations with variable bathymetry. We extend in this paper the subtraction method (resulting in well-balancing) to the case of unstaggered central finite volume methods that computes the numerical solution on a single grid. In addition, the proposed scheme avoids solving Riemann problems occurring at cell boundaries as it employs intermediately a layer of ghost-staggered cells. The proposed numerical scheme is then implemented and validated. We successfully manage to solve classical SWE problems from the literature featuring steady states and other equilibria. The results of the study are consistent with previous research, which supports the use of the proposed method to solve shallow water equations.

Keywords: Shallow Water Equations, Subtraction method, Unstaggered central methods, Well-Balanced discretizations.

1. Introduction

The Shallow Water Equations are a set of partial differential equations describing the flow of an incompressible shallow layer of fluid with a free surface over a bottom topography under gravity. The bottom topography can either be flat or variable. Shallow water equations are only admissible when the flow being studied is that of a shallow layer of fluid, i.e. when the fluid depth is much smaller than the wavelength of the waves being studied. They can model many natural phenomena such as tsunamis, tides, ocean currents, river floods, etc.

Numerous numerical techniques have been employed to tackle shallow water problems with flat and varying bottom topographies. For instance, numerical techniques based on Riemann solvers were developed in [Greenberg and LeRoux (1996); Kim (2003)], and

*Lebanese American University, Byblos, Lebanon.
recently, in [Aleksyuk et al. (2022)], an exact Riemann solver for the SWE’s was developed. On another hand, well-balanced schemes were established in [Touma (2009); Touma and Khankan (2012); Botta et al. (2004); Greenberg and LeRoux (1996)]. Additionally, authors in [Touma (2016); Audusse et al. (2004)] addressed the drying of domains, such as Audusse et al. [Audusse et al. (2004)] who used a hydrostatic reconstruction to obtain a well-balanced scheme that handles dry states ($h = 0$). There are also [Bryson et al. (2011)] whose authors developed central upwind schemes, as well as [Shu and Xing (2005); Lu and Qiu (2011); Xing and Shu (2006); Noelle et al. (2007)] whose authors used finite difference and finite volume WENO schemes to tackle such problems. In [Guo et al. (2022)], the authors developed a new well-balanced finite volume CWENO scheme for the shallow water equations. Also, the authors of [Ciallella et al. (2022)] demonstrated a well-balanced finite volume WENO method combined with the modified Patankar Deferred Correction time integration method. Schemes based on relaxation models were presented in [Delis and Katsaounis (2005); Delis and Katsaounis (2003)]. And recently, in [Gaburro et al. (2018); Desveaux et al. (2016); Castro and Pares (2020); Arpaia and Ricchiuto (2020); Busto and Dumbser (2022)], well-balanced schemes to solve the shallow water equations were presented. In [Kent et al. (2023)], the authors used an iterated semi-implicit time-stepping scheme along with a finite-volume transport method, and the cubed sphere grid, hence a mixed finite element discretization of the SWE’s. Whereas in [Kaptsov et al. (2022)], the authors used conservative invariant finite difference schemes for the modified shallow water equations in Lagrangian coordinates.

On another hand, many schemes and methods have been employed recently on solving nonlinear PDE’s, and due to which many useful results have been found such as in [Gao et al. (2023); Chen et al. (2023a); Zhao et al. (2022); Chen et al. (2023b); Liu et al. (2023); Liu et al. (2023); Yin and Xing (2023); Yin et al. (2022); Yin et al. (2021); Lu et al. (2021); Lu and Chen (2021); Dong (2023); Xin et al. (2021); Buachart et al. (2014); Sepehrirahnama et al. (2020)].

As previously stated in [Touma and Khankan (2012)], the NT-scheme developed in [Nessyahu and Tadmor (1990)] is a non-oscillatory central scheme that approximates solutions to hyperbolic systems such as SWE systems. Its requirement for two staggered grids is a disadvantage, though. Jiang et al. [Jiang et al. (1998)] created the first unstaggered outlook on NT-schemes. We can therefore think of the UCS (unstaggered central schemes) developed in [Touma (2009)] as being inspired by the previously mentioned schemes, with the advantages of avoiding Riemann solvers and staggered grids, working with only one original grid, a staggered ghost grid, and projecting back the numerical solution onto the original grid.

In this paper, we present a new numerical method for solving shallow water equation (SWE) problems in one dimension. It is second-order accurate, well-balanced, unstaggered, and uses the subtraction method. Its extra advantage is the ability to apprehend all types of steady states of the system at the discrete level.

The new method uses the same approach as the reconstruction technique previously employed in [Kanbar et al. (2020)]. Instead of resolving for the unknown solution $U$, here, we will be evolving the error function $\Delta U$ between $U$ and a given steady state of the system $\bar{U}$, such that $\Delta U = U - \bar{U}$. The steady state we will use is the lake at rest steady state.

