Shallow-water equations are widely used to model water flow in rivers, lakes, reservoirs, coastal areas, and other situations in which the water depth is much smaller than the horizontal length scale of motion. The classical shallow-water equations, the Saint-Venant system, were originally proposed about 150 years ago and still are used in a variety of applications. For many practical purposes, it is extremely important to have an accurate, efficient and robust numerical solver for the Saint-Venant system and related models. As their solutions are typically non-smooth and even discontinuous, finite-volume schemes are among the most popular tools. In this paper, we review such schemes and focus on one of the simplest (yet highly accurate and robust) methods: central-upwind schemes. These schemes belong to the family of Godunov-type Riemann-problem-solver-free central schemes, but incorporate some upwinding information about the local speeds of propagation, which helps to reduce an excessive amount of numerical diffusion typically present in classical (staggered) non-oscillatory central schemes. Besides the classical one- and two-dimensional Saint-Venant systems, we will consider the shallow-water equations with friction terms, models with moving bottom topography, the two-layer shallow-water system as well as general non-conservative hyperbolic systems.
1. Introduction

Shallow-water equations arise in modelling of water flows in rivers, lakes, reservoirs, coastal areas, and other situations in which the water depth is much smaller than the horizontal length scale of motion. Therefore, shallow-water and closely related equations are widely used in oceanography and atmospheric sciences to model among others such hazardous phenomena as hurricanes/typhoons and tsunamis. Besides scientific applications, shallow-water models are used for flood mitigation as well as in coastal, hydraulic and civil engineering to design harbour areas, develop urban coastal areas, construct coastal protection systems, etc.

The classical shallow-water equations (the Saint-Venant system) were originally proposed by de Saint-Venant (1871), but are still widely used in a variety of applications. In the two-dimensional (2-D) case, the simplest version of the Saint-Venant system (with the viscosity and bottom friction terms being neglected) reads as

\[
\begin{align*}
  h_t + (hu)_x + (hv)_y &= 0, \\
  (hu)_t + \left( hu^2 + \frac{1}{2}gh^2 \right)_x + (huv)_y &= -ghB_x, \\
  (hv)_t + (huv)_x + \left( hv^2 + \frac{1}{2}gh^2 \right)_y &= -ghB_y,
\end{align*}
\]  

(1.1)

where \(x\) and \(y\) are the horizontal spatial coordinates, \(t\) is time, \(h(x, y, t)\) is the water depth, \(u(x, y, t)\) and \(v(x, y, t)\) are the \(x\)- and \(y\)-velocity components, respectively, \(g\) is the constant gravitational acceleration, and \(B(x, y)\) is the bottom topography, which is prescribed and time-independent.

The system (1.1) is a nonlinear hyperbolic system of partial differential equations (PDEs), which admits very complicated, generically non-smooth
solutions that may contain shock and rarefaction waves, and – in the case of discontinuous bottom topography – also contact discontinuities. Except for very simple initial data, no analytical solution of (1.1) is available and thus one has to solve the Saint-Venant system numerically. It is well known (see e.g. LeVeque 2002 and references therein) that solutions of (1.1) may breakdown even when the initial data are smooth and the bottom topography is flat ($B \equiv \text{const.} \Rightarrow B_x \equiv B_y \equiv 0$), and thus designing stable and accurate numerical methods for (1.1) is a highly non-trivial task.

Another difficulty in the development of numerical methods for (1.1) is related to the fact that the system (1.1) is a system of balance laws, and a good numerical method must respect a delicate balance between the flux and source terms. This means that the scheme should be able to exactly preserve initial data that correspond to (at least some physically relevant) steady-state solutions. Such schemes are called well-balanced and their advantage over non-well-balanced schemes can be clearly observed when a (relatively) coarse computational grid is used to capture steady-state or quasi-steady-state solutions (note that in applications, many important solutions are, in fact, small perturbations of steady-state solutions), as a magnitude of the truncation error of the non-well-balanced scheme in such situations may be larger than the magnitude of the waves to be captured. Provided the non-well-balanced method is converging, one can obviously achieve a high resolution of the low magnitude waves by refining the mesh, but this approach may be computationally expensive or even unaffordable, especially in large-scale simulations. We refer the reader to LeVeque (1998), Jin (2001), Kurganov and Levy (2002), Gallouët, Hérad and Seguin (2003), Russo (2005), Li and Chen (2006), Noelle, Pankratz, Puppo and Natvig (2006), Castro, Pardo Milánés and Parés (2007), Lukáčová-Medvid’ová, Noelle and Kraft (2007), Noelle, Xing and Shu (2007), Russo and Khe (2010), Fjordholm, Mishra and Tadmor (2011), Bernstein, Chertock and Kurganov (2016) and Cheng and Kurganov (2016) for several well-balanced schemes, some of which will be reviewed in detail in Section 5.

Besides the requirement to be well-balanced, there is another important property a good numerical method should possess: it should be positivity-preserving, that is, the method should guarantee that the computed water depth remains non-negative at all times. This is extremely practically important in situations when some parts of the computational domain are dry ($h = 0$) or almost dry ($0 < h \ll 1$). Even though the validity of the Saint-Venant system in the presence of dry areas is questionable, dealing with shore areas and islands and thus tracking wet/dry fronts is unavoidable in both scientific and engineering applications. For several well-balanced positivity-preserving schemes, we refer the reader to Perthame and Simeoni (2001), Audusse et al. (2004), Tang, Tang and Xu (2004), Gallardo, Parés and Castro (2007), Kurganov and Petrova (2007), Berthon

In this paper, we review finite-volume schemes for the Saint-Venant system (1.1) and related models. In general, finite-volume schemes (see e.g. the monographs by Godlewski and Raviart 1996, Kröner 1997 and LeVeque 2002) are a popular tool for numerically solving hyperbolic systems of balance laws, which in the 2-D case read as

\[ U_t + F(U)_x + G(U)_y = S(U). \]  

(1.2)

Here, \( U(x, y, t) \in \mathbb{R}^N \) is a vector of unknown functions, \( F \) and \( G \) are flux functions, and \( S \) is a source term. Since solutions of (1.2) are generically discontinuous, the system (1.2) is understood in a weak or integral sense. Namely, we take a certain spatial domain \( C \) and integrate (1.2) over the space–time control volume \( C \times [t, t + \Delta t] \) to obtain the following integral equation:

\[
\int_C U(x, y, t + \Delta t) \, dx \, dy = \int_C U(x, y, t) \, dx \, dy \\
- \int_t^{t+\Delta t} \int_{\partial C} (n_x F(U(x, y, \tau)) + n_y G(U(x, y, \tau))) \, ds \, d\tau \\
+ \int_t^{t+\Delta t} \int_C S(U(x, y, \tau)) \, dx \, dy \, d\tau,
\]  

(1.3)

where \( \partial C \) is a boundary of \( C \) and \( n = (n_x, n_y)^\top \) is its outer unit normal.

Note that for smooth solutions (1.3) is equivalent to (1.2) assuming that the former is satisfied for all spatial domains \( C \) and all \( t \geq 0 \) and \( \Delta t > 0 \). The advantage of the integral formulation (1.3), however, is that it is valid for piecewise smooth solutions as well. Unlike classical finite-difference methods (Richtmyer and Morton 1994), which are designed based on the classical PDE formulation (1.2), finite-volume schemes are constructed using the integral formulation (1.3) and thus finite-volume schemes are an appropriate tool for capturing weak, non-smooth solutions of nonlinear hyperbolic PDEs. Instead of computing the point values of \( U \), which may be undefined at the discontinuities, the computed quantities in finite-volume schemes are the averages over the spatial domains,

\[
\overline{U}_C(t) = \frac{1}{|C|} \int_C U(x, y, t) \, dx \, dy,
\]
which are well-defined quantities. In order to evolve them from time $t$ to $t + \Delta t$ according to (1.3), one would need to evaluate the flux and source integrals on the right-hand side of (1.3). This may be highly non-trivial due to the complicated structure of the solutions of (1.2).

One of the major features of hyperbolic systems of balance laws is propagation (with a finite speed) of the information along the characteristic surfaces. This brings the idea of upwinding, which may stabilize the computation of the flux integrals in (1.3). The Godunov scheme (Godunov 1959) is the first finite-volume upwind scheme designed for one-dimensional (1-D) hyperbolic systems of conservation laws (system (1.2) with $S \equiv 0$). The main idea behind the construction of the Godunov scheme is a global approximation of the solution using a piecewise constant function (the computational domain is split into the cells and the solution is approximated by a constant piece in each of the cells) followed by the upwind evaluation of the flux integrals. As we explain in Section 3.1, the latter requires solving the Riemann problem, which may be quite complicated and computationally expensive. The Lax–Friedrichs scheme (Friedrichs 1954, Lax 1954) is a prototype of non-oscillatory central schemes which offer a Riemann-problem-solver-free alternative to the upwind schemes. In central schemes, the solution is evolved in time still using the same integral equation (1.3), but the control volume is selected in such a way that the location of discontinuities in the piecewise constant approximant of the solution at time level $t$ does not coincide with $\partial C$, which helps to avoid the necessity to solve any Riemann problem (see Sections 3.2 for details). The drawback of the central schemes, however, is their relatively high numerical dissipation, which can be decreased by taking into account the local speeds of propagation and thus introducing some upwinding information into the central schemes. This leads to a class of central-upwind schemes, which were introduced in Kurganov and Tadmor (2000) and Kurganov, Noelle and Petrova (2001); see also Rusanov (1961) and Harten, Lax and van Leer (1983) and the discussion in Sections 3.3 and 4, where 1-D and 2-D central-upwind schemes are described, respectively.

Unfortunately, both the Godunov and Lax–Friedrichs schemes are only first-order accurate and thus they typically require the use of very fine (often impractically fine) meshes to achieve a high resolution of discontinuous solutions. It is well known (see e.g. Godlewski and Raviart 1996 Kröner 1997, LeVeque 2002) that sharp approximated solutions can be obtained using higher, at least second-order finite-volume schemes. In order to construct such schemes, one needs to replace the piecewise constant approximation of the computed solution with a piecewise polynomial one, which is more accurate, but makes the upwinding substantially more complicated. For pioneering work on second-order upwind and central schemes, which employ piecewise linear approximations, we refer the reader to van Leer (1979) and Nessyahu and Tadmor (1990), respectively.
In Section 5, we turn to a description of the extension of central-upwind schemes to shallow-water equations, which was first presented in Kurganov and Levy (2002) and then further developed in Kurganov and Petrova (2007), Bryson et al. (2011), Bollermann et al. (2013), Beljadid et al. (2016), Bernstein et al. (2016), Shirikhani et al. (2016), Cheng and Kurganov (2016) and Liu et al. (2018). The well-balanced property of the central-upwind schemes is enforced by reconstructing equilibrium variables (those that remain constant at the relevant steady states) and designing special well-balanced discretizations of the geometric source terms as described in Sections 5.1.1, 5.1.2 and 5.2.1. The designed well-balanced central-upwind schemes are made positivity-preserving with the help of several techniques described in detail in Sections 5.1.1, 5.1.3–5.1.5, 5.2.2 and 5.2.3. We then continue in Section 5.3 with a description of the well-balanced positivity-preserving central-upwind scheme for the Saint-Venant system with friction terms. This scheme, which was developed in Chertock, Cui, Kurganov and Wu (2015b), is a non-trivial extension of the central-upwind scheme from Kurganov and Petrova (2007).

When the bottom topography is discontinuous, the Saint-Venant system would contain the non-conservative product terms $hB_x$ and $hB_y$. In order to design robust and accurate central-upwind schemes for non-conservative hyperbolic system, we first rewrite the central-upwind scheme from Kurganov et al. (2001) in a different form and then take into account the jump of the non-conservative product terms across the cell interfaces. This results in path-conservative central-upwind schemes, which have recently been developed in Castro Díaz, Kurganov and Morales de Luna (2018) and are described here in Section 6. The path-conservative central-upwind schemes can be made well-balanced by modifying the numerical viscosity of the original central-upwind schemes as described in Section 6.1.

Finally, in Section 7, we present an extension of the central-upwind and path-conservative central-upwind schemes to two related shallow-water models: the Saint-Venant system with time-dependent bottom topography (Section 7.1), and the two-layer shallow-water system (Section 7.2).

### 2. Hyperbolic systems of conservation and balance laws

A general 1-D system of balance laws reads as

$$ U_t + F(U)_x = S(U), \quad (2.1) $$

where $U(x, t) \in \mathbb{R}^N$ is a vector of unknown functions, $F$ is a flux function, and $S$ is a source term. If $S \equiv 0$, (2.1) reduces to a system of conservation laws:

$$ U_t + F(U)_x = 0. \quad (2.2) $$
The systems (2.1) and (2.2) are hyperbolic if the Jacobian $\partial F/\partial U$ has real eigenvalues

$$\lambda_1 \left( \frac{\partial F}{\partial U} \right) \leq \cdots \leq \lambda_N \left( \frac{\partial F}{\partial U} \right).$$

For example, for the 1-D Saint-Venant system of shallow-water equations,

$$h_t + (hu)_x = 0, \quad (hu)_t + \left( hu^2 + \frac{1}{2} gh^2 \right)_x = -ghB_x,$$

(2.3)

$U = (h, q)^\top$ with $q := hu$ denoting the discharge,

$$F(h, q) = \left( q, \frac{q^2}{h} + \frac{gh^2}{2} \right)^\top, \quad S(h; B) = (0, -ghB_x)^\top,$$

the Jacobian is

$$\frac{\partial F}{\partial U}(h, q) = \begin{pmatrix} 0 & 1 \\ -u^2 + gh & 2u \end{pmatrix},$$

and its eigenvalues are $\lambda_1 = u - \sqrt{gh}$ and $\lambda_2 = u + \sqrt{gh}$.

One of the main features of the hyperbolic systems is a finite speed of propagation: any change in the solution propagates at a speed bounded by the lower and upper bounds on $\lambda_1$ and $\lambda_N$, respectively. This property is the key point used in the construction of finite-volume methods for hyperbolic PDEs as explained below.

Another important feature of hyperbolic systems is that they admit non-smooth (discontinuous) solutions. Moreover, the solutions of nonlinear hyperbolic systems may break down even when the initial data are infinitely smooth. This fact should be taken into account when such numerical techniques as solution approximation and its global (in space) reconstruction are designed to be used by a finite-volume scheme.

A general 2-D system of balance laws is given by (1.2) and if $S \equiv 0$, it reduces to a system of conservation laws:

$$U_t + F(U)_x + G(U)_y = 0.$$  

(2.4)

The systems (1.2) and (2.4) are hyperbolic if both the $x$- and $y$-directional Jacobians $\partial F/\partial U$ and $\partial G/\partial U$ have real eigenvalues

$$\lambda_1 \left( \frac{\partial F}{\partial U} \right) \leq \cdots \leq \lambda_N \left( \frac{\partial F}{\partial U} \right) \quad \text{and} \quad \lambda_1 \left( \frac{\partial G}{\partial U} \right) \leq \cdots \leq \lambda_N \left( \frac{\partial G}{\partial U} \right),$$

respectively.
For example, for the 2-D Saint-Venant system of shallow-water equations (1.1), \( \mathbf{U} = (h, q^x, q^y) \) with \( q^x := hu \) and \( q^y := hv \) denoting the \( x \)- and \( y \)-discharges, respectively,

\[
\mathbf{F}(h, q^x, q^y) = \left( q^x, \frac{(q^x)^2}{h} + \frac{gh^2}{2}, \frac{q^x q^y}{h} \right)^	op,
\]

\[
\mathbf{G}(h, q^x, q^y) = \left( q^y, \frac{q^x q^y}{h}, \frac{(q^y)^2}{h} + \frac{gh^2}{2} \right)^	op,
\]

\[
\mathbf{S}(h; B) = (0, -ghB_x, -ghB_y)^	op,
\]

the corresponding Jacobians are

\[
\frac{\partial \mathbf{F}}{\partial \mathbf{U}}(h, q^x, q^y) = \begin{pmatrix}
0 & 1 & 0 \\
-u^2 + gh & 2u & 0 \\
-uv & v & u
\end{pmatrix},
\]

and

\[
\frac{\partial \mathbf{G}}{\partial \mathbf{U}}(h, q^x, q^y) = \begin{pmatrix}
0 & 0 & 1 \\
-uv & v & u \\
-v^2 + gh & 0 & 2v
\end{pmatrix},
\]

and their eigenvalues are

\[
\lambda_1 \left( \frac{\partial \mathbf{F}}{\partial \mathbf{U}} \right) = u - \sqrt{gh}, \quad \lambda_2 \left( \frac{\partial \mathbf{F}}{\partial \mathbf{U}} \right) = u, \quad \lambda_3 \left( \frac{\partial \mathbf{F}}{\partial \mathbf{U}} \right) = u + \sqrt{gh}
\]

and

\[
\lambda_1 \left( \frac{\partial \mathbf{G}}{\partial \mathbf{U}} \right) = v - \sqrt{gh}, \quad \lambda_2 \left( \frac{\partial \mathbf{G}}{\partial \mathbf{U}} \right) = v, \quad \lambda_3 \left( \frac{\partial \mathbf{G}}{\partial \mathbf{U}} \right) = v + \sqrt{gh}.
\]

The structure of 2-D solutions may be much more complicated than the structure of their 1-D counterparts. However, the main features related to the solution breakdown and finite speed of propagation are still the same and they can be used in the construction of 2-D finite-volume methods.

