Overcoming numerical shockwave anomalies using energy balanced numerical schemes. Application to the Shallow Water Equations with discontinuous topography.

A. Navas-Montilla , J. Murillo [1]

anavas@unizar.es, Javier.Murillo@unizar.es, Fluid Mechanics-LIFTEC, CSIC-Universidad de Zaragoza. Zaragoza, Spain

Abstract

When designing a numerical scheme for the resolution of conservation laws, the selection of a particular source term discretization (STD) may seem irrelevant whenever it ensures convergence with mesh refinement, but it has a decisive impact on the solution. In the framework of the Shallow Water Equations (SWE), well-balanced STD based on quiescent equilibrium are unable to converge to physically based solutions, which can be constructed considering energy arguments. Energy based discretizations can be designed assuming dissipation or conservation, but in any case, the STD procedure required should not be merely based on ad hoc approximations. The STD proposed in this work is derived from the Generalized Hugoniot Locus obtained from the Generalized Rankine Hugoniot conditions and the Integral Curve across the contact wave associated to the bed step. In any case, the STD must allow energy-dissipative solutions: steady and unsteady hydraulic jumps, for which some numerical anomalies have been documented in the literature. These anomalies are the incorrect positioning of steady jumps and the presence of a spurious spike of discharge inside the cell containing the jump. The former issue can be addressed by proposing a modification of the energy-conservative STD that ensures a correct dissipation rate across the hydraulic jump, whereas the latter is of greater complexity and cannot be fixed by simply choosing a suitable STD, as there are more variables involved. The problem concerning the spike of discharge is a well-known problem in the scientific community, also known as slowly-moving shock anomaly, it is produced by a non-linearity of the Hugoniot locus connecting the states at both sides of the jump. However, it seems that this issue is more a feature than a problem when considering steady solutions of the SWE containing hydraulic jumps. The presence of the spurious spike in the discharge has been taken for granted and has become a feature of the solution. Even though it does not disturb the rest of the solution in steady cases, when considering transient cases it produces a very undesirable shedding of spurious oscillations downstream that should be circumvented. Based on spike-reducing techniques (originally designed for homogeneous Euler equations) that propose the construction of interpolated fluxes in the untrustworthy regions, we design a novel Roe-type scheme for the SWE with discontinuous topography that reduces the presence of the aforementioned spurious spike. The resulting spike-reducing method in combination with the proposed STD ensures an accurate positioning of steady jumps, provides convergence with mesh refinement, which was not possible for previous methods that cannot avoid the spike.

Keywords: Roe solver, Energy balanced, Shallow water, Source terms, Hydraulic jump, Postshock oscillations

---

1. Introduction

There is a wide variety of physical problems modelled by non-homogeneous hyperbolic systems of conservation laws that are dominated by source terms. For such problems, the treatment of the source terms when designing a numerical scheme is of utmost importance in order to provide realistic and physically feasible solutions. Depending on the nature of the source term, different numerical techniques may be required. In this work, we focus on a certain type of source term, called geometric source term, present in many physical one–dimensional (1D) problems. This kind of source makes the conserved quantities account for the variation in space of a geometric variable, which is provided in the problem. Examples of mathematical models including geometric source terms are, for instance, the SWE with discontinuous topography, which is the object of study in the present work, the 1D Euler equations in a duct of variable cross section [1] and 1D flow in collapsible vessels [2].

Most popular methods for the resolution of homogeneous hyperbolic problems are within the framework of finite volume Godunov's numerical schemes [3], which aim to provide a numerical solution to the problem by means of a prior discretization of the domain into volume cells and integration of the information and governing equations inside these cells. After integration, simple algebraic evolution equations for the conserved variables, that depend upon the same variables at a previous time step and the fluxes at cell interfaces, arise. The keystone in Godunov's schemes is the computation of the numerical fluxes at cell interfaces, which is carried out by means of the resolution of the so-called Riemann Problems (RPs). RPs are initial value problems defined at cell interfaces, whose initial data is piecewise constant data given by the cell-averaged variables at each side of the discontinuity. They may be regarded as first order approach to the more general Cauchy problem [4].

When dealing with geometric source terms, it is necessary to account for the jump of the geometric quantity across cell interfaces when defining numerical fluxes at cell interfaces. To this end, augmented solvers were introduced [5, 6, 7]. When using augmented solvers, the source term is accounted for in the solution of the RP as an extra stationary wave at the interface. Due to the presence of the new wave, two solutions appear now at each side of the initial discontinuity instead of having a single homogeneous solution. Augmented versions of the traditional Roe [8] (ARoe) and HLLC [9, 10] solvers were presented by Murillo in [11] and [12] respectively. An extense review of the ARoe method can be found in [13].

If examining the system of equations in the so-called non-conservative form, the contribution of the source term is modelled as an additional stationary wave at the interface, which allows to include the effect of the source term in the eigenstructure of the system. This way, it can be noticed that the presence of a jump in the geometric variable gives rise to a contact wave and furthermore, that Riemann invariants are not necessarily conserved across such a wave, as pointed out by Rosatti et al. [14]. This issue will be recalled when designing the numerical scheme.

In the early stages of the design of numerical schemes for hyperbolic problems with source terms, the main effort was put on how to modify the original schemes, initially designed for homogeneous equations, so that they maintain the discrete equilibrium between fluxes and source term under steady state. When considering realistic applications, such goal was translated into the preservation of physical steady situations of quiescent equilibrium. For instance, in the framework of the SWE, the preservation of the steadiness of the solution for still water at rest. Numerical schemes satisfying this property were called well-balanced schemes [15, 16, 17, 18, 19].

When considering steady states with moving water over a irregular bed profile, the preservation of the C-property (exact conservation property) [16] is also of utmost importance in order to provide an exact equilibrium between fluxes and source terms. Numerical methods preserving the C-property are able to ensure a uniform discharge under steady conditions and can be constructed using flux-type definitions of the source terms [20, 6].

We can still take the well-balanced and C-property a step further by considering the conservation of the discrete specific mechanical energy in the scheme, enhancing in this way the performance of the numerical method. When friction is not considered in the SWE, mechanical energy is conserved under steady conditions in absence of hydraulic jumps. Such idea of energy conservation can be integrated in the numerical scheme, allowing the extension of well-balanced methods to exactly well-balanced methods [21, 22, 23, 24, 25, 26], hereafter referred to as E-schemes. Numerical methods defined as E-schemes will always satisfy the energy conservation property in the discrete level, hereafter referred to as E-property. Arbitrary order augmented Roe and HLL schemes preserving the E-property, called AR-ADER and HLLS-ADER E-schemes respectively, were presented by the authors of this work in [27, 28] and applied to the SWE. As a result of preserving the E-property, the aforementioned schemes were able to provide the exact solution in transient cases with independence of the grid and also to converge to the exact solution in transient problems at a high rate as the grid is refined.

For transient problems in the framework of the SWE, different approaches can be found in the literature regarding the treatment of the source term contact discontinuity. Two main tendencies are observed in the literature: one is based on energy and mass conservation and the other one based on mass and momentum conservation. For instance, some authors [29, 30] claim that energy must always be conserved since the bed step discontinuity is a contact wave and Riemann invariants, namely mass and energy for the bed step discontinuity, are conserved across contact waves. Alcrudo et al. [31] also state that the use of the mass-energy approach is necessary, specially when the slope of the bed profile becomes infinite (e.g. in the bed step), however, they allow for the possibility of some dissipation across the bed step, due to recirculation. On the other hand, Bernetti et al. [32] hold that the relation among variables across a bed discontinuity must be calculated by means of the Generalized Rankine-Hugoniot (GRH) conditions for the full system of equations. As an effort to unify all the previous approaches, Rosatti et al. [14] proposed a novel technique, based on the GRH conditions and using energy as a constraint to rule out solutions that are not physically admissible. They show that in nonconservative systems, such as the SWE, unlike in standard conservative systems, Riemann invariants are generally not constant across a contact discontinuity whose relevant eigenvalue is independent from the problem variables, and use this statement to design a numerical scheme that allows for the presence of dissipation due to recirculation at the bed step.

In the present work, the authors are faithful to the original SW system and do not include any dissipation mechanism (e.g. recirculation at bed step), as the original equations do not consider friction terms. Dissipation will only take place in certain conditions, such as a sudden change of flow regime (hydraulic jump), according to the physical behavior described by the equations. A theoretical study on the relations among states across the bed step contact wave is included in the text, leading to the particular conditions that ensure conservation of energy across the step: the Generalized Hugoniot Locus (GHL) derived from the GRH must coincide with the Integral Curve (IC). In other words, not only the GRH conditions must be fulfilled but also Riemann invariants should be conserved, as the specific mechanical energy is one of the relevant invariants for the characteristic field of the contact wave.

The AR-ADER and HLLS-ADER methods in [28], proposed by the authors of this work, are based on a particular energy conservative STD which is computed as a linear combination of a differential and integral approximation of the integral of the source term at cell interfaces. The method was presented in [25] for the first time and allowed to enhance the capabilities of augmented solvers in the framework of the SWE. Very high order methods are truly desirable as they have the ability of reducing dramatically numerical diffusion, allowing to provide predictions that would not be affordable by first order numerical schemes [33]. This can be done at the cost of replacing time derivatives by spatial derivatives. As a result, the strengths and also the weaknesses of the approximate solver used are enhanced.

E-schemes in [28] have desirable properties: they provide the exact solution for steady cases and are convergent to the exact solution with arbitrary order for transient cases including non-resonant and resonant cases. But there is still room for improvement. A recent study on the convergence of several schemes, including first order ARoe E-scheme, to steady shocks (hydraulic jumps) [34] proved that this scheme leads to a displacement of the hydraulic jump. When moving to very high order, integration of the source term must be done using a quadrature rule that matches with the order of convergence of the numerical scheme. This could be seen as an opportunity to improve numerical results regarding the positioning of the hydraulic jump, but contrary to intuition, the same issue observed in the first order scheme is repeated when using the high order methods in [28]. This issue is deeply studied and addressed here, proposing a STD that makes the scheme unequivocally identify the position of the hydraulic jump and dissipate the exact amount of energy across it. This technique will be referred to as selective energy balanced formulation (SEBF) of the integral of the source term and is applied to the ARoe and HLLS solvers, and their high order versions.

High order also preserves the effect of undesirable numerical shockwave anomalies. The utilization of high order numerical schemes in presence of spurious oscillations prevents numerical diffusion from dissipating those oscillations as fast as they would be dissipated if a first order scheme was used. It has been widely reported in the literature that significant numerical anomalies arise in presence of shock waves. An example of such problems are the Carbuncle [35, 36], the slowly-moving shock [37, 40] and the wall-heating phenomenon [41], all of them leading to spurious numerical solutions. Another major point addressed in this work is the study of such anomalies in the framework of SWE with and without bed variations and the extension of a spike-reducing scheme for non-homogeneous systems that avoids the presence of spurious oscillations due to numerical shocks. Shockwaves are typical solutions for nonlinear hyperbolic systems of conservation laws and their numerical treatment is of utmost importance to provide accurate solutions. As mentioned by Zaide and Roe [42], physical shockwaves have a finite width which is determined by the physical dissipation processes, however, when considering numerical shockwaves, a numerical width, usually much greater than the physical width, is enforced. This leads to the appearance of intermediate states which cannot be given a direct physical interpretation. Such states cannot be removed even when refining the grid, therefore we find in the literature that a special emphasis is put on this issue when designing a numerical scheme. Up to the present time, most studies have been carried out in the framework of Euler equations. In this work we will focus on the SWE.

Some of the problems related to numerical shockwave anomalies were first identified by Cameron and Emery [43, 44], who proposed some improvements based on the addition of artificial viscosity and modification of the grid. Here, we focus on the slowly-moving shock problem, which is associated to hydraulic jumps in the SWE. The slowly-moving shock problem was first investigated by Roberts in [37], who defined it as numerical noise generated in the discrete shock transition layer which is transported downstream. Such noise will be hereafter referred to as post-shock oscillations. In [37], the schemes of Godunov, Roe, and Osher were examined and the source of this error as also provided by using the Hugoniot locus. It was also observed that the slowly-moving shock problem only appears for systems of equations and not for scalar equations, where such schemes perform correctly. It is worth pointing out that even for non-linear systems, the slowly-moving shock problem does not appear if the Hugoniot curves are linear [38], as happens in the system in [39]. Later on, Arora and Roe [40] carried out a thorough study on this problem and evidenced that it can be ruinous when, for instance, making calculations of shock-sound interaction.