Our paper will be sorted as follows. In section 2, we will present our UCS with the subtraction method. In section 3, several common problems from the literature will be solved using the well-balanced UCS method presented in [Touma and Khankan (2012)] to compare their results to results obtained using UCS with subtraction. Some problems will also be compared to different methods found in the literature. And finally, we will have our concluding remarks in section 4.
2. UCS with subtraction method

The shallow water equations system in one space dimension writes as:

\[
\begin{align*}
\frac{\partial}{\partial t} U + \frac{\partial}{\partial x} f(U) &= S(U), \\
U(x,0) &= U_0(x) 
\end{align*}
\]  

(1)

where \( x \in \mathcal{D} \subset (-\infty, \infty] \) and \( t \in \mathbb{R}^* \). The unknown vector solution, the flux function and the source term in (1) are, respectively, as follows

\[
U = \begin{pmatrix} h \\ hu \end{pmatrix}, \quad f(U) = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \end{pmatrix}, \quad S(U) = \begin{pmatrix} 0 \\ -gh \frac{db}{dx} \end{pmatrix}. 
\]  

(2)

In system (1)-(2), \( h(x,t), u(x,t), \) and \( b(x) \) denote the height of the water column, the water velocity, and the topography (waterbed) function respectively. The constant \( g = 9.812 \) denotes the gravitational constant.

The derivation involves two main physical principles, mainly the conservation of mass and the conservation of momentum.

By the definition of shallow waters, we assume that the wavelength is much bigger than its average height. This allows us to assume that the water velocity is only horizontal, where the vertical component of the velocity vector can be assumed to be zero. We can also assume that the horizontal component is the same throughout the fluid depth. Another assumption to be taken is that the density of the fluid \( \rho \) is constant. The fluid to be studied is in-compressible.

Finally, we can write the shallow water equations as follows:

\[
\begin{align*}
\left( \begin{array}{c} h \\ hu \end{array} \right)_t + \left( \begin{array}{c} hu \\ hu^2 + \frac{1}{2}gh^2 \end{array} \right)_x &= \left( \begin{array}{c} 0 \\ -gh \frac{db}{dx} \end{array} \right). 
\end{align*}
\]  

(3)

Note that this system becomes homogeneous if we have a flat bottom topography, \( b(x) = \text{constant} \).

Furthermore, the resulting homogeneous system of equations is hyperbolic, and the eigenvalues and eigenvectors are real and linearly independent.

We consider system (1)-(2) explicitly shown in system (3), and we assume that \( \tilde{U} \) is the lake at rest stationary solution of system (1). Then, the error between the unknown solution \( U \) and this steady state \( \tilde{U} \) is \( \Delta U = U - \tilde{U} \), and so \( U = \Delta U + \tilde{U} \).

Since \( \tilde{U} \) is a stationary solution, then

\[
\tilde{U}_t = 0 \quad \text{and so} \quad f(\tilde{U})_x = S(\tilde{U}).
\]  

(4)

Now, substitute \( U \) with \( \Delta U + \tilde{U} \) in system (1) along with the fact that \( \tilde{U}_t = 0, \)

\[
\Delta U_t + f(\Delta U + \tilde{U})_x = S(\Delta U + \tilde{U}).
\]  

(5)

Performing a term by term subtraction in equations (4) and (5), we obtain

\[
\Delta U_t + [f(\Delta U + \tilde{U}) - f(\tilde{U})]_x = S(\Delta U + \tilde{U}) - S(\tilde{U}).
\]

The right-hand side simplifies as follows,

\[
S(\Delta U + \tilde{U}) - S(\tilde{U}) = \left( \begin{array}{c} 0 \\ -g \Delta h \frac{db}{dx} \end{array} \right) - \left( \begin{array}{c} 0 \\ -g \tilde{h} \frac{db}{dx} \end{array} \right) = \left( \begin{array}{c} 0 \\ -g \Delta h \frac{db}{dx} \end{array} \right).
\]

Accordingly, instead of solving system (1) now, we will be solving the following initial value problem:

\[
\begin{align*}
\partial_t \Delta U + \partial_x \left[ f(\Delta U + \tilde{U}) - f(\tilde{U}) \right] &= S(\Delta U), \quad x \in \mathcal{D} \subset \mathbb{R}, t > 0 \\
\Delta U(x,0) &= \Delta U_0(x).
\end{align*}
\]  

(6)
We divide the computational domain $\mathcal{D}$ in system (6) into control cells $C_i = [x_{i-1/2}, x_{i+1/2}]$ and staggered dual cells $D_{i+1/2} = [x_i, x_{i+1}]$. The control cells are centered at the nodes $x_i$ and have length $\Delta x$. The staggered dual cells are centered at the nodes $x_{i+1/2} = x_i + \frac{\Delta x}{2}$ and also have length $\Delta x$. The time step is labeled as $\Delta t$ and is used to increment the time variable following the rule $t^{n+1} = t^n + \Delta t$. $\Delta t$ is obtained dynamically to ensure the stability of the numerical schemes.

Given the numerical solution $U^n_i$, an approximation to the exact solution $U(x_i, t^n)$, of system (1), and knowing a priori a steady state $\tilde{U}$ of system (1), we define the error $\Delta U^n_i = U^n_i - \tilde{U}_i$ as a piecewise linear function at time $t^n$ over the cell $C_i$. Thus, the exact solution $\Delta U(x_i, t^n)$ will be approximated by $\Delta U^n_i$. Our goal is then to generate the values of the numerical solution $\Delta U^{n+1}_i$ at $t^{n+1}$ on the control cells $C_i$.