3. One-dimensional finite-volume schemes

In this section we will review finite-volume schemes for the 1-D hyperbolic systems of conservation laws (2.2).

For the sake of simplicity, we consider a uniform finite-volume mesh consisting of the cells \( C_j := [x_{j-1/2}, x_{j+1/2}] \) of size \( |C_j| \equiv \Delta x \), centred at \( x_j = (x_{j-1/2} + x_{j+1/2})/2 \), and assume that at a certain time level \( t^n \) the cell averages of the solution,

\[
\overline{\mathbf{U}}_j^n \approx \frac{1}{\Delta x} \int_{C_j} \mathbf{U}(x, t^n) \, dx,
\]
Finite-volume schemes for shallow-water equations

are available. We then approximate $U(x, t^n)$ using a piecewise polynomial interpolant

$$\tilde{U}^n(x) = \sum_j \mathcal{P}_j^n(x) \chi_j(x),$$  \hspace{1cm} (3.1)

where $\mathcal{P}_j^n(x)$ are polynomial pieces satisfying the conservation,

$$\frac{1}{\Delta x} \int_{C_j} \mathcal{P}_j^n(x) \, dx = \tilde{U}_j^n,$$

accuracy (for smooth $U$),

$$\mathcal{P}_j^n(x) = U(x, t^n) + O((\Delta x)^p) \quad \text{for } x \in C_j,$$

where $p$ is the desired order of spatial approximation, and non-oscillatory requirements. The latter property of the reconstruction can be defined in many alternative ways. For example, one may bound the total variation of each component of $U$,

$$TV[(\tilde{U}^n(x))^{(i)}] \leq TV[(\tilde{U}_j^n)^{(i)}] + O(\Delta x), \quad i = 1, \ldots, N,$$

where

$$TV[(\tilde{U}_j^n)^{(i)}] := \sum_j [(\tilde{U}^n_{j+1})^{(i)} - (\tilde{U}^n_j)^{(i)}],$$

or require the number-of-extrema non-increasing property to be satisfied in a component-wise manner. Milder non-oscillatory and essentially non-oscillatory criteria can be also imposed, especially on higher-order interpolants.


After the interpolant (3.1) is constructed, we integrate (2.2) over a certain space–time control volume $C \times [t^n, t^{n+1}]$ (where $t^{n+1} := t^n + \Delta t^n$ and the time step $\Delta t^n$ is selected using an appropriate Courant–Friedrichs–Lewy
(CFL) condition), to obtain the cell averages $\mathbf{U}_C^{n+1}$ at the new time level $t = t^{n+1}$. Depending on a particular choice of $C$, one may design either upwind (Section 3.1) or central (Section 3.2) finite-volume schemes.

### 3.1. Upwind schemes

In the classical approach dating back to the original Godunov scheme (1959), the space–time control volume is selected to be $C_j \times [t^n, t^{n+1}]$. This results in

$$
\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t^n}{\Delta x} \left[ \mathcal{F}_{j+1/2}^{n+1/2} - \mathcal{F}_{j-1/2}^{n+1/2} \right],
$$

where

$$
\mathcal{F}_{j+1/2}^{n+1/2} \approx \frac{1}{\Delta t^n} \int_{t^n}^{t^{n+1}} \mathbf{F}(\mathbf{U}(x_{j+1/2}, t)) \, dt
$$

is a numerical flux, which needs to be computed based on the data (3.1) available at time level $t = t^n$. In order to evaluate the time integral in (3.3), one has to (approximately) solve the following initial value problem (IVP) with the initial data prescribed at time $t = t^n$:

$$
\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = 0, \quad t \in (t^n, t^{n+1}],
$$

$$
\mathbf{U}(x, t^n) = \begin{cases}
\mathcal{P}_j^n(x), & x < x_{j+1/2}, \\
\mathcal{P}_{j+1}^n(x), & x > x_{j+1/2}.
\end{cases}
$$

In the case of the first-order piecewise constant reconstruction, that is, when $\mathcal{P}_j^n(x) = \mathbf{U}_j^n$ for all $j$, the IVP (3.4) is a Riemann problem, whose solution, as is well known, is self-similar:

$$
\mathbf{U}(x, t) = \mathbf{U}_{j+1/2}^{n+1} (\xi), \quad \xi := \frac{x - x_{j+1/2}}{t - t^n}.
$$

Therefore it is constant at $x = x_{j+1/2}$ for all $t \in (t^n, t^{n+1}]$ and the numerical flux (3.3) becomes

$$
\mathcal{F}_{j+1/2}^{n+1/2} = \mathbf{F}(\mathbf{U}_{j+1/2}^{n+1}(0)).
$$

In order to complete the construction of the first-order Godunov scheme one then needs to analytically solve the Riemann problem (3.4) to find $\mathbf{U}_{j+1/2}^{n+1}(0)$. For some hyperbolic systems of conservation laws this can be done; see, for example, Godlewski and Raviart (1996), Kröner (1997) and LeVeque (2002). Alternatively, the Riemann problem (3.4) can be solved approximately; a variety of approximate Riemann problem solvers can be found in Godlewski and Raviart (1996), Kröner (1997), LeVeque (2002) and Toro (2009), for example.

If (3.1) is a piecewise linear function used in the construction of second-order upwind schemes, the IVP (3.4) is a generalized Riemann problem.
Finite-volume schemes for shallow-water equations

(0.1) GRPs, whose solution is no longer self-similar. GRPs are much harder to solve analytically. Although this was done for some hyperbolic systems of conservation laws in Ben-Artzi (1989), Ben-Artzi and Falcozv (1984, 1986, 2003) and Ben-Artzi, Li and Warnecke (2006), approximate GRP solvers are more popular as they are much easier to construct and implement; see, for example, Godlewski and Raviart (1996), LeVeque (2002) and Lukáčová-Medvid’ová, Morton and Warnecke (2004).

A simpler and thus more popular approach for constructing high-order upwind schemes is by using the semi-discrete formulation of (0.2), which is obtained by integrating (0.2) over the cell $C_j$, that is,

$$
\frac{d}{dt} \bar{U}_j(t) = -\frac{1}{\Delta x} \left[ F_{j+1/2}(t) - F_{j-1/2}(t) \right],
$$

(0.5)

where

$$
\bar{U}_j(t) \approx \frac{1}{\Delta x} \int_{C_j} U(x,t) \, dx
$$

are cell averages assumed to be available at a certain time level $t$ and the numerical fluxes $F_{j+1/2}(t)$ are obtained by (approximately) solving the following Riemann problem:

$$
U_t + F(U)_x = 0, \quad t \in (t, t + \tau],
$$

$$
U(x,t) = \begin{cases} 
U_{j+1/2}(t), & x < x_{j+1/2}, \\
U_{j+1/2}(t), & x > x_{j+1/2}.
\end{cases}
$$

Here, $\tau$ is a small positive number and $U_{j+1/2}^+(t)$ and $U_{j+1/2}^-(t)$ are the right- and left-sided values of the piecewise polynomial interpolant (reconstructed at time $t$ from the cell averages $\bar{U}_j(t)$)

$$
\tilde{U}(x;t) = \sum_j \mathcal{P}_j(x;t) \chi_j(x),
$$

(0.6)

namely,

$$
U_{j+1/2}^+(t) = \mathcal{P}_{j+1}(x_{j+1/2};t) \quad \text{and} \quad U_{j+1/2}^-(t) = \mathcal{P}_j(x_{j+1/2};t).
$$

(0.7)

The semi-discrete scheme is implemented by numerically solving the ODE system (0.5) using an appropriate ODE solver. A popular choice of such solvers is that of strong stability preserving (SSP) Runge–Kutta and multi-stage methods; see, for example, Gottlieb, Ketcheson and Shu (2011) and Gottlieb, Shu and Tadmor (2001). These methods were originally designed to ensure that the total variation of the computed solution does not increase at the time evolution stage and thus they were originally referred to as total-variation-diminishing (TVD) methods in Shu (1988) and Shu and Osher (1988).
Remark 3.1. The order of the resulting scheme is determined by the orders of the piecewise polynomial reconstruction (3.6), (3.7) and the ODE solver.

Remark 3.2. We would like to emphasize that the fully discrete upwind schemes designed based on solving the GRPs typically achieve higher resolution than the semi-discrete upwind schemes. The latter schemes are, however, substantially simpler and computationally less expensive.

3.2. Central schemes

Central schemes offer a simpler, Riemann-problem-solver-free alternative to upwind schemes. They are obtained by selecting shifted space–time control volumes that contain the corresponding Riemann fans. The simplest choice of such control volumes, \([x_j, x_{j+1}] \times [t^n, t^{n+1}]\), was suggested in Nessyahu and Tadmor (1990), where a second-order staggered central scheme (the Nessyahu–Tadmor scheme) was derived. The cell averages at time level \(t = t^{n+1}\) are then obtained over the staggered grid (compare with (3.2)):

\[
\bar{U}_{j+1/2}^{n+1} = \frac{1}{\Delta x} \int_{x_j}^{x_{j+1}} \bar{U}^n(x) \, dx - \Delta t^n \frac{\Delta x}{\Delta x} [\mathcal{F}_{j+1/2}^{n+1} - \mathcal{F}_{j}^{n+1/2}],
\]

(3.8)

where the numerical fluxes are

\[
\mathcal{F}_{j}^{n+1/2} \approx \frac{1}{\Delta t^n} \int_{t^n}^{t^{n+1}} F(U(x,t)) \, dt.
\]

(3.9)

We note that the spatial integral on the right-hand side of (3.8) can be explicitly computed for any piecewise polynomial reconstruction \(\bar{U}^n\). Further, the time integrals in (3.9) can be easily computed using an appropriate quadrature because the solution of the IVP

\[
U_t + F(U)_x = 0, \quad t \in (t^n, t^{n+1}],
\]

\[
U(x, t^n) = \bar{U}^n(x)
\]

remains smooth at the cell centres \(x = x_j\), provided the CFL number is taken to be less than or equal to 1/2, in order to prevent the nonlinear (possibly discontinuous) waves generated at cell interfaces at time level \(t = t^n\) from reaching the cell centres before \(t = t^{n+1}\).

The second-order Nessyahu–Tadmor scheme is designed by taking the second-order piecewise linear reconstruction, for which

\[
\mathcal{P}^n_j(x) = \bar{U}^n_j + (U_x)^n_j (x - x_j),
\]

where \((U_x)^n_j\) are the slopes that approximate \(U_x(x_j, t^n)\) in a non-oscillatory manner using a nonlinear slope limiter, and the midpoint quadrature for
the temporal integrals in (3.9). This results in
\[ \overline{U}_{j+1/2}^{n+1} = \frac{1}{2} (\overline{U}_j^n + \overline{U}_{j+1}^n) + \frac{\Delta x}{8} \left[ (U_x)_j^n - (U_x)_{j+1}^{n+1} \right] \]
\[ - \frac{\Delta t^n}{\Delta x} \left[ F(U_{j+1/2}^{n+1/2}) - F(U_{j}^{n+1/2}) \right], \tag{3.10} \]
where the midpoint values \( U_{j}^{n+1/2} \) are predicted using the Taylor expansion as follows:
\[ U_{j}^{n+1/2} = U_j^n + \frac{\Delta t^n}{2} (U_t)_j^n = \overline{U}_j^n - \frac{\Delta t^n}{2} (F(U)_x)_j^n. \tag{3.11} \]
Here, the derivatives \( (F(U)_x)_j^n \) can be computed with the help of the same limiter, which was used in the computation of \( (U_x)_j^n \).

**Remark 3.3.** Note that if all of the slopes \( (U_x)_j^n \) and \( (F(U)_x)_j^n \) are set to zero, the Nessyahu–Tadmor scheme (3.10), (3.11) will reduce to the first-order staggered Lax–Friedrichs scheme
\[ \overline{U}_{j+1/2}^{n+1} = \frac{1}{2} (\overline{U}_j^n + \overline{U}_{j+1}^n) - \frac{\Delta t^n}{\Delta x} \left[ F(\overline{U}_{j+1}^{n+1/2}) - F(\overline{U}_j^n) \right], \tag{3.12} \]
which is a Godunov-type version of the classical Lax–Friedrichs scheme from Friedrichs (1954) and Lax (1954).


**Remark 3.5.** The staggered central schemes provide a ‘black-box’ tool for solving general (multidimensional) hyperbolic systems of conservation laws. However, the amount of numerical diffusion present in staggered central schemes is quite large and it significantly increases when the time step is taken to be small or, alternatively, when a long-term (steady-state) computation is performed. In order to clarify this point, let us consider the first-order staggered Lax–Friedrichs scheme (3.12) and rewrite it in the following equivalent form:
\[ \frac{\overline{U}_{j+1/2}^{n+1} - \overline{U}_{j+1/2}^n}{\Delta t^n} + \frac{F(\overline{U}_{j+1}^{n+1/2}) - F(\overline{U}_j^n)}{\Delta x} = \frac{(\Delta x)^2}{8\Delta t^n} \cdot \frac{\overline{U}_{j+1}^n - 2\overline{U}_{j+1/2}^n + \overline{U}_j^n}{(\Delta x/2)^2}. \tag{3.13} \]
As one can see, the terms on the left-hand side of (3.13) approximate \( U_t \) and \( F(U)_x \), respectively, while the term on the right-hand side of (3.13) represents the numerical viscosity/diffusion present in the scheme. Notice
that the numerical viscosity coefficient is proportional to $(\Delta x)^2/\Delta t^n$ and its size depends on the ratio of $\Delta t^n$ and $\Delta x$. In the hyperbolic regime, one typically takes $\Delta t^n \sim \Delta x$, and then the viscosity coefficient is proportional to $\Delta x$ as it is supposed to be for first-order schemes. If, however, $\Delta t^n$ is for some reason taken to be small then the numerical dissipation increases so significantly that the scheme may become inconsistent (e.g. if $\Delta t^n \sim (\Delta x)^2$). In addition, the excessive numerical dissipation present in staggered central schemes may badly affect the quality of solutions computed at large times (this makes the schemes inappropriate for capturing time-independent steady-state solutions, as pointed out in Kurganov and Tadmor 2000).

In order to overcome this drawback, we have developed a new class of Riemann-problem-solver-free methods: central-upwind schemes, which are briefly described in Section 3.3. We refer the reader to Kurganov (2016), Kurganov and Lin (2007) and Kurganov et al. (2017) for details of their derivation.

3.3. Central-upwind schemes

Central-upwind schemes are derived using the Godunov-type central approach described in Section 3.2, but using a different set of non-uniform space–time control volumes whose size is determined based on the one-sided local speeds of propagation $a_{j+1/2}^\pm$, which are determined as follows. We first note that the piecewise polynomial reconstruction (3.6) is generically discontinuous at the cell interfaces $x = x_{j+1/2}$ and the (non-smooth) waves generated there propagate with local speeds whose upper and lower bounds can be estimated using the largest and smallest eigenvalues of the Jacobian $\partial F/\partial U$. In most cases, reliable estimates are given by

$$a_{j+1/2}^+(t) = \max \left\{ \lambda_N \left( \frac{\partial F}{\partial U} (U^-_{j+1/2}(t)) \right), \lambda_N \left( \frac{\partial F}{\partial U} (U^+_{j+1/2}(t)) \right), 0 \right\};$$

$$a_{j+1/2}^-(t) = \min \left\{ \lambda_1 \left( \frac{\partial F}{\partial U} (U^-_{j+1/2}(t)) \right), \lambda_1 \left( \frac{\partial F}{\partial U} (U^+_{j+1/2}(t)) \right), 0 \right\}. \quad (3.14a)$$

However, as was recently pointed out in Guermond and Popov (2016), this estimate may be inaccurate even in the case of convex flux $F$. For the estimates on one-sided local speeds in the case of non-convex fluxes, we refer the reader to Kurganov, Petrova and Popov (2007), for example.

Equipped with the set of non-uniform space–time control volumes, the solution is evolved to the new time level at which it is given as a set of cell averages of $U$ over a new (strictly) non-uniform mesh containing twice as many cells as the original (uniform) mesh. The obtained solution is then projected onto the original grid, which makes the resulting fully discrete central-upwind scheme practically feasible; see Kurganov (2016) and
Kurganov and Lin (2007) for details of their derivation. The fully discrete central-upwind scheme is, however, quite complicated (especially its 2-D version recently developed in Kurganov et al. 2017).