The spike-reducing techniques presented in this work are of first order of accuracy and one could think that by increasing the order of the scheme the slowly-moving shock problem could be circumvented. However, as mentioned by other authors [38, 45, 46], the slowly-moving shock problem will only be accentuated when increasing the accuracy of the scheme. Such an increase of accuracy will be translated into a longer preservation of post-shock oscillations as they provide a better resolution of the spurious physics. When using a high order scheme, the order is reduced to first order in the vicinity of the shock and the numerical solution within this region will behave according to what is expected from a first order scheme [47, 48]. Away from the shock, the order of accuracy is higher and therefore the spurious oscillations will be better resolved, preventing them from vanishing as one would desire. It must be borne in mind that even when using high order interpolations with limiting techniques, such as Total Variation Diminishing (TVD) interpolations and Essentially Non-oscillatory (ENO) schemes, the slowly-moving shock problem is accentuated [46].

The slowly-moving shock problem has been deeply studied for homogeneous systems of equations (e.g. the Euler equations) but scarcely studied for systems dominated by source terms. In [46], numerical results for the computation of a 1D compressible flow through a divergent nozzle by means of different first and high order schemes were presented, showing the inability of all schemes to converge to the exact solution in presence of shocks. The authors outline that this is due to the appearance of a spike in the momentum and the shedding of spurious oscillations downstream. This is the slowly-moving shock problem in the limit when shock speed is nil. The SWE are analogous to the 1D compressible flow with varying area, hence the slowly-moving shock problem is also likely to appear.

Here we focus on the slowly-moving shock problem in the SWE. To this end, we identify the conditions for the aforementioned problem to appear by studying the Hugoniot locus of the SWE and by seeking slowly-moving shock-type waves. We notice that they are only produced when dealing with a kind of transcritical shocks called hydraulic jumps, characterized by a change of sign of the relevant eigenvalue across them. A complete description of such kind of waves is provided and a thorough study on the shock structure, comparing exact and Godunov type solutions, is carried out in phase space. The slowly-moving shock problem in the SWE is a well-known problem in the scientific community, characterized by a spike in the discharge at the cell where the hydraulic jump is contained. In fact, it seems that this problem is more a feature than a problem when considering steady solutions of the SWE containing hydraulic jumps. The presence of the spurious spike in the discharge has been taken for granted as it does not perturb the rest of the solution. However, when considering transient cases, it produces a very undesirable shedding of spurious oscillations downstream that should be avoided.

When designing numerical schemes for the computation of slowly-moving shocks, the addition of extra artificial viscosity seems to be the most preferred technique in the scientific community [43, 44, 37, 40, 45, 49, 50]. If we want to avoid extra diffusion, another suitable possibility is the use of interpolation of fluxes, which avoids using the evaluation of the physical fluxes in the untrustworthy intermediate cells corresponding to the shock discontinuity. This idea of flux interpolation was first presented by Zaide and Roe [42], who proposed to find the fluxes in the intermediate cells by extrapolation from trustworthy neighbors. The authors claim that, by enforcing a linear shock structure and unambiguous sub-cell shock position, numerical shock-wave anomalies are dramatically reduced. It could be said that their method is also based on the addition of artificial viscosity, as their flux functions can be regarded as the traditional Roe flux plus a viscosity term. However, the flux interpolation functions use dissipation to control shock structure rather than to approach the true viscous solution and therefore they do not expand the shock profile [38].

In this work, we use the approach in [42] to propose a novel spike-reducing flux function for the SWE with varying bed. Prior to the presentation of the proposed technique, the flux functions in [42] are applied to the SWE with flat bed, showing their spike-reducing nature. The proposed technique is assessed in a variety of situations, including steady and transient cases, with continuous and discontinuous bed profiles, proving the expected spike-reducing behavior. The analogous SWE problem of the 1D nozzle problem in [46], which is the steady flow over a hump, is reproduced in this work, showing that the proposed scheme leads to a convergent solution, even when measured with L_∞ error norm.

The outline of the paper is next presented. In section 2, an introduction to nonlinear systems of conservation laws with source terms is provided and the definition of geometric source terms and derivation of the GRH conditions for such systems are recalled. In this section, the description of non-conservative systems and the treatment of contact waves in this kind of systems is also recalled following [14]. In Section 3, we briefly describe Godunov type finite volume schemes. Section 4 is devoted to the description of the SWE, both in conservative and non-conservative form, including a thorough study on the bed step contact wave. In this section, the numerical treatment of the source term in the SWE is also described and the novel SEBF discretization method is presented. At the end of this section, numerical results for the computation of steady flows are displayed. Section 5 is entirely devoted to the study of numerical shockwave anomalies in the SWE. A thorough description of the slowly-moving shock problem arising from the hydraulic jump, using the phase-space representation, is presented. In Section 6, numerical fixes addressing the aforementioned problem are studied. First, numerical results for the computation of several homogeneous test cases using the flux functions A and B in [42] are shown. Then, the novel spike-reducing technique for the SWE with source term is presented and a set of tests are carried out to evidence the capabilities of the proposed method. Finally, in Section 7 we present a summary of the work and the concluding remarks.

2. Nonlinear systems of equations with source term

The basic ideas underlying this work can be illustrated by examining hyperbolic nonlinear systems of equations with source terms in 1D, that can be expressed in integral form as

$$\frac{\partial}{\partial t}\int_{x_1}^{x_2}\mathbf{U}dx + \mathbf{F}|_{x_2} - \mathbf{F}|_{x1} - \int_{x_1}^{x_2}\mathbf{S}dx = 0\,, \tag{1}$$

where x_1, x_2 are the limits of a generic control volume and with N_λ equations. Such systems arise naturally from the conservation laws for certain physical quantities in nature. The differential formulation is obtained when assuming a smooth variation of the variables and an infinitesimal width of the control volume, yielding

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}\,, \tag{2}$$

where $\mathbf{U} = \mathbf{U}(x,t) \in \mathcal{C} \subset \mathbb{R}^{N_\lambda}$ is the vector of conserved quantities that takes values on $\mathcal{C}$, the set of admissible states of $\mathbf{U}$, $\mathbf{F} = \mathbf{F}(\mathbf{U})$ is the flux function that represents a nonlinear mapping of the conserved quantities from $\mathcal{C}$ to $\mathbb{R}^{N_\lambda}$ and $\mathbf{S}$ is the source term, that will be considered a function of the conserved quantities and spatial coordinate as $\mathbf{S} = \mathbf{S}(\mathbf{U}, x)$. In this work, we put a special emphasis on the so-called *geometric source terms*, that are expressed as

$$\mathbf{S}(\mathbf{U}, x) = \mathbf{S}_s(\mathbf{U})\frac{d}{dx}\mathbf{S}_g(x)\,, \tag{3}$$

with $\mathbf{S}_s(\mathbf{U})$ a function of the conserved quantities and $\mathbf{S}_g(x)$ the geometric function that depends upon the position x and can be discontinuous [28].

From (2), the Jacobian matrix of the convective part is defined as

$$\mathbf{J} = \frac{d\mathbf{F}(\mathbf{U})}{d\mathbf{U}}\,. \tag{4}$$

Assuming that the convective part in (2) is strictly hyperbolic, with N_λ real eigenvalues $\lambda^1, ..., \lambda^{N_\lambda}$ and eigenvectors $\mathbf{e}^1, ..., \mathbf{e}^{N_\lambda}$, it is possible to define the matrices $\mathbf{P} = (\mathbf{e}^1, ..., \mathbf{e}^{N_\lambda})$ and $\mathbf{P}^{-1}$ with the property that they diagonalize the Jacobian as

$$\mathbf{J} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}\,. \tag{5}$$

2.1. Conservative vs non-conservative form

For the sake of simplicity, dependency of variables upon the conserved quantities is hereafter omitted. A generic homogeneous conservative system is written as

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = 0\,, \tag{6}$$

where $\mathbf{U}$ is the vector of conserved quantities and $\mathbf{F}$ the vector of conservative fluxes. It can be expressed in its quasilinear form as

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{J}\frac{\partial \mathbf{U}}{\partial x} = 0\,, \tag{7}$$

where the Jacobian matrix $\mathbf{J} = d\mathbf{F}/d\mathbf{U}$ can be diagonalized with N_λ eigenvalues by means of N_λ linearly independent eigenvectors. The following relation is worth being shown

$$\mathbf{J} \cdot \mathbf{e}^m - \lambda^m \mathbf{e}^m = 0\,, \tag{8}$$

where λ^m and $\mathbf{e}^m$ are the eigenvalues and right eigenvectors of matrix $\mathbf{J}$.

Non-homogeneous hyperbolic conservation laws (2) cannot be expressed in the strict conservative form of (6) due to the presence of the source term. When having geometric source terms of the type of (3), they can be expressed in non-conservative form as

$$\frac{\partial \hat{\mathbf{U}}}{\partial t} + \frac{\partial \hat{\mathbf{F}}(\hat{\mathbf{U}})}{\partial x} + \mathbf{H}\frac{\partial \hat{\mathbf{U}}}{\partial x} = 0\,, \tag{9}$$

where $\hat{\mathbf{U}} \in \mathcal{C} \subset \mathbb{R}^{N_\lambda + N_S}$ is the new vector of variables composed of the N_λ conserved variables in (2) plus additional N_S variables related to the source term, $\hat{\mathbf{F}}(\hat{\mathbf{U}}) : \mathcal{C} \longrightarrow \mathbb{R}^{N_\lambda + N_S}$ is the vector of conservative fluxes and $\mathbf{H}$ the matrix of non-conservative fluxes.

In this work, we will focus on physical problems (e.g. the shallow water model with bed topography) with a geometric source term like (3) that only involves a single geometric quantity, $s_g(x)$, as follows

$$\mathbf{S}_g(x) = (0,\ \dots\ , s_g(x),\ \dots\ , 0)^T\,. \tag{10}$$

In this case, the new vector of variables will be constructed as $\hat{\mathbf{U}} = (\mathbf{U}\,, s_g)^T$, hence $N_S = 1$, with $\lambda^s = 0$, the speed of the wave associated to the source equal to zero as the geometric quantity does not evolve in time. This is depicted in Figure 1, for an arbitrary system with $N_\lambda = 3$ and a single geometric variable, that is $N_S = 1$.

Also notice that the evolution equation corresponding to the geometric quantity, s_g, reads

$$\frac{\partial s_g}{\partial t} = 0\,, \tag{11}$$

which stands for the conservation of this quantity in time, as it only depends upon the spatial position x.

The non-conservative system in (9) can be more compactly expressed as

$$\frac{\partial \hat{\mathbf{U}}}{\partial t} + \mathbf{A}\frac{\partial \hat{\mathbf{U}}}{\partial x} = 0\,, \tag{12}$$

where $\mathbf{A} = \mathbf{J} + \mathbf{H}$ and with $\mathbf{J} = d\hat{\mathbf{F}}/d\hat{\mathbf{U}}$. Relation in (8) is now written as

$$\mathbf{J} \cdot \hat{\mathbf{e}}^m - \hat{\lambda}^m \hat{\mathbf{e}}^m = -\mathbf{H} \cdot \hat{\mathbf{e}}^m\,, \tag{13}$$

where $\hat{\lambda}^m$ and $\hat{\mathbf{e}}^m$ are the eigenvalues and right eigenvectors of matrix $\mathbf{A}$.

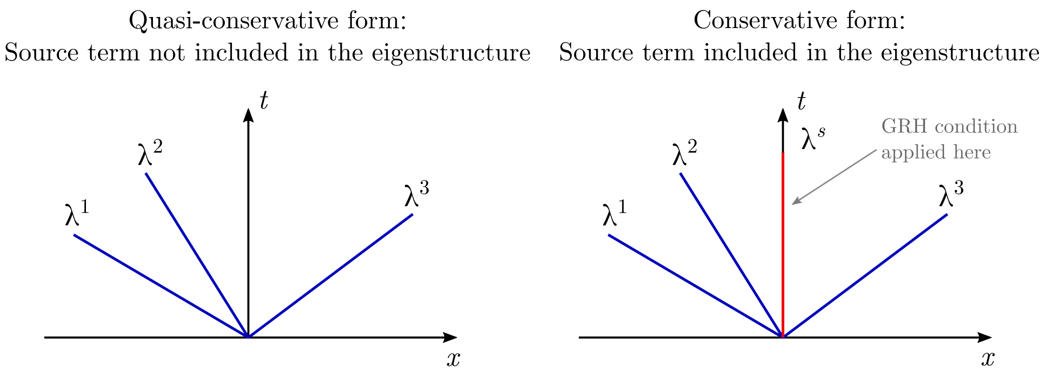

Figure 1: Difference in eigenstructure between the quasi-conservative system (2) and the non-conservative system (9).

For the sake of clarity, it is worth recalling that the system in (6) will be hereafter referred to as conservative system, the system in (2) as quasi-conservative system and the system in (12) as non-conservative system. This work focuses on the study of hyperbolic equations with source term, therefore (6) will be useless in what follows.