We define the piecewise linear function $P_i(x, t)$ approximating the exact unknown error function $\Delta U(x, t)$ on the cells $C_i$:

$$\Delta U^n_i = \frac{1}{\Delta x} \int_{C_i} P_i(x, t) \, dx \approx \frac{1}{\Delta x} \int_{C_i} \Delta U_1(x, t) \, dx,$$

and then,

$$P_i(x, t^n) = \Delta U^n_i + (x - x_i) (\Delta U^n_i)', \quad \forall x \in C_i$$

where $(\Delta U^n_i)'$ is an estimate to the partial derivative $\frac{\partial \Delta U}{\partial x}(x_i, t^n)$ obtained by applying a limiting procedure of the numerical derivatives with the aid of the MC-limiter and given by:

$$(\Delta U^n_i)' = \text{MinMod} \left[ \frac{\Delta U^n_i - \Delta U^n_{i-1}}{\Delta x}, \frac{\Delta U^n_{i+1} - \Delta U^n_{i-1}}{2\Delta x}, \theta \frac{\Delta U_{i+1} - \Delta U^n_i}{\Delta x} \right],$$

where $\theta \in (1, 2)$, and the MinMod function is defined in [Touma (2009)].

From system (6), we get the following balance law:

$$\Delta U_i + \left[ f \left( \Delta U + \tilde{U} \right) - f \left( \tilde{U} \right) \right]_x = S(\Delta U). \quad (7)$$
Now, we integrate the balance law (7) over the domain \( R_{i+1/2}^{n+1} = [x_i, x_{i+1}] \times [t^n, t^{n+1}] \):
\[
\int_{R_{i+1/2}^{n+1}} \Delta U_t + \left[ f \left( \Delta U + \hat{U} \right) - f \left( \hat{U} \right) \right] dt = \int_{R_{i+1/2}^{n+1}} \mathbf{S} \left( \Delta U \right) \, dR.
\]
We apply Green’s formula to the double integral on the left-hand side. After a few simplifications, fixing and rearranging the integrals, we obtain:
\[
\int_{x_i}^{x_{i+1}} \left( -\Delta U(x, t^n) \right) dx \]
\[
+ \int_{t^n}^{t^{n+1}} \left[ f \left( (\Delta U + \hat{U})_i, t \right) - f \left( \hat{U}_i, t \right) \right] dt
\]
\[
+ \int_{x_i}^{x_{i+1}} \left( -\Delta U(x, t^{n+1}) \right) dx + \int_{t^n}^{t^{n+1}} \left[ f \left( (\Delta U + \hat{U})_i, t \right) - f \left( \hat{U}_i, t \right) \right] dt = \int_{t^n}^{t^{n+1}} \int_{x_i}^{x_{i+1}} \mathbf{S} \left( \Delta U \right) \, dxdt. \tag{8}
\]
We estimate the integrals of the error function in the above equation by following the mean value theorem for the integral with respect to space. This can be indeed performed since \( \Delta U(x, t) \) is supposed to be a piecewise linear function constructed at the center of the control cells. The integrals become:
\[
\int_{x_i}^{x_{i+1}} \left( -\Delta U(x, t^n) \right) dx = -\Delta x \Delta U_{i+1/2}^n
\]
and
\[
\int_{x_i}^{x_{i+1}} \left( -\Delta U(x, t^{n+1}) \right) dx = \Delta x \Delta U_{i+1/2}^{n+1}. \tag{9}
\]
Next we approximate the integrals of the flux function with 2nd-order accuracy by applying the midpoint quadrature rule. This leads to:
\[
\int_{t^n}^{t^{n+1}} \left[ f \left( (\Delta U + \hat{U})_i, t \right) - f \left( \hat{U}_i, t \right) \right] dt 
\approx \Delta t \left[ f \left( (\Delta U + \hat{U})_i, t^{n+1/2} \right) - f \left( \hat{U}_i, t^{n+1/2} \right) \right]
\]
and
\[
\int_{t^n}^{t^{n+1}} \left[ f \left( (\Delta U + \hat{U})_i, t \right) - f \left( \hat{U}_i, t \right) \right] dt 
\approx -\Delta t \left[ f \left( (\Delta U + \hat{U})_i, t^{n+1/2} \right) - f \left( \hat{U}_i, t^{n+1/2} \right) \right]. \tag{10}
\]
Plugging equations (9) and (10) into equation (8), dividing by \( \Delta x \), and rearranging, we obtain the forward evolution of \( \Delta U_i^n \) onto the staggered grid at time \( t^{n+1} \):
\[
\Delta U_{i+1/2}^{n+1} = \Delta U_{i+1/2}^n - \frac{\Delta t}{\Delta x} \int_{t^n}^{t^{n+1}} f \left( (\Delta U + \hat{U})_{i+1}, t \right) - f \left( \hat{U}_{i+1}, t \right) dt 
\]
\[
- f \left( (\Delta U + \hat{U})_i, t \right) + f \left( \hat{U}_i, t \right) \right] + \frac{\Delta x}{\Delta t} \int_{x_i}^{x_{i+1}} \mathbf{S} \left( \Delta U \right) \, dxdt. \tag{11}
\]
The term \( \Delta U_{i+1/2}^{n+1} \) occurring in the above equation is the forward projected value of \( \Delta U_i^n \) onto the cells of the staggered grid. It is generated by expanding \( \Delta U(x, t^n) \) in a Taylor series in space and then projecting the series onto the staggered grid:
\[
\Delta U_{i+1/2}^{n+1} = \frac{1}{2} \left( \Delta U_i^n + \Delta U_{i+1}^n \right) + \frac{\Delta x}{8} \left( (\Delta U_i^n)' - (\Delta U_{i+1}^n)' \right)
\]
where \((\Delta U^n_i)')\) is the numerical derivative estimating the spatial derivative of \(\Delta U(x_i, t^n)\) calculated using a derivative limiting step.

