Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 206 (2005) 311–333
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

A well-balanced positivity preserving "second-order" scheme for shallow water flows on unstructured meshes

Emmanuel Audusse *, Marie-Odile Bristeau

INRIA, Project Bang, Domaine de Voluceau, 78153 Le Chesnay, France

Received 18 May 2004; received in revised form 23 November 2004; accepted 7 December 2004
Available online 24 March 2005

Abstract

We consider the solution of the Saint-Venant equations with topographic source terms on 2D unstructured meshes by a finite volume approach. We first present a stable and positivity preserving homogeneous solver issued from a kinetic representation of the hyperbolic conservation laws system. This water depth positivity property is important when dealing with wet–dry interfaces. Then, we introduce a local hydrostatic reconstruction that preserves the positivity properties of the homogeneous solver and leads to a well-balanced scheme satisfying the steady-state condition of still water. Finally, a formally second-order extension based on limited reconstructed values on both sides of each interface and on an enriched interpretation of the source terms satisfies the same properties and gives a noticeable accuracy improvement. Numerical examples on academic and real problems are presented.

PACS: 02.30.Jr; 02.60.Cb; 47.11.+j
Keywords: Saint-Venant system; Shallow water flow; Finite volumes; Kinetic solver; Hydrostatic reconstruction; Well-balanced scheme; Positivity preserving scheme; Second-order extension

1. Introduction

We consider in this article the 2D Saint-Venant system with topographic source term. This hyperbolic system of balance laws was introduced in [42] and is very commonly used for the numerical simulation of various geophysical shallow-water flows, such as rivers, lakes or coastal areas, or even oceans, atmosphere or avalanches [8,24] when completed with appropriate terms. It can be derived as a formal first-order approximation of the three-dimensional free surface incompressible Navier–Stokes equations, using the

* Corresponding author. Tel.: +33 1 39 63 58 29; fax: +33 1 39 63 58 82.
E-mail addresses: emmanuel.audusse@inria.fr (E. Audusse), marie-odile.bristeau@inria.fr (M.-O. Bristeau).

0021-9991/$ - see front matter © 2005 Published by Elsevier Inc.
doi:10.1016/j.jcp.2004.12.016

so-called shallow water assumption [16,21]. Usually, several other terms are added in order to take into account frictions on the bottom and the surface and other physical features. One can also describe the evolution of a temperature (or a concentration of a pollutant) advected by the flow by adding a third equation to the system [4,14].

The difficulty to define accurate numerical schemes for such hyperbolic systems is related to their deep mathematical structure. The first existence proof of weak solutions after shocks in the large is due to Lions et al. [34] in 1996. It is based on the kinetic interpretation of the system which is also a method to derive numerical schemes with good stability properties, such as positivity of the water depth, ability to compute dry areas, and eventually to satisfy a discrete entropy inequality.

In the context of the discretization of hyperbolic systems of balance laws another fundamental point is to get schemes that satisfy the preservation of steady-states such as still water equilibrium in the context of the Saint-Venant system. The difficulty to build such schemes was pointed out by several authors and led to the notion of *well-balanced* schemes. Different approaches to satisfy the well-balanced property have been proposed. The Roe solver [40] has been modified in order to preserve steady-states by Bermudez and Vasquez [6]. A two-dimensional extension is performed by Bermudez et al. [5] and recent extensions to other types of homogeneous solvers can be found in [12,13]. Leveque [33] and Jin [29] propose other ways to adapt exact or approximate Riemann solver to the non-homogeneous case. Following the idea of Leroux and Greenberg [20] for the scalar case, Gosse [18,19] and Gallouët et al. [17] construct numerical schemes based on the solution of the – exact or approximate – Riemann problem associated with a larger system where a third equation on the variable describing the bottom topography is added. Approaches based on central schemes are used in [31] or [41] and Perthame and Simeoni [39] propose a kinetic method that includes the source terms in the kinetic formulation (see also [11]).

Most of the works mentioned above are entirely devoted to the well-balancing property and do not treat the stability problems. Among the Riemann solvers, only the Godunov method [20] preserves positivity but it leads to quite complex and time consuming algorithms. Some other approaches are more successful, but to the best of our knowledge, they are restricted to one-dimensional problems (kinetic interpretation of the source terms [39]) or to cartesian two-dimensional grids (central schemes [31]).

In this paper, we propose a new finite volume method for the 2D Saint-Venant system with source terms that ensures well-balancing and positivity of the water depth and that allows us to deal with unstructured meshes. Our method is based on a *kinetic solver* and a *hydrostatic reconstruction* procedure.

Considering the homogeneous Saint-Venant equations we first present their kinetic interpretation and how this can be used to deduce a macroscopic finite volume *kinetic solver*. The solver has good stability properties as the inherent preservation of the water depth positivity even when applications with dry areas are considered. The kinetic schemes were first developed in the context of the Euler equations [36,30,38] and we present their adaptation to the Saint-Venant system. We refer to Perthame [37] for a complete survey of the theoretical properties of the kinetic schemes. We also present many original developments in order to take into account the boundary conditions, reduce the diffusion of the scheme or increase the efficiency of the method.

Second we consider the Saint-Venant system with source terms and we present a *hydrostatic reconstruction* strategy which allows us to extend any positivity preserving homogeneous scheme to a positivity preserving *well-balanced* scheme. The idea is to use the relation associated to the equilibrium to define new interface values that will be used in the definition of the finite volume fluxes and of the source terms. This part is an extension to two-dimensional problems of what is presented in [3] for the 1D case.

We finally present and describe a conservative and positivity preserving *formally second-order extension* on unstructured two-dimensional meshes, based on linear reconstruction procedures. By introducing an enriched discretization of the source terms we construct a stable and well-balanced ''second-order'' scheme. Some academic numerical tests give precise information about the improvement on the accuracy of the

results. Some numerical tests on real problems show that the method can be applied to very complex problems.

We present in this paper all the points mentioned above. We especially concentrate on the implementation of the method but we also (at least briefly) describe all the theoretical points that are necessary to make this article self-contained. The outline is as follows. We present the 2D shallow water equations and their main properties in Section 2. Thereafter, we introduce the homogeneous two-dimensional *kinetic solver* in Section 3, including 2D finite volume formalism, 2D kinetic interpretation of the Saint-Venant system and a precise description of the numerical implementation (boundary conditions, upwinding for the tangential velocity). Section 4 introduces the *hydrostatic reconstruction* and we present a positivity preserving *well-balanced* kinetic scheme which is adapted to the still water steady-state. We develop formally second-order extension of the scheme and we show its property in Section 5. Section 6 presents numerical results on idealized and real tests. We present in Appendix A the detailed proofs concerning the water depth positivity and the well-balancing property.

2. The Saint-Venant system

2.1. Equations

We consider the 2D Saint-Venant system, written in conservative form – see [21,16]

$$\frac{\partial h}{\partial t} + \operatorname{div}(h\mathbf{u}) = 0, \tag{2.1}$$

$$\frac{\partial h\mathbf{u}}{\partial t} + \operatorname{div}(h\mathbf{u}\otimes\mathbf{u}) + \nabla\left(\frac{g}{2}h^2\right) + gh\nabla Z = 0, \tag{2.2}$$

where $h(t,x,y) \geqslant 0$ is the water depth, $\mathbf{u}(t,x,y) = (u,v)^{\mathrm{T}}$ the flow velocity, g the acceleration due to gravity and $Z(x,y)$ the bottom topography, and therefore $h + Z$ is the water surface level (see Fig. 1). We denote also $\mathbf{q}(t,x,y) = (q_x,q_y)^{\mathrm{T}} = h(t,x,y)\mathbf{u}(t,x,y)$ the flux of water.

To obtain a well-posed problem we add to this system some initial conditions

$$h(0,x,y) = h^0(x,y), \quad \mathbf{u}(0,x,y) = \mathbf{u}^0(x,y), \tag{2.3}$$

and boundary conditions. In this paper, we consider only two types of boundaries: solid walls on which we prescribe a slip condition $\mathbf{u}\cdot\mathbf{n} = 0$ with $\mathbf{n}$ the unit outward normal to the boundary, and fluid boundaries on which we prescribe zero, one or two of the following conditions depending on the type of the flow (fluvial or torrential): water level $h + Z$ given, flux $\mathbf{q}$ given.

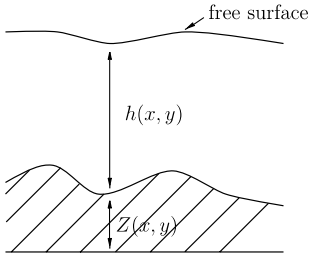

Fig. 1. Shallow water description of a free surface flow.

2.2. Properties of the system

The system (2.1), (2.2) is a first-order hyperbolic system of balance laws and can be written in the general form

$$\frac{\partial \mathbf{U}}{\partial t} + \operatorname{div} \mathbf{F}(\mathbf{U}) = \mathbf{B}(\mathbf{U}), \tag{2.4}$$

with $\mathbf{U} = (h, q_x, q_y)^{\mathrm{T}}$ and

$$\mathbf{F}(\mathbf{U}) = \begin{pmatrix} q_x & q_y \\ \frac{q_x^2}{h} + \frac{g}{2}h^2 & \frac{q_x q_y}{h} \\ \frac{q_x q_y}{h} & \frac{q_y^2}{h} + \frac{g}{2}h^2 \end{pmatrix}, \quad \mathbf{B}(\mathbf{U}) = \begin{pmatrix} 0 \\ -gh\partial_x Z \\ -gh\partial_y Z \end{pmatrix}. \tag{2.5}$$

We mention now some important properties of the system (2.1), (2.2) that have to be taken into account to construct a well-adapted numerical solver.

This system is *strictly hyperbolic* for $h > 0$ (see [9]). Moreover, it admits an *invariant region* $h(t,x) \geqslant 0$, the water depth h can indeed vanish (flooding zones, dry regions, tidal flats) and the system loses hyperbolicity at $h = 0$ which generates theoretical and numerical difficulties.

Another important property is related to the source terms: the Saint-Venant system admits *non-trivial steady-states*. They are characterized by

$$\operatorname{div}(h\mathbf{u}) = 0, \quad \nabla P - \operatorname{curl} \mathbf{u} \begin{pmatrix} v \\ u \end{pmatrix} = 0,$$

where

$$P(x,y) = \frac{|\mathbf{u}|^2}{2} + g(h + Z).$$

It follows in particular that P is constant along streamlines and in the irrotational areas. Due to the complexity of the general equilibria and because of its importance in the applications we are interested particularly in one of these equilibria: the so-called *still water steady-state*

$$\mathbf{u} = 0, \quad h + Z = H, \tag{2.6}$$

where H is a constant.