2.2. Integral relations in discontinuous solutions

It is of utmost importance to mention that there exists a certain relation between the wave speed and the jump of conserved quantities and fluxes across the discontinuities carried by the waves. This relation is called *Rankine-Hugoniot (RH) condition* or *jump condition*. When dealing with non-homogeneous systems of equations, such condition must be extended to account for the contribution of the source term, leading to the *Generalized Rankine-Hugoniot (GRH) condition*.

Initial system in (2) is composed of N_λ waves, nevertheless, none of these waves are related to the source term and only conventional RH conditions

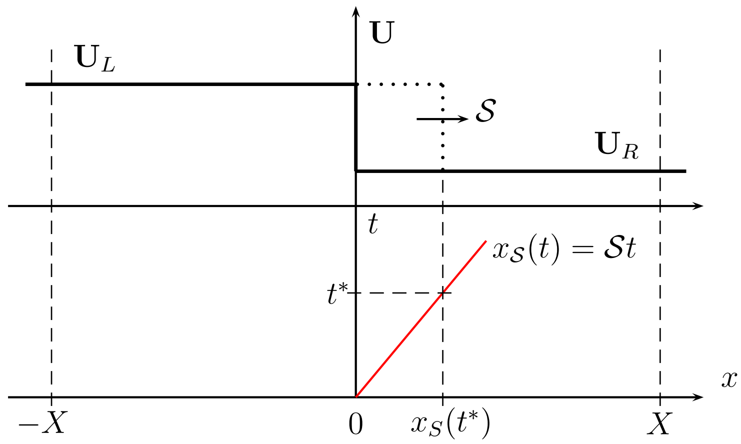

Figure 2: Discontinuity propagation in a non-linear system. The integration domain for the derivation of the Rankine-Hugoniot condition is depicted.

could be defined across them. In order to study the more general case, where GRH can be defined, it is necessary to express the system in (2) in its non-conservative form according to Equation (9). In this way, the system is not only characterized by the N_λ eigenvalues associated to the conservative fluxes but also by other N_S eigenvalues, related to extra variables modelling the source term, as the dynamics of the source term is included, in some way, in the set of characteristic fields. For the sake of simplicity, N_S is hereafter set to 1.

The derivation of the GRH condition for the system in (2) with a geometric source term, or (9) equivalently, can be derived in two different ways. The first one would be using equation (2) and considering the source term as a Dirac delta that moves with the wave [51]. The second option, the one we use here, is to derive the GRH condition from the non-conservative system of equations in (9). It is done by integrating (9) over an arbitrary domain $[-X, X]$ with X sufficiently large, as depicted in Figure 2. Notice that the displacement of the discontinuity represented in Figure 2 is done from $t = t_0$ to $t = t^* = t_0 + \delta t$, with δt of differential size. For each λ^m wave defining a characteristic field, the left and right states of the solution at each side of the discontinuity carried by wave λ^m are denoted by $\mathbf{U}_L$ and $\mathbf{U}_R$, and the speed of the discontinuity is denoted by $\mathcal{S}^m$. The integral of (9) over $[-X, X]$ reads

$$\int_{-X}^{X} \frac{\partial \hat{\mathbf{U}}}{\partial t} dx + \int_{-X}^{X} \frac{\partial \hat{\mathbf{F}}}{\partial x} dx + \int_{-X}^{X} \mathbf{H} \frac{\partial \hat{\mathbf{U}}}{\partial x} dx = 0\,. \tag{14}$$

Considering that the integration domain does not change in time, Equation (14) is rewritten as

$$\frac{d}{dt} \int_{-X}^{X} \hat{\mathbf{U}} dx + \left[\hat{\mathbf{F}} \right]_{-X}^{X} + \int_{-X}^{X} \mathbf{H} \frac{\partial \hat{\mathbf{U}}}{\partial x} dx = 0\,. \tag{15}$$

If separating the first term on the left hand side of Equation (15) as

$$\frac{d}{dt} \left(\int_{-X}^{x_S(t)} \hat{\mathbf{U}} dx + \int_{x_S(t)}^{X} \hat{\mathbf{U}} dx \right) = \frac{d}{dt} \left(\hat{\mathbf{U}}_L (X + \mathcal{S}^m t) + \hat{\mathbf{U}}_R (X - \mathcal{S}^m t) \right) \tag{16}$$

and taking the time derivative of the previous result, Equation (16) is rewritten as

$$\frac{d}{dt} \int_{-X}^{X} \hat{\mathbf{U}} dx = \mathcal{S}^m \left(\hat{\mathbf{U}}_L - \hat{\mathbf{U}}_R \right)\,. \tag{17}$$

When combining the results obtained in (15) and (17), the following condition for the jump is obtained

$$\hat{\mathbf{F}}_R - \hat{\mathbf{F}}_L - \hat{\mathbf{D}} = \mathcal{S}^m \left(\hat{\mathbf{U}}_R - \hat{\mathbf{U}}_L \right)\,, \tag{18}$$

where

$$\hat{\mathbf{D}} = - \int_{-X}^{X} \mathbf{H} \frac{\partial \hat{\mathbf{U}}}{\partial x} dx \tag{19}$$

is a suitable approximation of the integral of the source term. Notice that the case $\mathbf{D} = 0$ corresponds to the traditional RH condition.

When using this formulation, it must be borne in mind that the geometric variable is known and is considered to only change at fixed positions, that is to say, discontinuities on the geometric variable remain at a fixed location. This helps to understand the conditions for the application of the GRH condition.

Let us consider a discontinuity traveling at speed $\mathcal{S}^m \neq 0$. Application of the GRH condition in (18) for the geometric variable yields

$$\mathcal{S}^m \left([s_g]_R - [s_g]_L \right) = 0\,, \tag{20}$$

according to (11). It is observed that $[s_g]_R = [s_g]_L$ for any $\mathcal{S}^m \neq 0$, which agrees with the aforementioned consideration saying that variations on the geometric variable only take place at fixed positions. This implies that

$$\hat{\mathbf{D}} = 0\,, \tag{21}$$

recovering the traditional RH condition

$$\mathbf{F}_R - \mathbf{F}_L = \mathcal{S}^m \left(\mathbf{U}_R - \mathbf{U}_L \right) \tag{22}$$

for all $\mathcal{S}^m \neq 0$. Notice that the vectors of fluxes and variables in (22) do not include the source term as its contribution is nil at this point.

On the other hand, if $\mathcal{S}^m = 0$, application of the GRH condition in (18) for the geometric variable yields

$$0 \cdot \left([s_g]_R - [s_g]_L \right) = 0\,, \tag{23}$$

which holds for any combination of $[s_g]_R$ and $[s_g]_L$. Therefore, for $\mathcal{S}^m = 0$, the GRH condition always applies and is written as

$$\hat{\mathbf{F}}_R - \hat{\mathbf{F}}_L = \hat{\mathbf{D}}\,. \tag{24}$$

Here, the last component of the equation, corresponding to the source variable, is useless again, therefore we can rewrite (24) as

$$\mathbf{F}_R - \mathbf{F}_L = \mathbf{D}\,, \tag{25}$$

with $\hat{\mathbf{D}} = (\mathbf{D}, 0)^T$ and due to the nature of the source in (3), the integral of this source can be expressed as

$$\mathbf{D} = \int_{[\mathbf{S}_g]_L}^{[\mathbf{S}_g]_R} \mathbf{S}_s d\mathbf{S}_g\,, \tag{26}$$

with $\delta \left[\mathbf{S}_g\right]_L^R$ the jump in the geometric variable across the wave.

It is worth recalling that the set of right (left) states that can be connected to a given left (right) state by means of a discontinuous solution describe a curve in the phase space called Hugoniot Locus (HL), or Generalized Hugoniot Locus (GHL).

2.3. Integral curves and Riemann invariants

Let us consider a hyperbolic system expressed in non-conservative form as (9)

$$\frac{\partial \hat{\mathbf{U}}}{\partial t} + \mathbf{A}\frac{\partial \hat{\mathbf{U}}}{\partial x} = 0\,, \tag{27}$$

where matrix $\mathbf{A}$ can be diagonalized with $N_\lambda + N_S$ eigenvalues by means of $N_\lambda + N_S$ linearly independent eigenvectors. For the sake of clarity, hat symbol in vectors standing for the extended vectors that include the equation of the source term is hereafter ommited. Each eigenvalue $\lambda^m(\mathbf{U})$, or eigenvector $\mathbf{e}^m(\mathbf{U})$ equivalently, defines a *characteristic field* associated to it, for $m = 1, ..., N_\lambda + N_S$. The properties of the characteristic fields will provide useful information about the solution. Prior to the analysis of the characteristic fields, it is worth introducing the concepts of Integral Curves and state space. The state space, or phase plane, is the representation of a component of the state vector with respect to the other components. For instance, if considering a system of $N_\lambda + N_S = 2$ equations, with $\mathbf{U} = (u_1, u_2)$, the state space representation will be given by the representation of u_1-u_2 in a Cartesian coordinate system.

Definition 1. *(Integral Curve). Let $\mathbf{U}(\xi)$ be a smooth curve through state space parametrized by the scalar ξ. This curve is said to be an Integral Curve (IC) of the vector field $\mathbf{e}^m$ if at each point, the tangent vector to the curve, $d\mathbf{U}(\xi)/d\xi$ is an eigenvector of $\mathbf{J}(\mathbf{U}(\xi))$ corresponding to the eigenvalue $\lambda^m(\mathbf{U}(\xi))$. When considering a particular set of eigenvectors, the integral curve for $\mathbf{e}^m$ field is given by*

$$\frac{d\mathbf{U}(\xi)}{d\xi} = \nu(\xi) \cdot \mathbf{e}^m(\mathbf{U}(\xi))\,, \tag{28}$$

with $\nu(\xi)$ a constant parameter that depends on the normalization of the eigenvectors [51].

When analyzing the solution of hyperbolic systems of conservation laws, it is observed that the wave pattern present in the solution is related to the variation of the characteristic speed, $\lambda^m(\mathbf{U})$, along the integral curve of the vector field $\mathbf{e}^m$. This variation can be expressed as the directional derivative of $\lambda^m(\mathbf{U})$ in the direction of the eigenvector [51]

$$\frac{d}{d\xi}\lambda^m(\mathbf{U}(\xi)) = \nabla_u \lambda^m(\mathbf{U}(\xi)) \cdot \mathbf{e}^m(\mathbf{U}(\xi)) \,. \tag{29}$$

When $\lambda^m(\mathbf{U})$ is constant along the integral curve, that is (29) is equal to zero, the characteristic field is said to be *linearly degenerate*. On the other hand, if $\lambda^m(\mathbf{U})$ varies along the integral curve, which means that the characteristic curves are compressing or expanding, the characteristic field is said to be *genuinely nonlinear*.

Along each integral curve, there are certain quantities that remain constant. Such quantities are called Riemann invariants.

Definition 2. *(Riemann invariant). The scalar w^m is said to be a m-Riemann invariant when*

$$\nabla_u w^m(\mathbf{U}) \cdot \mathbf{e}^m(\mathbf{U}) \neq 0 \,, \quad \forall \mathbf{U} \in \mathcal{C} \,, \tag{30}$$

with $\mathcal{C} \subseteq \mathbb{R}^{N_\lambda}$ and where ∇_u stands for the gradient with respect to the components of vector $\mathbf{U}$.

2.4. The solution of non-linear hyperbolic systems

Non-linear hyperbolic systems of the type of (2) admit complex solutions including shocks, rarefaction waves or contact waves. For the sake of brevity, the latter are only described here, as they have important implications in the design of numerical schemes in presence of geometric source terms. A more detailed study on shocks and rarefactions can be found in [52]. Contact waves in conservative and non-conservative systems are described below:

- **Contact wave in conservative (homogeneous) systems:** If λ^m defines a *linearly degenerate field* and the following conditions apply:
  - RH condition:

$$\mathbf{F}(\mathbf{U}_L) - \mathbf{F}(\mathbf{U}_R) = \mathcal{S}^m \left(\mathbf{U}_L - \mathbf{U}_R\right) \tag{31}$$

  - Parallel characteristic condition:

$$\lambda^m(\mathbf{U}_L) = \mathcal{S}^m = \lambda^m(\mathbf{U}_R) \tag{32}$$

  - Conservation of the Riemann Invariants across the discontinuity.

then left and right states $\mathbf{U}_L$ and $\mathbf{U}_R$ will be connected by a single jump discontinuity wave of speed $\mathcal{S}^m$ called contact wave.