The predicted values \(\Delta U^{n+1/2}_i\) in equation (11) are calculated at the intermediate time \(t^{n+1/2}\) by expanding \(\Delta U(x_i, t^n)\) in a Taylor series in time and then using the first-order term:

\[
\Delta U \left( x_i, t^{n+1/2} \right) = \Delta U \left( x_i, t^n + \frac{\Delta t}{2} \right) \approx \Delta U \left( x_i, t^n \right) + \frac{\Delta t}{2} \Delta U_i \left( x_i, t^n \right) .
\]

Using the balance law (7), we obtain:

\[
\Delta U^{n+1/2}_i \approx \Delta U^n_i + \frac{\Delta t}{2} (-T_1 + T_2)
\]

where the terms \(T_1\) and \(T_2\) are given by

\[
T_1 = \left[ f \left( \Delta U + \tilde{U} \right) - f \left( \tilde{U} \right) \right] \left|_{(x_i, t^n)} \right.,
\]

\[
T_2 = \left[ S (\Delta U) \right] \left|_{(x_i, t^n)} \right.,
\]

Hence we obtain

\[
\Delta U^{n+1/2}_i = \Delta U^n_i + \frac{\Delta t}{2} \left[ -(f^n)' + f_i + S^n_i \right] \quad (12)
\]

where \((f_i)'\) along with \(f_i\) denote the estimates to the derivatives in space at the point \((x_i, t^n)\) of \(f = f (\Delta U + \tilde{U})\) and \(\tilde{f} = f (\tilde{U})\), respectively.

Note that \((f^n)'\) and \(f_i\) can be calculated using the MC-limiter or by using their respective Jacobians, such that

\[
(f_i)' = \left[ \frac{\partial f}{\partial U} \cdot \frac{\partial U}{\partial x} \right] = J(f) \cdot (U_i)',
\]

Next we use the following discretization of the source term at time \(t^n\) :

\[
S^n_i = S (\Delta U) \left|_{(x_i, t^n)} \right. = \left( 0 \quad -g \Delta h^n_i \frac{\partial b_i}{dx} \right).
\]

On the other hand, we estimate the integral on the right-hand side of equation (11) with \(2^{nd}\)–order accuracy. For that we employ the trapezoidal quadrature rule along with centered differences, and the midpoint quadrature rule to obtain:

\[
\int_{\tau}^{t^{n+1}} \int_{x_i}^{x_{i+1}} S (\Delta U) \, dx \, dt \approx \Delta t \Delta x S \left( \Delta U^{n+1/2}_i, \Delta U^{n+1/2}_{i+1} \right).
\]

The term to the right-hand side of the above equation is given by

\[
S \left( \Delta U^{n+1/2}_i, \Delta U^{n+1/2}_{i+1} \right) = \left( 0 \quad c_2 \right),
\]

with \(c_2\) is given by \(c_2 = -g \left( \frac{b_{i+1} - b_i}{\Delta x} \right) \frac{\Delta h^{n+1/2}_i + \Delta h^{n+1/2}_{i+1}}{2} \).

Finally, the term \(\Delta U^{n+1}_{i+1}\) denotes the backward projected value of the generated solution \(\Delta U^{n+1}_{i+1}\) on the cells of the original mesh. It is calculated by expanding \(\Delta U (x, t^{n+1})\) in a Taylor series in space and then projecting the series onto the original grid

\[
\Delta U^{n+1}_i = \frac{1}{2} \left( \Delta U^{n+1}_{i-1/2} + \Delta U^{n+1}_{i+1/2} \right) + \frac{\Delta x}{8} \left( \left( \Delta U^{n+1}_{i-1/2} \right)' - \left( \Delta U^{n+1}_{i+1/2} \right)' \right) \quad (14)
\]

where \((\Delta U^{n+1}_{i+1/2})'\) is the limited numerical derivative of \(\Delta U^{n+1}_{i+1/2}\) calculated using the MC-limiter.
So, now that we have $\Delta U_{i}^{n+1}$, we can find the solution $U_{i}^{n+1}$ at the cell-center of $C_{i}$ at the next time step $t^{n+1}$ as follows:

$$U_{i}^{n+1} = \Delta U_{i}^{n+1} + \hat{U}_{i}.$$  

We will summarize our scheme steps in order:

- Knowing the numerical solution $U_{n}^{n}$ at time $t^{n}$ and the lake at rest state $\hat{U}_{i}$, we set $\Delta U_{i}^{n} = U_{n}^{n} - \hat{U}_{i}$.