Note that for 1D flows, general steady-states are characterized by the simpler relations: $hu(x) = C_1$, $P(x) = C_2$, where C_1 and C_2 are two constants, and can be exactly computed by adapted numerical schemes – see [17,19]. These relations are used in Section 6 to compute the exact solution of the 1D flows in a 2D channel with a bump at the bottom.

3. Homogeneous scheme: the kinetic solver

In this section, we present a *positivity preserving* numerical discretization for the homogeneous system of Eq. (2.4) with $\mathbf{B} = 0$. A classical approach for solving hyperbolic systems consists of using finite volume schemes (see [22,32,7]) which are defined by the fluxes computed at the control volume interfaces. We recall the 2D formalism of this method in Section 3.1. Then, we present in Section 3.2 a kinetic equation deduced from a kinetic interpretation of the Saint-Venant system. In Section 3.3, we show how the fluxes of the finite volume kinetic solver are deduced from the discretization of this kinetic equation. In Sections 3.4–3.6, we

describe the numerical implementation and in Section 3.7 we prove that this scheme preserves the positivity of the water depth.

3.1. 2D finite volume formalism

We recall here the general formalism of finite volumes. Let Ω denote the computational domain with boundary Γ, which we assume is polygonal. Let T_h be a triangulation of Ω for which the vertices are denoted by P_i with S_i the set of interior nodes and G_i the set of boundary nodes. The dual cells C_i are obtained by joining the centers of mass of the triangles surrounding each vertex P_i. We use the following notations (see Fig. 2):

- K_i, set of subscripts of nodes P_j surrounding P_i,
- $|C_i|$, area of C_i,
- Γ_{ij}, boundary edge between the cells C_i and C_j,
- L_{ij}, length of Γ_{ij},
- $\mathbf{n}_{ij}$, unit normal to Γ_{ij}, outward to C_i ($\mathbf{n}_{ji}=-\mathbf{n}_{ij}$).

If P_i is a node belonging to the boundary Γ, we join the centers of mass of the triangles adjacent to the boundary to the middle of the edge belonging to Γ (see Fig. 3) and we denote

- Γ_i, the two edges of C_i belonging to Γ,
- L_i, length of Γ_i (for sake of simplicity we assume in the following that $L_i = 0$ if P_i does not belong to Γ),
- $\mathbf{n}_i$, the unit outward normal defined by averaging the two adjacent normals.

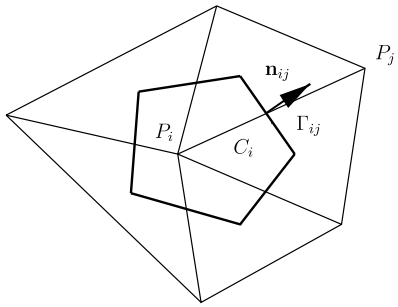

Fig. 2. Dual cell C_i.

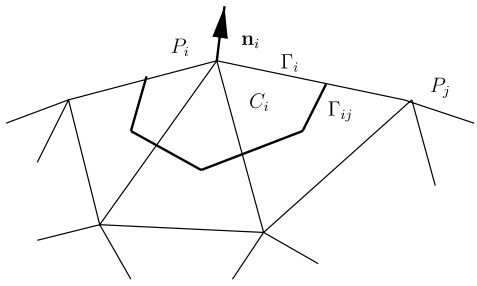

Fig. 3. Boundary cell C_i.

Let Δt be the timestep, we set $t^n = n\Delta t$. We denote by $\mathbf{U}_i^n$ the approximation of the cell average of the exact solution at time t^n

$$\mathbf{U}_i^n \simeq \frac{1}{|C_i|}\int_{C_i} \mathbf{U}(t^n, \mathbf{x})\, d\mathbf{x}. \tag{3.1}$$

We integrate in space and time, Eq. (2.4) on the set $C_i \times (t^n, t^{n+1})$ and integrating by parts the divergence term, we obtain

$$\int_{C_i} \mathbf{U}(t^{n+1}, \mathbf{x})\, d\mathbf{x} - \int_{C_i} \mathbf{U}(t^n, \mathbf{x})\, d\mathbf{x} + \int_{t^n}^{t^{n+1}} \int_{\partial C_i} \mathbf{F}(\mathbf{U}) \cdot \mathbf{n}\, d\mathbf{x}\, dt = 0. \tag{3.2}$$

So we can write

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \sum_{j \in K_i} \sigma_{ij} F(\mathbf{U}_i^n, \mathbf{U}_j^n, \mathbf{n}_{ij}) - \sigma_i F(\mathbf{U}_i^n, \mathbf{U}_{e,i}^n, \mathbf{n}_i), \tag{3.3}$$

with

$$\sigma_{ij} = \frac{\Delta t L_{ij}}{|C_i|}, \quad \sigma_i = \frac{\Delta t L_i}{|C_i|}. \tag{3.4}$$

In (3.3) the term $F(\mathbf{U}_i, \mathbf{U}_j, \mathbf{n}_{ij})$ denotes an interpolation of the normal component of the flux $\mathbf{F}(\mathbf{U}) \cdot \mathbf{n}_{ij}$ along the edge Γ_{ij}. This interpolation is usually performed using a one-dimensional solver since locally the problem looks like a planar discontinuity. In the next subsections we define $F(\mathbf{U}_i, \mathbf{U}_j, \mathbf{n}_{ij})$ using the kinetic interpretation of the system. The computation of the value $\mathbf{U}_{e,i}$, which denotes a value outside C_i defined such that the boundary conditions are satisfied, and the definition of the boundary flux $F(\mathbf{U}_i, \mathbf{U}_{e,i}, \mathbf{n}_i)$ are described in Section 3.6.

3.2. Kinetic interpretation of the Saint-Venant system

Here, we introduce a kinetic approach to the homogeneous version of the system (2.1), (2.2) and in the next subsection we deduce from the discretization of the corresponding kinetic equation, a kinetic solver for this system.

Let $\chi(w)$ be a positive, even and compactly supported function defined on IR^2, i.e.

$$\chi(\mathbf{w}) = \chi(-\mathbf{w}) \geqslant 0, \tag{3.5}$$

$$\exists w_M \in \mathbb{R}, \text{ such that } \chi(\mathbf{w}) = 0 \quad \text{for } \|\mathbf{w}\| \geqslant w_M. \tag{3.6}$$

We also assume that

$$\int_{\mathbb{R}^2} \begin{pmatrix} 1 \\ w_i w_j \end{pmatrix} \chi(\mathbf{w})\, d\mathbf{w} = \begin{pmatrix} 1 \\ \delta_{ij} \end{pmatrix} \tag{3.7}$$

with δ_{ij} the Kronecker symbol.

An example of function χ satisfying these properties is

$$\chi(\mathbf{w}) = \frac{1}{12} 1\mathrm{I}_{|w_i| \leqslant \sqrt{3}}, \quad i = 1, 2. \tag{3.8}$$

We introduce a microscopic density of particles $M(t, \mathbf{x}, \xi)$ at time t, in position $\mathbf{x}$ and with kinetic velocity ξ. It is defined by a so-called *Gibbs equilibrium*

$$M(t, \mathbf{x}, \xi) = M(h, \xi - \mathbf{u}) = \frac{h(t, \mathbf{x})}{\tilde{c}^2} \chi\left(\frac{\xi - \mathbf{u}(t, \mathbf{x})}{\tilde{c}}\right), \tag{3.9}$$

with $\tilde{c}$ defined by

$$\tilde{c}^2 = \frac{gh}{2}. \tag{3.10}$$

With these definitions we can write a kinetic interpretation of the system (2.1), (2.2). Indeed the Saint-Venant system can be seen as an integration in ξ against $(1,\xi)^{\mathrm{T}}$ of the following *kinetic equation*:

$$\frac{\partial M}{\partial t} + \xi \cdot \nabla_{\mathbf{x}} M = Q(t,\mathbf{x},\xi), \tag{3.11}$$

where the "collision term" $Q(t,\mathbf{x},\xi)$ satisfies for a.e. $(t,\mathbf{x})$,

$$\int_{\mathbb{R}^2} \begin{pmatrix} 1 \\ \xi \end{pmatrix} Q \, \mathrm{d}\xi = 0. \tag{3.12}$$

Indeed from the definition of M (3.9), the properties of χ (3.5)–(3.7) and from (3.12), we deduce:

$$\begin{pmatrix} h \\ \mathbf{q} \\ \frac{\mathbf{q}\otimes\mathbf{q}}{h} + \frac{g}{2} h^2 \mathbf{Id} \end{pmatrix} = \int_{\mathbb{R}^2} \begin{pmatrix} 1 \\ \xi \\ \xi \otimes \xi \end{pmatrix} M(\xi) \, \mathrm{d}\xi. \tag{3.13}$$

This equivalency produces a very useful consequence: the nonlinear system (2.1), (2.2) can be viewed as a linear transport equation (3.11) on a nonlinear quantity M, for which it is easier to find a simple numerical scheme with good theoretical properties. Since the Saint-Venant system results from an integration in ξ of the kinetic equation, an integration of the numerical scheme for the kinetic equation (3.11) provides a solver that is consistent with the Saint-Venant system and that presents interesting stability properties.

Remark 3.1. It is also possible to introduce a kinetic interpretation of the 2D Saint-Venant system (2.1), (2.2) including the topographic source terms. In [39] the authors use such a 1D kinetic interpretation to derive their kinetic solver. It is a very elegant way to discretize the full Saint-Venant system in once but it leads to time consuming algorithms which are not adapted to compute real 2D problems.

3.3. Kinetic solver

In this subsection, we will use 2D finite volume formalism described in Section 3.1 to discretize the kinetic equation (3.11). This will lead to a consistent solver for the homogeneous Saint-Venant system after integration.