- **Contact wave in non-conservative systems** (with geometric source term) where the relevant eigenvalue does not depend upon $\mathbf{U}$ [14]:

  The presence of contact discontinuities in RPs given by non-homogeneous systems of conservation laws has to be taken into account when constructing augmented solvers. In this work, we consider contact waves whose relevant eigenvalue does not depend upon $\mathbf{U}$. This would be the case of a system like (9) where $\mathbf{H}\mathbf{U}_x$ includes the contribution of the geometric source term (3). For such case, given a initial left state, $\mathbf{U}_L$, the right state, hereafter denoted by $\mathbf{U}(\xi)$, does not necessarily lie on the integral curve, while it will always be related to the left state by means of the GRH condition [4, 14], as all discontinuous solutions do satisfy this relation. Recall that $\mathbf{U}_L = \mathbf{U}(\xi = 0)$.

  Let us consider the non-conservative system in (9) and assume that the m-th characteristic field, associated to eigenvalue λ^m and eigenvector $\mathbf{e}^m$, is linearly degenerate. Then, the associated contact wave is given by

$$\mathbf{U}(x,t) = \begin{cases} \mathbf{U}_L & x < \mathcal{S}^m t \\ \mathbf{U}(\xi) & x > \mathcal{S}^m t \end{cases} \tag{33}$$

  with constant speed $\mathcal{S}^m = \lambda^m(\mathbf{U}(\xi)) = \lambda^m(\mathbf{U}_L)$. All possible $\mathbf{U}(\xi)$ states can be found by means of the GHL. From (18) we have

$$\mathbf{F}(\mathbf{U}(\xi)) - \mathbf{F}(\mathbf{U}_L) - \mathcal{S}^m(\mathbf{U}(\xi) - \mathbf{U}_L) = \mathbf{D}\,. \tag{34}$$

  In this way, $\mathbf{U}(\xi)$ will satisfy the GRH condition, however, we have not imposed yet any condition for the conservation of the relevant m-Riemann invariants across the contact discontinuity, hence IC and GHL may not coincide. To find the condition so that such sets of states coincide, following [14], let us consider the differential form of (34)

$$\frac{d}{d\xi}\left[\mathbf{F}(\mathbf{U}(\xi)) - \mathcal{S}^m\mathbf{U}(\xi)\right] = \frac{d}{d\xi}\mathbf{D} \tag{35}$$

  that can be rewritten as

$$\frac{d\mathbf{F}}{d\mathbf{U}}\frac{d\mathbf{U}(\xi)}{d\xi} - \mathcal{S}^m \frac{d\mathbf{U}(\xi)}{d\xi} = \frac{d}{d\xi}\mathbf{D}\,. \tag{36}$$

To enforce the solution to lie on both the IC and the GHL, we set $\mathbf{U} = \mathbf{U}^m(\xi)$ to be the set of states lying on the IC according to (28), yielding

$$\mathbf{J}\frac{d\mathbf{U}^m(\xi)}{d\xi} - \mathcal{S}^m \frac{d\mathbf{U}^m(\xi)}{d\xi} = \frac{d}{d\xi}\mathbf{D}\,, \tag{37}$$

where $d\mathbf{U}^m(\xi)/d\xi$ can be substituted by $\mathbf{e}^m$ as the solution follows the IC, and $\mathcal{S}^m$ by λ^m, leading to

$$\mathbf{J}\cdot\mathbf{e}^m - \lambda^m\cdot\mathbf{e}^m = \frac{d}{d\xi}\mathbf{D}\,, \tag{38}$$

that can be rewritten by means of (13) as

$$-\mathbf{H}\cdot\mathbf{e}^m = \frac{d}{d\xi}\mathbf{D}\,. \tag{39}$$

Only when relation in (39) is satisfied, the IC and GHL coincide and the Riemann invariants are conserved across the contact wave. This property will be used later to design an E-scheme for the SWE.

3. Finite volume discretization

In the present framework, problems of interest are defined as initial value boundary problems (IVBP) that can be expressed as

$$\begin{cases} \text{PDEs:} & \dfrac{\partial\mathbf{U}}{\partial t} + \dfrac{\partial\mathbf{F}(\mathbf{U})}{\partial x} = \mathbf{S} \\[2ex] \text{IC:} & \mathbf{U}(x,0) = \mathring{\mathbf{U}}(x) \\[2ex] \text{BC:} & \mathbf{U}(a,t) = \mathbf{U}_a(t) \quad \mathbf{U}(b,t) = \mathbf{U}_b(t) \end{cases} \tag{40}$$

defined inside the domain $[a,b]\times[0,T]$, with $\mathring{\mathbf{U}}(x)$ the initial condition and $\mathbf{U}_a(t)$ and $\mathbf{U}_b(t)$ the left and right boundary conditions. When using a first order finite volume approach, the domain is discretized in computational cells and the conserved variables and governing equations are integrated inside those cells, leading to algebraic equations that depend upon piecewise constant data. In this work, the following computational grid composed of N cells is used

$$a = x_{\frac{1}{2}} < x_{\frac{3}{2}} < ... < x_{N-\frac{1}{2}} < x_{N+\frac{1}{2}} = b\,, \tag{41}$$

as shown in Figure 3, with cells and cell sizes defined as

$$\Omega_i = \left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}\right]\,, \qquad \Delta x_i = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}\,, \qquad i = 1, ..., N \tag{42}$$

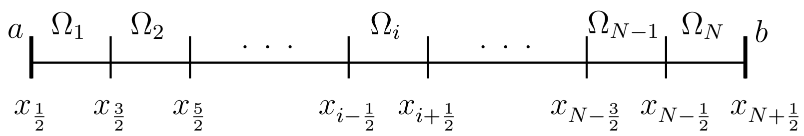

Figure 3: Mesh discretization

Inside each cell, conserved quantities at time t^n are defined as cell averages as

$$\mathbf{U}_i^n = \frac{1}{\Delta x_i}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{U}(x, t^n)dx\,, \qquad i = 1, ..., N\,. \tag{43}$$

Following the approach proposed by Godunov, the finite volume discretization of the system in (2) inside $[x_{i-1/2}, x_{i+1/2}] \times [t^n, t^{n+1}]$ is straightforward derived from integration of (2) in this volume, leading to

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x}[\mathbf{F}_{i+1/2}^- - \mathbf{F}_{i-1/2}^+]\,, \tag{44}$$

where $\mathbf{F}_{i+1/2}^-$ and $\mathbf{F}_{i-1/2}^+$ are the numerical fluxes, which are computed solving the Riemann Problems (RPs) at the interfaces by means of a suitable Riemann solver.

Analogously, equation (44) can be rewritten in terms of fluctuations, generally denoted by $\delta\mathbf{M}$, leading to

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x}[\delta\mathbf{M}_{i+1/2}^- + \delta\mathbf{M}_{i-1/2}^+]\,, \tag{45}$$

where

$$\delta \mathbf{M}^-_{i+1/2} = \mathbf{F}^-_{i+1/2} - \mathbf{F}_i\,,$$
$$\delta \mathbf{M}^+_{i-1/2} = \mathbf{F}_i - \mathbf{F}^+_{i-1/2}\,, \tag{46}$$

represent the contribution of the incoming waves to the right and left edges, respectively. The Riemann solver selected here is called the augmented Roe Riemann solver (ARoe) and is detailed in Appendix A.

4. Application to the Shallow Water Equations (SWE)

The SWE can be expressed in matrix form as

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = \mathbf{S}\,. \tag{47}$$

where

$$\mathbf{U} = \begin{pmatrix} h \\ hu \end{pmatrix}, \; \mathbf{F} = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \end{pmatrix}, \; \mathbf{S} = \begin{pmatrix} 0 \\ S_z \end{pmatrix}, \tag{48}$$

where h is the water depth, u is the depth averaged velocity, hu the discharge and g is the acceleration of gravity. The source term S_z involves the variations in bed geometry S_z

$$S_z = -gh\frac{dz}{dx}\,, \tag{49}$$

where z stands for the bed elevation.

In order to design a suitable numerical scheme that mimics the physical behavior of (47), these equations must be thoroughly analyzed. In physics, invariance of certain quantities is usually present in systems. In the SWE, the mechanical energy is an example. From the analysis of (47) under steady regime and considering a smooth solution, we obtain that

$$\frac{\partial}{\partial x}\left(\frac{u^2}{2g} + h + z\right) = 0\,, \tag{50}$$

where $E = \frac{u^2}{2g} + h + z$ is the specific mechanical energy. By looking at this quantity when designing the numerical scheme, the well-balanced property can be extended to the so-called energy-balanced property, which allows the numerical scheme to provide the exact solution in steady cases with moving water.

It is worth pointing out that, unlike in previous publications [14], the authors in this work are faithful to the original system in (47) and do not include any dissipation mechanism (for instance, across shocks), as the original equations do not consider extra friction terms. When neglecting shear stress, dissipation will only take place in certain conditions, such as a sudden change of flow regime, according to the physical behavior described by the original equations.

For system in (47), the discretization of the source term is not a trivial task and additional information must be taken into account in order to construct a trustworthy numerical solution and eventually obtain an energy-balanced scheme. The analysis of the system of equations in non-conservative form is useful to this end as it provides information on the physical nature of the additional wave associated to the source term.

4.1. Characteristic analysis of the SWE system in its non-conservative form

According to Equation (9), system in (47) can be expressed in non-conservative form

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} + \mathbf{H}(\mathbf{U})\frac{\partial \mathbf{U}}{\partial x} = 0\,, \tag{51}$$

where

$$\mathbf{U} = \begin{pmatrix} h \\ hu \\ z \end{pmatrix}, \ \mathbf{F} = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \\ 0 \end{pmatrix}, \ \mathbf{H} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & gh \\ 0 & 0 & 0 \end{pmatrix}. \tag{52}$$

The Jacobian matrix of the flux reads

$$\mathbf{J} = \begin{pmatrix} 0 & 1 & 0 \\ c^2 - u^2 & 2u & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{53}$$

and it can be used to construct the following matrix

$$\mathbf{A} = \mathbf{J} + \mathbf{H} = \begin{pmatrix} 0 & 1 & 0 \\ c^2 - u^2 & 2u & gh \\ 0 & 0 & 0 \end{pmatrix}, \tag{54}$$

allowing to express the system in quasilinear form. The eigenvalues and eigenvectors that diagonalize $\mathbf{A}$ are given by

$$\lambda^1 = u - c, \qquad \lambda^S = 0\,, \qquad \lambda^2 = u + c \tag{55}$$

and

$$\mathbf{e}^1 = \begin{pmatrix} 1 \\ \lambda^1 \\ 0 \end{pmatrix}, \qquad \mathbf{e}^S = \begin{pmatrix} 1 \\ 0 \\ u^2/gh - 1 \end{pmatrix}, \qquad \mathbf{e}^2 = \begin{pmatrix} 1 \\ \lambda^2 \\ 0 \end{pmatrix}. \tag{56}$$

For the sake of clarity and consistency throughout the text, the characteristic field corresponding to the source variable, z, is denoted by S while the two other fields are denoted by 1 (for the left moving wave) and 2 (for the right moving wave). The nature of each characteristic field can be studied as pointed out in Section 2.3. Following definition in (29), for this particular case we have

$$\begin{array}{l} \nabla_u \lambda^1(\mathbf{U}) \cdot \mathbf{e}^1(\mathbf{U}) = -\dfrac{\sqrt{g}}{2\sqrt{h}}\,, \\ \nabla_u \lambda^S(\mathbf{U}) \cdot \mathbf{e}^S(\mathbf{U}) = 0\,, \\ \nabla_u \lambda^2(\mathbf{U}) \cdot \mathbf{e}^2(\mathbf{U}) = \dfrac{\sqrt{g}}{2\sqrt{h}}\,, \end{array} \tag{57}$$

noticing that the S-characteristic field associated to the bed step is linearly degenerate as the eigenvalue λ^S is zero $\forall \mathbf{U}$ (the step is regarded as a stationary discontinuity) while the 1 and 2-characteristic fields are genuinely nonlinear.

The integral curve for each of the characteristic fields can be derived from equation (28). The integral curve associated to the 1-characteristic field, parametrized by ξ and starting at $(h, hu, z) = (h^*, (hu)^*, z^*)$, reads

$$\mathbf{U}^1(\xi) = \begin{pmatrix} h(\xi) \\ hu(\xi) \\ z(\xi) \end{pmatrix} = \begin{pmatrix} h^* + \xi \\ (h^* + \xi)\left[u^* - 2(\sqrt{g(h^*+\xi)} - \sqrt{gh^*})\right] \\ z^* \end{pmatrix}. \tag{58}$$

Similarly, the integral curve for the 2-characteristic field can be calculated, obtaining the conjugated of (58). It is more interesting to analyze the result for the S-characteristic field, that reads

$$\mathbf{U}^S(\xi) = \begin{pmatrix} h(\xi) \\ hu(\xi) \\ z(\xi) \end{pmatrix} = \begin{pmatrix} h^* + \xi \\ (hu)^* \\ \frac{u^{*2}}{2g} + z^* - \frac{(hu)^{*2}}{2g(h^*+\xi)^2} - \xi \end{pmatrix}, \tag{59}$$

as it can be given a physical meaning. One can realize that the third equation in vector (59), in combination with the first and second equations, stands for the conservation of the specific mechanical energy across the contact wave. Such an idea can be more generally conveyed by saying that the Riemann invariants of the S-characteristic field are the discharge and the mechanical energy. In Table 1, the Riemann invariants for all waves are presented.