- Find the forward projected values
  $$\Delta U_{i}^{n+1/2} = \frac{1}{2} (\Delta U_{i}^{n} + \Delta U_{i+1}^{n}) + \frac{\Delta x}{8} \left( (\Delta U_{i}^{n})' - (\Delta U_{i+1}^{n})' \right)$$
  where $(\Delta U_{i}^{n})'$ is found using MC-limiter.

- Find the estimated values at the fractional time step $t^{n+1/2}$
  $$\Delta U_{i}^{n+1/2} \triangleq \Delta U_{i}^{n} + \frac{\Delta t}{2} \left[- (f_{i}^{n})' + \hat{f}_{i}^{n} + S_{i}^{n} \right]$$
  where $(f_{i}^{n})'$ and $\hat{f}_{i}^{n}$ denote estimates to the spatial partial derivatives of $f = f \left( \Delta U + \hat{U} \right)$ and $f = f \left( \hat{U} \right)$, respectively.

- Find the solution on the staggered dual cells at time $t^{n+1}$
  $$\Delta U_{i+1/2}^{n+1} = \Delta U_{i+1/2}^{n} - \frac{\Delta t}{\Delta x} \left[ f_{i+1/2}^{n+1} - \hat{f}_{i+1}^{n+1/2} - f_{i}^{n+1/2} + \hat{f}_{i}^{n+1/2} \right] + \Delta t S \left( \Delta U_{i+1/2}^{n+1/2}, \Delta U_{i+1}^{n+1/2} \right)$$
  where
  \begin{enumerate}
  \item\label{1} $f_{i+1/2}^{n+1} = f \left( \Delta U + \hat{U} \right)$
  \item\label{2} $\hat{f} = f \left( \hat{U} \right)$
  \item\label{3} $S \left( \Delta U_{i+1/2}^{n+1/2}, \Delta U_{i+1}^{n+1/2} \right) = \begin{pmatrix} 0 \\ c_{2} \end{pmatrix}$.
  \end{enumerate}

- Find the solution needed on the original grid at time $t^{n+1}$ by back-projecting $\Delta U_{i+1/2}^{n+1}$
  $$\Delta U_{i}^{n+1} = \frac{1}{2} \left( \Delta U_{i+1/2}^{n+1} + \Delta U_{i-1/2}^{n+1} \right) + \frac{\Delta x}{8} \left( (\Delta U_{i+1/2}^{n+1})' - (\Delta U_{i-1/2}^{n+1})' \right)$$
  where $(\Delta U_{i+1/2}^{n+1})'$ is found using the MC-limiter.

Any numerical method used to solve the shallow water equations (SWE) must have the C-property, which means that it must preserve rest. In other words, a lake at rest in the real world must also be at rest in the numerical model. This means that the steady state of a lake at rest must be preserved at the discrete level, i.e. $h + b = \text{constant and } u = 0$. This new scheme we are developing, UCS with subtraction, satisfies the C-property because it preserves all steady states up to machine accuracy.

In fact, if $U_{n}^{n}$ is a stationary solution of system (1), $U_{n}^{n} = \bar{U}_{i}$, i.e. $\Delta U_{i}^{n} = 0$, we want to show that the updated numerical solution $U_{n}^{n+1}$ at time $t^{n+1}$ does not change over time, meaning that it is a stationary solution of the system of equations (1), i.e. $\Delta U_{i}^{n+1} = 0$. To show that, we must show that:
(1) \( \Delta U_{i}^{n+1/2} = 0 \),
(2) \( \Delta U_{i+1/2}^{n+1} = 0 \),
(3) \( \Delta U_{i}^{n+1} = 0 \).

Proof:
(1) Using equation (12) and the fact that \( \Delta U_{i}^{n} = 0 \),
\[
\Delta U_{i}^{n+1/2} = \frac{\Delta t}{2} \left[ -f \left( U_{i}^{n} + 1 \right) - f \left( U_{i}^{n} \right) \right] + S \left( \Delta U_{i}^{n} \right) = 0.
\]

(2) Similarly, using equations (11,13) and the fact that \( \Delta U_{i}^{n+1/2} = 0 \),
\[
\Delta U_{i+1/2}^{n+1} = 0.
\]

(3) Similarly, using equation (14) and the fact that \( \Delta U_{i+1/2}^{n+1} = 0 \),
\[
\Delta U_{i}^{n+1} = 0.
\]

3. Numerical Experiments

The partial differential equation in (7) can be written as:
\[
\Delta U_{i} + F(\Delta U)_{x} = S(\Delta U),
\]
where \( F(\Delta U) = f(\Delta U + 1) - f(\Delta U) \).