In this subsection, we assume that P_i is an interior point. Given the solution $\mathbf{U}_i^n$ at time t^n for each cell, we compute $\mathbf{U}_i^{n+1}$ by the following algorithm with three steps:

- *Definition of the discrete kinetic density.* We define $M_i^n = M(h_i^n, \xi - \mathbf{u}_i^n)$ with M defined by (3.9).
- *Advection scheme.* We use the microscopic equation (3.11) with $Q = 0$. Since this equation is linear, we can apply a simple upwind scheme [22] which defines a density function $f_i^{n+1}(\xi)$

$$f_i^{n+1}(\xi) - M_i^n(\xi) + \frac{\Delta t}{|C_i|} \sum_{j\in K_i} L_{ij} \xi \cdot \mathbf{n}_{ij} M_{ij}^n(\xi) = 0, \tag{3.14}$$

with the fluxes $M_{ij}^n(\xi)$ computed by the upwind formula

$$M_{ij}^n(\xi) = \begin{cases} M_i^n(\xi) & \text{for } \xi \cdot \mathbf{n}_{ij} \geqslant 0, \\ M_j^n(\xi) & \text{for } \xi \cdot \mathbf{n}_{ij} \leqslant 0. \end{cases} \tag{3.15}$$

Note however that the density function $f(\xi)$ is not an equilibrium (see Remark 3.2).

- *Computation of the macroscopic solution.* Nevertheless, by analogy with the computations in (3.13), we can recover the macroscopic quantities $\mathbf{U}_i^{n+1}$ at time t^{n+1} by integration

$$\mathbf{U}_i^{n+1} = \int_{\mathbb{R}^2} \begin{pmatrix} 1 \\ \xi \end{pmatrix} f_i^{n+1}(\xi)\, \mathrm{d}\xi. \tag{3.16}$$

Remark 3.2. The interpretation is that, as usual, the collision term Q, which forces the relaxation of f to Gibbs equilibrium M, is neglected in the advection scheme (3.14). And at each timestep we deduce $M_i^{n+1}(\xi)$ from $\mathbf{U}_i^{n+1}$ which is a way to perform all collisions at once and to recover the Gibbs equilibria without computing them explicitly.

The numerical consistency of the kinetic solver relies on the fact that if we consider the exact solution of the homogeneous kinetic transport equation (3.11) with $\nabla Z = 0$ and $Q = 0$ – we can prove that the macroscopic quantities obtained through the integration process described previously are first-order approximations in time of the solutions of the Saint-Venant system – see [22].

The numerical feasibility of the kinetic solver relies on two facts. First the possibility to neglect the collision term in the microscopic advection scheme. Second the possibility to write directly a finite volume formula, which therefore avoids using the extra variable ξ in the actual implementation. Indeed, Eq. (3.16) can be written with the form (3.3) with

$$F(\mathbf{U}_i, \mathbf{U}_j, \mathbf{n}_{ij}) = \mathbf{F}^+(\mathbf{U}_i, \mathbf{n}_{ij}) + \mathbf{F}^-(\mathbf{U}_j, \mathbf{n}_{ij}), \tag{3.17}$$

and

$$\mathbf{F}^+(\mathbf{U}_i, \mathbf{n}_{ij}) = \int_{\xi\cdot\mathbf{n}_{ij}\geqslant 0} \xi\cdot\mathbf{n}_{ij} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M_i(\xi)\, \mathrm{d}\xi, \tag{3.18}$$

$$\mathbf{F}^-(\mathbf{U}_j, \mathbf{n}_{ij}) = \int_{\xi\cdot\mathbf{n}_{ij}\leqslant 0} \xi\cdot\mathbf{n}_{ij} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M_j(\xi)\, \mathrm{d}\xi. \tag{3.19}$$

From the definition (3.9) of M and the property (3.5), M remains non-negative and (3.18), (3.19) imply

$$F_h^+(\mathbf{U}_i, \mathbf{n}_{ij}) \geqslant 0, \quad F_h^-(\mathbf{U}_j, \mathbf{n}_{ij}) \leqslant 0, \tag{3.20}$$

where $F_h^\pm$ are the components of the flux related to the water depth h.

If we choose the χ function on the form (3.8) we can compute the integrals in (3.18) and (3.19) analytically and thus the kinetic velocity ξ does not appear in the resulting kinetic solver that finally looks like a classical macroscopic flux vector splitting solver. We refer to the next subsection – where we derive a slightly different scheme – for a presentation of the exact macroscopic formula corresponding to (3.18), (3.19).

3.4. Numerical implementation

We give here some details on the implementation of the kinetic scheme defined by (3.3), (3.17)–(3.19). For the efficiency of the method, we code in fact a variant where the choice of the function χ depends on the interface under consideration. For an interface with unit normal $\mathbf{n} = (n_x, n_y)^{\mathrm{T}}$, we define a local basis (n, τ) associated to the normal direction and to the tangential one. We denote $\hat{\mathbf{U}}_{\mathbf{n}} = (h, q_n, q_\tau)^{\mathrm{T}}$, the vector deduced from $\mathbf{U}$ by the rotation in this new basis and $\hat{\mathbf{u}} = (u_n, u_\tau)^{\mathrm{T}} = (\frac{q_n}{h}, \frac{q_\tau}{h})^{\mathrm{T}}$. So we have $\hat{\mathbf{U}}_{\mathbf{n}}$ defined by

$$\hat{\mathbf{U}}_{\mathbf{n}} = \mathbf{R}_{\mathbf{n}}\mathbf{U} \quad \text{with } \mathbf{R}_{\mathbf{n}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & n_x & n_y \\ 0 & -n_y & n_x \end{pmatrix}, \tag{3.21}$$

and

$$\mathbf{F}^+(\mathbf{U},\mathbf{n}) = \mathbf{R}_{\mathbf{n}}^{-1}\hat{\mathbf{F}}^+(\hat{\mathbf{U}}_{\mathbf{n}}). \tag{3.22}$$

Using (3.18), we give the detailed expression of $\hat{\mathbf{F}}^+(\hat{\mathbf{U}}_i)$ related to the interface Γ_{ij}

$$\hat{\mathbf{F}}^+(\hat{\mathbf{U}}_{i,\mathbf{n}_{ij}}) = \frac{h_i}{\tilde{c}_i^2}\int_{\{\xi_n \geqslant 0\}\times\mathbb{R}} \xi_n \begin{pmatrix} 1 \\ \xi \end{pmatrix} \chi\left(\frac{\xi - \hat{\mathbf{u}}_i}{\tilde{c}_i}\right) \mathrm{d}\xi, \tag{3.23}$$

or, after the change of variables $\mathbf{w} = \frac{\xi - \hat{\mathbf{u}}_i}{\tilde{c}_i}$,

$$\hat{\mathbf{F}}^+(\hat{\mathbf{U}}_{i,\mathbf{n}_{ij}}) = \begin{pmatrix} \hat{F}_h^+ \\ \hat{F}_{q_n}^+ \\ \hat{F}_{q_\tau}^+ \end{pmatrix} = h_i \int_{\{w_n \geqslant \frac{-u_{i,n}}{\tilde{c}_i}\}\times\mathbb{R}} (u_{i,n} + w_n\tilde{c}_i) \begin{pmatrix} 1 \\ u_{i,n} + w_n\tilde{c}_i \\ u_{i,\tau} + w_\tau\tilde{c}_i \end{pmatrix} \chi(\mathbf{w})\, \mathrm{d}\mathbf{w}. \tag{3.24}$$

Now if we choose the χ function of the form (3.8) for the interface under consideration, easy computations lead to

$$\hat{\mathbf{F}}^+(\hat{\mathbf{U}}_{i,\mathbf{n}_{ij}}) = \frac{\tilde{c}_i}{6g\sqrt{3}} \begin{pmatrix} 3(M_+^2 - M_-^2) \\ 2(M_+^3 - M_-^3) \\ 3u_{i,\tau}(M_+^2 - M_-^2) \end{pmatrix}, \tag{3.25}$$

where

$$M_\pm = (u_{i,n} \pm \tilde{c}_i\sqrt{3})_+ \tag{3.26}$$

and the kinetic interpretation leads in this case to a quite simple macroscopic scheme. Note that other choices of χ functions with the same property are possible.

3.5. Upwind kinetic scheme

Note that due to the fact that $\chi(\mathbf{w})$ is even, the term with w_τ disappears in (3.24). We obtain an analogous property for $\mathbf{F}^-$ and so the flux related to the tangential component looks like

$$\hat{F}_{q_\tau}(\hat{\mathbf{U}}_{i,\mathbf{n}_{ij}}, \hat{\mathbf{U}}_{j,\mathbf{n}_{ij}}) = u_{i,\tau}\, \hat{F}_h^+(\hat{\mathbf{U}}_{i,\mathbf{n}_{ij}}) + u_{j,\tau}\, \hat{F}_h^-(\hat{\mathbf{U}}_{j,\mathbf{n}_{ij}}). \tag{3.27}$$

In order to reduce the diffusion of the scheme, we slightly modify the computation of this flux. For the computation of $q_{i,\tau}^{n+1}$ we replace (3.27) by the following expression:

$$\hat{F}_{q_\tau}(\hat{\mathbf{U}}_{i,\mathbf{n}_{ij}}, \hat{\mathbf{U}}_{j,\mathbf{n}_{ij}}) = u_{ij,\tau}\, \hat{F}_h(\hat{\mathbf{U}}_{i,\mathbf{n}_{ij}}, \hat{\mathbf{U}}_{j,\mathbf{n}_{ij}}), \tag{3.28}$$

with

$$u_{ij,\tau} = \begin{cases} u_{i,\tau} & \text{for } \hat{F}_h \geqslant 0, \\ u_{j,\tau} & \text{for } \hat{F}_h \leqslant 0. \end{cases} \tag{3.29}$$

Formula (3.29) introduces some upwinding depending on the sign of the total flux. In [10] a numerical result shows the efficiency of (3.28), (3.29).