The central-upwind schemes may be significantly simplified if one passes to a semi-discrete limit by taking $\max_n(\Delta t^n) \to 0$. This leads to a class of semi-discrete central-upwind schemes developed in Kurganov and Lin (2007), Kurganov et al. (2001) and Kurganov and Tadmor (2000), which in the 1-D case can be put into the flux form (3.5) with the central-upwind numerical flux

$$
\mathcal{F}_{j+1/2} = \frac{a_{j+1/2}^+ F(U_{j+1/2}^-) - a_{j+1/2}^- F(U_{j+1/2}^+)}{a_{j+1/2}^+ - a_{j+1/2}^-} + \frac{a_{j+1/2}^+ a_{j+1/2}^-}{a_{j+1/2}^+ - a_{j+1/2}^-} \left[ U_{j+1/2}^+ - U_{j+1/2}^- - \delta U_{j+1/2} \right].
$$

(3.15)

Here,

$$
\delta U_{j+1/2} := \minmod(U_{j+1/2}^+ - U_{j+1/2}^*, U_{j+1/2}^* - U_{j+1/2}^-)
$$

(3.16)

is the built-in anti-diffusion term (see Kurganov 2016, Kurganov and Lin 2007 for details of its derivation) and

$$
U_{j+1/2}^* = \frac{a_{j+1/2}^+ U_{j+1/2}^+ - a_{j+1/2}^- U_{j+1/2}^- - \{ F(U_{j+1/2}^+) - F(U_{j+1/2}^-) \}}{a_{j+1/2}^+ - a_{j+1/2}^-}
$$

(3.17)

are the intermediate values. Finally, the minmod function in (3.16) is defined as follows:

$$
\minmod(z_1, z_2, \ldots) = \begin{cases} 
\min(z_1, z_2, \ldots), & \text{if } z_i > 0, \text{ for all } i, \\
\max(z_1, z_2, \ldots), & \text{if } z_i < 0, \text{ for all } i, \\
0, & \text{otherwise}. 
\end{cases}
$$

Note that all of the indexed quantities in (3.15)–(3.17) are time-dependent, but we omit this dependence for the sake of brevity.

**Remark 3.6.** The semi-discrete central-upwind scheme (3.5), (3.15)–(3.17) should be implemented using an appropriate ODE solver. In the purely hyperbolic regime, we usually use the three-stage third-order SSP Runge–Kutta method; see, for example, Gottlieb, Ketcheson and Shu (2011) and Gottlieb, Shu and Tadmor (2001).

**Remark 3.7.** We note that both the fully discrete schemes presented in Kurganov and Lin (2007) and Kurganov et al. (2001) and the semi-discrete scheme (3.5), (3.15)–(3.17) belong to the class of Godunov-type central
schemes, as the solution is evolved in time using the integral form of conservation laws without (approximately) solving (generalized) Riemann problems. On the other hand, when all of the Jacobian eigenvalues are of the same sign, the schemes reduce to the corresponding upwind schemes. This is the reason why starting from Kurganov et al. (2001) we identify these central schemes as central-upwind schemes.

**Remark 3.8.** In earlier works on central-upwind schemes, the projection step was not performed in the sharpest manner, so the numerical diffusion was not completely minimized and the anti-diffusion term $\delta U_{j+1/2}$ in (3.15) was, in fact, set to zero for all $j$; see Kurganov et al. (2001) and Kurganov and Tadmor (2000).

**Remark 3.9.** In the first work on central-upwind schemes, by Kurganov and Tadmor (2000), the space–time control volumes were taken to be symmetric with respect to $x = x_{j+1/2}$. This leads to $a_{j+1/2}^+ \equiv -a_{j+1/2}^-$ for all $j$, and the resulting central-upwind schemes, often called non-staggered central schemes, are more diffusive.

**Remark 3.10.** The order of the semi-discrete central-upwind schemes is determined formally only by the order of the piecewise polynomial reconstruction (3.6), used to compute the values $U_{j+1/2}^{\pm}$, and the order of the ODE solver.

**Remark 3.11.** We would like to point out that the first-order versions of the central-upwind schemes from Kurganov and Tadmor (2000) and Kurganov et al. (2001) coincide with the Rusanov scheme from Rusanov (1961) and the Harten–Lax–van Leer scheme from Harten et al. (1983), respectively.

4. Two-dimensional central-upwind schemes

A fully discrete 2-D central-upwind scheme for (2.4) has recently been rigorously derived in Kurganov et al. (2017) using the 2-D Godunov-type central approach developed in Arminjon and Viallon (1995), Arminjon et al. (1997) and Jiang and Tadmor (1998). The fully discrete central-upwind scheme is capable of achieving very sharp resolution, but it is quite complicated. Much simpler (though a little more diffusive) semi-discrete central-upwind schemes can be derived using either the dimension-by-dimension Kurganov and Levy (2000) and Kurganov and Tadmor (2000) or genuinely multidimensional approach. The latter one was implemented on Cartesian (Chertock et al. 2018, Kurganov and Lin 2007, Kurganov et al. 2001, Kurganov and Petrova 2001, Kurganov and Tadmor 2002), triangular (Kurganov and Petrova 2005), quadrilateral (Shirkhani et al. 2016) and general polygonal (Beljadid et al. 2016) grids.
We now briefly describe the second-order semi-discrete central-upwind scheme on Cartesian grids. To this end, we split the computational domain into the Cartesian cells \( C_{j,k} := [x_{j-1/2}, x_{j+1/2}] \times [y_{k-1/2}, y_{k+1/2}] \) of size \( |C_{j,k}| = \Delta x \Delta y \) centred at \((x_j, y_k) = ((x_{j-1/2} + x_{j+1/2})/2, (y_{k-1/2} + y_{k+1/2})/2)\) and assume that at a certain time level \( t \) the cell averages of the solution,

\[
\overline{U}_{j,k}(t) \approx \int_{C_{j,k}} U(x,y,t) \, dx \, dy,
\]

are available. We then perform a piecewise linear reconstruction

\[
\tilde{U}(x,y;t) = \sum_j \mathcal{P}_{j,k}(x,y;t) \chi_{j,k}(x,y),
\]

where \( \chi_{j,k} \) is a characteristic function of the cell \( C_{j,k} \) and \( \mathcal{P}_{j,k} \) is a linear piece

\[
\mathcal{P}_{j,k}(x,y;t) = \overline{U}_{j,k}(t) + (U_x)_{j,k}(x-x_j) + (U_y)_{j,k}(y-y_k).
\]

As in the 1-D case, the slopes in (4.2) are to be computed using a nonlinear limiter to prevent oscillations. The reconstruction (4.1), (4.2) is used to evaluate the following four point values in each cell, that is,

\[
U_{j,k}^W(t) = \overline{U}_{j,k}(t) - \frac{\Delta x}{2}(U_x)_{j,k}, \quad U_{j,k}^E(t) = \overline{U}_{j,k}(t) + \frac{\Delta x}{2}(U_x)_{j,k},
\]

\[
U_{j,k}^S(t) = \overline{U}_{j,k}(t) - \frac{\Delta y}{2}(U_y)_{j,k}, \quad U_{j,k}^N(t) = \overline{U}_{j,k}(t) + \frac{\Delta y}{2}(U_y)_{j,k},
\]

and then the directional one-sided local speeds of propagation can be estimated as follows:

\[
a_{j+1/2,k}^+(t) = \max \left\{ \lambda_N \left( \frac{\partial F}{\partial U}(U_{j,k}^E(t)) \right), \lambda_N \left( \frac{\partial F}{\partial U}(U_{j+1,k}^W(t)) \right), 0 \right\}, \quad (4.3a)
\]

\[
a_{j+1/2,k}^-(t) = \min \left\{ \lambda_1 \left( \frac{\partial F}{\partial U}(U_{j,k}^E(t)) \right), \lambda_1 \left( \frac{\partial F}{\partial U}(U_{j+1,k}^W(t)) \right), 0 \right\}, \quad (4.3b)
\]

\[
b_{j,k+1/2}^+(t) = \max \left\{ \lambda_N \left( \frac{\partial G}{\partial U}(U_{j,k}^N(t)) \right), \lambda_N \left( \frac{\partial G}{\partial U}(U_{j,k+1}^S(t)) \right), 0 \right\}, \quad (4.3c)
\]

\[
b_{j,k+1/2}^-(t) = \min \left\{ \lambda_1 \left( \frac{\partial G}{\partial U}(U_{j,k}^N(t)) \right), \lambda_1 \left( \frac{\partial G}{\partial U}(U_{j,k+1}^S(t)) \right), 0 \right\}. \quad (4.3d)
\]

Equipped with the reconstructed point values (4.2) and one-sided local speeds (4.3), the cell averages are evolved in time by solving the following system of ODEs:

\[
\frac{d}{dt} \overline{U}_{j,k} = -\frac{\mathcal{F}_{j+1/2,k} - \mathcal{F}_{j-1/2,k}}{\Delta x} - \frac{\mathcal{G}_{j,k+1/2} - \mathcal{G}_{j,k-1/2}}{\Delta y}, \quad (4.4)
\]
where $\mathcal{F}$ and $\mathcal{G}$ are central-upwind numerical fluxes. There are several versions of the central-upwind fluxes. For example, the $x$- and $y$-fluxes used in Chertock et al. (2018) are

\[
\mathcal{F}_{j+1/2,k} = \frac{a_{j+1/2,k}^+ F(U_{j,k}^E) - a_{j+1/2,k}^- F(U_{j+1,k}^W)}{a_{j+1/2,k}^+ - a_{j+1/2,k}^-} \\
+ \frac{a_{j+1/2,k}^+ a_{j+1/2,k}^-}{a_{j+1/2,k}^+ - a_{j+1/2,k}^-} \left[ U_{j+1,k}^W - U_{j,k}^E - \delta U_{j+1/2,k} \right]
\]

(4.5)

and

\[
\mathcal{G}_{j,k+1/2} = \frac{b_{j,k+1/2}^+ G(U_{j,k}^N) - b_{j,k+1/2}^- G(U_{j,k+1}^S)}{b_{j,k+1/2}^+ - b_{j,k+1/2}^-} \\
+ \frac{b_{j,k+1/2}^+ b_{j,k+1/2}^-}{b_{j,k+1/2}^+ - b_{j,k+1/2}^-} \left[ U_{j,k+1}^S - U_{j,k}^N - \delta U_{j,k+1/2} \right]
\]

(4.6)

respectively. Here,

\[
\delta U_{j+1/2,k} := \text{minmod}( U_{j+1,k}^W - U_{j+1/2,k}^*, U_{j+1/2,k}^* - U_{j,k}^E )
\]

(4.7)

and

\[
\delta U_{j,k+1/2} := \text{minmod}( U_{j,k+1}^S - U_{j,k+1/2}^*, U_{j,k+1/2}^* - U_{j,k}^N )
\]

(4.8)

are the built-in anti-diffusion terms, and

\[
U_{j+1/2,k}^* = \frac{a_{j+1/2,k}^+ U_{j+1,k}^W - a_{j+1/2,k}^- U_{j,k}^E - \{ F(U_{j+1,k}^W) - F(U_{j,k}^E) \}}{a_{j+1/2,k}^+ - a_{j+1/2,k}^-}
\]

(4.9)

and

\[
U_{j,k+1/2}^* = \frac{b_{j,k+1/2}^+ U_{j,k+1}^S - b_{j,k+1/2}^- U_{j,k}^N - \{ G(U_{j,k+1}^S) - G(U_{j,k}^N) \}}{b_{j,k+1/2}^+ - b_{j,k+1/2}^-}
\]

(4.10)

are the corresponding intermediate values. Note that as in the 1-D case, all of the indexed quantities in (4.4)–(4.10) are time-dependent, but for the sake of brevity we omit this dependence here and in the rest of this paper.

\textbf{Remark 4.1.} We would like to stress that the numerical fluxes (4.5) and (4.6) are only second-order accurate (even if the piecewise polynomial reconstruction (4.1) is of a higher order). Fourth-order central-upwind fluxes were developed in Kurganov et al. (2001) and Kurganov and Petrova (2001) and higher-order fluxes can be derived in a similar way.
5. Well-balanced positivity-preserving semi-discrete central-upwind schemes for the Saint-Venant system

In this section, we describe semi-discrete central-upwind schemes for the 1-D and 2-D Saint-Venant systems (2.3) and (1.1) and their modifications obtained by taking into account the bottom friction.

5.1. Central-upwind schemes for the one-dimensional Saint-Venant system

We first consider the 1-D Saint-Venant system (2.3), which is the 1-D system of balance laws (2.1) with

\[ U = (h, q)^\top, \quad F(h, q) = \left( q, \frac{q^2}{h} + \frac{gh^2}{2} \right)^\top \text{ and } S(h; B) = (0, -ghB_x)^\top. \]

A semi-discrete scheme for the system of balance laws (2.2) reads as

\[
\frac{d}{dt} \bar{U}_j = -\frac{1}{\Delta x} \left[ F_{j+1/2} - F_{j-1/2} \right] + \bar{S}_j, \tag{5.1}
\]

where

\[
\bar{S}_j \approx \frac{1}{\Delta x} \int_{C_j} S(U) \, dx \tag{5.2}
\]

is an approximation of the source term cell average.

When the semi-discrete central-upwind scheme for the system of conservation laws (2.2) is extended to the system of balance laws (2.1), the fluxes are still given by (3.15)–(3.17) and one only needs to select an appropriate quadrature for the integral on the right-hand side of (5.2). This seems to be quite easy, but the problem here is that the use of a standard quadrature, for example the midpoint rule that results in

\[
\bar{S}_j^{(2)} = -ghB_x(x_j), \tag{5.3}
\]

leads to a non-well-balanced scheme, which does not respect the balance between the flux and source terms in (2.3). As demonstrated in Kurganov and Levy (2002, Example 1), the resulting scheme (5.1)–(5.3), (3.15)–(3.17) is incapable of exactly preserving steady-state solutions of (2.3) and, as a result, it fails to accurately capture small perturbations of steady-state (quasi-steady-state) solutions, as the truncation error of the scheme may be larger than the magnitude of the waves to be captured.

In general, a good well-balanced numerical scheme for the 1-D Saint-Venant system should be able to exactly preserve smooth steady-state solutions of (2.3), which are given by

\[ q \equiv \hat{q} = \text{const.}, \quad E := \frac{u^2}{2} + g(h + B) \equiv \hat{E} = \text{const.} \tag{5.4} \]
In fact, the most important (from the practical point of view) steady state out of those in (5.4) is a so-called ‘lake at rest’ state,
\[ q \equiv 0, \quad w = h + B \equiv \bar{w} = \text{const.}, \]
which corresponds to still water with a flat water surface. We note that in many applications the water waves that must be captured are, in fact, small perturbations of the ‘lake at rest’ steady state.

5.1.1. Still-water equilibria-preserving central-upwind scheme
In order to design a still-water equilibria-preserving central-upwind scheme, we follow the main ideas presented in Kurganov and Levy (2002) and Kurganov and Petrova (2007) and proceed in the following way.

Step 1: piecewise linear reconstruction of the bottom topography. We first replace the original bottom topography function with its continuous piecewise linear interpolant consisting of the linear pieces that connect the points \( (x_j+1/2, B_j+1/2) \):
\[ \tilde{B}(x) = B_{j-1/2} + (B_{j+1/2} - B_{j-1/2}) \cdot \frac{x - x_{j-1/2}}{\Delta x}, \quad x \in C_j, \] (5.6)
schematically shown in Figure 5.1. Here,
\[ B_{j+1/2} := \frac{B(x_{j+1/2} + 0) + B(x_{j+1/2} - 0)}{2}, \]
which reduces to \( B_{j+1/2} = B(x_{j+1/2}) \) if \( B \) is continuous at \( x = x_{j+1/2} \).

We note that since \( \tilde{B} \) is a piecewise linear function, its point value at \( x = x_j \) coincides with its cell average over the cell \( C_j \) and is also equal to
the average of the values of $\tilde{B}$ at the endpoints of $C_j$, namely,
\[ B_j := \tilde{B}(x_j) = \overline{B}_j = \frac{1}{\Delta x} \int_{C_j} \tilde{B}(x) \, dx = \frac{B_{j+1/2} + B_{j-1/2}}{2}. \] (5.7)

**Step 2: reconstruction of equilibrium variables.** One of the key components of well-balanced schemes is performing a piecewise polynomial reconstruction of the equilibrium variables, $w$ and $q$, rather than the conservative ones, $h$ and $q$. To explain this, we notice that if at some time level $\tilde{w}_j := \tilde{h}_j + B_j \equiv \tilde{w}$ and $\tilde{q}_j \equiv 0$ for all $j$, then all of the point values $w^\pm_{j+1/2}$ will assuredly have exactly the same value $\tilde{w}$ only if $w$ (which is constant at ‘lake at rest’ steady states) is reconstructed. If, instead of $w$, the water depth $h$ is reconstructed, then it may happen that $h^+_{j+1/2} \neq h^-_{j+1/2}$ and hence $w^+_{j+1/2} = h^+_{j+1/2} + B_{j+1/2} \neq w^-_{j+1/2} = h^-_{j+1/2} + B_{j+1/2}$, which will make the resulting reconstruction non-well-balanced.