It was shown in [Rogers et al. (2003)] that the jacobian matrix \( J(F) = \frac{\partial F}{\partial U} \) is the same as the jacobian matrix \( J(f) = \frac{\partial f}{\partial U} \) corresponding to the pde in system (1); and this will be used in the CFL condition below: In all the following examples, the numerical results shown are made with a CFL=0.485 such that
\[
\Delta t = CFL \times \frac{\Delta x}{\alpha}
\]
where \( \alpha = \max_{i} |\lambda_{i}| \) where \( \lambda_{i} \) are the eigenvalues of the jacobian matrix \( J_{i}(f_{i}) = \frac{\partial f_{i}}{\partial U_{i}} \) over the cell \( C_{i} \):
\[
\lambda_{i} = u_{i} \pm \sqrt{gh_{i}}.
\]

In this section, we will be solving some problems common in the literature. We will mainly compare our computed solution with those previously obtained with a well-balanced unstaggered central scheme presented in [Touma and Khankan (2012)]. We will also compare our solutions to other results presented in the literature.

The steady state \( \bar{U} \) taken is that of a lake at rest:
\[
\bar{U}^{n} = \left( \begin{array}{c}
\bar{H}^{n} \\
\bar{b}^{n} \\
\bar{u}^{n}
\end{array} \right) \quad \text{such that} \quad \bar{H}^{n} + \bar{b}^{n} = \text{constant and} \quad \bar{u}^{n} = 0.
\]

We will use the following abbreviations: ‘IC’ stands for initial condition, ‘WL’ stands for water level \( H = h + b \), \( u_{0} \) stands for the initial water velocity \( u(x, 0) \), \( h_{0} \) stands for initial water height \( h(x, 0) \), \( b \) stands for bottom topography \( b(x) \), \( H_{0} \) stands for initial water level \( H_{0} = h_{0} + b \), \( t_{f} \) stands for final time, \( t_{0} \) stands for initial time, ‘WB-UCS’ stands for the well-balanced unstaggered central scheme, and ‘UCS-Sub’ stands for the unstaggered central scheme with the subtraction method.
3.1. One-dimensional Toro’s problem

We consider a simplified version of Toro’s dam break problem, where the flow is only in the $x$-direction as defined in [Touma and Khankan (2012)], originally from [Toro (2001)].

The computational domain, which is the region of space that is being simulated, is divided into equal-sized segments using 200 grid points. The total length of the computational domain is 40 units.

The riverbed is flat ($b = 0$), $u_0 = 0$, $H_0 = h_0 + b = h_0$ where

\[
h_0 = \begin{cases} 
2.5, & 17.5 < x < 22.5, \\
0.5, & \text{otherwise}. 
\end{cases}
\]

The final time is $t_f = 1.4$.

3.2. Lake at rest test case

We consider this one-dimensional water flow over variable bottom topography problem. The computational domain is the range of $x$-values from 0 to 10. It is divided into 100 grid points. The waterbed function $b = b(x)$ is defined by:

\[
b = \begin{cases} 
0.25, & 0 \leq x < 1, \\
g(x), & 1 \leq x \leq 9, \\
0.30, & 9 < x \leq 10, 
\end{cases}
\]

where

\[
g(x) = \left(10e^{-x^2} + 15e^{-(x-2.5)^2} + 10e^{-(x-5)^2/2} + 6e^{-2(x-7.5)^2} + 16e^{-(x-10)^2}\right)/20.
\]

The initial conditions are $H_0 = 1$ and $u_0 = 0$. The final time is $t_f = 1$. 
Fig. 3. Toro’s problem: Water level obtained using WB-UCS (solid line) and UCS-Sub(dotted line) at different times before the final time.

Fig. 4. Toro’s problem: Water level obtained using WB-UCS (solid line) and UCS-Sub(dotted line) at $t_f = 1.4$. 
Figure (5) shows the water level initially at time $t_0 = 0$. Figure (6) shows the water level obtained at the end of the simulation, at time $t_f = 1$ using WB-UCS (solid line) and UCS-Sub (dotted line). The UCS-Sub method was able to produce a numerical solution that perfectly met the steady state requirement. This is a strong indication that the method is accurate and reliable. The numerical solution is in perfect agreement with the results obtained using the WB-UCS method and with the analytic equation solution.
3.3. **Water flow test case over variable waterbed**

We consider next the water flow test case on variable bathymetry as presented earlier in [Touma (2009)]. The computational domain is set to be the region on the $x$-axis satisfying $|x| \leq 10$ which we mesh using 2000 grid points. The waterbed bathymetry $b = b(x)$ is piecewise described by

$$b = \begin{cases} 
\frac{1}{5} \left(1 - \left(\frac{x}{2}\right)^2\right), & -2 \leq x \leq 2, \\
0, & \text{otherwise}.
\end{cases}$$

The initial water level $H_0 = 1$ and $u_0 = 1$. The final time is $t_f = 20$.