3.6. Boundary conditions

The treatment of the boundary conditions is presented in detail in [9]. Here, we just recall some main features about the computation of the boundary flux $F(\mathbf{U}_i^n, \mathbf{U}_{e,i}^n, \mathbf{n}_i)$ appearing in (3.3). The variable $\mathbf{U}_{e,i}^n$ can be interpreted as an approximation of the solution in a ghost cell adjacent to the boundary. As before we introduce the local coordinates and define $\hat{\mathbf{U}}_{e,i}^n = (h_{e,i}^n, q_{e,i,n}^n, q_{e,i,\tau}^n)^{\mathrm{T}}$. Then, we can use the local flux vector splitting form associated to the kinetic formulation

$$\hat{F}(\hat{\mathbf{U}}_{i,\mathbf{n}_i}^n, \hat{\mathbf{U}}_{e,i}^n) = \hat{\mathbf{F}}^+(\hat{\mathbf{U}}_{i,\mathbf{n}_i}^n) + \hat{\mathbf{F}}^-(\hat{\mathbf{U}}_{e,i}^n).$$

On the solid wall we prescribe a continuous slip condition – see Section 2. In the numerical scheme we prescribe it weakly by defining $\hat{\mathbf{U}}_{e,i}^n = (h_i^n, -q_{i,n}^n, q_{i,\tau}^n)^{\mathrm{T}}$. It follows that finally

$$\hat{F}(\hat{\mathbf{U}}_{i,\mathbf{n}_i}^n, \hat{\mathbf{U}}_{e,i}^n) = \left(0, \frac{g h_i^{n2}}{2}, 0\right)^{\mathrm{T}}. \tag{3.30}$$

On the fluid boundaries, the type of the flow and then the number of boundary conditions depend on the Froude number. Here, we consider a local Froude number associated to the normal component of the velocity. For the fluvial cases, we define completely $\mathbf{U}_e$ by adding to the given boundary condition, the assumption that the *Riemann invariant* that is related to the outgoing characteristic is constant along this characteristic (see [9]).

3.7. Properties of the scheme

It is clear from (3.3), (3.17)–(3.19) that the scheme is conservative. We established also in Section (3.3) that the scheme is consistent. Now the CFL condition for the explicit scheme (3.14) applied to the linear microscopic equation writes

$$\Delta t \leqslant \min \frac{|C_i|}{(L_i + \sum_{j\in K_i} L_{ij})(|\mathbf{u}_i^n| + w_M \tilde{c}_i^n)}, \tag{3.31}$$

where w_M is defined in (3.6). It is obvious from (3.14) to (3.15) that under this CFL condition the non-negativity of the density function is preserved by the advection scheme. Some computations allow to prove that this stability property extends to the macroscopic water depth – see the complete proof in Appendix A.

Theorem 3.1. *Under the CFL condition* (3.31), *the kinetic scheme* (3.3), (3.17), (3.22), (3.24) *preserves the water depth positivity.*

Note also that the computation of the dry areas does not need any special feature.

4. Well-balanced scheme: the hydrostatic reconstruction

In order to be able to compute realistic flows we consider now the case $\nabla Z \neq 0$ and introduce a numerical discretization for the source terms. As motivated in the introduction the fundamental point is to satisfy the *well-balanced* requirement, i.e. to preserve a local discrete equivalent of the continuous still water steady-state (2.6)

$$\begin{pmatrix} \forall j \in K_i & h_j^n + Z_j = h_i^n + Z_i = H \\ & \mathbf{u}_j^n = \mathbf{u}_i^n = 0 \end{pmatrix} \Rightarrow \begin{matrix} h_i^{n+1} + Z_i = H \\ \mathbf{u}_i^{n+1} = 0 \end{matrix}. \tag{4.1}$$

In this section, we present a very general way to do that, starting from any consistent homogeneous solver. As we construct in the previous section a homogeneous solver which is able to ensure the non-negativity of the water depth, a crucial requirement is to be able to preserve this stability property. The method we present is an extension to the two-dimensional flows of the *interface hydrostatic reconstruction* method we developed in [3] in the 1D framework.

Given the solution $\mathbf{U}_i^n$ at time t^n for each cell, we compute $\mathbf{U}_i^{n+1}$ by the following algorithm with four steps:

- *Interface topography.* We first construct a piecewise constant approximation of the bottom topography $Z(x)$

$$Z_i = \frac{1}{|C_i|}\int_{C_i} Z(x)\,\mathrm{d}x.$$

We define an interface topography

$$Z_{ij}^* = Z_{ji}^* = \max(Z_i, Z_j). \tag{4.2}$$

- *Interface water depth.* From the discrete form of the well-balanced requirement (4.1) we define new interface values by $\mathbf{U}_{ij}^* = (h_{ij}^*, h_{ij}^*\mathbf{u}_i)^{\mathrm{T}}$ where h_{ij}^* is the *hydrostatic reconstructed* water depth

$$h_{ij}^* = (h_i + Z_i - Z_{ij}^*)_+. \tag{4.3}$$

The fact that $h_{ij}^* \leqslant h_i$, that is an obvious consequence of the definitions (4.2), (4.3), is a crucial point to ensure the positivity preserving property for the well-balanced scheme (see the proof of Theorem 4.1 in Appendix A).

- *Source term.* From the hydrostatic balance

$$\nabla\left(\frac{g}{2}h^2\right) = -gh\nabla Z,$$

we introduce an adapted discretization of the source terms

$$S(\mathbf{U}_i, \mathbf{U}_{ij}^*, \mathbf{n}_{ij}) = \begin{pmatrix} 0 \\ \frac{g}{2}(h_{ij}^{*2} - h_i^2)\mathbf{n}_{ij} \end{pmatrix}. \tag{4.4}$$

This definition allows to ensure the well-balancing property and is consistent with the continuous source term (see proof of Theorem 4.1 in Appendix A).

- *Computation of the solution.* Finally, we deduce the well-balanced scheme from the previous homogeneous solver by using the interface values introduced in (4.3) instead of the in-cell values in the definition of the fluxes (3.25), (3.26)

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \sum_{j\in K_i}\sigma_{ij}F(\mathbf{U}_{ij}^{*,n}, \mathbf{U}_{ji}^{*,n}, \mathbf{n}_{ij}) - \sigma_i F(\mathbf{U}_i^n, \mathbf{U}_{e,i}^n, \mathbf{n}_i) + \sum_{j\in K_i}\sigma_{ij}S(\mathbf{U}_i^n, \mathbf{U}_{ij}^{*,n}, \mathbf{n}_{ij}). \tag{4.5}$$

Remark 4.1. The first step of the algorithm is needed to be done only one time at the beginning of the computation. It is no more the case for the second order case (see next section).

We can now prove that the hydrostatic reconstruction strategy allows us to preserve the still water steady-state while preserving the positivity property of the homogeneous solver (see the proof in Appendix A).

Theorem 4.1. *The scheme defined by (4.4), (4.5) with (3.17)–(3.19) satisfies the following properties*:

(i) *it is consistent with the Saint-Venant system with source terms*,

(ii) *it preserves the water depth positivity under the CFL condition*

$$\Delta t \leqslant \min \frac{|C_i|}{\sum_{j\in K_i}\left[L_{ij}\ (|\mathbf{u}_i^n| + w_M \tilde{c}_{ij}^{*,n})\right] + L_i\ (|\mathbf{u}_i^n| + w_M \tilde{c}_i^n)}, \tag{4.6}$$

a fortiori if Δt *satisfies (3.31)*,

(iii) *it preserves the still water steady-state.*

Note that this extension of a positivity preserving homogeneous solver to a positivity preserving well-balanced one does not increase the complexity of the algorithms. Furthermore, only the *input values* and not the solver itself are modified.

Remark 4.2. From the definitions (4.2)–(4.4), it is obvious that for $Z = Cst$, we recover the original homogeneous scheme.

Remark 4.3. It appears in the proof of Theorem 4.1 that to construct the scheme on the interface values $\mathbf{U}_{ij}^*$ instead of the cell values $\mathbf{U}_i$ allows us to numerically preserve at each interface the balance between the hydrostatic pressure and the influence of the topographic source terms that is associated to the still water steady-state. This explains the name *interface hydrostatic reconstruction method*. For further details refer to [3].

5. "Second-order" extension

In order to improve the accuracy of the results the first-order scheme defined in Sections 3 and 4 can be extended to a formally second-order one using a MUSCL like extension (see [44]). In Section 5.1, we define limited reconstructed variables and in Section 5.2, we introduce a "second-order" well-balanced scheme that preserves the positivity and equilibrium properties of the first-order scheme. See Remark 5.2 about the quotation marks.

5.1. Second-order reconstructions

In the definition of the flux (3.17), we replace the piecewise constant values $\mathbf{U}_i, \mathbf{U}_j$ by more accurate reconstructions deduced from piecewise linear approximations, namely the values $\mathbf{U}_{ij}, \mathbf{U}_{ji}$ reconstructed on both sides of the interface. More precisely, we are looking for piecewise linear approximation of the primitive variable $\hat{\mathbf{W}} = (h, u_n, u_\tau)^{\mathrm{T}}$, actually the detailed expression of the flux given in (3.24) uses the primitive variables.

We divide each cell C_i in subtriangles obtained by joining each edge Γ_{ij} to the node P_i, we denote T_{ij} the subtriangle related to Γ_{ij} (see Fig. 4). We denote $|C_{ij}|$ the area of T_{ij}. Let M be the middle point of the interface Γ_{ij}, we define $\hat{\mathbf{W}}_{ij} = (h_{ij}, u_{ij,n}, u_{ij,\tau})^{\mathrm{T}}$ as an approximation of $\hat{\mathbf{W}}$ at point M in two steps. First we deduce $\hat{\mathbf{W}}_{ij}^{(1)} = (h_{ij}^{(1)}, u_{ij,n}, u_{ij,\tau})^{\mathrm{T}}$ from a piecewise linear reconstruction on the subtriangle T_{ij}:

$$\hat{\mathbf{W}}_{ij}^{(1)} = \hat{\mathbf{W}}_i + \overrightarrow{P_i M} \cdot \nabla \hat{\mathbf{W}}_{ij}, \tag{5.1}$$

with $\nabla \hat{\mathbf{W}}_{ij}$ defined here as follows (see [25]).

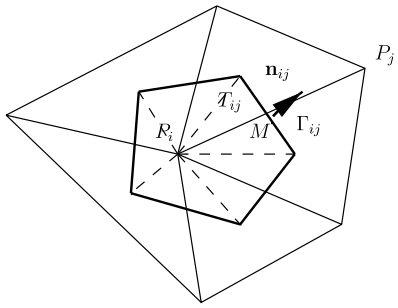

Fig. 4. Subcells T_{ij}.