**Step 3: well-balanced discretization of the source term.** After $w$ and $q$ are reconstructed, the point values of $q$ at the ‘lake at rest’ steady state (5.5) are $q^\pm_{j+1/2} \equiv 0$ and the point values of $h$ are obtained from $h^\pm_{j+1/2} = w^\pm_{j+1/2} - B_{j+1/2} = \tilde{w} - B_{j+1/2}$. The fluxes at $x = x_{j+1/2}$ are then
\[ \mathbf{F}(h^\pm_{j+1/2}, q^\pm_{j+1/2}) = \left( 0, \frac{g}{2}(\tilde{w} - B_{j+1/2})^2 \right)^T. \]
Thus, the numerical fluxes (3.15)–(3.17) at the ‘lake at rest’ steady state are
\[ \mathcal{F}^{(1)}_{j+1/2} = 0 \quad \text{and} \quad \mathcal{F}^{(2)}_{j+1/2} = \frac{g}{2}(\tilde{w} - B_{j+1/2})^2, \] (5.8)
and the scheme (5.1) reduces to
\[ \frac{d}{dt} \tilde{h}_j = 0, \quad \frac{d}{dt} \tilde{q}_j = g(\tilde{w} - B_j) \cdot \frac{B_{j+1/2} - B_{j-1/2}}{\Delta x} + \overline{S}^{(2)}_j, \] (5.9)
where
\[ \overline{S}^{(2)}_j \approx -\frac{1}{\Delta x} \int_{C_j} ghB_x \, dx. \]

It is now clear that the right-hand side of (5.9) will vanish (and hence the designed central-upwind scheme will be well-balanced) if the following well-balanced quadrature is used to evaluate $\overline{S}^{(2)}_j$:
\[ \overline{S}^{(2)}_j = -g \tilde{h}_j \cdot \frac{B_{j+1/2} - B_{j-1/2}}{\Delta x}. \] (5.10)
Note that this well-balanced quadrature can be, in fact, obtained by replacing $B_x(x_j)$ in the midpoint quadrature (5.3) with its finite-difference approximation $(B_{j+1/2} - B_{j-1/2})/\Delta x$. 

Available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0962492918000028

Downloaded from https://www.cambridge.org/core. University of Maryland College Park, on 05 Jan 2019 at 17:01:26, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0962492918000028
Remark 5.1. We would like to point out that the quadrature (5.10) is well-balanced not only for the central-upwind scheme, but for any semi-discrete scheme with the numerical flux satisfying (5.8) at ‘lake at rest’ steady states. This was first noticed in Audusse et al. (2004).

5.1.2. Moving-water equilibria-preserving central-upwind scheme
Designing moving-water equilibria-preserving central-upwind schemes is a more complicated task, which has recently been accomplished in Cheng and Kurganov (2016) and Cheng et al. (2018) in two different ways. We now briefly describe the moving-water equilibria-preserving scheme from Cheng and Kurganov (2016), which was designed following the key steps from Xing, Shu and Noelle (2011).

Step 1: computation of $\tilde{E}_j$. Given the cell averages $\tilde{h}_j$ and $\tilde{q}_j$ at a certain time level, we first compute $u_j = \frac{\tilde{q}_j}{\tilde{h}_j}$ and $E_j = \frac{u_j^2}{2} + g(\tilde{h}_j + B_j)$.

Note that at the moving-water steady state (5.4), $E_j \equiv \tilde{E}$, so that in this case, the equilibrium variables are $q$ and $E$.

Step 2: reconstruction of equilibrium variables. We now reconstruct $q$ and $E$ instead of $q$ and $h$ (or $w$). This guarantees that if $E_j \equiv \tilde{E}$ and $\tilde{q}_j \equiv 0$ for all $j$, then all of the reconstructed point values satisfy $E^\pm_{j+1/2} \equiv \tilde{E}$ and $q^\pm_{j+1/2} \equiv 0$.

Step 3: evaluation of point values of $h$. After computing $q^\pm_{j+1/2}$ and $E^\pm_{j+1/2}$, we recover the corresponding point values of $h$ by solving the following cubic equation:

$$E^\pm_{j+1/2} = \frac{(q^\pm_{j+1/2})^2}{2(h^\pm_{j+1/2})^2} + g(h^\pm_{j+1/2} + B_{j+1/2})$$

for $h^\pm_{j+1/2}$. Details on how to find the unique physically relevant solution of (5.11) can be found in Cheng and Kurganov (2016).

Step 4: well-balanced discretization of the source term. We now assume that the reconstructed point values satisfy $q^\pm_{j+1/2} \equiv \tilde{q}$ and $E^\pm_{j+1/2} \equiv \tilde{E}$ for all $j$ (note that this implies $h^+_{j+1/2} = h^-_{j+1/2} =: h_{j+1/2}$ for all $j$), and evaluate the corresponding central-upwind numerical fluxes (3.15)–(3.17):

$$\mathcal{F}^{(1)}_{j+1/2} = \tilde{q} \quad \text{and} \quad \mathcal{F}^{(2)}_{j+1/2} = \frac{\tilde{q}^2}{h_{j+1/2}} + \frac{g}{2} h_{j+1/2}^2.$$
We then take into account that $E_{j+1/2} = E_{j-1/2}$, that is,

$$g(h_{j+1/2} - h_{j-1/2}) = -\frac{\hat{q}^2}{2} \left( \frac{1}{h_{j+1/2}^2} - \frac{1}{h_{j-1/2}^2} \right) - g(B_{j+1/2} - B_{j-1/2}),$$

and substitute this together with (5.12) into the semi-discrete scheme (5.1) to obtain

$$\frac{d}{dt} h_j = 0,$$

$$\frac{d}{dt} \tilde{q}_j = g \left( \frac{h_{j+1/2} + h_{j-1/2}}{2} \cdot \frac{B_{j+1/2} - B_{j-1/2}}{\Delta x} - \frac{h_{j+1/2} - h_{j-1/2}}{4\Delta x} \left( \frac{\hat{q}}{h_{j+1/2}} - \frac{\hat{q}}{h_{j-1/2}} \right)^2 + \tilde{S}_j^{(2)}. \right)$$

Once again, we complete the construction of a well-balanced, moving-water equilibria-preserving central-upwind scheme by designing the well-balanced quadrature for $\tilde{S}_j^{(2)}$, which will guarantee that the right-hand side of (5.13) vanishes:

$$\tilde{S}_j^{(2)} = g \left( \frac{h_{j+1/2}^+ + h_{j-1/2}^-}{2} \cdot \frac{B_{j+1/2} - B_{j-1/2}}{\Delta x} + \frac{h_{j+1/2}^- - h_{j-1/2}^+}{4\Delta x} (u_{j+1/2}^- - u_{j-1/2}^+)^2 \right).$$

Note that the obtained quadrature (5.14) is second-order accurate since the term $(u_{j+1/2}^- - u_{j-1/2}^+)^2 = O((\Delta x)^2)$ for smooth solutions.

### 5.1.3. Positivity-preserving central-upwind scheme

When the water surface $w$ is reconstructed (as in the still-water equilibria-preserving scheme described in Section 5.1.1), it may happen that at some cell $C_j$ either $w_{j+1/2}^- < B_{j+1/2}$ or $w_{j-1/2}^+ < B_{j-1/2}$ even when $\bar{w}_j \geq B_j$. Such a possibility is shown schematically in Figure 5.2, in which the first-order reconstruction $\tilde{w}(x)$ is clearly below $\bar{B}(x)$ for some values of $x$ including several cell interfaces. Obviously, the use of conventional nonlinear limiters designed to minimize oscillations will not help to prevent the appearance of negative point values of $h$ in (almost) dry areas.

In order to ensure that the reconstructed values of $h$ are non-negative, we take the following steps.

**Step 1: positivity correction of the water surface reconstruction $\tilde{w}$.** When a part of the linear piece of $\tilde{w}$ in cell $C_j$ is below the corresponding linear piece of $\bar{B}$, we need to modify the slope $(w_x)_j$ to prevent the appearance of any negative values of $h$. This can be done in several ways. In Kurganov
and Petrova (2007), we proposed the following simple correction algorithm (an alternative correction procedure will be described in Section 5.3 below):

If \( w_{j+1/2}^- < B_{j+1/2} \), then take \( (w_x)_j = \frac{B_{j+1/2} - \bar{w}_j}{\Delta x/2} \)

\[ \Rightarrow \quad w_{j+1/2}^- = B_{j+1/2}, \quad w_{j-1/2}^+ = 2\bar{w}_j - B_{j+1/2}; \quad (5.15a) \]

If \( w_{j-1/2}^+ < B_{j-1/2} \), then take \( (w_x)_j = \frac{\bar{w}_j - B_{j-1/2}}{\Delta x/2} \)

\[ \Rightarrow \quad w_{j+1/2}^- = 2\bar{w}_j - B_{j-1/2}, \quad w_{j-1/2}^+ = B_{j-1/2}. \quad (5.15b) \]

It is obvious that this correction procedure guarantees that the resulting reconstruction \( \tilde{w} \) will remain conservative and stay above the piecewise linear approximant of the bottom topography function \( \tilde{B} \); see Figure 5.3(b). Therefore, all of the corrected values of \( h_{j+1/2}^\pm \) will be non-negative.

**Step 2: velocity desingularization.** After the water surface correction performed in Step 1, all of the values of \( h \) will be non-negative. However, they may be very small or even zero. This will not allow one to (accurately) compute the velocities \( u_{j+1/2}^\pm \), required in the computation of numerical flux and local speeds of propagation. In order to avoid the division by very small numbers, one should desingularize the velocity computation. This can be
done in one of the following ways (for simplicity we omit the ± and \( j \pm 1/2 \) indexes):

\[
    u = \begin{cases} 
        q/h & \text{if } h \geq \varepsilon, \\
        0 & \text{otherwise,} 
    \end{cases} \quad (5.16)
\]

\[
    u = \frac{2hq}{h^2 + \max(h^2, \varepsilon^2)}, \quad (5.17)
\]

\[
    u = \frac{\sqrt{2}hq}{\sqrt{h^4 + \max(h^4, \varepsilon^4)}}, \quad (5.18)
\]

where \( \varepsilon \) is a small \textit{a priori} chosen positive number (it should be selected based on some practical considerations that would suggest what values of \( h \) may be considered negligibly small for each problem at hand). Each of the desingularization formulae (5.16)–(5.18) has its advantages and disadvantages; we refer the reader to Chertock \textit{et al.} (2015b) and Kurganov and Petrova (2007) for discussions on this matter.

As one can easily see, (5.16)–(5.18) reduce to \( u = q/h \) for large values of \( h \), but when \( h \) is small, the entire scheme will remain consistent only if we recompute the discharge using

\[
    q := h \cdot u,
\]

where \( u \) is computed by one of (5.16), (5.17) or (5.18).

**Remark 5.2.** Instead of desingularizing the computation of the velocity point values \( u_{j+1/2}^\pm \), one can implement an alternative approach: desingularize the computation of \( u_j = \tilde{q}_j/\tilde{h}_j \) and then reconstruct a piecewise linear interpolant \( \tilde{u} \). In this case, the point values of \( u \) are obtained by

\[
    u_{j+1/2}^+ = u_{j+1} - \frac{\Delta x}{2}(u_x)_{j+1}, \quad u_{j+1/2}^- = u_j + \frac{\Delta x}{2}(u_x)_j,
\]
where the slopes \((u_x)_j\) are computed using a nonlinear limiter. The discharge point values are then computed using the reconstructed values of \(h\) and \(u\), namely, by \(q_{j+1/2}^\pm = h_{j+1/2}^\pm \cdot u_{j+1/2}^\pm\).

We note that the resulting method will still be capable of exactly preserving ‘lake at rest’ steady states at which both \(q \equiv 0\) and \(u \equiv 0\). For moving-water equilibria, however, \(q \equiv \text{const.}\) but \(u \not\equiv \text{const.}\), and thus one cannot reconstruct \(u\) instead of \(q\) while designing a moving-water equilibria-preserving schemes.

**Step 3: adaptive time-step control.** As was proved in Kurganov and Petrova (2007, Theorem 2.1), if the resulting system of ODEs (5.1) is numerically solved using the forward Euler method, then all of the evolved cell averages of \(h\) will be non-negative provided the CFL number is taken to be smaller than 1/2:

\[
\Delta t \leq \Delta t^* := \frac{\Delta x}{2a}, \quad a = \max_j \left\{a^+_{j+1/2}, -a^-_{j-1/2}\right\},
\]

(5.19)

where \(a^\pm_{j+1/2}\) are the local propagation speeds defined in (3.14). A similar result will be true for higher-order explicit SSP methods. However, as pointed out in Chertock et al. (2015b), the time steps should be selected adaptively in the following way (described here for the three-stage third-order SSP Runge–Kutta method).

While the solutions of (5.1) is evolved from time \(t\) to \(t + \Delta t\) by the three-stage third-order SSP Runge–Kutta method, there will be two intermediate solutions, which we will denote by \(\overline{U}^I\) and \(\overline{U}^II\), respectively. We would like to emphasize that Theorem 2.1 from Kurganov and Petrova (2007) directly applies to the first stage of the three-stage third-order SSP Runge–Kutta method and hence the time-step restriction (5.19) guarantees the positivity of \(\overline{h}^I_j\) for all \(j\) provided \(h_j(t) \geq 0\) for all \(j\). The positivity of \(\overline{h}^II_j\) and \(\overline{h}(t + \Delta t)\) will then be ensured by the same theorem provided \(\Delta t \leq \Delta t^*_I\) and \(\Delta t \leq \Delta t^*_II\), where \(\Delta t^*_I\) and \(\Delta t^*_II\) are computed using (5.19) applied to the intermediate solutions \(\overline{U}^I\) and \(\overline{U}^II\), respectively. In order to satisfy all of the aforementioned time-step restrictions, the following adaptive strategy should be implemented.

(i) Given the solution \(\overline{U}(t)\), set \(\Delta t := \kappa \Delta t^*_*,\) where \(\kappa \in (0, 1)\) and \(\Delta t^*_*\) is given by (5.19).

(ii) Use \(\Delta t\) to compute \(\overline{U}^I\) at the first stage of the three-stage third-order SSP Runge–Kutta method.

(iii) Given the intermediate solution \(\overline{U}^I\), compute \(\Delta t^I_*\) by (5.19).

(iv) If \(\Delta t^I_* < \Delta t\), set \(\Delta t := \kappa \Delta t^I_*\) and go back to step (ii).
(v) Use $\Delta t$ to compute $\overline{U}^I$ at the second stage of the three-stage third-order SSP Runge–Kutta method.

(vi) Given the intermediate solution $\overline{U}^I$, compute $\Delta t^I_* \equiv \Delta t$ by (5.19).

(vii) If $\Delta t^I_* < \Delta t$, set $\Delta t := \kappa \Delta t^I_*$ and go back to step (ii).

(viii) Use $\Delta t$ to compute $\overline{U}(t + \Delta t)$ at the third stage of the three-stage third-order SSP Runge–Kutta method.

Note that in Chertock et al. (2015b), we have used $\kappa = 0.9$.

5.1.4. Capturing wet/dry fronts

We would like to point out that when $h = 0$, the ‘lake at rest’ steady state (5.5) reduces to

$$q \equiv 0, \quad h \equiv 0,$$

which can be viewed as a ‘dry lake’. A good numerical scheme may be considered ‘truly’ well-balanced when it is capable of exactly preserving both ‘lake at rest’ and ‘dry lake’ steady states, as well as their combinations corresponding to the situations, in which the spatial domain $X$ is split into two non-overlapping parts $X_1$ (wet area) and $X_2$ (dry area) and the solution satisfies (5.5) in $X_1$ and (5.20) in $X_2$.

Unfortunately, the central-upwind schemes presented above are not ‘truly’ well-balanced. The problem occurs at wet/dry fronts, where the water depth is very small and the reconstruction of $w$ is corrected. One way to design a ‘truly’ well-balanced central-upwind scheme is to modify the reconstruction of $w$ in partially flooded cells, as in Bollermann et al. (2013). We now briefly describe this special well-balanced wet/dry reconstruction of the water surface.

Assuming at a certain time level $t$, $\overline{w}_j \geq B_j$ for all $j$, we define the following three types of computational cells.

- **Fully flooded cell.** If $\overline{w}_j \geq \max(\hat{B}_{j-1/2}, \hat{B}_{j+1/2})$ and $\overline{h}_j := \overline{w}_j - B_j > 0$, the cell is fully flooded.

- **Partially flooded cell.** If $B_j < \overline{w}_j < \max(\hat{B}_{j-1/2}, \hat{B}_{j+1/2})$, the cell is partially flooded.

- **Dry cell.** If $\overline{w}_j = B_j$, the cell is dry.

The main idea of the well-balanced reconstruction in partially flooded cells is to allow sub-cell resolution there, that is, to allow the interpolant in these cells to consist of two linear pieces instead of just one as in fully flooded cells. For example, the first-order flat water surface reconstructions $w_j(x)$...
in fully and partially flooded cells, schematically presented in Figure 5.4, can be written as

\[ w_j(x) = \begin{cases} 
  w_j & \text{in the ‘wet’ part of the cell,} \\
  \tilde{B}(x) & \text{in the ‘dry’ part of the cell.} 
\end{cases} \] (5.21)

Here, \( w_j = \overline{w}_j \) in fully flooded cells, and in partially flooded cells the value of \( w_j \) is determined from the water conservation requirement stating that the area of the shaded triangle in Figure 5.4(b) is equal to \( \Delta x \cdot \overline{h}_j \), which uniquely determines both \( w_j \) and the location of the wet/dry interface \( x_w^* \); see Bollermann et al. (2013) for details.

Before proceeding with the description of the second-order well-balanced wet/dry reconstruction of the water surface, we note that the positivity correction presented in Section 5.1.3 does not preserve ‘lake at rest’ steady states near wet/dry interfaces. In order to illustrate this, we show a typical situation in Figure 5.5. Let us assume that the actual water surface is flat so that in the fully flooded cell \( C_{j+1} \), \( w_{j+1/2}^+ = \overline{w}_j \) = \( \hat{w} \), but one can clearly see that the corrected value of \( w_{j+1/2}^- \) is different from \( \hat{w} \) and hence some artificial waves will be generated at \( x = x_{j+1/2} \).