![Figure 7. Water flow over variable bottom topography: Initial water level at $t_0 = 0$.](image1)

![Figure 8. Water Flow over variable bottom topography problem: Water level obtained using WB-UCS (solid line) and UCS-Sub (dotted line) at different times before the final time.](image2)

Figure (7) shows the water level initially at time $t_0 = 0$. Figures (8,9) show the water level obtained at different times and at the final time respectively using WB-UCS (solid line) and UCS-Sub (dotted line) at different times before the final time.
Fig. 9. Water Flow over variable bottom topography problem: Water level obtained using WB-UCS (solid line) and UCS-Sub (dotted line) at $t_f = 20$.

line) and UCS-Sub (dotted line). The perfect alignment of results from the two schemes confirms the robustness and potential of the proposed UCS-Sub.

Fig. 10. Water flow over variable bottom topography: Water level obtained using WB-UCS and UCS-Sub compared with results obtained using the NT+SB method.

Figure (10) shows the water level at the final time obtained using WB-UCS and UCS-Sub in comparison with results obtained using NT+SB; the results were those presented in [Touma (2009)]. The results show a great match between all solutions.

3.4. Dam break flow, no bottom topography

We consider now the one-dimensional dam break problem previously presented in [Delis and Katsaounis (2003)]. Consider a channel of Length $L = 2000$. A dam is located at
At the time $t_0 = 0$, the dam collapses. The goal is to find the solution at time $t_f = 50$. The computational domain is the range of $x$-values from 0 to 2000. It is divided into 101 grid points to get $\Delta x = 20$.

The riverbed is flat ($b = 0$), $u_0 = 0$, $H_0 = h_0 + b = h_0$ where

$$h_0 = \begin{cases} h_1, & x \leq 1000, \\ h_2, & x > 1000, \end{cases}$$

with $h_1 = 10 > h_2$. The solution to this problem features a shock wave propagating downstream and a rarefaction wave propagating upstream. The type of flow changes between subcritical, supercritical, and strongly supercritical according to the water height downstream, as follows:

- if $h_2/h_1 \geq 0.5$: subcritical flow,
- if $h_2/h_1 < 0.5$: subcritical upstream, supercritical downstream,
- if $h_2/h_1 << 0.5$: strongly supercritical downstream.

Figure (11) shows the water level initially at time $t_0 = 0$. Subcritical stands for $h_0/h_1 = 0.5$, Super- critical stands for $h_0/h_1 = 0.05$, and Strongly Supercritical stands for $h_0/h_1 = 0.005$.

Figure (11) shows the water level initially at time $t_0 = 0$. Figure (12) shows the water level obtained at the final time $t_f = 50$ using WB-UCS (solid line) and UCS-Sub (dotted line). The numerical results are consistent across all three cases: subcritical, supercritical, and strongly supercritical.

In figure (13), the initial conditions are adjusted and we compare our results with those shown in [Kim (2003)]. Now, the channel length is $L = 100$ and the dam is located at its center. Upstream water height is $h_1 = 1$ and downstream water height is $h_2 = 0.1$.

Figure (13) shows the water level at time $t_f = 8$. On the left, we show the water level computed by WB-UCS and UCS-Sub as compared with the analytic solution ([Kim (2003)], [Toro (2001)]). On the right, we show the water level computed by WB-UCS and UCS-Sub as compared with the solutions obtained using the HLLC approximate Riemann solver [Kim (2003)]. The precise results reveal excellent agreement across all of the solutions, confirming the potential of our scheme.

3.5. Steady flow over a bump

We consider now steady flow over a bump previously considered in [Xing and Shu (2006)]. This problem aims to study convergence in time toward steady flow over a bump. The
Fig. 12. Dam break flow problem: Subcritical, Supercritical, Strongly Supercritical respectively. Water level obtained using WB-UCS (solid line) and UCS-Sub (dotted line) at $t_f = 50$.

Fig. 13. Dam break flow: Water level obtained using WB-UCS and UCS-Sub compared with analytical solution (left) and HLLC approximate Riemann solver (right).

Numerical solution is computed on the region $x \in [0, 25]$ of the $x$-axis, which we mesh using 201 grid points leading to $\Delta x = 0.125$. The bottom topography function $b = b(x)$
Table 1. $L_1$ error and order of convergence of steady flow over a bump, subcritical case.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$L_1$ error $h$</th>
<th>$L_1$ order</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>1.60e - 03</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>3.85e - 04</td>
<td>2.06</td>
</tr>
<tr>
<td>800</td>
<td>7.73e - 05</td>
<td>2.32</td>
</tr>
</tbody>
</table>

is given by

$$b = \begin{cases} 
\frac{1}{5} - \frac{1}{20} (x - 10)^2, & 8 \leq x \leq 12, \\
0, & \text{otherwise.} 
\end{cases}$$

The initial conditions are $u_0 = 0$ for the initial velocity and $H_0 = H_d$ for the initial water level, where $H_d$ is the water level downstream. The final time is $t_f = 200$. We can get 3 different cases: subcritical, transcritical without shock, and transcritical with shock. They vary with different boundary conditions.

- Subcritical flow: $(hu)_{up} = 4.42, H_d = 2$,
- Transcritical flow without shock: $(hu)_{up} = 1.53, H_d = 0.66(*)$,
- Transcritical flow with shock: $(hu)_{up} = 0.18, H_d = 0.33$,

where $(hu)$ is the discharge, $up$ stands for upstream, and (*): $H_d = 0.66$ is imposed only when the flow is subcritical.