Characteristic field	1-Riemann invariant	2-Riemann invariant
1	$u + 2\sqrt{gh}$	z
S	hu	$\frac{u^2}{2g} + h + z$
2	$u - 2\sqrt{gh}$	z

Table 1: Summary of Riemann invariants for the non-homogeneous SWE.

4.2. Conservation of energy across the bed-step contact wave

As outlined in the previous section, the S-characteristic field in the non-conservative SWE in (52) is a linearly degenerate field. This kind of field arises from the presence of the bed step and is characterized by a contact wave of zero celerity, $\lambda^S = 0$, since the bed elevation does not vary in time.

Discontinuous solutions describing a contact wave are generally expressed by (33). For this particular case, the right state will be denoted by $\mathbf{U}_R$, hence (33) is rewritten as

$$\mathbf{U}(x,t) = \begin{cases} \mathbf{U}_L & x < 0 \\ \mathbf{U}_R & x > 0 \end{cases} \tag{60}$$

where $\mathbf{U}_L = (h_L, (hu)_L, z_L)^T$ and $\mathbf{U}_R = (h_R, (hu)_R, z_R)^T$ are the left and right states respectively. Notice that we may write $(hu)_L = h_L u_L$ for the sake of clarity and recall that this quantity represents the first Riemann invariant of the S-characteristic field, hence $h_L u_L = h_R u_R$. The second Riemann invariant is the specific mechanical energy, hence $u_L^2/2 + g(h+z)_L = u_R^2/2 + g(h+z)_R$.

Across the contact wave in (60), the Generalized Rankine-Hugoniot (GRH) condition in (24) must hold for all variables. For this particular case, it reads

$$
\begin{gathered}
h_R u_R - h_L u_L = 0\,, \\
\left(g\frac{h_R^2}{2} + h_R u_R^2\right) - \left(g\frac{h_L^2}{2} + h_L u_L^2\right) = D\,,
\end{gathered}
\tag{61}
$$

with D a suitable approximation of the integral of the source term across the bed step

$$
D = -\int_{z_L}^{z_R} ghdz\,, \tag{62}
$$

that can be rewritten as

$$
D = -\int_{x_L}^{x_R} gh\frac{dz}{dx}dx\,. \tag{63}
$$

As outlined before, GRH condition in (61) must be ensured so that (60) is a weak solution of the problem, hence the right state $(h_R, h_R u_R, z_R)$ must lie on the Generalized Hugoniot Locus (GHL) for a given left state $(h_L, h_L u_L, z_L)$. However, this condition does not ensure the conservation of Riemann invariants across the contact wave. Only when condition in (39) holds, Riemann invariants are conserved and the IC coincide with the GHL. In other words, we can state that the Integral Curve (IC) coincide with the GHL if (61) holds and the Riemann invariants of the S-field in Table 1 are conserved.

It seems clear that the election of a suitable discretization of the integral of the source term in (63) is crucial. In [14], a particular STD based on physical considerations that accounts for the dissipation of energy across the step was chosen. Under this assumption, they showed that equation (39) is not always satisfied and proved that the Riemann invariant associated to the specific mechanical energy was not anymore conserved across the step. In this way, they provided a coherent mathematical framework for the physically-based dissipative discretization of the bed step and they constructed a Riemann solver based on such ideas.

Unlike [14], in the present work the authors do not include any additional energy dissipation mechanism. Here, an energy-conservative STD is sought, hence both the GRH condition and Equation (39) must hold, as Riemann invariants have to be conserved across the contact wave. Following [14], equation (39) is rewritten as

$$-\int_0^{\hat{\xi}} \mathbf{H} \cdot \mathbf{e}^S d\xi = \mathbf{D}\,, \tag{64}$$

where $\hat{\xi} = h_R - h_L$ is the value of ξ on the right state. We define

$$h(\hat{\xi}) = h_R \qquad u(\hat{\xi}) = u_R \qquad z(\hat{\xi}) = z_R\,. \tag{65}$$

Our goal here is to find the expression for $\mathbf{D}$ satisfying (64) and to this end, we have to manipulate (64) using extra relations among left and right states. It is worth recalling that for the derivation of condition (64) (originally (39)), $\mathbf{U}(\xi)$ was imposed to lie on the IC, given by Equation (59). Here we will work under the same assumption, hence $\mathbf{U}(\hat{\xi}) = \mathbf{U}_R = (h_R, h_R u_R, z_R)$ lies on the IC for a given left state. Water depth along the IC can be expressed as

$$h(\hat{\xi}) = h_L + \hat{\xi} = h_R \tag{66}$$

and in the same way, the velocity along the IC is

$$u(\hat{\xi}) = \frac{h_L u_L}{h_L + \hat{\xi}} = \frac{h_R u_R}{h_R} = u_R\,, \tag{67}$$

with a constant discharge

$$q = hu(\hat{\xi}) = h_L u_L = h_R u_R\,, \tag{68}$$

also denoted by q, and a variable bed elevation along the IC

$$z(\hat{\xi}) \equiv z_R = z_L + h_L - h_R + \frac{u_L^2}{2g} - \frac{u_R^2}{2g}\,. \tag{69}$$

In the following derivation, condition (64) will be combined with the relations between left and right states in (66)-(69), allowing to find the expression of $\mathbf{D}$ satisfying the RI and the GRH conditions. The product $\mathbf{H} \cdot \mathbf{e}^S$ reads

$$\mathbf{H} \cdot \mathbf{e}^S = \begin{pmatrix} 0 \\ u^2(\xi) - gh(\xi) \\ 0 \end{pmatrix} \tag{70}$$

and using (67) in (70), the latter yields

$$
-\int_0^{\hat{\xi}} \left(\begin{array}{c} 0 \\ \left(\dfrac{h_L u_L}{h_L+\xi}\right)^2 - g(h_L+\xi) \\ 0 \end{array} \right) d\xi = \left(\begin{array}{c} 0 \\ D \\ 0 \end{array} \right) . \tag{71}
$$

From (71), only the second component will be considered

$$
-\int_0^{\hat{\xi}} \left(\frac{h_L u_L}{h_L+\xi}\right)^2 d\xi + \int_0^{\hat{\xi}} g(h_L+\xi) d\xi = D\,. \tag{72}
$$

Integrating (72) and using the relation $h_L u_L = h_R u_R$ in (68) when required, it yields

$$
\left(g\frac{h_R^2}{2} + h_R u_R^2\right) - \left(g\frac{h_L^2}{2} + h_L u_L^2\right) = D\,, \tag{73}
$$

with the right state laying on the IC in (59). It can be noticed that equation (73) coincides with the GRH condition for the conservation of momentum.

Now, combination of equation (73) with (69) allows to derive the particular STD, D, that under the assumed hypotheses will ensure the conservation of the Riemann invariants and lead to an energy-conservative scheme. For the sake of clarity, equation (73) is rewritten as

$$
\delta\left(g\frac{h^2}{2} + hu^2\right)_{L,R} = D \tag{74}
$$

and so is (69), the equation for the conservation of energy

$$
\delta\left(\frac{u^2}{2} + g(h+z)\right)_{L,R} = 0 \tag{75}
$$

where $\delta(\cdot)_{L,R} = (\cdot)_R - (\cdot)_L$ is a difference operator. From (74), it is straightforward to obtain

$$
\left(g\bar{h}\delta h + \bar{u}\delta(hu) + \overline{hu}\delta u\right)_{L,R} = D\,, \tag{76}
$$

where

$$
\bar{(\cdot)}_{L,R} = \frac{(\cdot)_L + (\cdot)_R}{2} \tag{77}
$$

is an average operator. For the sake of simplicity, subscript $(\cdot)_{L,R}$ is dropped in Equations (78)-(82) as they always refer to the left and right states of the contact wave in this derivation. Noticing that $\delta(hu)_{L,R} = h_R u_R - h_L u_L = 0$, Equation (76) yields

$$g\bar{h}\delta h + \overline{hu}\delta u = D\,. \tag{78}$$

The equation for the conservation of energy in (75) is multiplied by $\bar{h}$ and rewritten as

$$\bar{h}\bar{u}\delta u + g\bar{h}\delta h + g\bar{h}\delta z = 0\,, \tag{79}$$

from where the term $g\bar{h}\delta h$ can be expressed as

$$g\bar{h}\delta h = -\bar{h}u\delta u - g\bar{h}\delta z \tag{80}$$

and can be inserted in (78), leading to

$$D = -g\bar{h}\delta z + (\overline{hu} - \bar{h}\bar{u})\delta u\,. \tag{81}$$

It is straightforward to show that (81) can be rewritten as

$$D = -g\bar{h}\delta z + \delta(hu^2) - \bar{u}\delta(hu) - \bar{h}\delta\left(\frac{1}{2}u^2\right)\,, \tag{82}$$

with $\delta(hu) = 0$ according to the GRH conditions, hence

$$D = -g\bar{h}\delta z + \delta(hu^2) - \bar{h}\delta\left(\frac{1}{2}u^2\right)\,. \tag{83}$$

As outlined before, weak solutions for the bed step contact wave are always required to satisfy the GRH condition. That is to say, for a given left state, the right state is calculated using (61). When the discretization of the source term in (63), D, is undefined, there are infinite solutions for the right state and only when choosing a particular discretization, the right state can be determined. Unlike the approach proposed in [14] where the authors impose a particular STD based on energy dissipation hypothesis, here the expression for the discretization of the source term is derived imposing the equivalence between GHL and IC. To this end, apart from the GRH condition, we require an extra condition given by (39) in order to ensure the constancy of Riemann invariants across the wave. Notice that such a condition consists of the equation for the conservation of energy provided by the IC.

4.3. Numerical discretization of the source term at cell interfaces for augmented solvers

When using augmented solvers, such as the HLLS and ARoe solvers, numerical approximations over the integral of the source term at cell interfaces are required. The approximation of the spatial integral of the source term at cell interface $i+1/2$, that is inside $[x_i, x_{i+1}]$, will be referred to as

$$\int_{x_i}^{x_{i+1}} -g\,h\,\frac{dz}{dx}dx \approx \bar{S}_{i+1/2}\,. \tag{84}$$

We can find in the literature different numerical approaches for Equation (84), however, this choice is not trivial since most of such approaches are not able to ensure a numerical solution that converges to a physically based solution with mesh refinement, even when using high order schemes. This problem is put into evidence when looking, for instance, at the discrete energy level or at the shock positioning given by the numerical scheme. In this section, four different source term discretizations are described. Two of them, the differential formulation (DF) and the integral formulation (IF), are traditional approaches, which are easy to program and exhibit an overall acceptable performance but they are not able to ensure conservation of energy. Moreover, the IF does not allow the numerical scheme to converge to the exact shock position, for steady shocks, with mesh refinement. The other two STDs described here, in contrast, are energy balanced discretizations, that is to say, they allow the numerical scheme to preserve the discrete level of energy (when required) and to dissipate the exact amount of energy in presence of hydraulic jumps. Such techniques are called weighted energy balanced formulation (WEBF) and the selective energy balanced method (SEBF) and whereas the former is still not able to make the scheme converge to the exact position of the hydraulic jump under steady regime, the latter does, as it will be shown in the following section. Therefore, among the four techniques described here, only the SEBF which is presented here for the first time, is well suited for both energy conservation and accurate shock capturing.

One possibility is to compute it considering a smooth variation of the variables across the interface, as

$$\bar{S}^{DF}_{i+1/2} = -g\bar{h}\delta z\,, \tag{85}$$

which will be referred to as differential formulation (DF) and where

$$\bar{h} = \frac{1}{2}(h_{i+1} + h_i)\,, \qquad \delta z = z_{i+1} - z_i\,. \tag{86}$$

The second possibility is the so-called integral formulation (IF), derived from the integration of the pressure along the bottom step for a piecewise constant data reconstruction of the bed elevation, z. If assuming that the pressure distribution is hydrostatic over the step and depends only on the free-surface level on the side of the discontinuity where the bottom elevation is lower, the source term is evaluated explicitly at $t = 0$ as [11]

$$\bar{S}^{IF}_{i+1/2} = -g\left(h_j - \frac{|\delta z'|}{2}\right)_{i+\frac{1}{2}} \delta z'_{i+\frac{1}{2}}\,, \tag{87}$$

where z is the bed level surface, and j and $\delta z'$ are given by

$$j = \begin{cases} i & \text{if } \delta z_{i+\frac{1}{2}} \geq 0 \\ i+1 & \text{if } \delta z_{i+\frac{1}{2}} < 0 \end{cases} \qquad \delta z' = \begin{cases} h_i & \text{if } \delta z_{i+\frac{1}{2}} \geq 0 \text{ and } d_i < z_{i+1} \\ -h_{i+1} & \text{if } \delta z_{i+\frac{1}{2}} < 0 \text{ and } d_{i+1} < z_i \\ \delta z & \text{otherwise} \end{cases} \tag{88}$$

and $d = h + z$ is the water level surface.