If the point M belongs to the triangle T_k, we denote $\nabla\hat{\mathbf{W}}_M = \nabla\hat{\mathbf{W}}|_{T_k}$ where $\nabla\hat{\mathbf{W}}|_{T_k}$ is the constant gradient of $\hat{\mathbf{W}}$ deduced from a P1 approximation on the triangle T_k. We denote by $\nabla\hat{\mathbf{W}}_i$ an approximate gradient at node P_i computed by a weighted average of the gradients on the surrounding triangles

$$\nabla\hat{\mathbf{W}}_i = \frac{\sum_{k\in T_i}|C_k|\ \nabla\hat{\mathbf{W}}|_{T_k}}{\sum_{k\in T_i}|C_k|} \tag{5.2}$$

and

$$\nabla\hat{\mathbf{W}}_{mi} = (1+\beta)\nabla\hat{\mathbf{W}}_i - \beta\nabla\hat{\mathbf{W}}_M, \quad 0 \leqslant \beta \leqslant 1, \tag{5.3}$$

where T_i is the set of triangles surrounding the node P_i.

Then, we use an appropriate slope limiter to deduce $\nabla\hat{\mathbf{W}}_{ij}$

$$\nabla\hat{\mathbf{W}}_{ij} = \lim(\nabla\hat{\mathbf{W}}_M, \nabla\hat{\mathbf{W}}_{mi}). \tag{5.4}$$

In the following computations we have used either the minmod limiter defined by

$$\lim(a,b) = \begin{cases} 0 & \text{if } \operatorname{sign}(a) \neq \operatorname{sign}(b), \\ \operatorname{sign}(a)\min(|a|,|b|) & \text{otherwise} \end{cases}$$

or the Van Albada limiter defined by

$$\lim(a,b) = \begin{cases} 0 & \text{if } \operatorname{sign}(a) \neq \operatorname{sign}(b), \\ \frac{a(b^2+\varepsilon)+b(a^2+\varepsilon)}{a^2+b^2+2\varepsilon} & \text{otherwise} \end{cases}$$

with $0 \leqslant \varepsilon \ll 1$.

But this first linear reconstruction does not ensure the conservativity of the water depth. Thus, we define h_{ij} from $h_{ij}^{(1)}$ and h_i in a correction step (see [38])

$$h_{ij} = h_i + \beta_i^+(h_{ij}^{(1)} - h_i)_+ - \beta_i^-(h_{ij}^{(1)} - h_i)_-, \quad \beta_i^\pm = \min\left(1, \frac{\sum_{j\in K_i}|C_{ij}|(h_{ij}^{(1)} - h_i)_\mp}{\sum_{j\in K_i}|C_{ij}|(h_{ij}^{(1)} - h_i)_\pm}\right).$$

This second step ensures at the same time the conservation of the water depth

$$\sum_{j\in K_i}|C_{ij}|h_{ij} = \left(\sum_{j\in K_i}|C_{ij}|\right)h_i = |C_i|h_i, \tag{5.5}$$

and some control of the reconstructed values

$$h_{ij} \in \left[\min(h_{ij}^{(1)}, h_i), \max(h_{ij}^{(1)}, h_i)\right]. \tag{5.6}$$

It is well known that the linear reconstruction procedure associated to the use of some limiter preserves the positivity of the water depth. The relation (5.6) shows that our method to ensure the conservation property (5.5) preserves this positivity property. Thus, the second-order reconstruction is conservative and positivity preserving.

In the case where $\mathbf{B} = 0$, the formally second-order scheme is obtained by replacing (3.3) by

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \sum_{j \in K_i} \sigma_{ij} F(\mathbf{U}_{ij}^n, \mathbf{U}_{ji}^n, \mathbf{n}_{ij}) - \sigma_i F(\mathbf{U}_i^n, \mathbf{U}_{e,i}^n, \mathbf{n}_i). \tag{5.7}$$

This means that, once the reconstructed values are computed at the middle of the interface, we assume that the variables are constant on each side of the interface and we apply again a locally 1D solver.

5.2. "Second-order" well-balanced scheme

For the cases where $\mathbf{B} \neq 0$ we consider also piecewise linear approximation of the variable Z and we reconstruct values Z_{ij}, Z_{ji} on both sides of the interface as it is done before for the primitive variables. In fact, so that the formally second-order scheme preserves the *well-balanced* property, it is necessary that the formally second-order reconstruction preserves an interface equilibrium. It means that if

$$h_i + Z_i = h_j + Z_j = H, \quad \mathbf{u}_i = \mathbf{u}_j = 0, \tag{5.8}$$

then the "second-order" reconstructed values have to satisfy

$$h_{ij} + Z_{ij} = h_{ji} + Z_{ji} = H, \quad \mathbf{u}_{ij} = \mathbf{u}_{ji} = 0. \tag{5.9}$$

The velocity part is obvious but, as we require also that the "second-order" reconstruction preserves the positivity of the water depth, it is proved in [3] that the right way to build a *well-balanced* formally second-order scheme is to reconstruct and correct the variables $h + Z$ and h and then to deduce the interface values for Z – see [3] for further explanations, especially for the case of dry/wet interface.

Given the solution $\mathbf{U}_i^n$ at time t^n for each cell, we thus compute $\mathbf{U}_i^{n+1}$ by the following algorithm with five steps:

- "*Second-order*" *reconstruction.* We define "second-order" reconstructions of the primitive variables as it is described in the previous subsection. We also apply the same techniques to obtain a "second-order" reconstruction of the free surface.
- *Interface topography.* From these "second-order" reconstructed values we derive "second-order" reconstructed values Z_{ij} for the topography. Then, we define an interface topography

$$Z_{ij}^* = Z_{ji}^* = \max(Z_{ij}, Z_{ji}). \tag{5.10}$$

- *Interface water depth.* We define new hydrostatic reconstructed interface values by $\mathbf{U}_{ij}^* = (h_{ij}^*,\ h_{ij}^* \mathbf{u}_{ij})^{\mathrm{T}}$ with

$$h_{ij}^* = (h_{ij} + Z_{ij} - Z_{ij}^*)_+. \tag{5.11}$$

- *Source term.* We define an interface source term

$$S(\mathbf{U}_{ij}, \mathbf{U}_{ij}^*, \mathbf{n}_{ij}) = \begin{pmatrix} 0 \\ \frac{g}{2}(h_{ij}^{*2} - h_{ij}^2)\mathbf{n}_{ij} \end{pmatrix}. \tag{5.12}$$

By contrast to the first-order scheme, we also introduce a centered source term to satisfy the consistency of the numerical scheme

$$S^c(\mathbf{U}_i^n, \mathbf{U}_{ij}^n, Z_i, Z_{ij}, \mathbf{n}_{ij}) = \begin{pmatrix} 0 \\ -\frac{g}{2}(h_{ij} + h_i)(Z_{ij} - Z_i)\mathbf{n}_{ij} \end{pmatrix}. \tag{5.13}$$

- *Computation of the solution.* Finally, we write the formally second-order well-balanced scheme by introducing the new interface values (5.11) in the fluxes (3.25), (3.26)

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \sum_{j\in K_i} \sigma_{ij} F(\mathbf{U}_{ij}^{*,n}, \mathbf{U}_{ji}^{*,n}, \mathbf{n}_{ij}) - \sigma_i F(\mathbf{U}_i^n, \mathbf{U}_{e,i}^n, \mathbf{n}_i) + \sum_{j\in K_i} \sigma_{ij}[S(\mathbf{U}_{ij}^n, \mathbf{U}_{ij}^{*,n}, \mathbf{n}_{ij}) + S^c(\mathbf{U}_i^n, \mathbf{U}_{ij}^n, Z_i, Z_{ij}, \mathbf{n}_{ij})]. \tag{5.14}$$

Remark 5.1. Note that there is now two sets of interface values: first the "second-order" reconstructed values deduced from the in-cell values, second the hydrostatic reconstructed values deduced from the "second-order" reconstructed values.

As in the first-order case we can now prove that the hydrostatic reconstruction strategy allows us to preserve the still water steady-state while preserving the positivity property of the homogeneous solver – see the proof in Appendix A.

Theorem 5.1. *The formally second-order scheme defined by (5.14), (5.13) with (3.17)–(3.19) satisfies the following properties*:

(i) *it preserves the water depth positivity under the CFL condition*

$$\Delta t \leqslant \min\left[\min_{i\in\mathbf{S}_i}\min_{j\in K_i}\frac{|C_{ij}|}{L_{ij}\left(|\mathbf{u}_{ij}^n| + w_M\tilde{c}_{ij}^{*,n}\right)}, \min_{i\in\mathbf{G}_i}\max_{0\leqslant\alpha\leqslant 1}\left(\alpha\frac{|C_i|}{L_i\left(|\mathbf{u}_i^n| + w_M\tilde{c}_i^n\right)}, (1-\alpha)\min_{j\in K_i}\frac{|C_{ij}|}{L_{ij}\left(|\mathbf{u}_{ij}^n| + w_M\tilde{c}_{ij}^{*,n}\right)}\right)\right], \tag{5.15}$$

(ii) *it preserves the still water steady-state.*

Remark 5.2. Note that we do not prove here that the scheme is second-order accurate. A complete proof would suppose more sophisticated reconstructions as ENO or WENO reconstructions – see [3,1,28]. We do not use them here for two reasons. First our main goal is to derive a stable and accurate but simple scheme that can be used for industrial purposes. Second we show in the next section that our simpler "second-order" reconstruction has already a great impact on the accuracy of the results. Moreover, we consider some tests for which analytical solutions exist and we exhibit that the second-order accuracy is obtained. Let us also precise that second-order accuracy in time is obtained as usual by a Runge–Kutta method (the CFL condition need not be modified).