In order to make sure that the water surface interpolant in partially flooded cells respects the ‘lake at rest’ steady state, we proceed as follows. Let us assume that cell \( C_j \) is partially flooded, cell \( C_{j+1} \) is fully flooded, and that \( B_{j-1/2} > \overline{w}_j > \tilde{B}_{j+1/2} \) (the case \( B_{j+1/2} > \overline{w}_j > B_{j-1/2} \) can be treated similarly; in the case when cell \( C_{j+1} \) is partially flooded, we simply use the
Figure 5.5. Conservative and positivity-preserving, but not well-balanced piecewise linear reconstruction described in Section 5.1.3.

first-order flat water surface reconstruction (5.21)), as shown in Figure 5.6. We then set the value of $w$ at $x = x_{j+1/2}$ to be $w_{j+1/2}^- := w_{j+1/2}^+$, and then use the water conservation requirement to uniquely determine one of the following.

(i) The water surface value $w_{j-1/2}^+$ in the case when the total amount of water is quite large (as in Figure 5.6(a)), which leads to the reconstruction in cell $C_j$ consisting of just one linear piece:

$$\tilde{w}(x) = w_{j-1/2}^+ + (w_{j+1/2}^- - w_{j-1/2}^+) \cdot \frac{x - x_{j-1/2}}{\Delta x}, \quad x \in C_j.$$  

(ii) The point $x^*_R \in C_j$ such that the total amount of water is quite small and can be placed in the interval $[x^*_R, x_{j+1/2}]$ (as in Figure 5.6(b)), which leads to the reconstruction in cell $C_j$ consisting of the following two linear pieces:

$$\tilde{w}(x) = \begin{cases} 
\tilde{B}(x^*_R) + (w_{j+1/2}^- - \tilde{B}(x^*_R)) \cdot \frac{x - x^*_R}{x_{j+1/2} - x^*_R} & \text{if } x \in [x^*_R, x_{j+1/2}], \\
\tilde{B}(x) & \text{if } x \in [x_{j-1/2}, x^*_R]. 
\end{cases}$$

We refer the reader to Bollermann et al. (2013) for details.

5.1.5. Positivity-preserving via ‘draining’ time-step technique

While the original still-water equilibria-preserving central-upwind scheme from Kurganov and Petrova (2007) described in Sections 5.1.1 and 5.1.3 is proved to be positivity-preserving, neither the moving-water equilibria-preserving central-upwind schemes from Cheng et al. (2018) and Cheng and Kurganov (2016) (one of which is described in Section 5.1.2 above) nor the ‘truly’ well-balanced central-upwind scheme from Bollermann et al. (2013) described in Section 5.1.4 are positivity-preserving. However, the positivity of the computed water depth can be enforced using the ‘draining’ time-step technique introduced in Bollermann et al. (2011) and briefly described below.
Figure 5.6. Well-balanced second-order water surface reconstruction in cases (i) (a) and (ii) (b). $x_R^*$ is the reconstructed location of the wet/dry interface. In both cases, the dashed line represents the corresponding well-balanced first-order water surface reconstructions.

The key idea of the ‘draining’ time-step technique is to limit the outgoing fluxes in draining cells according to the following algorithm, which we present in the case of forward Euler time discretization, which for the first equation in (5.1) results in

$$\tilde{h}_{j}^{n+1} = \tilde{h}_{j}^{n} - \frac{\Delta t^{n}}{\Delta x} (F_{j+1/2}^{(1)} - F_{j-1/2}^{(1)}). \tag{5.22}$$

Here, the numerical fluxes are evaluated at the time level $t = t^n$ and $\Delta t^n$ is the time step restricted by (5.19).

In order to guarantee the positivity of $\tilde{h}_{j}^{n+1}$ (assuming that $\tilde{h}_{j}^{n} \geq 0$ for all $j$), we first introduce the so-called ‘draining’ time step

$$\Delta t_{j}^{\text{drain}} := \frac{\tilde{h}_{j}^{n} \Delta x}{\max(0, F_{j+1/2}^{(1)}) + \max(0, -F_{j-1/2}^{(1)})},$$

which describes the time when the water contained in cell $C_j$ at the beginning of the time step would have left via the outflow fluxes. We then replace the evolution step (5.22) by

$$\tilde{h}_{j}^{n+1} = \tilde{h}_{j}^{n} - \frac{\Delta t^{n}}{\Delta x} F_{j+1/2}^{(1)} - \frac{\Delta t^{n}}{\Delta x} F_{j-1/2}^{(1)},$$

where we set the effective time steps at the cell interfaces to be

$$\Delta t_{j+1/2}^{n} = \min(\Delta t^{n}, \Delta t_{i}^{\text{drain}}), \quad i = j + \frac{1}{2} - \frac{\text{sgn}(F_{j+1/2}^{(1)})}{2}.$$  

This guarantees that $\tilde{h}_{j}^{n+1} \geq 0$ for all $j$. 

---

Available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0962492918000028
Remark 5.3. We note that the ‘draining’ time-step technique can be applied to any Godunov-type finite-volume methods, not necessarily central-upwind schemes.

Remark 5.4. An extension of the ‘draining’ time-step technique to the 2-D case is quite straightforward.

5.2. Central-upwind schemes for the two-dimensional Saint-Venant system

In this section, we consider the 2-D Saint-Venant system (1.1), which is the 2-D system of balance laws (1.2) with

$$
U = \begin{pmatrix}
  h \\
  q^x \\
  q^y
\end{pmatrix}, \quad F(U) = \begin{pmatrix}
  q^x \\
  \frac{(q^x)^2}{h} + \frac{gh^2}{2} \\
  q^x q^y
\end{pmatrix}, \quad G(U) = \begin{pmatrix}
  q^y \\
  q^x q^y \\
  \frac{(q^y)^2}{h} + \frac{gh^2}{2}
\end{pmatrix},
$$

and $S(h; B) = (0, -ghB_x, -ghB_y)^T$.

For the sake of simplicity, we will only consider central-upwind schemes on Cartesian grids as we have done in Section 4 above. We refer the reader to Bryson et al. (2011) and Liu et al. (2018), Shirkhani et al. (2016) and Beljadid et al. (2016) for the central-upwind schemes for the 2-D Saint-Venant system on triangular, quadrilateral and general polygonal grids, respectively.

A semi-discrete scheme for the system of balance laws (1.2) reads as

$$
\frac{d}{dt} \overline{U}_{j,k} = -\frac{1}{\Delta x} \left[ F_{j+1/2,k} - F_{j-1/2,k} \right] - \frac{1}{\Delta y} \left[ G_{j,k+1/2} - G_{j,k-1/2} \right] + \overline{S}_{j,k},
$$

where

$$
\overline{S}_{j,k} \approx \frac{1}{\Delta x \Delta y} \int_{C_{j,k}} S(U) \, dx \, dy
$$

is an approximation of the source term cell average.

When the semi-discrete central-upwind scheme for the system of conservation laws (2.4) is extended to the system of balance laws (1.2), the fluxes are still given by (4.5)–(4.10), and one only needs to select an appropriate quadrature for the integral on the right-hand side of (5.24).

As in the 1-D case, a good well-balanced numerical scheme for the 2-D Saint-Venant system should be able to exactly preserve smooth steady-state solutions of (1.1), which generally have a much more complicated structure than their 1-D counterparts. However, the ‘lake at rest’ steady state can still be easily found explicitly and it is given by

$$
q^x \equiv q^y \equiv 0, \quad w = h + B \equiv \hat{w} = \text{const.}
$$

In the rest of Section 5.2, we will refer to a central-upwind scheme as well-balanced if it is capable of exactly preserving (5.25).
5.2.1. Well-balanced central-upwind scheme

In order to design a well-balanced central-upwind scheme, we follow the main ideas presented in Kurganov and Levy (2002) and Kurganov and Petrova (2007), and proceed as follows.

Step 1: piecewise bilinear reconstruction of the bottom topography. We start by replacing the original bottom topography function $B$ with its continuous piecewise bilinear approximation $\tilde{B}$, which at each cell $C_{j,k}$ is given by the bilinear form:

$$
\tilde{B}(x, y) = B_{j-1/2,k-1/2} + (B_{j+1/2,k-1/2} - B_{j-1/2,k-1/2}) \cdot \frac{x - x_{j-1/2}}{\Delta x} 
+ (B_{j-1/2,k+1/2} - B_{j-1/2,k-1/2}) \cdot \frac{y - y_{k-1/2}}{\Delta y}
+ (B_{j+1/2,k+1/2} - B_{j+1/2,k-1/2} - B_{j-1/2,k+1/2} + B_{j-1/2,k-1/2})
\times \frac{(x - x_{j-1/2})(y - y_{k-1/2})}{\Delta x \Delta y}, \quad (x, y) \in C_{j,k}.
$$

Here, $B_{j\pm 1/2,k\pm 1/2}$ are the values of $\tilde{B}$ at the corners of the cell $C_{j,k}$, computed according to the following formula:

$$
B_{j\pm 1/2,k\pm 1/2} := \frac{1}{2} \left( \max_{\xi^2+\eta^2=1} \lim_{h,\ell \to 0} B(x_{j\pm 1/2} + h\xi, y_{k\pm 1/2} + \ell\eta) 
+ \min_{\xi^2+\eta^2=1} \lim_{h,\ell \to 0} B(x_{j\pm 1/2} + h\xi, y_{k\pm 1/2} + \ell\eta) \right),
$$

which reduces to

$$
B_{j\pm 1/2,k\pm 1/2} = B(x_{j\pm 1/2}, y_{k\pm 1/2})
$$

if the function $B$ is continuous at $(x_{j\pm 1/2}, y_{k\pm 1/2})$.

Note that the restriction of the interpolant $\tilde{B}$ along each of the lines $x = x_j$ or $y = y_k$ is a continuous piecewise linear function, and, as in the 1-D case (see (5.7)), the cell average of $\tilde{B}$ over the cell $C_{j,k}$ is equal to its value at the centre of the cell and is also equal to the average of the values of $\tilde{B}$ at the midpoints of the edges of $C_{j,k}$. That is, we have

$$
B_{j,k} := \tilde{B}(x_j, y_k) = \overline{B}_{j,k} = \frac{1}{\Delta x \Delta y} \iint_{C_{j,k}} \tilde{B}(x, y) \, dx \, dy
= \frac{1}{4} (B_{j+1/2,k} + B_{j-1/2,k} + B_{j,k+1/2} + B_{j,k-1/2}),
$$

where

$$
B_{j+1/2,k} := \tilde{B}(x_{j+1/2}, y_k) = \frac{1}{2} (B_{j+1/2,k+1/2} + B_{j+1/2,k-1/2})
$$
and

\[ B_{j,k+1/2} := \tilde{B}(x_j, y_{k+1/2}) = \frac{1}{2}(B_{j+1/2,k+1/2} + B_{j-1/2,k+1/2}). \]

**Remark 5.5.** We would like to point out that when a triangular grid is used (as in Bryson et al. 2011, Liu et al. 2018), the bottom topography function should be approximated using the continuous piecewise linear interpolant uniquely determined by the point values of \( B \) at the triangular cell vertices. In the case of general quadrilateral (Shirkhani et al. 2016) or polygonal (Beljadid et al. 2016) grids, the finite-volume cell may be split into several triangular sub-cells and then the bottom topography function can be approximated using a continuous piecewise linear interpolant as well.

**Step 2: reconstruction of equilibrium variables.** As in the 1-D case, we reconstruct the equilibrium variables, \( w, q^x \) and \( q^y \), rather than the conservative ones, \( h, q^x \) and \( q^y \).

**Step 3: well-balanced discretization of the source term.** After \( w, q^x \) and \( q^y \) are reconstructed, the point values of \( q^x \) and \( q^y \) at the ‘lake at rest’ steady state (5.25) are

\[
(q^x_{j,k})^E \equiv (q^x_{j,k})^N \equiv (q^x_{j,k})^S \equiv 0,
\]

\[
(q^y_{j,k})^E \equiv (q^y_{j,k})^N \equiv (q^y_{j,k})^S \equiv 0,
\]

and the point values of \( h \) are given by

\[
h_{j,k}^E = w_{j,k}^E - B_{j+1/2,k} = \hat{w} - B_{j+1/2,k},
\]

\[
h_{j,k}^W = w_{j,k}^W - B_{j-1/2,k} = \hat{w} - B_{j-1/2,k},
\]

\[
h_{j,k}^N = w_{j,k}^N - B_{j,k+1/2} = \hat{w} - B_{j,k+1/2},
\]

\[
h_{j,k}^S = w_{j,k}^S - B_{j,k-1/2} = \hat{w} - B_{j,k-1/2}.
\]

The fluxes at \((x_{j+1/2}, y_k)\) and \((x_j, y_{k+1/2})\) are then

\[
F(U_{j,k}^E) = F(U_{j+1,k}^W) = \left(0, \frac{g}{2}(\hat{w} - B_{j+1/2,k})^2, 0\right)^\top,
\]

\[
G(U_{j,k}^N) = G(U_{j+1,k}^S) = \left(0, 0, \frac{g}{2}(\hat{w} - B_{j,k+1/2})^2\right)^\top.
\]

Thus the numerical fluxes (4.5)–(4.10) at the ‘lake at rest’ steady state are

\[
F_{j+1/2,k}^{(1)} = F_{j+1/2,k}^{(3)} = 0, \quad F_{j+1/2,k}^{(2)} = \frac{g}{2}(\hat{w} - B_{j+1/2,k})^2,
\]

\[
G_{j+1/2,k}^{(1)} = G_{j+1/2,k}^{(2)} = 0, \quad G_{j+1/2,k}^{(3)} = \frac{g}{2}(\hat{w} - B_{j,k+1/2})^2.
\]
and the scheme (5.23) reduces to

\[
\frac{d}{dt} \bar{h}_{j,k} = 0, \tag{5.26a}
\]

\[
\frac{d}{dt} \bar{q}^x_{j,k} = g(\bar{w} - B_{j,k}) \cdot \frac{B_{j+1/2,k} - B_{j-1/2,k}}{\Delta x} + \bar{S}^{(2)}_{j,k}, \tag{5.26b}
\]

\[
\frac{d}{dt} \bar{q}^y_{j,k} = g(\bar{w} - B_{j,k}) \cdot \frac{B_{j,k+1/2} - B_{j,k-1/2}}{\Delta y} + \bar{S}^{(3)}_{j,k}, \tag{5.26c}
\]

where

\[
\bar{S}^{(2)}_{j,k} \approx - \frac{1}{\Delta x \Delta y} \int_{C_{j,k}} ghB_x \, dx \, dy, \quad \bar{S}^{(3)}_{j,k} \approx - \frac{1}{\Delta x \Delta y} \int_{C_{j,k}} ghB_y \, dx \, dy.
\]

As in the 1-D case, the right-hand side of (5.26) must vanish to ensure that the designed central-upwind scheme is well-balanced, and therefore the following well-balanced quadratures are used to evaluate \( \bar{S}^{(2)}_{j,k} \) and \( \bar{S}^{(3)}_{j,k} \):

\[
\bar{S}^{(2)}_{j,k} = -gh_{j,k} \cdot \frac{B_{j+1/2,k} - B_{j-1/2,k}}{\Delta x}, \quad \bar{S}^{(3)}_{j,k} = -gh_{j,k} \cdot \frac{B_{j,k+1/2} - B_{j,k-1/2}}{\Delta y}.
\]

**Remark 5.6.** We stress that development of a well-balanced quadrature for the source term discretization on triangular grids is a more delicate task. One first needs to apply Green’s formula to the vector field \((\frac{1}{2}(w - B)^2, 0)^\top\), and only then does it become clear how to identify the desired quadrature; see Bryson et al. (2011) and Liu et al. (2018) for details.

5.2.2. Positivity-preserving central-upwind scheme

As in the 1-D case, the original water surface reconstruction may produce negative values of \( h = w - B \). In order to ensure that the reconstructed values of \( h \) are non-negative, we take the same three steps as in Section 5.1.3.