In cases of still water with a continuous water level surface, both (85) and (87) do ensure quiescent equilibrium. In this particular case hydrostatic forces are exactly balanced.

In order to extend the well-balanced property for static equilibrium to the energy-balanced property, that ensures the exact conservation of energy in steady cases with moving water, it is necessary to impose extra conditions in the discretization of the source term. Generally, under the assumption of conservation of energy across the bed step contact wave, the best choice for the discretization of the bed source term seems to be Equation (81). However, such a discretization does not allow to construct an explicit scheme as it depends upon the intermediate states at both sides of the bed step, $\mathbf{U}_i^-$ and $\mathbf{U}_{i+1}^+$.

Under steady conditions and considering no change in flow regime across the RP, it is straightforward to prove that $\mathbf{U}_i = \mathbf{U}_i^-$ and $\mathbf{U}_{i+1} = \mathbf{U}_{i+1}^+$, hence (81) can be rewritten in terms of the initial data as

$$D = \begin{aligned} & -g\left(\frac{h_{i+1}+h_i}{2}\right)(z_{i+1}-z_i)+ \\ & \left[\left(\frac{(hu)_{i+1}+(hu)_i}{2}\right)-\left(\frac{h_{i+1}+h_i}{2}\right)\left(\frac{u_{i+1}+u_i}{2}\right)\right](u_{i+1}-u_i)\,. \end{aligned} \tag{89}$$

For the sake of clarity, notation for Equation (89) is simplified, considering variations and averages across the interface $i+1/2$, that is, the left and right states of the RP. By doing this, (89) is rewritten as

$$D = \left\{-g\bar{h}\delta z + (\overline{hu} - \bar{h}\bar{u})\delta u\right\}_{i+1/2}\,. \tag{90}$$

In shallow flows, there are physically feasible situations where energy is dissipated, such as hydraulic jumps. Ideally, such a shock would be considered as a pure discontinuity where energy is suddenly dissipated, however, when using a finite volume formulation, the shock width is of the size of a cell, since the discretization considers constant values in each cell and the discontinuity cannot be kept anymore as a discontinuity inside a cell. As a consequence, energy dissipation must be imposed at the interfaces of the cell containing the shock, as it is not possible to explicitly carry out the dissipation of energy inside the cell.

Murillo [25] proposed a novel approach for the discretization of the source term that allows to construct an exactly energy balanced scheme. This approximation is based on the principle of conservation of mechanical energy and is only applied to the leading term, since higher order terms become nil in steady state when energy is conserved, as mentioned above.

Considering the IF and DF approaches for the discretization of the source term, it is possible to evaluate $\bar{S}_{i+1/2}$ as a combination of them as

$$\bar{S}_{i+1/2} = (1-\mathcal{A})S^{DF}_{i+1/2} + \mathcal{A}S^{IF}_{i+1/2}\,, \tag{91}$$

where $0 \leq \mathcal{A} \leq 1$. This formulation will be referred to as weighted energy balanced formulation (WEBF). In order to satisfy both energy and momentum conservation under steady conditions, a value $\mathcal{A}_E$ is defined as

$$\mathcal{A}_E = \frac{\delta(hu^2) - \bar{h}\delta\left(\frac{u^2}{2}\right)}{S^{IF}_{i+1/2} - S^{DF}_{i+1/2}}\,, \tag{92}$$

according to [25]. Coefficient $\mathcal{A}_E$ can be used in (91) to ensure the conservation of energy for smooth solutions. On the other hand, when considering transcritical jumps, energy must be dissipated, hence the value of weight coefficient $\mathcal{A}$ in (91) is set to 1. Considering these situations, the complete algorithm for the calculation of $\mathcal{A}$ reads [25]

$$\mathcal{A} = \begin{cases} 1 & \text{if } u_{i+1}u_i > 0 \text{ and } u_i > 0 \text{ and } |Fr_{i+1}| < 1 \text{ and } |Fr_i| > 1 \\ 1 & \text{if } u_{i+1}u_i > 0 \text{ and } u_i < 0 \text{ and } |Fr_{i+1}| > 1 \text{ and } |Fr_i| < 1 \\ \mathcal{A}_E & \text{otherwise} \end{cases} \tag{93}$$

where Fr_i and Fr_{i+1} are the Froude numbers on the left and right sides of the interface. It is worth pointing out that $\mathcal{A}_E$ can be straightforward obtained from Equation (90).

On the other hand, instead of imposing the exact amount of dissipation of energy across the shock by means of a tailored STD at that point, in this work we propose to add an additional degree of freedom to the equations by means of using a traditional discretization of the source term at the interfaces surrounding the hydraulic jump while maintaining the energy conservative formulation in (90) for the rest. The differential discretization of the source term is chosen at those interfaces. This technique allows the numerical scheme to converge to the exact position of the shock while recovering the exact solution in both the subcritical and supercritical regions connected by the transcritical shock, with independence of the grid refinement.

The proposed approach is next explained. We propose to use Roe celerities, $\tilde{\lambda}^m$ to identify the cell containing the hydraulic jump, since it is known that both celerities at the left interface are positive (supercritical flow entering the cell) while the celerities at the right interface correspond to subcritical conditions (one negative and the other one positive). Let us consider the cells, Ω_i, as single cells contained in the computational domain Ω such that $\Omega = \{\Omega_i \mid i \in [1, ..., N]\}$. Considering the possibility of multiple hydraulic jumps within the domain, we denote the set of cells containing a positive-flow hydraulic jump as

$$\mathcal{D}^+ = \left\{ \Omega_i \mid \Omega_i \in \Omega \wedge \tilde{\lambda}^1_{i-1/2} \cdot \tilde{\lambda}^1_{i+1/2} < 0 \wedge h_{i-1} < h_{i+1} \right\} \tag{94}$$

and the set of cells containing a negative-flow hydraulic jump as

$$\mathcal{D}^- = \left\{ \Omega_i \mid \Omega_i \in \Omega \wedge \tilde{\lambda}^2_{i-1/2} \cdot \tilde{\lambda}^2_{i+1/2} < 0 \wedge h_{i-1} > h_{i+1} \right\} \tag{95}$$

and the set of Riemann Problems at the left and right interfaces of cells $\Omega_i \in \mathcal{D}^+ \cup \mathcal{D}^-$

$$\mathcal{R}_1 = \left\{ \mathrm{RP}_{i+1/2} \mid i \in \mathbb{N} \wedge \Omega_i \in \mathcal{D}^+ \cup \mathcal{D}^- \right\} \tag{96}$$
$$\mathcal{R}_2 = \left\{ \mathrm{RP}_{i-1/2} \mid i \in \mathbb{N} \wedge \Omega_i \in \mathcal{D}^+ \cup \mathcal{D}^- \right\} \tag{97}$$

respectively, where $\mathrm{RP}_{i-1/2}$ stands for the Riemann Problem at left interface and $\mathrm{RP}_{i+1/2}$ at right interface. The whole set of RPs is given by

$$\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \,. \tag{98}$$

By using the previous definitions, the approximation of the integral of the source term at any interface is defined as follows

$$\bar{S}_{i+1/2} = \begin{cases} -g\bar{h}\delta z + (\overline{hu} - \bar{h}\bar{u})\delta u & \text{if} \quad \mathrm{RP}_{i+1/2} \notin \mathcal{R} \\ -g\bar{h}\delta z & \text{if} \quad \mathrm{RP}_{i+1/2} \in \mathcal{R} \end{cases} \tag{99}$$

and the method will be hereafter referred to as selective energy balanced formulation (SEBF).

4.4. The ARoe scheme for the SWE

When applied to the Shallow Water Equations, the Augmented Roe solver provides a linearized solution that can be straightforward expanded from the homogeneous case. The approximate Jacobian $\widetilde{\mathbf{J}}$ of the homogeneous part is given by [8]

$$\widetilde{\mathbf{J}}_{i+1/2} = \begin{pmatrix} 0 & 1 \\ \widetilde{c}^2 - \widetilde{u}^2 & 2\widetilde{u} \end{pmatrix}_{i+1/2}, \qquad \delta\mathbf{F}_{i+1/2} = \widetilde{\mathbf{J}}_{i+1/2}\delta\mathbf{U}_{i+1/2} \,, \tag{100}$$

where

$$\begin{gathered} \widetilde{\lambda}^1 = \widetilde{u} - \widetilde{c}\,, \qquad \widetilde{\lambda}^2 = \widetilde{u} + \widetilde{c} \\ \widetilde{\mathbf{e}}^1 = \begin{pmatrix} 1 \\ \widetilde{u} - \widetilde{c} \end{pmatrix}, \qquad \widetilde{\mathbf{e}}^2 = \begin{pmatrix} 1 \\ \widetilde{u} + \widetilde{c} \end{pmatrix} \end{gathered} \tag{101}$$

with

$$\widetilde{c} = \sqrt{g\frac{h_i + h_{i+1}}{2}}\,, \qquad \widetilde{u} = \frac{u_{i+1}\sqrt{h_{i+1}} + u_i\sqrt{h_i}}{\sqrt{h_{i+1}} + \sqrt{h_i}}\,. \tag{102}$$

4.5. Test case 1: steady shock capturing for the SWE with bed topography

In this test case, steady solutions for the flow over the following bed elevation profile

$$z(x) = \begin{cases} 0 & \text{if} \quad x < 8 \\ 0.05(x-8) & \text{if} \quad 8 \leq x \leq 12 \\ 0.2 - 0.05\,(x-12)^2 & \text{if} \quad 12 \leq x \leq 14 \\ 0 & \text{if} \quad x > 12 \end{cases} \tag{103}$$

are computed using the ARoe solver in combination with the different discretization techniques for the source term outlined before. The computational domain is $[0, 20]$ and the solution is computed for $t = 600$ s. CFL number is set to 0.45 for all cases. The discharge is imposed to 0.6 m^2/s upstream to obtain the critical point at the cell with maximum bed elevation, that is $z_{max} = 0.2$. Downstream, the water depth is also imposed to $h = 0.621$ m in order to generate a hydraulic jump downstream. Different computational grids, composed of 100, 200, 400, 800 and 1600 cells respectively, are used to compute the numerical solution.

Numerical solutions provided by the ARoe solver when using the different approximations of the source term presented before, namely the differential formulation (DF), the integral formulation (IF), the weighted energy balanced formulation (WEBF) and the novel selective energy balanced method (SEBF), are presented and compared with the exact solution in Figures 4, 5. In Figure 4, the numerical solutions for $h+z$ and q computed by the ARoe solver in combination with all the previous techniques on two grids of 100 and 400 cells are plotted together and compared with the exact solution. To study the effect of mesh refinement in the accuracy of the numerical solution and convergence to the exact position of the shock, a detailed plot of the solution provided by each one of the methods is presented in Figure 5 for three different grids composed of 200, 400 and 800 cells respectively. Numerical results evidence that those approximations based on the integral discretization of the source term, such as the energy balanced approach from [25] and the integral discretization itself, do not accurately capture the position of the shock, with independence of the grid. In any case, the former strategy provides much better results than the latter, as it is energy-conservative. On the other hand, it is evidenced that both the differential formulation and the selective energy balanced formulation do accurately capture the shock position for any grid.

It is also noticed that a spurious spike in the numerical discharge appears for all methods and what is of utmost relevance, that the amplitude of this spike is not reduced with mesh refinement, as observed in Figure 4.