Figure (14) shows the water level initially at time $t_0 = 0$. Figures (15-17) show the water level obtained in different cases using WB-UCS (solid line) and UCS-Sub (dotted line) at the final time $t_f = 200$. We have excellent superposing solutions, which demonstrate the validity of our approach.

We see in figure (18) a comparison between the solutions obtained using WB-UCS and UCS-Sub compared with the analytic solutions [Xing and Shu (2006)] (left) and with the solutions obtained using WENO [Xing and Shu (2006)] (right). The figures show excellent agreement between the results which proves how promising our scheme is.
Fig. 15. Steady flow over a bump: Case of a subcritical flow. Water level obtained using WB-UCS (solid line) and UCS-Sub (dotted line) at $t_f = 200$.

Fig. 16. Steady flow over a bump: Case of a transcritical flow without shock. Water level obtained using WB-UCS (solid line) and UCS-Sub (dotted line) at $t_f = 200$. 
3.6. **Dam break over a rectangular bump**

We now study a one-dimensional problem with a fast-moving flow over an uneven riverbed as discussed in [Touma and Khankan (2012)]. The computational domain is $[0, 1500]$ and it is divided into equal-sized segments using 600 grid points. The numerical solution is resolved at the terminal time $t_f = 15$.

The bottom topography function $b = b(x)$ is given by

$$b = \begin{cases} 
8, & |x - 750| < 375, \\
0, & \text{otherwise},
\end{cases} \quad (15)$$
and the initial conditions are

\[ H_0 = \begin{cases} 
20, & x < 750, \\
15, & \text{otherwise},
\end{cases} \quad (16) \]

for water level so we can deduce the water height function at \( t_0 = 0 \) and the velocity \( u_0 \) is set to be zero.

Fig. 19. Dam break over a rectangular bump: Initial water level at \( t_0 = 0 \).

Fig. 20. Dam break over a rectangular bump: Water level obtained using WB-UCS (solid line) and UCS-Sub (dotted line) at different times before the final time.
Figure (19) shows the water level initially at time $t_0 = 0$. Figures (20, 21) show the water level obtained at different times and at the final time respectively using WB-UCS (solid line) and UCS-Sub (dotted line). The results show excellent agreement between both schemes.

### 3.7. Dam break over an inclined plane

We consider this one-dimensional problem featuring a dam break over inclined planes with different angles of inclination as presented in [Touma (2016)]. We mesh the computational domain $[-15, 15]$ using 200 nodes and compute the numerical solution at the terminal time $t_f = 2$.

The waterbed function is defined by $b = b(x) = x \tan(\beta)$, where $\beta$ represents the angle of inclination. The initial conditions are $u_0 = 0$ and

$$
H_0 = \begin{cases} 
1, & x < 0, \\
0, & \text{otherwise}. 
\end{cases}
$$

In figure (22), we see 3 rows; each row stands for a different angle of inclination. First row, $\beta = 0$, second row, $\beta = \pi/60$, and third row, $\beta = -\pi/60$. We see in each row the initial condition and next to it, the water level obtained at the final time $t_f = 2$ using WB-UCS with the wet and dry states (solid line) and UCS-Sub (dotted line). WB-UCS results were extracted from [Touma (2016)]. The results show excellent agreement between both schemes.

### 4. Conclusion

In this paper, we developed a new numerical method for solving the one-dimensional shallow water equations (SWEs) on variable topographies. The method is second-order
Fig. 22. Dam break problem over an inclined plane: Horizontally; the initial water level at time $t_0 = 0$ next to the water level obtained using WB-UCS + wet dry states (solid line) and UCS-Sub (dotted line) at $t_f = 2$.

accurate, unstaggered, and central, and it uses the subtraction approach to ensure well-balanced solutions. The approach has the benefits of UCS methods in avoiding Riemann solvers and staggered grid methods, as well as properly resolving any steady state of the considered system. The key concept is to evolve the error between the vector solution and any steady state of the system, rather than the unknown vector solution. We used the lake at rest stationary solution as the steady state in the error function in our work. The problems we resolved produced very encouraging outcomes, and our numerical re-
sults are consistent with those found in other studies, thus validating the accuracy and the capability of the blended unstaggered central method with the subtraction method. In addition, the proposed numerical scheme showed a reduction in the computational time of about 15% as compared to the time required by other well-balanced techniques which usually necessitate building on special discretizations of the source term according to the discretization of the water height derivative, whereas UCS-Sub does not need to apply this as it has the C-property without any extra treatment. Not only does this make it faster as mentioned, but it also makes it easier to implement. Generalizations of the presented one-dimensional technique to the two-dimensional context are currently being investigated. We also would like to test our scheme on different problems including friction. Even though we validated our scheme on a problem including a dry domain, but future work might check the treatment of the dry/wet states in all types of problems with discontinuous bottom topography and bowls for instance.

5. Funding

The second author received funds from LAU-PIRF project number I0015.

References