First, we correct the reconstruction of \( w \) by modifying the slopes \((w_x)_{j,k}\) and \((w_y)_{j,k}\) as follows:

If \( w^E_{j,k} < B_{j+1/2,k} \), then take \((w_x)_{j,k} = \frac{B_{j+1/2,k} - \bar{w}_{j,k}}{\Delta x/2}\)

\[\Rightarrow w^E_{j,k} = B_{j+1/2,k}, \quad w^W_{j,k} = 2\bar{w}_{j,k} - B_{j+1/2,k};\]

If \( w^W_{j,k} < B_{j-1/2,k} \), then take \((w_x)_{j,k} = \frac{\bar{w}_{j,k} - B_{j-1/2,k}}{\Delta x/2}\)

\[\Rightarrow w^E_{j,k} = 2\bar{w}_{j,k} - B_{j-1/2,k}, \quad w^W_{j,k} = B_{j-1/2,k};\]

If \( w^N_{j,k} < B_{j,k+1/2} \), then take \((w_y)_{j,k} = \frac{B_{j,k+1/2} - \bar{w}_{j,k}}{\Delta y/2}\)

\[\Rightarrow w^N_{j,k} = B_{j,k+1/2}, \quad w^S_{j,k} = 2\bar{w}_{j,k} - B_{j,k+1/2};\]
If \( w_{j,k}^S < B_{j,k-1/2} \), then take \((w_y)_{j,k} = \frac{\overline{w}_{j,k} - B_{j,k-1/2}}{\Delta y/2} \)

\[ \implies w_{j,k}^N = 2\overline{w}_{j,k} - B_{j,k-1/2}, \quad w_{j,k}^S = B_{j,k-1/2}. \]

This correction procedure guarantees that the resulting reconstruction \( \tilde{w} \) will remain conservative and its restrictions on the lines \( y = y_k \) and \( x = x_j \) will be above \( \overline{B}(x, y_k) \) and \( \overline{B}(x_j, y) \), respectively. Hence, all of the corrected values \( h_{j,k}^E = w_{j,k}^E - B_{j+1/2,k} \), \( h_{j,k}^W = w_{j,k}^W - B_{j-1/2,k} \), \( h_{j,k}^N = w_{j,k}^N - B_{j,k+1/2} \), and \( h_{j,k}^S = w_{j,k}^S - B_{j,k-1/2} \) will be non-negative. Notice that unlike the 1-D case, this does not guarantee the non-negativity of \( \tilde{w} - \overline{B} \) in the entire cell \( C_{j,k} \). However, this is not a problem since our goal is to preserve positivity of the cell averages of \( h \) and its point values used in the scheme \( (h_{j,k}^E, h_{j,k}^W, h_{j,k}^S, h_{j,k}^N) \).

**Remark 5.7.** In the case of a triangular grid, the positivity correction is performed in a different way: one should make sure that the point values of \( w \) at the cell vertices are above the values of \( \overline{B} \) there, and this guarantees that the entire corrected linear piece of \( w \) will stay above the corresponding linear piece of \( B \). We refer the reader to Bryson et al. (2011) and Liu et al. (2018) for details.

Second, we desingularize the velocity computation using \( u = q^x/h \) and \( v = q^y/h \) (this is done exactly as in the 1-D case described in Section 5.1.3). Third, we implement the same adaptive time-step control described in Section 5.1.3. However, the basic time-step bound that guarantees non-negativity of \( h \) is more restrictive (one has to use the CFL number to be less than or equal to 1/4 instead of 1/2 used in the 1-D case), and one has to choose

\[ \Delta t \leq \Delta t_* := \min \left\{ \frac{\Delta x}{4a}, \frac{\Delta y}{4b} \right\}, \]

where

\[ a = \max_{j,k} \{a^+_{j+1/2,k}, -a^-_{j+1/2,k} \}, \quad b = \max_{j,k} \{b^+_{j,k+1/2}, -b^-_{j,k+1/2} \}, \]

and \( a^\pm_{j+1/2,k} \) and \( b^\pm_{j,k+1/2} \) are the local propagation speeds defined in (4.3).

### 5.2.3. Capturing wet/dry fronts

Extension of the 1-D ‘truly’ well-balanced central-upwind scheme described in Section 5.1.4 to the 2-D case is highly non-trivial. In fact, it is unclear whether this can be done using a Cartesian grid and piecewise bilinear approximation of the bottom topography. Liu et al. (2018) have constructed a 2-D ‘truly’ well-balanced central-upwind scheme on triangular grids. Compared to the 1-D case, additional degrees of freedom need to be taken into...
account and more types of partially flooded cells, in which sub-cell resolution is required, are to be considered to design a special wet/dry reconstruction of the water surface. Moreover, when this reconstruction is used, an alternative discretization for the source term has to be derived in order to maintain the well-balanced property of the resulting central-upwind scheme. We refer the reader to Liu et al. (2018) for details.

5.3. Central-upwind schemes for the Saint-Venant systems with friction terms

In this section, we consider a more general shallow-water system, which is obtained by taking into account the bottom friction terms. In the 2-D case, the studied system reads as

\[
\begin{align*}
    h_t + (hu)_x + (hv)_y &= 0, \\
    (hu)_t + \left( hu^2 + \frac{g}{2} h^2 \right)_x + (huv)_y &= -ghB_x - ghI_x, \\
    (hv)_t + (huv)_x + \left( hv^2 + \frac{g}{2} h^2 \right)_y &= -ghB_y - ghI_y,
\end{align*}
\]

where the \( I_x \) and \( I_y \) are the components of the bottom friction slope. There are many ways to model friction terms; see, for example, Gerbeau and Perthame (2001) and Kellerhals (1967). One of the most popular choices is the classical Manning formulation (see e.g. Flamant 1891, Darcy 1857, Gauckler 1867, Manning 1891):

\[
I_x = \frac{n^2}{h^{4/3}} \frac{u}{\sqrt{u^2 + v^2}}, \quad I_y = \frac{n^2}{h^{4/3}} \frac{v}{\sqrt{u^2 + v^2}},
\]

where \( n \) is the Manning coefficient. Notice that if \( h \approx 0 \), the friction terms (5.28) become stiff damping terms, and this increases the level of complexity in the development of efficient numerical methods for the system (5.27), (5.28). Another difficulty is related to the fact that this system admits not only ‘lake at rest’ steady states (5.25), but other steady-state solutions, which may become more relevant in many practical situations, for example, when drainage of rain water in urban areas is simulated; see, for example, Cea, Garrido and Puertas (2010) and Cea and Vázquez-Cendón (2012) and references therein.

Let us first consider the simplest 1-D case, in which the system (5.27), (5.28) reduces to

\[
\begin{align*}
    h_t + (hu)_x &= 0, \\
    (hu)_t + \left( hu^2 + \frac{g}{2} h^2 \right)_x &= -ghB_x - g \frac{n^2}{h^{1/3}} |u|u.
\end{align*}
\]
As one can easily see, the system (5.29) admits non-trivial ($u \neq 0$) steady states in the form

$$hu \equiv \text{const.}, \quad h \equiv \text{const.}, \quad B_x \equiv \text{const.}$$

(5.30)

This solution corresponds to the situation when the water flows over a slanted infinitely long surface with a constant slope. The structure of 2-D steady states is substantially more complicated. However, the quasi-1-D steady-state solutions

$$h \equiv \text{const.}, \quad hu \equiv \text{const.}, \quad B_x \equiv \text{const.}, \quad B_y \equiv 0$$

(5.31)

and

$$h \equiv \text{const.}, \quad hu \equiv 0, \quad hv \equiv \text{const.}, \quad B_x \equiv 0, \quad B_y \equiv \text{const.}$$

(5.32)

are still physically relevant.

A well-balanced Roe-type numerical scheme which is capable of exactly preserving the steady states (5.30), (5.31) and (5.32) was proposed in Cea and Vázquez-Cendón (2012). However, to maintain the positivity of the water depth $h$, the scheme in Cea and Vázquez-Cendón (2012) may require one to use very small time steps and thus may not be robust in certain settings. Another Godunov-type scheme for the 1-D system (5.29) was proposed in Berthon, Marche and Turpault (2011). Although this method does not suffer from restrictive time-stepping, it is capable of preserving ‘lake at rest’ steady states only.

In Chertock \textit{et al.} (2015b), we have developed well-balanced positivity-preserving central-upwind schemes for both 1-D (5.29) and 2-D (5.27) systems. For the sake of brevity, we will now describe the 1-D scheme only.

The 1-D semi-discrete central-upwind scheme for (5.29) is still given by (5.1), but the discrete source term now consists of approximations of the following two integrals:

$$\bar{S}^{(2)}_j \approx -\frac{1}{\Delta x} \int_{C_j} ghB_x \, dx - \frac{1}{\Delta x} \int_{C_j} gn^2 \frac{2\bar{h}_j}{h^{1/3}} |u|u \, dx.$$  

(5.33)

One can easily show that the well-balanced quadrature for the first integral on the right-hand side of (5.33) is still given by (5.10) and the well-balanced quadrature for the second integral on the right-hand side of (5.33) is obtained using the midpoint rule and one of the desingularization formulae (5.16), (5.17) or (5.18). In Chertock \textit{et al.} (2015b), the desingularization formula (5.17) has been used and the resulting quadrature is

$$\frac{1}{\Delta x} \int_{C_j} gn^2 \frac{2\bar{h}_j}{h^{1/3}} |u|u \, dx \approx gn^2 \left( \frac{2\bar{h}_j}{h^{1/3} + \max(h_j^2, \varepsilon^2)} \right)^{7/3} |\bar{q}_j| \bar{q}_j.$$
In order to complete the construction of the semi-discrete central-upwind scheme, we proceed along the lines of Sections 3.3 and 5.1. The resulting scheme, however, will have two major drawbacks. First, it will be capable of preserving ‘lake at rest’ steady states only. Second, it will be very inefficient in the case when some (almost) dry areas are present, as the friction term may become very stiff there and explicit SSP Runge–Kutta methods will have very severe time-step stability restriction. We improve the central-upwind scheme by taking the following two steps.

**Step 1: modified positivity correction of the water surface reconstruction \( \tilde{w} \).** As one can easily see, the water surface positivity correction procedure described in Section 5.1.3 does not preserve the water surface profile that correspond to the steady state (5.30); see Figure 5.3(b). Therefore, in Chertock et al. (2015b) we have proposed an alternative positivity correction procedure presented in Figure 5.3(c) and described here.

In the cells, where the original limited water surface reconstruction produces negative values of \( h \), we make the slopes of \( w \) equal to the corresponding slopes of \( B \). Namely, we proceed as follows:

\[
\text{If } w_{j+1/2}^- < B_{j+1/2} \text{ or } w_{j-1/2}^+ < B_{j-1/2}, \text{ then take } (w_x)_j = (B_x)_j
\]

\[
\implies w_{j+1/2}^- = \bar{h}_j + B_{j+1/2}, \quad w_{j-1/2}^+ = \bar{h}_j + B_{j-1/2}. \quad (5.34)
\]

This correction (unlike the correction procedure described in Sections 5.1.3 and its more sophisticated modification described in Sections 5.1.4) will not only guarantee the positivity of \( h_{j+1/2}^\pm \) but will also be able to exactly reconstruct the steady-state solution (5.30).

**Step 2: steady state and sign preserving semi-implicit Runge–Kutta methods.** As mentioned earlier, in the presence of dry and/or almost dry areas, the explicit treatment of the friction terms imposes a severe time-step restriction, which may be several orders of magnitude smaller than a typical time step used for the corresponding friction-free Saint-Venant system.

An attractive alternative to explicit methods is that of implicit–explicit (IMEX) SSP Runge–Kutta solvers, which treat the stiff part of the underlying ODE system implicitly and thus typically have the stability domains based on the non-stiff term only; see, for example, Ascher, Ruuth and Spiteri (1997), Higueras and Roldán (2006), Hundsdorfer and Ruuth (2007), Higueras, Happenhofer, Koch and Kupka (2014) and Pareschi and Russo (2001, 2005). However, a straightforward implementation of these methods may break the discrete balance between the fluxes, geometric source and the friction terms maintained by the derived semi-discrete central-upwind scheme, and the resulting fully discrete method will not be able to preserve the relevant steady states and the positivity of the computed water depth.
In order to overcome this difficulty, we have recently developed a family of second-order semi-implicit time integration methods for systems of ODEs with stiff damping term; see Chertock, Cui, Kurganov and Tong (2015a). In these methods, only a portion of the stiff term is implicitly treated, and therefore the evolution equation is very easy to solve and implement compared to fully implicit or IMEX methods. The important feature of the ODE solvers introduced in Chertock et al. (2015a) resides in the fact that they are capable of exactly preserving the steady states as well as maintaining the sign of the computed solution under the time step restriction determined by the non-stiff part of the system only. These semi-implicit methods are based on the modification of explicit SSP Runge–Kutta methods and are proven to have a formal second order of accuracy, $A(\alpha)$-stability and stiff decay. We now briefly describe the application of the second-order semi-implicit ODE solver SI-RK3 from Chertock et al. (2015a) to the ODE system (5.1), (3.15)–(3.17), (5.33).

We first introduce the grid function of the numerical solution $\overline{U} := \{\overline{U}_j\}$.

We then denote the discretization of the sum of fluxes and geometric source term on the right-hand side of (5.1) by

$$L^{(1)}[\overline{U}]_j := -\frac{F^{(1)}_{j+1/2} - F^{(1)}_{j-1/2}}{\Delta x},$$

$$L^{(2)}[\overline{U}]_j := -\frac{F^{(2)}_{j+1/2} - F^{(2)}_{j-1/2}}{\Delta x} - g \overline{h}_j \frac{B_{j+1/2} - B_{j-1/2}}{\Delta x},$$

and introduce the discrete friction coefficient

$$\mathcal{M}(\overline{U}_j) := -gn^2 \left( \frac{2\overline{h}_j}{\overline{h}_j^2 + \max(\overline{h}_j^2, \varepsilon^2)} \right)^{7/3} |\overline{q}_j|,$$

so that the ODE system (5.1) can be rewritten as

$$\frac{d}{dt} \overline{h}_j = L^{(1)}[\overline{U}]_j,$$

$$\frac{d}{dt} \overline{q}_j = L^{(2)}[\overline{U}]_j + \mathcal{M}(\overline{U}_j) \overline{q}_j. \quad (5.35)$$

We now implement the SI-RK3 method to the system (5.35). (The SI-RK3 method is a second-order semi-implicit Runge–Kutta method based on the three-stage third-order SSP Runge–Kutta method; for details, see Chertock et al. (2015a, Section 3).) The resulting fully discrete scheme can be written as

$$\overline{h}^I_j = \overline{h}_j(t) + \Delta t L^{(1)}[\overline{U}(t)]_j, \quad (5.36a)$$

$$\frac{d}{dt} \overline{q}_j(t) = \frac{\overline{q}_j(t) + \Delta t L^{(2)}[\overline{U}(t)]_j}{1 - \Delta t \mathcal{M}(\overline{U}_j(t))}, \quad (5.36b)$$
A. Kurganov

\[
\tilde{h}^I_j = \frac{3}{4} h^I_j(t) + \frac{1}{4} \left( \tilde{h}^I_j + \Delta t \mathcal{L}^{(1)}(U^I_j) \right), \\
\tilde{q}^I_j = \frac{3}{4} q^I_j(t) + \frac{1}{4} \cdot \frac{\tilde{q}^I_j + \Delta t \mathcal{L}^{(2)}(U^I_j)}{1 - \Delta t \mathcal{M}(U^I_j)},
\]

(5.36c)

(5.36d)

\[
\tilde{h}^III_j = \frac{1}{3} \tilde{h}^I_j(t) + \frac{2}{3} \left( \tilde{h}^III_j + \Delta t \mathcal{L}^{(1)}(U^{III}_j) \right), \\
\tilde{q}^III_j = \frac{1}{3} \tilde{q}^I_j(t) + \frac{2}{3} \frac{\tilde{q}^III_j + \Delta t \mathcal{L}^{(2)}(U^{III}_j)}{1 - \Delta t \mathcal{M}(U^{III}_j)},
\]

(5.36e)

(5.36f)

\[
\tilde{h}_j(t + \Delta t) = \tilde{h}^{III}_j, \\
\tilde{q}_j(t + \Delta t) = \tilde{q}^{III}_j - \frac{(\Delta t)^2 \mathcal{L}^{(2)}(U^{III}_j) \mathcal{M}(U^{III}_j)}{1 + (\Delta t \mathcal{M}(U^{III}_j))^2},
\]

(5.36g)

(5.36h)

where

\[
\begin{align*}
U^I &= (h^I, q^I)^\top, \\
U^{III} &= (h^{III}, q^{III})^\top \quad \text{and} \quad U^{III} = (h^{III}, q^{III})^\top.
\end{align*}
\]

As has been proved in Chertock et al. (2015a), the fully discrete scheme (5.36) is well-balanced (in the sense that it is capable of preserving both ‘lake at rest’ (5.25) and ‘slanted surface’ (5.30) steady states) and positivity-preserving. The latter is, in fact, enforced by implementing an adaptive time-step control similar to the one described in Section 5.1.3.

6. Non-conservative hyperbolic systems

There are many shallow-water related models that contain non-conservative product terms. For instance, the original Saint-Venant system will contain such term(s) if the bottom topography \( B \) is discontinuous. Non-conservative product terms also arise in many multilayer/multiphase models as momentum/energy exchange terms and in many other situations.

For the simplicity of presentation, we consider a general non-conservative 1-D hyperbolic system

\[
U_t + F(U)_x = N(U)U_x,
\]

(6.1)

where \( N \in \mathbb{R}^{N \times N} \). The presence of non-conservative terms on the right-hand side of (6.1) makes both analysis and development of numerical methods for the system (6.1) very difficult tasks. In fact, when the solutions are discontinuous, which is a common feature of nonlinear hyperbolic systems, these non-conservative terms are not well defined in the distributional framework and the usual concept of weak solution cannot be used. Instead they should be understood as the Borel measures, as in Dal Maso, LeFloch and Murat (1995); see also LeFloch (2002, 2004). This concept has been

In Castro Díaz et al. (2018) we have recently developed path-conservative central-upwind schemes, in which the concept of path-conservative schemes has been incorporated into the framework of simple and robust central-upwind schemes. We are now going to describe this approach.