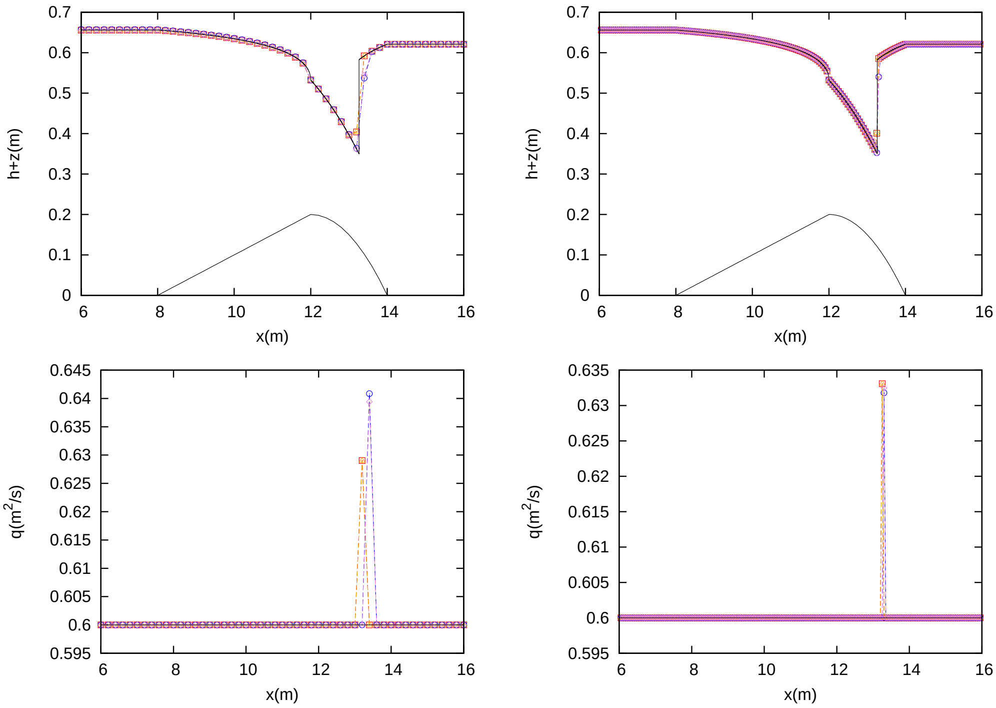

Figure 4: Test case 1. Exact (−) and numerical solution for $h + z$ (top) and q (bottom) computed by the ARoe solver in combination with the DF (− △ −), IF (− ○ −), SEBF (−□−) and WEBF (− ◇ −), using 100 (left) and 400 cells (right).

The numerical solution for the specific mechanical energy, computed using the aforementioned techniques in the grids of 100 and 400 cells, is presented in Figure 6 left and right respectively. It is observed that only when using an energy-balanced STD (E-scheme), such as the ARoe solver in combination with the SEBF or WEBF formulations, energy is conserved. On the other hand, when using the DF and IF formulations of the source term, energy is not conserved though it converges with mesh refinement. Among the assessed methods, the SEBF is the one providing the best performance, as it ensures the conservation of energy when required and accurately captures the position of the hydraulic jump. This method provides the exact solutions in all cells but the one containing the shock, with independence of the grid.

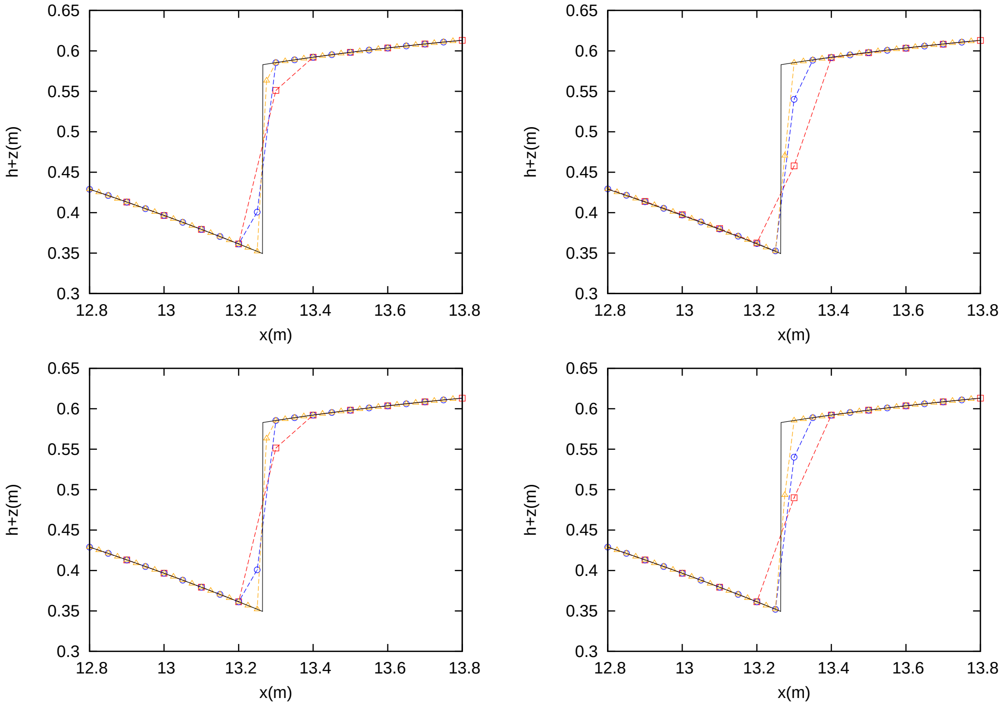

Figure 5: Test case 1. Exact ($-$) and numerical solution for $h+z$ computed by the ARoe solver in combination with the DF (top left), IF (top right), SEBF (bottom left) and WEBF (bottom right) using 200 ($-\square-$), 400 ($-\circ-$) and 800 ($-\triangle-$) cells.

5. Numerical shockwave anomalies in the SWE: computation of the hydraulic jump

It has been widely reported in the literature that significant numerical anomalies arise in presence of shock waves. An example of such problems are the Carbuncle, the slowly-moving shock and the wall-heating phenomenon, all of them leading to spurious numerical solutions. The aforementioned problems have been deeply studied in the framework of Euler equations and some authors have proposed different numerical techniques to address them. Here, we will focus on the numerical anomalies present when computing steady and moving hydraulic jumps, which are a particular type of shock waves in the framework of the Shallow Water Equations (SWE). Specifically, our interest lies in the reduction of the spike in the discharge, reported in the previous section.

The hydraulic jump occurs when a supercritical flow suddenly changes to

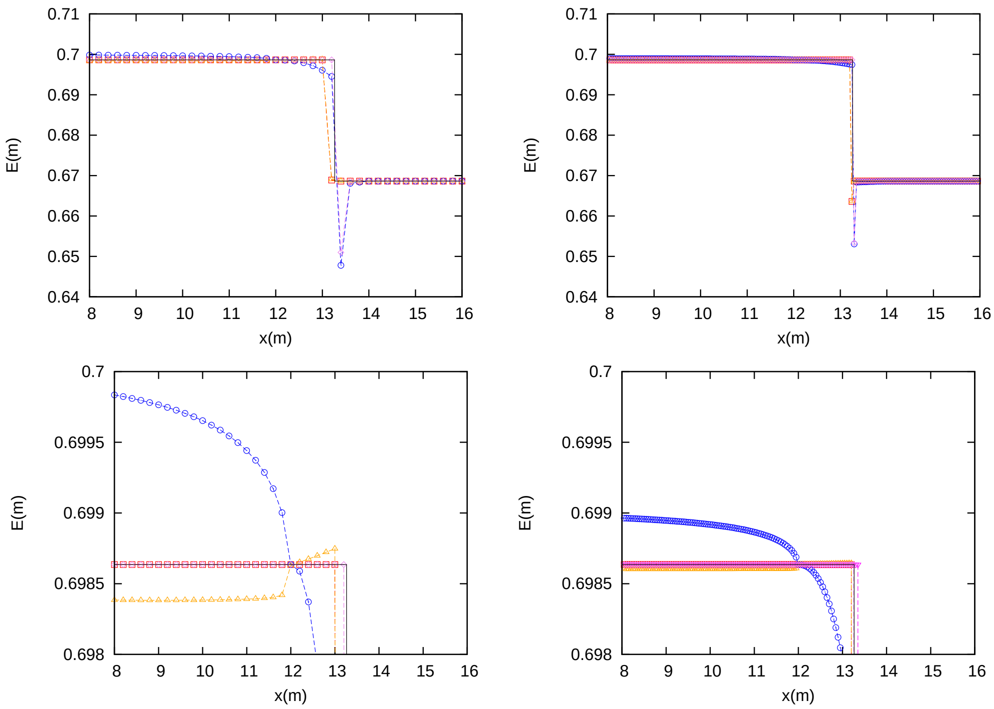

Figure 6: Test case 1. Numerical solution for the specific mechanical energy computed by the ARoe solver in combination with the DF (— △ —), IF (— ○ —), SEBF (—□—) and WEBF (— ◇ —) (top) and detail of the solution (bottom), using 100 (left) and 400 (right) cells.

subcritical conditions, generating a steep free surface elevation where intense mixing takes place and a large amount of mechanical energy is dissipated. Mathematically, hydraulic jumps are modelled by a discontinuity corresponding to a shock wave and the relation between the states at each side of the discontinuity is provided by the RH conditions.

5.1. Hugoniot locus of the hydraulic jump

To understand the mathematical treatment of the hydraulic jump and the numerical anomalies arising from such a wave, it is worth studying first the analytical solution of this type of wave under the simplest conditions, that is over flat bed. From Rankine-Hugoniot (RH) conditions, all possible values connecting the left and right states can be determined and represented in phase space as $(h(\xi), hu(\xi))$ by means of the so-called Hugoniot locus

$$\mathbf{U}(\xi) = \begin{pmatrix} h(\xi) \\ hu(\xi) \end{pmatrix} = \begin{pmatrix} h_L + \xi \\ (hu)_L + \xi \left(u_L \pm \sqrt{gh_L + \frac{1}{2} g\xi \left(3 + \frac{\xi}{h_L}\right)} \right) \end{pmatrix}, \tag{104}$$

where $\xi = h - h_L$, with h the independent variable used for the parametrization. From (104), we notice that two families of curves are possible, denoted by Ψ^1 and Ψ^2, which are associated to the 1-wave and 2-wave respectively. Such curves are defined by

$$\Psi^1(\xi) = \begin{pmatrix} \psi_1^1(\xi) \\ \psi_2^1(\xi) \end{pmatrix} = \begin{pmatrix} h_L + \xi \\ (hu)_L + \xi \left(u_L - \sqrt{gh_L + \frac{1}{2} g\xi \left(3 + \frac{\xi}{h_L}\right)} \right) \end{pmatrix}, \tag{105}$$

$$\Psi^2(\xi) = \begin{pmatrix} \psi_1^2(\xi) \\ \psi_2^2(\xi) \end{pmatrix} = \begin{pmatrix} h_L + \xi \\ (hu)_L + \xi \left(u_L + \sqrt{gh_L + \frac{1}{2} g\xi \left(3 + \frac{\xi}{h_L}\right)} \right) \end{pmatrix}. \tag{106}$$

Figure 7 depicts different curves obtained for different left-reference states using (105) in red and (106) in blue, for $\Psi^1, \Psi^2 \in \mathbb{R}^+ \times \mathbb{R}^+$. Also the curve $hu(h) = \sqrt{gh^3}$ that represents the transition between supercritical (white background) and subcritical region (green background) is depicted in the figure. For any given set of two points laying on a curve, a weak solution of the PDEs in the form of a shock wave is mathematically possible. It is worth pointing out that further representations of the aforementioned curves will be carried out by the parametrization of ψ_2^m, which is the discharge hu, in terms of ψ_1^m, which is h, so that their representation in the phase space h, hu is straightforward.

It must be borne in mind that not every choice of subcritical state that is connected to a given supercritical state represents a hydraulic jump. For instance, let us consider a left supercritical state given by $h_L = 0.85$ and $hu_L = 3.411764705882353$ and let us find two possible right states connected to it, each of them laying on each branch of the Hugoniot locus. This is depicted in Figure 8, where the original left state is denoted by F, the right state lying on the 1-curve, Ψ^1, is denoted by G and the right state lying on

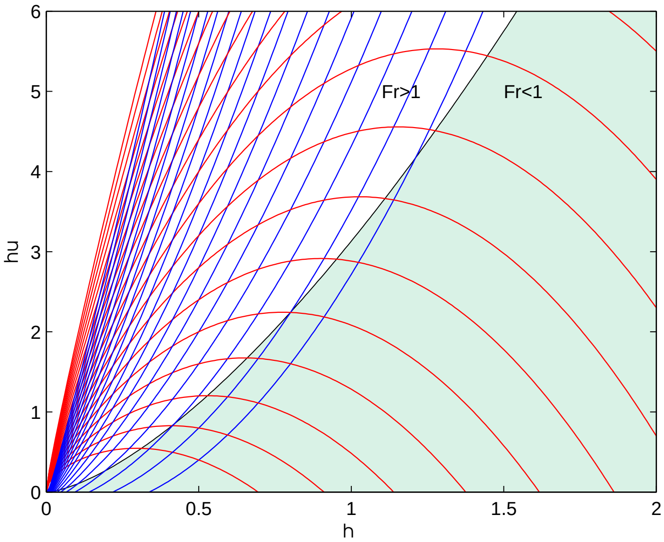

Figure 7: Phase space $(h, hu) \in \mathbb{R}^+ \times \mathbb{R}^+$ with the subcritical region depicted in green background and the supercritical region in white background, showing the Hugoniot locus Ψ^1 in red and Ψ^2 in blue, obtained for different left-reference states using (105) and (106) respectively.