6. Numerical results

We present here the numerical results of different test problems. We begin with the two-dimensional version of a classical ideal test problem extracted from [23] and commonly used (see, e.g. [12,17]): a stationary flow over a parabolic bump for which an exact solution is known. We consider a rectangular channel of length 20. and width 2. (we assume a non-dimensionalized problem), the bottom is defined by

$$Z(x,y) = \begin{cases} 0.2 - 0.05(x-10.)^2 & \text{if } 8. \leqslant x \leqslant 12. \quad \forall y, \\ 0. & \text{elsewhere.} \end{cases}$$

Depending on the values of the boundary conditions, we compute three different flow situations defined as follows:

- fluvial flow
  inflow: $\mathbf{q}_g = (4.42, 0)^{\mathrm{T}}$, outflow $h_g = 2$,

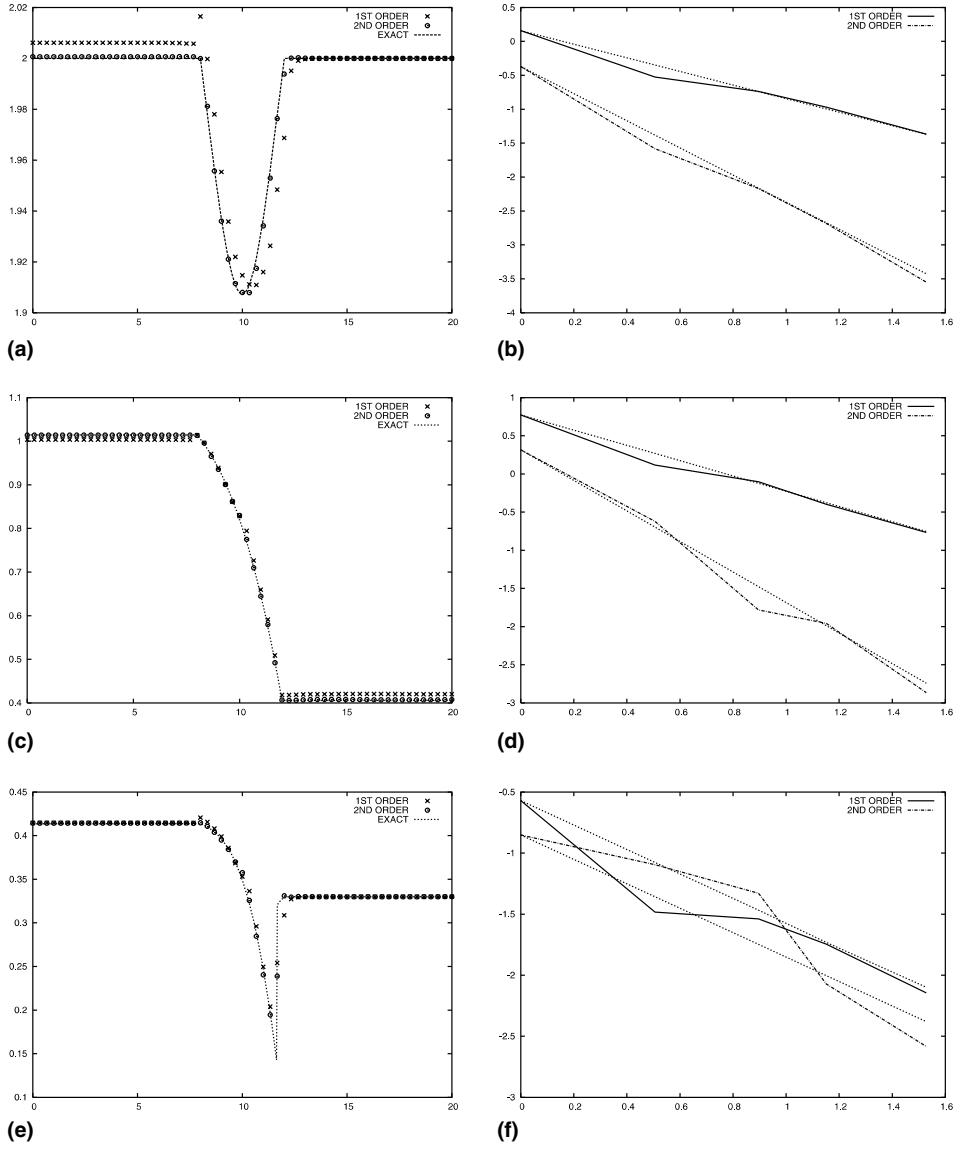

Fig. 5. Stationary flows over a bump: (a) fluvial flow – free surface; (b) fluvial flow – convergence rate; (c) transcritical flow – free surface; (d) transcritical flow – convergence rate; (e) transcritical flow with shock – free surface; (f) transcritical flow with shock – convergence rate.

- transcritical without shock (torrential outflow)
  inflow: $\mathbf{q}_g = (1.53, 0)^{\mathrm{T}}$, initial water depth $h^0 = 0.66$,
- transcritical with shock
  inflow: $\mathbf{q}_g = (0.18, 0)^{\mathrm{T}}$, outflow $h_g = 0.33$.

The given discharge is prescribed for each node of the inflow boundary. The initial solution is given by $\mathbf{q}^0 = \mathbf{q}_g$, $h^0 = h_g$. The three flows are computed on a rather coarse unstructured mesh of 510 nodes and 886 triangles (60 edges on the length and 6 edges on the width). In Figs. 5(a)–(c) and (e), the free surface profiles computed with the first-order and "second-order" schemes are compared to the exact solution. In these figures we have plotted only the points on the line $y = 0$ since the two-dimensional effects are negligible as shown in Fig. 6 where all the points of the free surface are plotted for the transcritical case. Results are quite good for such a coarse mesh. The improvement due to the formally second-order extension appears to be noticeable for all the cases. Note in particular the improvement on the computation of the free surface on the left side of the bump for the three cases and on the right side of the bump for the second case, i.e. where the water depth is not prescribed by the boundary conditions. Note also that the presence of a sonic point in the two last test cases does not need a special treatment.

In order to show the improvement due to the formally second-order reconstruction it appears interesting to look at the convergence rate of the error versus the space discretization for the three above problems. We have plotted in Figs. 5(b), (d) and (f), the $\log(L^1\text{-error})$ of the water depth versus $\log(h_{a_0}/h_a)$ for the first and the "second-order" scheme and they are compared to the theoretical order (we denote by h_a the average edge length and h_{a_0} the average edge length of the coarser mesh). These errors are computed on five meshes with 10, 20, 30, 40 and 60 edges on the length of the channel. These meshes are very coarse, nevertheless, it appears that the computed convergence rate are not far from the theoretical ones, the formally second-order scheme provides an effective convergence up to the second-order when the flow is sufficiently smooth and according to the estimations, the "second-order" scheme reduces to first-order near a discontinuity.

The second test problem is one of the tests of the Telemac code developed at EDF/LNHE [26], it concerns a water drop in a basin and we look at the solution after some reflections on the walls. The basin is a $20. \times 20.$ square box with flat bottom, the initial solution shown in Fig. 7(a), is defined by

$$h = 2.4(1. + \mathrm{e}^{-0.25[(x-10.05)^2+(y-10.05)^2]}), \quad \mathbf{u} = 0.$$

The solutions at $t = 1., 2., 3., 4.$ s obtained with the "second-order" approximation are given in Figs. 7(b)–(e) while the solution at $t = 4.$ s, damped by the first-order scheme is shown in Fig. 7(f). This result shows the accuracy improvement due to the "second-order" scheme, even for a problem with complex 2D interactions.

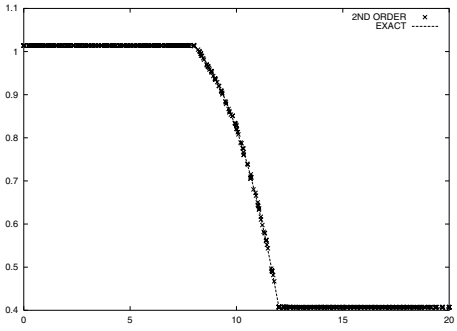

Fig. 6. Transcritical flow over a bump. Free surface.

The third test problem is a real life application, it concerns the Malpasset dam break. All the details on the data and a reference solution computed with the Telemac code are given in [27]. We present here in Fig. 8, the initial solution and the "second-order" solutions at $t = 1000$ s and $t = 2500$ s. These solutions are in good agreement with solutions obtained by other methods in [27]. The computation of this problem allows us to test, among others, the ability of the method to treat the still water (the sea area before the wave reaches it) and the wet–dry interfaces – which do not need any special treatment with our method.

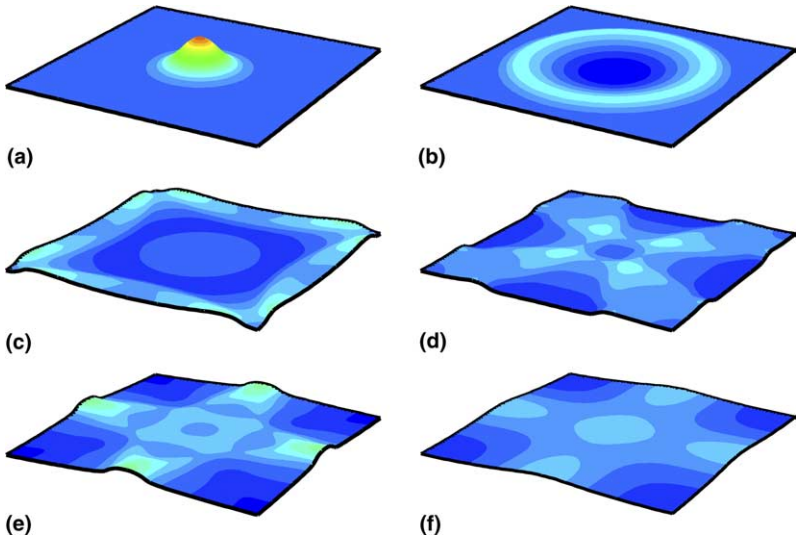

Fig. 7. Water drop in a basin: (a) $t = 0$ s; (b) "second-order", $t = 1$ s; (c) "second-order", $t = 2$ s; (d) "second-order", $t = 3$ s; (e) "second-order", $t = 4$ s; (f) first-order, $t = 4$ s.

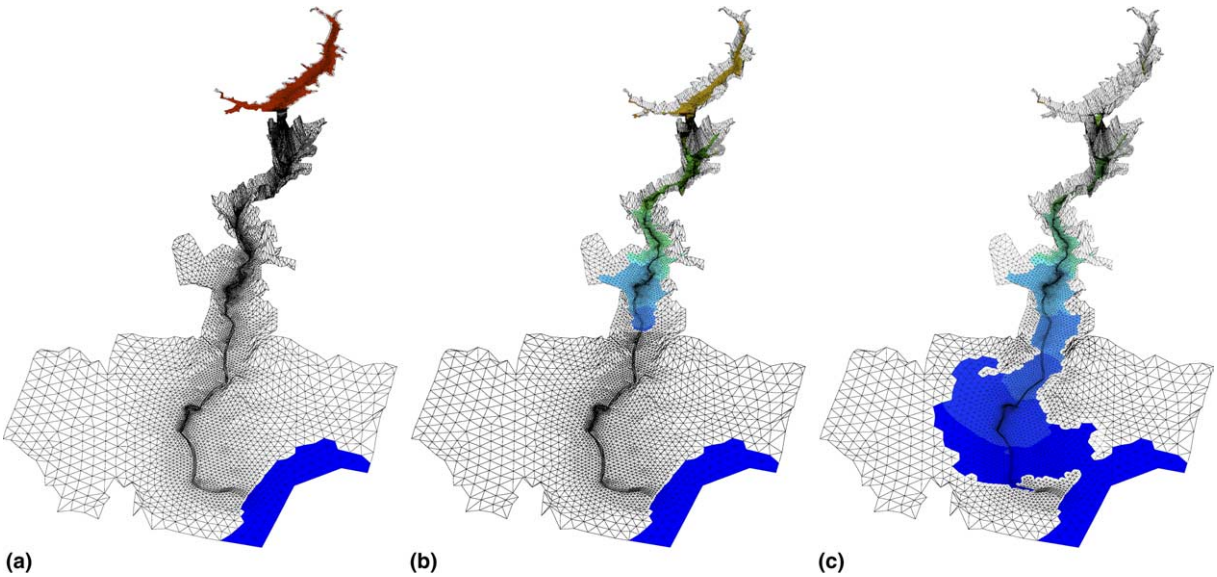

Fig. 8. Malpasset dam break: (a) initial solution; (b) $t = 1000$ s; (c) $t = 2500$ s.