Before presenting path-conservative central-upwind schemes, let us consider the conservative hyperbolic system (2.2) and rewrite the semi-discrete central-upwind scheme from Kurganov et al. (2001), which is given by (3.5)–(3.7), (3.14) and a simplified central-upwind flux

\[
\mathcal{F}_{j+1/2} = \frac{a^+_j F(U^-_{j+1/2}) - a^-_j F(U^+_{j+1/2})}{a^+_j - a^-_j} + \frac{a^+_j a^-_j}{a^+_j - a^-_j} \left[ U^+_{j+1/2} - U^-_{j+1/2} \right],
\]

in an alternative form. To this end, we first define two new coefficients

\[
\alpha^+_j := \frac{-2 a^+_j a^-_j}{a^+_j - a^-_j} \quad \text{and} \quad \alpha^-_j := \frac{a^+_j + a^-_j}{a^+_j - a^-_j},
\]

and the quantities

\[
D^\pm_{j+1/2} = \frac{1}{2} \left[ (1 \pm \alpha^+_j)(F(U^+_{j+1/2}) - F(U^-_{j+1/2})) \right. \]
\[\left. \pm \alpha^-_j (U^+_{j+1/2} - U^-_{j+1/2}) \right],
\]

which represent the differences between the numerical and physical fluxes at both sides of the cell interface.

Next, let \( \Psi_{j+1/2}(s) := \Psi(s; U^-_{j+1/2}, U^+_{j+1/2}) \) denote a sufficiently smooth path that connects \( U^-_{j+1/2} \) and \( U^+_{j+1/2} \):

\[
\Psi : [0, 1] \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N \quad \text{such that} \]
\[
\Psi(0; U^-_{j+1/2}, U^+_{j+1/2}) = U^-_{j+1/2}, \quad \Psi(1; U^-_{j+1/2}, U^+_{j+1/2}) = U^+_{j+1/2}. \]

Available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0962492918000028.
Equipped with (6.3), (6.4) and taking into account that
\[ F(U_{j+1/2}^-) - F(U_{j-1/2}^+) = \int_{C_j} A(\tilde{U}(x))(U_x)_j \, dx, \quad A(U) := \frac{\partial F}{\partial U}(U), \tag{6.5} \]
we rewrite the scheme (3.5)–(3.7), (3.14), (6.2) in the following form:
\[ \frac{d}{dt} U_j = -\frac{1}{\Delta x} \left( D^+_{j+1/2} + D^-_{j+1/2} + \int_{C_j} A(\tilde{U}(x))(U_x)_j \, dx \right), \tag{6.6} \]
with
\[ D^\pm_{j+1/2} = \frac{1 \pm \alpha^{j+1/2}}{2} \int_0^1 A(\Psi_{j+1/2}(s)) \frac{d\Psi_{j+1/2}}{ds} \, ds \pm \frac{\alpha^{j+1/2}}{2} (U^+_{j+1/2} - U^-_{j+1/2}). \tag{6.7} \]

In order to design a path-conservative central-upwind scheme for the non-conservative system (6.1), we first rewrite (6.1) in the quasi-linear form
\[ U_t + A(U)U_x = 0, \quad A(U) := \frac{\partial F}{\partial U}(U) - N(U). \tag{6.8} \]
The semi-discrete scheme (6.6), (6.7) can then be directly generalized to the non-conservative system (6.8) by replacing \( A(U) \) in (6.5)–(6.7) with \( A(U) \) from (6.8). This results in the following path-conservative central-upwind scheme for (6.1):
\[ \frac{d}{dt} \bar{U}_j = -\frac{1}{\Delta x} \left[ \mathcal{F}_{j+1/2} - \mathcal{F}_{j-1/2} - \mathcal{N}_j \right. \tag{6.9} \]
\[ \left. - \frac{a^+_{j-1/2}}{a^+_{j-1/2} - a^-_{j-1/2}} \mathcal{N}_{\Psi,j-1/2} + \frac{a^-_{j+1/2}}{a^+_{j+1/2} - a^-_{j+1/2}} \mathcal{N}_{\Psi,j+1/2} \right], \]
where
\[ \mathcal{N}_j := \int_{C_j} N(\tilde{U}(x))(U_x)_j \, dx, \tag{6.10a} \]
\[ \mathcal{N}_{\Psi,j+1/2} := \int_0^1 N(\Psi_{j+1/2}(s)) \frac{d\Psi_{j+1/2}}{ds} \, ds, \tag{6.10b} \]
and the numerical fluxes \( \mathcal{F}_{j+1/2} \) are given by (6.2).

In order to complete the construction of the path-conservative central-upwind scheme, one has to select a path in (6.4). The simplest path – though in many cases quite a robust one – is a linear path, which has been used in numerical examples reported in Castro Díaz et al. (2018). The linear path, however, is not necessarily an optimal one, and the impact of different paths depends on the problem at hand.
Remark 6.1. Notice that a straightforward discretization of the non-conservative term \( \mathcal{N}(U)U_x \) used, for example, in Chertock, Kurganov, Qu and Wu (2013), Kurganov (2006), Kurganov and Miller (2014) and Kurganov and Petrova (2009) leads to a very similar semi-discretization:

\[
\frac{d}{dt} \overline{U}_j = -\frac{1}{\Delta x} \left[ \mathcal{F}_{j+1/2} - \mathcal{F}_{j-1/2} - \mathcal{N}_j \right].
\] (6.11)

The only difference between (6.9) and (6.11) is in the last two terms on the right-hand side of (6.9), which account for the contribution of the jumps of the non-conservative products at the cell interfaces. Our numerical experiments conducted for several shallow-water models and reported in Castro Díaz et al. (2014, 2018) demonstrate that in many cases, the presence of the aforementioned terms in (6.9) helps the path-conservative central-upwind scheme to clearly outperform the original central-upwind scheme.

6.1. Well-balanced path-conservative central-upwind schemes

In this section we consider the slightly different non-conservative system

\[
U_t + F(U, B)_x = \mathcal{N}(U, B)U_x + S(U)B_x,
\] (6.12)

where \( B = B(x) \) is a given piecewise smooth function with a finite number of discontinuities. In such a case, the term \( S(U)B_x \) on the right-hand side of (6.12) may represent a geometric source term appearing, for example, in the Saint-Venant system (2.3) or the two-layer shallow-water system (7.4), which will be considered in Section 7.2.

It is possible to apply the above path-conservative central-upwind scheme to the system (6.12). To this end, we add the \((N + 1)\)st equation \( B_t = 0 \) to (6.12), introduce the extended vector of unknowns \( V := (U^T, B)^T \in \mathbb{R}^{N+1} \), and rewrite the system (6.12) in the following quasilinear form:

\[
V_t + \begin{pmatrix}
\frac{\partial F}{\partial U}(V) - \mathcal{N}(V) & -\hat{S}(V) \\
0 & 0
\end{pmatrix} V_x = 0, \quad \hat{S}(V) := S(U) - \frac{\partial F}{\partial B}(V).
\] (6.13)

The path-conservative central-upwind scheme (6.9), (6.2), (6.10) can now be directly applied to the system (6.13). However, the resulting scheme will have two major drawbacks. First, the numerical diffusion present in the path-conservative central-upwind scheme will, in general, affect the last equation so that the computed \( B \) will not stay time-independent. Second, the scheme will (most probably) not be well-balanced, in the sense that it will not be designed to preserve steady-state solutions of (6.12).

In order to overcome the first of the above difficulties, we apply the path-conservative central-upwind scheme to the first \( N \) equations of the system...
(6.13) only. This results in
\[
\frac{d}{dt} \overline{U}_j = -\frac{1}{\Delta x} (D^+_{j-1/2} + D^-_{j+1/2} + F(V^-_{j+1/2}) - F(V^+_{j-1/2}) - \mathcal{N}_j - S_j),
\]
where
\[
D^\pm_{j+1/2} = \frac{1 \pm \alpha_1^{j+1/2}}{2} (F(V^+_{j+1/2}) - F(V^-_{j+1/2}) - \mathcal{N}_{\Psi,j+1/2} - S_{\Psi,j+1/2})
\]
\[
\pm \frac{\alpha_0^{j+1/2}}{2} (U^+_{j+1/2} - U^-_{j+1/2})
\]
and
\[
\mathcal{N}_j := \int_{C_j} \mathcal{N}(\mathcal{P}_j(x)) \left( \frac{dP^{(1)}_j(x)}{dx}, \ldots, \frac{dP^{(N)}_j(x)}{dx} \right)^\top dx,
\]
\[
\mathcal{N}_{\Psi,j+1/2} := \int_0^1 \mathcal{N}(\Psi_{j+1/2}(s)) \left( \frac{d\Psi^{(1)}_{j+1/2}(s)}{ds}, \ldots, \frac{d\Psi^{(N)}_{j+1/2}(s)}{ds} \right)^\top ds,
\]
\[
S_j := \int_{C_j} S(\mathcal{P}_j(x)) \frac{dP^{(N+1)}_j(x)}{dx} dx,
\]
\[
S_{\Psi,j+1/2} := \int_0^1 S(\Psi_{j+1/2}(s)) \frac{d\Psi^{(N+1)}_{j+1/2}(s)}{ds} ds.
\]
Here, a piecewise polynomial reconstruction is applied to \( V \), that is, (3.6) and (3.7) are replaced with
\[
\overline{V}(x) = \sum_j \mathcal{P}_j(x) \chi_j(x), \quad \mathcal{P}_j = (P^{(1)}_j, \ldots, P^{(N)}_j, P^{(N+1)}_j)^\top
\]
and
\[
V^-_{j+1/2} = \mathcal{P}_j(x_{j+1/2}), \quad V^+_{j+1/2} = \mathcal{P}_j(x_{j+1/2}),
\]
respectively, a smooth path
\[
\Psi_{j+1/2}(s) = (\Psi^{(1)}_{j+1/2}, \ldots, \Psi^{(N)}_{j+1/2}, \Psi^{(N+1)}_{j+1/2})^\top := \Psi(s; V^-_{j+1/2}, V^+_{j+1/2})
\]
now connects the states \( V^-_{j+1/2} \) and \( V^+_{j+1/2} \), that is,
\[
\Psi : [0, 1] \times \mathbb{R}^{N+1} \times \mathbb{R}^{N+1} \to \mathbb{R}^{N+1} \quad \text{such that}
\]
\[
\Psi(0; V^-_{j+1/2}, V^+_{j+1/2}) = V^-_{j+1/2}, \quad \Psi(1; V^-_{j+1/2}, V^+_{j+1/2}) = V^+_{j+1/2},
\]
and the one-sided local speeds are calculated using the largest \( (\lambda_N) \) and smallest \( (\lambda_1) \) eigenvalues of
\[
\mathcal{A}(V) = \frac{\partial F}{\partial U}(V) - \mathcal{N}(V).
\]
Since the obtained scheme (6.14)–(6.16) is not guaranteed to preserve steady-state solutions of (6.12), it has to be modified further. In order to construct a well-balanced path-conservative central-upwind scheme, we follow the idea presented in Castro, Pardo, Parés and Toro (2010), Castro Díaz and Fernández-Nieto (2012) and Parés and Castro (2004), and add an additional term to $D^\pm_{j+1/2}$ so that (6.15) is replaced with

$$D^\pm_{j+1/2} = \frac{1 \pm \alpha_1^{j+1/2}}{2}(F(V^+_{j+1/2}) - F(V^-_{j+1/2}) - N_{\Psi,j+1/2} - S_{\Psi,j+1/2})$$

$$\pm \frac{\alpha_0^{j+1/2}}{2}(U^+_{j+1/2} - U^-_{j+1/2} - (A^*_{j+1/2})^{-1}\tilde{S}_{\Psi,j+1/2}).$$

Here, $A^*_{j+1/2}$ is an approximation of the Jacobian matrix $A(V)$ near $x = x_{j+1/2}$ (for example, one may use the Roe matrix (Roe 1981), but simpler strategies such as those studied in Masella, Faille and Gallouët (1999) may work as well), $A^*_{j+1/2}$ is its projection onto the subset of the state space containing the steady-state solutions to be preserved (for details see Castro et al. 2010; for a particular example see Section 7.2.1), and

$$\tilde{S}_{\Psi,j+1/2} := \int_0^1 \tilde{S}(\Psi_{j+1/2}(s)) \frac{d\Psi_{j+1/2}(s)}{ds} ds.$$

Finally, the scheme (6.14), (6.17) can be recast in the following form (compare with (6.9), (6.2)):

$$\frac{d}{dt} \overline{U}_j = -\frac{1}{\Delta x} \left[ \mathcal{F}_{j+1/2} - \mathcal{F}_{j-1/2} - N_j - S_j ight.$$  

$$- \frac{a^+_j}{a^+_j - a^-_j} (N_{\Psi,j-1/2} + S_{\Psi,j-1/2})$$

$$+ \frac{a^-_j}{a^+_j - a^-_j} (N_{\Psi,j+1/2} + S_{\Psi,j+1/2}) \right].$$

where the numerical flux $\mathcal{F}_{j+1/2}$ is given by

$$\mathcal{F}_{j+1/2} = \frac{a^+_j}{a^+_j - a^-_j} \left( F(V^-_{j+1/2}) - a^-_j F(V^+_{j+1/2}) \right)$$

$$+ \frac{a^+_j a^-_j}{a^+_j - a^-_j} \left[ U^+_{j+1/2} - U^-_{j+1/2} - (A^*_{j+1/2})^{-1}\tilde{S}_{\Psi,j+1/2} \right].$$

**Remark 6.2.** Notice that if the original central-upwind flux is used instead of the modified flux given by (6.20), then the scheme will be well-
balanced only in the case when

\[ U_{j+1/2}^{-} - U_{j+1/2}^{+} = 0 \quad \text{for all } j \]

at steady states. However, in a generic case, \( U_{j+1/2}^{+} - U_{j+1/2}^{-} \) does not necessarily vanish, while the corresponding term appearing on the right-hand side of (6.20),

\[ U_{j+1/2}^{+} - U_{j+1/2}^{-} - (A_{j+1/2}^*)^{-1} \hat{S}_{\Psi,j+1/2}, \]

is in fact an approximation of \( \Delta x (U_{x} - A(V)^{-1}\hat{S}(V)B_{x}) \) across the cell interface and it vanishes at steady-state solutions provided a proper path is selected in the evaluation of \( \hat{S}_{\Psi,j+1/2} \) in (6.18); see Castro et al. (2010), Castro Díaz and Fernández-Nieto (2012) and Parés and Castro (2004) for details. This guarantees a perfect balance between the source and flux terms as long as the reconstruction (3.6), (3.7) preserves the steady-state solution.

**Remark 6.3.** Also notice that the term we have added to make the scheme well-balanced appears in the numerical viscosity and thus does not affect the consistency of the resulting path-conservative central-upwind scheme (6.19), (6.20), (6.16), (6.18).

### 7. Some related shallow-water models

In this section we will briefly describe some related shallow-water models and their numerical approximations.

#### 7.1. Shallow-water models with time-dependent bottom topography

In many practically relevant situations, the bottom topography \( B \) is time-dependent due to erosion, sediment transport, dam breaks, floods and submarine landslides; see, for example, Grass (1981), Heinrich (1992), Hu, Cao, Pender and Tan (2012), Li and Duffy (2011), Morales de Luna, Castro Díaz, Parés Madroñal and Fernández Nieto (2009), Murillo and García-Navarro (2010), Pritchard and Hogg (2002), Simpson and Castelltort (2006), Tingsanchali and Chinnarasri (2001), Watts (2000), Wu and Wang (2007), Xia, Lin, Falconer and Wang (2010) and Yue, Cao, Li and Che (2008). Unfortunately, a straightforward application of the finite-volume methods described in Sections 3 and 4 may lead to very inaccurate results in the case of time-dependent bottom topography. This is due to the fact that in this case, the speed of water surface gravity waves is typically much larger than the speed at which the changes in the bottom topography occur. This imposes a severe stability restriction on the size of time steps, which, in turn, leads to excessive numerical diffusion that affects the computed bottom structure.
The simplest way to model the bottom topography propagation was proposed in Exner (1920). According to this approach, in the 2-D case, the bottom topography function satisfies the following equation:

\[ B_t + \zeta q^B(h, u, v)_x + \zeta p^B(h, u, v)_y = 0, \]  

(7.1)

where \( \zeta = 1/(1 - \gamma) \) is a constant with \( \gamma \) representing the porosity of the sediment layer. The sediment \( x \)- and \( y \)-discharges, \( q^B \) and \( p^B \), depend on various water and sediment properties. The simplest bottom topography transport fluxes, which were proposed in Grass (1981), are

\[ q^B = Au(u^2 + v^2)^{(m-1)/2}, \quad p^B = Av(u^2 + v^2)^{(m-1)/2}, \]  

(7.2)

where \( A \in [0, 1] \) is a non-dimensional constant which accounts for the effects of sediment grain size and kinematic viscosity. The values of \( A \) and a constant \( 1 \leq m \leq 4 \) are often obtained from the experimental data.

Efficient, accurate and robust numerical methods for the shallow-water system (1.1), (7.1), (7.2) and its 1-D version, the Saint-Venant system (2.3) coupled with the 1-D Exner equation,

\[ B_t + \frac{A}{1 - \gamma}(u^m)_x = 0, \]  

(7.3)

have recently been proposed in Bernstein, Chertock, Kurganov and Wu (2018) and will now be reviewed briefly.