7. Conclusion and outlook

In this article, we have introduced on one hand a stable homogeneous two-dimensional kinetic solver and on the other a hydrostatic reconstruction method to compute the source terms while preserving the stability properties of the homogeneous solver. We have also presented a formally second-order compatible extension. Thanks to these three ingredients we finally derived a positivity preserving well-balanced "second-order" scheme. We emphasize that this solution method seems to be a good compromise between efficiency, stability and accuracy. These properties are experimentally verified by using the algorithm to reproduce complex physical phenomena.

Moreover, let us note some extensions that can be derived. The stability properties of the kinetic solver can be used to derive stable schemes for avalanche flows [35] and can be extended to a multilayer Saint-Venant model [2]. An extension of the hydrostatic reconstruction that preserves all the 1D subcritic steady-states is under investigation and the same idea can be adapted to take into account the relation between the Darcy equation and the Saint-Venant system with strong friction coefficient. Note also that – if we consider a more sophisticated χ function – the semi-discrete version of the presented well-balanced scheme satisfies an in-cell entropy inequality (see [3] for a 1D proof). The extension of this property to the fully discrete scheme is under investigation.

Acknowledgments

The authors thank F. Bouchut, J.M. Hervouet, R. Klein and B. Perthame for fruitful discussions and helpful comments. This work was partially supported by EDF/LNHE and by HYKE European programme HPRN-CT-2002-00282 (http://www.hyke.org).

Appendix A

Proof of Theorem 3.1. (*Homogeneous kinetic solver*) Suppose that we have $h_i^n \geqslant 0$. We want to prove that $h_i^{n+1} \geqslant 0$. We give the proof for a general χ function. It includes of course the particular choice (3.8). It follows from the definitions (3.3), (3.17), (3.22), (3.24) that

$$h_i^{n+1} = h_i^n - \sum_{j\in K_i} \sigma_{ij}^n (\hat{F}_h^+(\hat{\mathbf{U}}_{i,\mathbf{n}_{ij}}^n) + \hat{F}_h^-(\hat{\mathbf{U}}_{j,\mathbf{n}_{ij}}^n)) - \sigma_i(\hat{F}_h^+(\hat{\mathbf{U}}_{i,\mathbf{n}_i}^n) + \hat{F}_h^-(\hat{\mathbf{U}}_{e,i}^n)). \tag{A.1}$$

The relations (3.20) and (3.22) imply

$$\hat{F}_h^-(\hat{\mathbf{U}}_{j,\mathbf{n}_{ij}}^n) \leqslant 0, \quad \hat{F}_h^-(\hat{\mathbf{U}}_{e,i}^n) \leqslant 0, \tag{A.2}$$

and using the expression for the flux (3.24), we have

$$h_i^{n+1} \geqslant h_i^n \left(1 - \sum_{j\in K_i} \sigma_{ij} \int_{\{w_n \geqslant \frac{-u_{i,n_{ij}}}{\tilde{c}_i}\}\times\mathbb{R}} (u_{i,n_{ij}} + w_n\tilde{c}_i)\chi(\mathbf{w})\,\mathrm{d}\mathbf{w} - \sigma_i \int_{\{w_n \geqslant \frac{-u_{i,n_i}}{\tilde{c}_i}\}\times\mathbb{R}} (u_{i,n_i} + w_n\tilde{c}_i)\chi(\mathbf{w})\,\mathrm{d}\mathbf{w}\right). \tag{A.3}$$

Since χ satisfies (3.5)–(3.7), we have for each n

$$\int_{\{w_n \geqslant \frac{-u_{i,n}}{\tilde{c}_i}\}\times\mathbb{R}} (u_{i,n} + w_n\tilde{c}_i)\chi(\mathbf{w})\,\mathrm{d}\mathbf{w} \leqslant |u_{i,n}| + \tilde{c}_i \int_{\{w_n \geqslant 0\}\times\mathbb{R}} w_n\chi(\mathbf{w})\,\mathrm{d}\mathbf{w}, \tag{A.4}$$

and from (3.5), (3.6) we deduce

$$\int_{\{w_n \geqslant 0\}\times\mathbb{R}} w_n \chi(\mathbf{w})\, d\mathbf{w} \leqslant \int_{\{0\leqslant w_n\leqslant 1\}\times\mathbb{R}} \chi(\mathbf{w})\, d\mathbf{w} + \int_{\{w_n\geqslant 1\}\times\mathbb{R}} w_n^2 \chi(\mathbf{w})\, d\mathbf{w} \leqslant 1. \tag{A.5}$$

Finally, using (A.4), (A.5) in (A.3), we obtain

$$h_i^{n+1} \geqslant h_i^n \left(1 - \frac{\Delta t}{|C_i|}\left[\sum_{j\in K_i}\left(L_{ij}\left(|u_{i,n_{ij}}| + \tilde{c}_i\right)\right) + L_i\left(|u_{i,n_i}| + \tilde{c}_i\right)\right]\right),$$

and it follows the positivity of h_i^{n+1} under the CFL condition (3.31) (from (3.7) we have $w_M \geqslant 1$). □

Proof of Theorem 4.1. (*First-order well-balanced scheme*) For (i) we suppose that the homogeneous solver is consistent. Then, the consistency property for the whole scheme is related to the following alternative form for the source term (4.4)

$$S(\mathbf{U}_i, \mathbf{U}_{ij}^*, \mathbf{n}_{ij}) = \begin{pmatrix} 0 \\ -g\frac{h_{ij}^* + h_i}{2}\widetilde{\Delta Z}_{ij}\mathbf{n}_{ij} \end{pmatrix},$$

where $\widetilde{\Delta Z}_{ij} = \min(h_i, (Z_j - Z_i)_+)$. Using this relation we prove in [3] that the scheme is consistent in the sense that is given in [7].

For the non-negativity property (ii), the key-point is that the definitions (4.2), (4.3) imply that $h_{ij}^* \leqslant h_i$. It follows that the out-fluxes are smaller than those of the corresponding homogeneous case. Thus, the positivity preserving property of the homogeneous solver is obviously preserved. More precisely, in the particular case of the kinetic solver, we can first prove as in Theorem 3.1 that the water depth positivity is preserved under the CFL condition (4.6). Then, we deduce from the relation $h_{ij}^* \leqslant h_i$ that the CFL condition (3.31) is more restrictive than (4.6).

To prove the preservation of the still water steady-state (iii), we assume that the solution at time t^n satisfies (4.1), then we have

$$\sum_{j\in K_i} \sigma_{ij} F(\mathbf{U}_{ij}^{*,n}, \mathbf{U}_{ji}^{*,n}, \mathbf{n}_{ij}) = \sum_{j\in K_i} \sigma_{ij} \begin{pmatrix} 0 \\ \frac{g}{2} h_{ij}^{*2}\mathbf{n}_{ij} \end{pmatrix}. \tag{A.6}$$

Concerning the boundary term, we assume also that the boundary conditions will preserve the steady-state (they can be either a slip condition, a given flux $\mathbf{q} = 0$ or a water depth given $h + Z = H$). Following the treatment of the boundary conditions developed in [9] the boundary term reduces to

$$\sigma_i F(\mathbf{U}_i^n, \mathbf{U}_{e,i}^n, \mathbf{n}_i) = \begin{pmatrix} 0 \\ \frac{g}{2} h_i^2\ \mathbf{n}_i \end{pmatrix}. \tag{A.7}$$

From (4.4) to (4.5), we obtain finally – a part of the source terms (4.4) is balanced by the fluxes (A.6)

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \sum_{j\in K_i} \sigma_{ij} \begin{pmatrix} 0 \\ \frac{g}{2} h_i^2 \mathbf{n}_{ij} \end{pmatrix} - \sigma_i \begin{pmatrix} 0 \\ \frac{g}{2} h_i^2\ \mathbf{n}_i \end{pmatrix}. \tag{A.8}$$

Using the property

$$\sum_{j\in K_i} L_{ij}\mathbf{n}_{ij} + L_i\mathbf{n}_i = 0,$$

and the definition (3.4) of σ_{ij} and σ_i, this proves the preservation of the still water steady-state. □

Proof of Theorem 5.1. (“*Second-order*” *well-balanced scheme*) To prove the non-negativity property (i), we follow the idea developed in [7] for the 1D problem. First we assume that P_i is an interior node. Using (3.17) and the definition (3.4) of σ_{ij}, the scheme (5.14) defines h_i^{n+1} by

$$h_i^{n+1} = \frac{1}{|C_i|}\left[|C_i|h_i^n - \frac{1}{\Delta t}\sum_{j\in K_i} L_{ij}(F_h^+(\mathbf{U}_{ij}^{*,n}, \mathbf{n}_{ij}) + F_h^-(\mathbf{U}_{ji}^{*,n}, \mathbf{n}_{ij}))\right]. \tag{A.9}$$

Using (5.5), we have

$$h_i^{n+1} = \frac{1}{|C_i|}\sum_{j\in K_i}\left[|C_{ij}|h_{ij}^n - \frac{L_{ij}}{\Delta t}(F_h^+(\mathbf{U}_{ij}^{*,n}, \mathbf{n}_{ij}) + F_h^-(\mathbf{U}_{ji}^{*,n}, \mathbf{n}_{ij}))\right]. \tag{A.10}$$

To verify the positivity of h_i^{n+1}, it is sufficient to have

$$|C_{ij}|h_{ij}^n - \frac{L_{ij}}{\Delta t}F_h^+(\mathbf{U}_{ij}^{*,n}, \mathbf{n}_{ij}) \geqslant 0 \quad \text{for } j \in K_i. \tag{A.11}$$

Using the relation $h_{ij}^* \leqslant h_{ij}$ and adapting the proof of Theorem 3.1, we obtain that the inequality (A.11) is satisfied under the condition (5.15).