In principle, one can apply a Godunov-type central or central-upwind scheme to the system (2.3), (7.3) in a straightforward manner. This, however, will lead to excessive smearing of the computed \( B \), as demonstrated in numerical examples reported in Bernstein et al. (2018). In order to understand this phenomenon, we first study the Jacobian of the system (2.3), (7.3), which has three real eigenvalues \( \lambda_1 < \lambda_2 < \lambda_3 \). As shown in Bernstein et al. (2018), \( \lambda_1 \) and \( \lambda_3 \) are close to the eigenvalues of the original 1-D Saint-Venant system (2.3), while \( \lambda_2 \) is close to zero and typically \( |\lambda_2| \ll \max(|\lambda_1|, |\lambda_3|) \). This corresponds to the bottom topography movement being much slower than the surface wave propagation, and it is well known that slow waves require a special treatment to be accurately resolved numerically.

In order to develop efficient and accurate numerical methods for the system (2.3), (7.3), we apply the second-order Strang operator splitting method (see e.g. Jia and Li 2011, Marchuk 1990, Strang 1968, Vabishchevich 2014): we split the Saint-Venant system (2.3) from the Exner equation (7.3). This allows one to treat slow and fast waves in a different manner and using different sizes of time steps. The size of splitting time steps is taken to be proportional to \( 1/|\lambda_2| \). We then follow the approach that was developed in the framework of the fast explicit operator splitting method, which we derived in Chertock, Doering, Kashdan and Kurganov (2010), Chertock, Kashdan and
Kurganov (2008), Chertock and Kurganov (2009) and Chertock, Kurganov and Petrova (2005, 2009): each Saint-Venant splitting sub-step consists of several smaller central-upwind steps. In this way we ensure the stability of the Saint-Venant sub-steps, while large Exner splitting sub-steps prevent excessive numerical dissipation, which may severely affect the resolution of the bottom topography, especially in the case when $B$ is discontinuous.

We note that at the Saint-Venant splitting sub-steps we ‘freeze’ the bottom topography, while at the Exner splitting sub-steps we ‘freeze’ the rest of the solution components.

Another difficulty related to the implementation of the central-upwind scheme is the fact that, as we have explained in Section 5.1.1, the evolution of the Saint-Venant solution uses the continuous piecewise linear interpolant of $B$ given in (5.6), which requires the point values of $B$ at the cell interfaces $x = x_{j+1/2}$. We therefore solve the Exner equation (7.3) on a staggered grid, that is, we evolve the point values $B_{j+1/2}$ rather than $B_j$ or the cell averages $\bar{B}_j$. To this end, we need to project the velocities $u_j = \tilde{q}_j/\bar{h}_j$ onto the staggered grid. This can be done in several different ways. We proceed as in Jiang et al. (1998): reconstruct a piecewise linear interpolant

$$\tilde{u}(x) = u_j + (u_x)_j(x - x_j), \quad x \in C_j$$

using a certain nonlinear limiter to evaluate $(u_x)_j$, then integrate it to obtain

$$\tilde{u}_{j+1/2} = \int_{x_j}^{x_{j+1}} \tilde{u}(x) \, dx = \frac{1}{2}((u_j + u_{j+1}) + \frac{\Delta x}{8}((u_x)_j - (u_x)_{j+1}).$$

Equipped with the staggered grid data, $\{B_{j+1/2}\}$ and $\{\tilde{u}_{j+1/2}\}$, we apply the central-upwind scheme to the Exner equation (7.3) using the local speeds proportional to $\lambda_2$, which is small and this guarantees that no excessive numerical diffusion is present at the central-upwind discretization of (7.3); see Bernstein et al. (2018) for details.

### 7.2. Two-layer shallow-water systems

A system of two-layer shallow-water equations, which models oceanographic water flows in straights and channels (see e.g. Castro, Macías and Parés 2001, Macías, Pares and Castro 1999), is conceptually more complicated than the single-layer Saint-Venant system since it contains non-conservative interlayer exchange terms. In the 1-D case the system reads as

\begin{align}
(h_1)_t + (q_1)_x &= 0, \\
(q_1)_t + \left(h_1 u_1^2 + \frac{1}{2}gh_1^2\right)_x &= -gh_1(B + h_2)_x,
\end{align}

(7.4a) (7.4b)
where \( h_1 \) and \( h_2 \) are the depths of the upper and lower layers, respectively, \( u_i \) and \( q_i = h_i u_i \), \( i = 1, 2 \) are their corresponding velocities and discharges, and \( r \leq 1 \) is the constant density ratio between the two layers. In the 2-D case, the two-layer shallow-water system reads

\[
\begin{align*}
(h_1)_t + (q_1^x)_x + (q_1^y)_y &= 0, \\
(q_1^x)_t + \left( h_1 u_1^2 + \frac{1}{2} g h_1^2 \right)_x + (h_1 u_1 v_1)_y &= -g h_1 (B + h_2)_x, \\
(q_1^y)_t + (h_1 u_1 v_1)_x + \left( h_1 v_1^2 + \frac{1}{2} g h_1^2 \right)_y &= -g h_1 (B + h_2)_y, \\
(h_2)_t + (q_2^x)_x + (q_2^y)_y &= 0, \\
(q_2^x)_t + \left( h_2 u_2^2 + \frac{1}{2} g h_2^2 \right)_x + (h_2 u_2 v_2)_y &= -g h_2 (B + r h_1)_x, \\
(q_2^y)_t + (h_2 u_2 v_2)_x + \left( h_2 v_2^2 + \frac{1}{2} g h_2^2 \right)_y &= -g h_2 (B + r h_1)_y,
\end{align*}
\]

where \( u_i, q_i^x = h_i u_i, i = 1, 2 \) are the \( x \)-velocities and discharges and \( v_i, q_i^y = h_i v_i, i = 1, 2 \) are the \( y \)-velocities and discharges.

The two-layer shallow-water systems (7.4) and (7.5) have been extensively studied and a number of numerical methods has been developed; see, for example, Abgrall and Karni (2009), Bouchut and Morales de Luna (2008), Bouchut and Zeitlin (2010), Castro, Macías and Parés (2001), Castro et al. (2008), Castro et al. (2004), Castro Díaz, Chacón Rebollo, Fernández-Nieto and Pares (2007) and Macías, Pares and Castro (1999). Unfortunately, most of these methods are not sufficiently robust since the computed solution heavily depends on the way the non-conservative product terms on the right-hand side of (7.4) and (7.5) are discretized.

In Kurganov and Petrova (2009), we have proposed a special way to treat these terms and developed a semi-discrete central-upwind scheme for (7.4) and (7.5). This method is based on the following reformulation of the two-layer shallow-water system. For simplicity, we will consider the 1-D system (7.4) and rewrite it in terms of the equilibrium variables \( h_1, q_1, w \equiv h_2 + B \) and \( q_2 \) (notice that at ‘lake at rest’ steady states \( h_1 \equiv \text{const.}, w \equiv \text{const.}, q_1 \equiv q_2 \equiv 0 \)) as

\[
\begin{align*}
(h_1)_t + (q_1)_x &= 0, \\
(q_1)_t + \left( \frac{q_1^2}{h_1} + g \varepsilon h_1 \right)_x &= g \varepsilon (h_1)_x,
\end{align*}
\]

\[ (7.6a) \]

\[ (7.6b) \]
Figure 7.1. Set-up for the two-layer shallow-water system (7.6).

\[ w_t + (q_2)_x = 0, \quad (q_2)_t + \left( \frac{q_2^2}{w - B} + \frac{g}{2} w^2 - \frac{g}{2} r h_1^2 - g \hat{\varepsilon} B \right)_x = -g \hat{\varepsilon} B_x - g r \varepsilon (h_1)_x, \]

where \( \varepsilon = h_1 + h_2 + B = h_1 + w \) is a water surface (see Figure 7.1) and \( \hat{\varepsilon} := rh_1 + w \).

Notice that while the two-layer shallow-water system (7.6) is equivalent (for smooth solutions) to the original two-layer system (7.4), it is suitable for a ‘favourable’ treatment of the non-conservative products. Our approach uses the fact that in the relevant applications, fluctuations of the total water level \( \varepsilon \) are relatively small, that is, after choosing a proper coordinate system, \( \varepsilon \sim 0 \); see Figure 7.1. Thus we reduce as much as possible the ‘influence’ of the particular choice of the non-conservative product discretization as the non-conservative product terms on the right-hand side of (7.6) are proportional to \( \varepsilon \).

The rewritten systems are then numerically solved by a two-layer version of the second-order well-balanced positivity-preserving central-upwind scheme presented in Sections 5.1.1 and 5.1.3. To this end, we (i) replace the bottom with its piecewise linear approximation \( \tilde{B} \); (ii) correct the reconstruction of \( w \) to ensure that \( \tilde{w} \geq \tilde{B} \); (iii) regularize the computed velocities to avoid division by \( h_i \sim 0 \); (iv) use the well-balanced quadrature for the geometric source term. The only difficulty with the implementation of the central-upwind scheme is related to the fact that one cannot find analytical expressions for the Jacobian eigenvalues needed to evaluate one-sided local speeds in (3.14). Indeed, as one can easily see, the eigenvalues are solutions of the following algebraic equation:

\[ (\lambda^2 - 2u_1 \lambda + u_1^2 - gh_1)(\lambda^2 - 2u_2 \lambda + u_2^2 - gh_2) = g^2 \hat{h}_1 h_2. \]

Although this equation cannot be explicitly solved, one can easily obtain an upper bound on the largest eigenvalue and a lower bound on the smallest eigenvalue using the Lagrange theorem; see, for example, Lagrange (1798) and Mignotte and Stefanescu (2002).
Remark 7.1. One can show that equation (7.7) may have either four or only two real solutions depending on the values of $h_1$, $h_2$, $u_1$, $u_1$ and $r$. This means that the two-layer shallow-water system (7.4) is only conditionally hyperbolic and, as observed in the literature, in the non-hyperbolic regime, its solution typically develops instabilities.

To study the hyperbolicity regions of (7.4), one can examine the first-order approximation of its Jacobian eigenvalues (see e.g. Castro et al. 2001, Schijf and Schonfeld 1953), which is given by

$$
\lambda_{1,2}(h_1, u_1, h_2, u_2) \approx U_m \pm \sqrt{g(h_1 + h_2)}, \tag{7.8a}
$$

$$
\lambda_{3,4}(h_1, u_1, h_2, u_2) \approx U_c \pm \sqrt{(1 - r)g\frac{h_1 h_2}{h_1 + h_2} \left(1 - \frac{(u_2 - u_1)^2}{(1 - r)g(h_1 + h_2)}\right)}, \tag{7.8b}
$$

where

$$
U_m = \frac{h_1 u_1 + h_2 u_2}{h_1 + h_2}, \quad U_c = \frac{h_1 u_2 + h_2 u_1}{h_1 + h_2}. \tag{7.9}
$$

From (7.8), (7.9) one expects that the two-layer shallow-water system (7.4) is hyperbolic as long as

$$(u_2 - u_1)^2 < (1 - r)g(h_1 + h_2),$$

and thus the hyperbolicity condition for the two-layer shallow-water system (7.4) depends on the relationship between $(u_2 - u_1)^2$ and $(1 - r)g(h_1 + h_2)$.

A detailed description of the resulting central-upwind scheme together with an illustration of its performance can be found in Kurganov and Petrova (2009). However, in examples in which the surface waves are not small, the method from Kurganov and Petrova (2009) may fail. Therefore, in Castro Díaz et al. (2018) we have developed a well-balanced path-conservative central-upwind scheme, which seems to be a robust numerical method for two-layer shallow-water system (7.4).

7.2.1. Well-balanced path-conservative central-upwind scheme for the two-layer shallow-water system

In order to design a well-balanced path-conservative central-upwind scheme for the two-layer shallow-water system (7.6), we proceed exactly as in Section 6.1 and add the fifth equation $B_t = 0$ to (7.6). We then rewrite the
system (7.6) in the vector form (6.13) with

\[ U = \begin{pmatrix} h_1 \\ q_1 \\ w \\ q_2 \\ B \end{pmatrix}, \quad V = \begin{pmatrix} h_1 \\ q_1 \\ w \\ q_2 \\ B \end{pmatrix}, \quad F(V) = \begin{pmatrix} \frac{q_1^2}{h_1} + g(h_1 + w)h_1 \\ \frac{q_1}{h_1} + g(h_1 + w)h_1 \\ \frac{q_2^2}{w-B} + \frac{g}{2}(w^2 - rh_1^2) - g(rh_1 + w)B \end{pmatrix}, \]

\[ B(V) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ g(h_1 + w) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -gr(h_1 + w) & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad S(U) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -g(rh_1 + w) \end{pmatrix}. \]

Notice that at the ‘lake at rest’ steady states,

\[ q_1 \equiv q_2 \equiv 0, \quad h_1 \equiv \text{const.}, \quad w = h_2 + B \equiv \text{const.}, \quad (7.10) \]

that is, \( U \equiv \text{const.} \), and thus

\[ U_{j+1/2}^+ - U_{j+1/2}^- \equiv 0 \quad \text{for all } j, \]

provided the piecewise polynomial reconstruction (3.6) exactly recovers the constant states (7.10). Therefore, a well-balanced central-upwind scheme will be obtained with the help of the simplified central-upwind flux (6.2) instead of a more complicated flux (6.20) derived in Section 6.1.

The path-conservative central-upwind scheme (6.19), (6.2), (6.16) is now applied to the two-layer shallow-water system in a straightforward manner. We use the piecewise linear reconstruction

\[ \mathcal{P}_j(x) = \bar{V}_j + (V_x)_j(x - x_j), \]

which leads to the following expressions for \( \mathcal{N}_j \) and \( \mathcal{S}_j \):

\[ \mathcal{N}_j = \begin{pmatrix} 0 \\ g((\bar{h}_1)_j + \bar{w}_j)((h_1)_j+1/2 - (h_1)_j-1/2) \\ 0 \\ -gr((\bar{h}_1)_j + \bar{w}_j)((h_1)_j-1/2 - (h_1)_j+1/2) \end{pmatrix}, \]

\[ \mathcal{S}_j = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -g(r(\bar{h}_1)_j + \bar{w}_j)(B_j+1/2 - B_j-1/2) \end{pmatrix}, \]

and the linear segment path

\[ \Psi_{j+1/2}(s) = V_{j+1/2}^- + s(V_{j+1/2}^+ - V_{j+1/2}^-), \]
which results in

\[ N_{\Psi,j+1/2} = \begin{pmatrix}
0 \\
g\phi_{j+1/2}/2 \\
-g\phi_{j+1/2}/2
\end{pmatrix}, \]

where

\[ \phi_{j+1/2} = ((h_1)_{j+1/2}^+ + w_{j+1/2}^+ + (h_1)_{j+1/2}^- + w_{j+1/2}^-)((h_1)_{j+1/2}^+ - (h_1)_{j+1/2}^-), \]

and

\[ S_{\Psi,j+1/2} = \begin{pmatrix}
0 \\
0 \\
0 \\
-g\psi_{j+1/2}/2
\end{pmatrix}, \]

where

\[ \psi_{j+1/2} = (r(h_1)_{j+1/2}^+ + w_{j+1/2}^+ + r(h_1)_{j+1/2}^- + w_{j+1/2}^-)(B_{j+1/2}^+ - B_{j+1/2}^-). \]

In order to design the scheme (6.19), (6.2) for the two-layer system (7.6), we use the matrix

\[ A(V) = \frac{\partial F}{\partial U}(V) - N(V) = \begin{pmatrix}
0 & 1 & 0 & 0 \\
h_1 - u_1^2 & 2u_1 & gh_1 & 0 \\
0 & 0 & 0 & 1 \\
g rh_2 & 0 & gh_2 - u_2^2 & 2u_2
\end{pmatrix}, \]

where \( h_2 = w - B \). The one-sided local speeds

\[ a_{j+1/2}^- = \min\{\lambda_1(A(V_{j+1/2}^-)), \lambda_1(A(V_{j+1/2}^+)), 0\}, \]
\[ a_{j+1/2}^+ = \max\{\lambda_4(A(V_{j+1/2}^-)), \lambda_4(A(V_{j+1/2}^+)), 0\} \]

are then obtained using the upper/lower bounds on the largest/smallest eigenvalues of \( A(V) \) using the Lagrange theorem as discussed above.

**Remark 7.2.** When the well-balanced path-conservative central-upwind scheme is designed, we use a generically discontinuous piecewise polynomial reconstruction of \( B \) (which is the fifth component of \( V \)) rather than the continuous piecewise linear one (5.6).

**Acknowledgements**

A large portion of the material covered in this review is based on the work of the author with various collaborators, and it is my pleasure to acknowledge Alina Chertock, Manuel Jesús Castro Díaz, Guoxian Chen, Yuanzhen...
Cheng, Shumo Cui, Xin Liu, Tomás Morales de Luna, Sebastian Noelle, Guergana Petrova, Eitan Tadmor and Tong Wu. I am indebted in particular to my friend and collaborator, Doron Levy, who co-authored the paper Kurganov and Levy (2002), which became the forerunner of a large body of subsequent work. Research was supported in part by NSF grant DMS-1521009 and NSFC grant 11771201.

REFERENCES


---

1 The URLs cited in this work were correct at the time of going to press, but the publisher and the authors make no undertaking that the citations remain live or are accurate or appropriate.


Finite-volume schemes for shallow-water equations


http://hal.archives-ouvertes.fr/hal-00129675/en/