If P_i is a boundary node, we have

$$h_i^{n+1} = \frac{1}{|C_i|}\left[|C_i|h_i^n - \frac{1}{\Delta t}\sum_{j\in K_i} L_{ij}(F_h^+(\mathbf{U}_{ij}^{*,n}, \mathbf{n}_{ij}) + F_h^-(\mathbf{U}_{ji}^{*,n}, \mathbf{n}_{ij})) - \frac{L_i}{\Delta t}(F_h^+(\mathbf{U}_i^n, \mathbf{n}_i) + F_h^-(\mathbf{U}_{e,i}^n, \mathbf{n}_i))\right]. \tag{A.12}$$

Then using (5.5), we can write

$$|C_i|h_i = \alpha|C_i|h_i + (1-\alpha)\sum_{j\in K_i}|C_{ij}|h_{ij}, \quad 0 \leqslant \alpha \leqslant 1, \tag{A.13}$$

and with the same arguments as previously, we obtain the positivity of the water depth under the condition (5.15).

The proof of the preservation of the still water steady-state (ii) is very similar to the proof for the first-order case. The computed fluxes (A.6) and (A.7) are unchanged and since we have the centered source term (5.13) we obtain

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \sum_{j\in K_i}\sigma_{ij}\begin{pmatrix}0\\ \frac{g}{2}h_{ij}^2\mathbf{n}_{ij}\end{pmatrix} - \sum_{j\in K_i}\sigma_{ij}\begin{pmatrix}0\\ \frac{g}{2}(h_{ij}+h_i)(Z_{ij}-Z_i)\mathbf{n}_{ij}\end{pmatrix} - \sigma_i\begin{pmatrix}0\\ \frac{g}{2}h_i^2\mathbf{n}_i\end{pmatrix}. \tag{A.14}$$

Since the "second-order" reconstruction preserves the still water steady-state, we have

$$h_{ij} + Z_{ij} = h_i + Z_i.$$

Thus, (A.14) reduces to (A.8) and the conclusion is the same. □

References

[1] R. Abgrall, On essentially non-oscillatory schemes on unstructured meshes: analysis and implementation, J. Comput. Phys. 114 (1) (1994) 45–58.

[2] E. Audusse, A multilayer Saint-Venant model, Discrete Cont. Dyn. Syst. Ser. B 5 (2) (2005) 189–214.

[3] E. Audusse, F. Bouchut, M.O. Bristeau, R. Klein, B. Perthame, A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows, SIAM J. Sci. Comp. 25 (6) (2004) 2050–2065.

[4] E. Audusse, M.O. Bristeau, Transport of pollutant in shallow water, a two time steps kinetic method, M2AN 37 (2) (2003) 389–416.

[5] A. Bermudez, A. Dervieux, J.A. Desideri, M.E. Vazquez, Upwind schemes for the two-dimensional shallow water equations with variable depth using unstructured meshes, Comput. Meth. Appl. Mech. Eng. 155 (1–2) (1998) 49–72.

[6] A. Bermudez, M.E. Vazquez, Upwind methods for hyperbolic conservation laws with source terms, Comput. Fluids 23 (8) (1994) 1049–1071.
[7] F. Bouchut, Nonlinear stability of finite volume methods for hyperbolic conservation laws, and well-balanced schemes for sources, Frontiers in Mathematics, Birkhäuser, 2004.
[8] F. Bouchut, A. Mangeney-Castelnau, B. Perthame, J.P. Vilotte, A new model of Saint-Venant and Savage-Hutter type for gravity driven shallow water flows, C.R. Math. Acad. Sci. Paris 336 (6) (2003) 531–553.
[9] M.O. Bristeau, B. Coussin, Boundary conditions for the shallow water equations solved by kinetic schemes, INRIA Report, 4282, 2001. Available from: <http://www.inria.fr/RRRT/RR-4282.html>.
[10] M.O. Bristeau, B. Perthame, Transport of pollutant in shallow water using kinetic schemes, in: ESAIM Proceedings, vol. 10, CEMRACS, 1999, pp. 9–21. Available from: <http://www.emath.fr/Maths/Proc/Vol.10>.
[11] R. Botchorishvili, B. Perthame, A. Vasseur, Equilibrium schemes for scalar conservation laws with stiff sources, INRIA Report, 3891, 2000. Available from: <http://www.inria.fr/RRRT/RR-3891.html>.
[12] T. Chacon Rebollo, A.D. Delgado, E.D.F. Nieto, An entropy-correction free solver for non homogeneous shallow water equations, M2AN 37 (2003) 363–390.
[13] T. Chacon Rebollo, A.D. Delgado, E.D.F. Nieto, A family of stable numerical solvers for the shallow water equations with source terms, Comput. Meth. Appl. Math. Eng. 192 (2003) 203–225.
[14] A. Chertock, A. Kurganov, On a hybrid final-volume-particle method, M2AN 38 (6) (2004) 1071–1091.
[16] S. Ferrari, F. Saleri, A new two-dimensional Shallow Water model including pressure effects and slow varying bottom topography, M2AN 38 (2) (2004) 211–234.
[17] T. Gallouët, J.M. Hérard, N. Seguin, Some approximate Godunov schemes to compute shallow-water equations with topography, Comput. Fluids 32 (2003) 479–513.
[18] L. Gosse, A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms, Comput. Math. Appl. 39 (2000) 135–159.
[19] L. Gosse, A well-balanced scheme using nonconservative products designed for hyperbolic systems of conservation laws with source terms, Math. Mod. Meth. Appl. Sci. 11 (2) (2001) 339–365.
[20] J.M. Greenberg, A.-Y. Leroux, A well-balanced scheme for the numerical processing of source terms in hyperbolic equations, SIAM J. Numer. Anal. 33 (1996) 1–16.
[21] J.-F. Gerbeau, B. Perthame, Derivation of viscous Saint-Venant system for laminar shallow water; numerical validation, Discrete Cont. Dyn. Syst. Ser. B 1 (1) (2001) 89–102.
[22] E. Godlewski, P.-A. Raviart, Numerical approximations of hyperbolic systems of conservation lawsApplied Mathematical Sciences, vol. 118, Springer, New York, 1996.
[23] N. Goutal, F. Maurel, in: Proceedings of the 2nd Workshop on Dam-break Simulation, Note Technique EDF, HE-43/97/016/B, 1997.
[24] J.M.N.T. Gray, S. Noelle, Y.C. Tai, Shock waves, dead zones and particle-free regions in rapid granular free-surface flows, J. Fluid Mech. 491 (2003) 161–181.
[25] H. Guillard, R. Abgrall, Modélisation Numérique Des Fluides Compressibles, Series in Applied Mathematics, Gauthier-Villars, Paris, 2001 (in French).
[26] J.M. Hervouet, Hydrodynamique Des écoulements à Surface Libre; Modélisation Numérique Avec La Méthode Des éléments Finis, Presses des Ponts et Chaussées, 2003 (in French).
[27] J.M. Hervouet, A high resolution 2-D dam-break model using parallelization, Hydrol. Process. 14 (2000) 2211–2230.
[28] C. Hu, C.W. Shu, Weighted essentially non-oscillatory schemes on triangular meshes, J. Comput. Phys. 150 (1999) 97–127.
[29] S. Jin, A steady-state capturing method for hyperbolic systems with geometrical source terms, M2AN 35 (2001) 631–645.
[30] B. Khobalatte, B. Perthame, Maximum principle of the entropy and second-order kinetic schemes, J. Math. Comp 205 (1994) 119–131.
[31] A. Kurganov, D. Levy, Central-upwind schemes for the Saint-Venant system, M2AN 36 (2002) 397–425.
[32] R.J. LeVeque, Numerical Methods for Conservation LawsLectures in Mathematics, Birkhäuser, ETH Zurich, 1992.
[33] R.J. LeVeque, Balancing source terms and flux gradients in high-resolution Godunov methods: the quasi-steady wave-propagation algorithm, J. Comput. Phys. 146 (1) (1998) 346–365.
[34] P.L. Lions, B. Perthame, P.E. Souganidis, Existence of entropy solutions for the hyperbolic systems of isentropic gas dynamics in Eulerian and Lagrangian coordinates, Commun. Pure Appl. Math. 49 (6) (1996) 599–638.
[35] A. Mangeney-Castelnau, J.P. Vilotte, M.O. Bristeau, F. Bouchut, B. Perthame, C. Simeoni, S. Yernini, A new kinetic scheme for Saint-Venant equations applied to debris avalanches, INRIA Report 4646, 2002. Available from: <http://www.inria.fr/RRRT/RR-4646.html>.
[36] B. Perthame, Second-order Boltzmann schemes for compressible Euler equations in one or two space dimensions, SIAM J. Numer. Anal. 29 (1) (1992) 1–19.
[37] B. Perthame, Kinetic Formulations of Conservation Laws, Oxford University Press, Oxford, 2002.

[38] B. Perthame, Y. Qiu, A variant of Van Leer's method for multidimensional systems of conservation laws, J. Comput. Phys. 112 (2) (1994) 370–381.
[39] B. Perthame, C. Simeoni, A kinetic scheme for the Saint-Venant system with a source term, Calcolo 38 (4) (2001) 201–231.
[40] P.L. Roe, Approximate Riemann solvers, parameter vectors and difference schemes, J. Comput. Phys. 43 (1981) 357–372.
[41] G. Russo, Central schemes for balance laws, Hyperbolic Problems: Theory, Numerics, ApplicationsInternat. Ser. Numer. Math., vols. I and II (Magdebourg, 2000), Birkhäuser, Basel, 2001, p. 140, 141, 821–829.
[42] A.J.C. de Saint-Venant, Théorie du mouvement non-permanent des eaux, avec application aux crues des rivières et à l'introduction des marées dans leur lit, C.R. Acad. Sci. Paris 73 (1871) 147–154 (in French).
[44] B. Van Leer, Towards the ultimate conservative difference schemes. V. A second-order sequel to the Godunov's method, J. Comput. Phys. 32 (1979) 101–136.