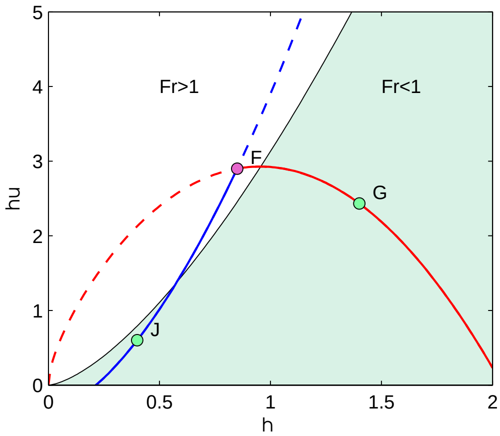

Figure 8: Phase space $(h, hu) \in \mathbb{R}^+ \times \mathbb{R}^+$ with the subcritical region depicted in green background and the supercritical region in white background, showing the Hugoniot locus Ψ^1 in red and Ψ^2 in blue.

the 2-curve, Ψ^2, is denoted by J. The entropically inadmissible region of the curves has been represented by dashed line. It is observed that both G and J lie on the subcritical region of the phase plane and they are both entropically admissible, however, only the combination of states F–G leads to a hydraulic jump, because G, unlike J, has a higher water depth than F and, what is decisive in this case, wave celerities of F and G have opposite sign. More generally, we can define an hydraulic jump as:

Definition 3. *(Hydraulic jump). Let the following discontinuous solution*

$$\mathbf{U}(x,t) = \begin{cases} (h,hu)_L & x < 0 \\ (h,hu)_R & x > 0 \end{cases} \tag{107}$$

be a weak solution of the SWE system, where $(h,hu)_L$ and $(h,hu)_R$ are two different states laying on Ψ^m and satisfying the entropy condition $\lambda^m(\mathbf{U}_L) > \mathcal{S}^m > \lambda^m(\mathbf{U}_R)$, with $\mathcal{S}^m$ the speed of the jump, that undergoes a flow transition as $Fr_L < 1 < Fr_R$ or $Fr_R < 1 < Fr_L$. Solution in (107) is termed as hydraulic jump if and only if $\lambda^m(\mathbf{U}_L) > 0 > \lambda^m(\mathbf{U}_R)$.

Notice that, according to the previous definition, hydraulic jumps admit that $\mathcal{S}^m$ be nil, hence they are the only shock-type solution for the SWE that can be stationary at a fixed position.

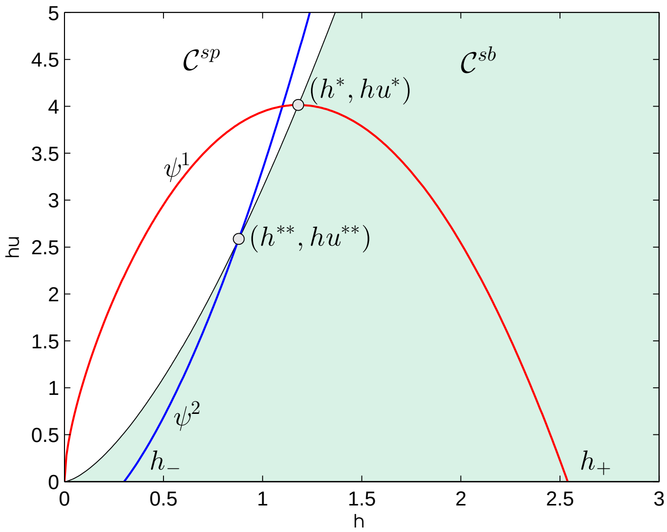

Figure 9: Phase space $(h,hu) \in \mathbb{R}^+ \times \mathbb{R}^+$ with the subcritical region, $\mathcal{C}^{sb}$, depicted in green and the supercritical, $\mathcal{C}^{sp}$, region in white, showing the Hugoniot locus Ψ^1 in red and Ψ^2 in blue and the corresponding intersection.

From the analysis of the Hugoniot locus considering $h, hu > 0$ and departing from a left reference point located in the supercritical region, we notice the following points:

- Curve $\sqrt{gh^3}$ is monotonically increasing and divides the space $\mathbb{R}^+ \times \mathbb{R}^+$ in two sets, $\mathcal{C}^{sp}$ and $\mathcal{C}^{sb}$, as follows

$$\mathcal{C}^{sp} = \left\{(h, hu) \in \mathbb{R}^2 \mid hu > \sqrt{gh^3} \wedge h > 0\right\} , \tag{108}$$

$$\mathcal{C}^{sb} = \left\{(h, hu) \in \mathbb{R}^2 \mid hu < \sqrt{gh^3} \wedge h > 0\right\} , \tag{109}$$

  such that $\mathcal{C}^{sp} \cup \mathcal{C}^{sb} \cup \mathcal{C}^{cr} = \mathbb{R}^+ \times \mathbb{R}^+$, where

$$\mathcal{C}^{cr} = \left\{(h, hu) \in \mathbb{R}^2 \mid hu = \sqrt{gh^3} \wedge h > 0\right\} . \tag{110}$$

- Curve $\sqrt{gh^3}$ is monotonically increasing.

- Curve ψ_2^1 has a global maximum at h_{max} such that $(h_{max}, hu_{max}) \in \mathbb{R}^+ \times \mathbb{R}^+$.

- Curve ψ_2^2 is monotonically increasing in $\mathbb{R}^+ \times \mathbb{R}^+$.

- Curves $\sqrt{gh^3}$ and ψ_2^1 intersect at a single point denoted by $(h^*, hu^*) \in \mathbb{R}^+ \times \mathbb{R}^+$, with $hu^* < hu_{max}$.

- Curves $\sqrt{gh^3}$ and ψ_2^2 intersect at a single point denoted by $(h^{**}, hu^{**}) \in \mathbb{R}^+ \times \mathbb{R}^+$.

- We can define two sets of h states, $\mathcal{H}^{sp,1} = (0, h^*)$ and $\mathcal{H}^{sb,1} = (h^*, h_+)$, with h_+ the value of h for which $\Psi^1 = (h_+, 0)$, such that $\Psi^1 \in \mathcal{C}^{sp} \ \ \forall h \in \mathcal{H}^{sp,1}$ and $\Psi^1 \in \mathcal{C}^{sb} \ \ \forall h \in \mathcal{H}^{sb,1}$.

- We can define two set of h states, $\mathcal{H}^{sp,2} = (h_-, h^{**})$ and $\mathcal{H}^{sb,2} = (h^{**}, \infty)$, with h_- the value of h for which $\Psi^2 = (h_-, 0)$, such that $\Psi^2 \in \mathcal{C}^{sb} \ \ \forall h \in \mathcal{H}^{sp,2}$ and $\Psi^2 \in \mathcal{C}^{sp} \ \ \forall h \in \mathcal{H}^{sb,2}$.

Definitions introduced in the previous statements are depicted in the top-left plot in Figure 10. From the previous points, the following observations are worth being mentioned:

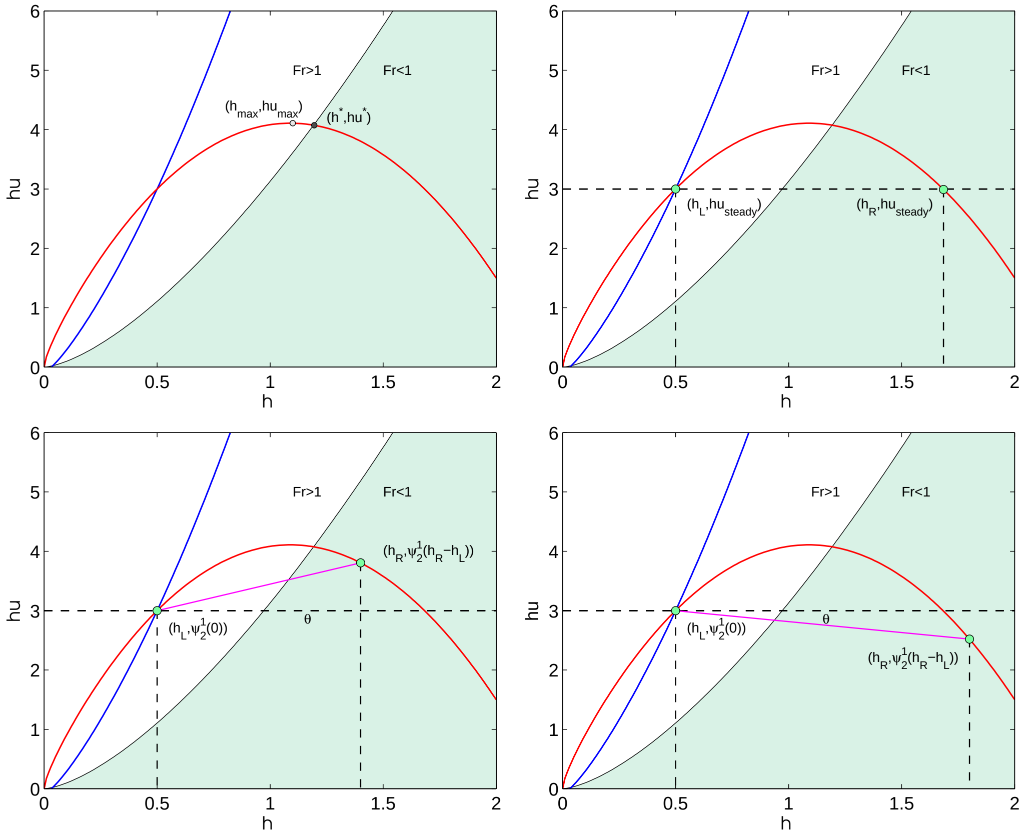

Figure 10: Hugoniot locus Ψ^1 in red and Ψ^2 in blue for the left state $(h, hu) = (0.5, 3)$, showing three possible solutions in the form of a hydraulic jump: a steady jump (top-right), a right-moving jump (bottom-left) and a left-moving jump (bottom-right).

- According to the two last points stated before, hydraulic jumps with $hu > 0$ only take place when $\Psi^m \in \mathcal{C}^{sp}$ $\forall h \in \mathcal{H}^{sp,m}$ and $\Psi^m \in \mathcal{C}^{sb}$ $\forall h \in \mathcal{H}^{sb,m}$, which is only possible for Ψ^1. Hence, any solution for RP$(\mathbf{U}_L, \mathbf{U}_R)$, with $\mathbf{U}_L = \Psi^1(0)$ and $\mathbf{U}_R = \Psi^1(h - h_L)$ $\forall h \in \mathcal{H}^{sb,1}, \forall h_L \in \mathcal{H}^{sp,1}$, evolves as a hydraulic jump.

- There exist two points $h_L \in \mathcal{H}^{sp,1}$ and $h_R \in \mathcal{H}^{sb,1}$ such that $\psi_2^1(0) = \psi_2^1(h_R - h_L) \equiv (hu)_{steady}$ and $\psi_2^1(0), \psi_2^1(h_R - h_L) \in (0, hu^*) \subset \mathbb{R}^+$. Such points correspond to the left and right states of the hydraulic jump under steady conditions with a constant discharge of $(hu)_{steady}$. This case is depicted in Figure 10 (top-right plot)

- There exist two other points $h_L \in \mathcal{H}^{sp,1}$ and $h_R \in \mathcal{H}^{sb,1}$ such that $\psi_2^1(0) \in (0, hu_{max}) \subset \mathbb{R}^+$ and $\psi_2^1(h_R - h_L) \in (0, hu^*) \subset \mathbb{R}^+$. If $\psi_2^1(0) < \psi_2^1(h_R - h_L)$ a right-moving transient shock will appear as depicted in Figure 10 (bottom-left plot). If $\psi_2^1(0) > \psi_2^1(h_R - h_L)$, a left-moving transient shock will appear as depicted in Figure 10 (bottom-right plot).

- Shock speed is equal to the slope of the magenta straight line in Figure 10, that is $\mathcal{S} = \tan\theta$.

- The previous statements apply to ψ_2^2 in the region $\mathbb{R}^+ \times \mathbb{R}^-$ when considering $hu < 0$.

5.2. Analytical study and comparison of the exact solution for 2 and 3-states hydraulic jumps.

Prior to analyzing the numerical solutions of Godunov's scheme to the hydraulic jump, it is worth studying the analytical solutions to this problem, which will help to understand the nature and characteristics of the numerical (discrete) solution to it. It is well known that an intermediate state appears in the numerical solution provided by Godunov's scheme, with independence of the solver [42]. The presence of this intermediate state, hereafter denoted by $\mathbf{U}_M$, is not of any physical relevance as it provides an unrealistic estimation of the average discharge in the intermediate cell (spike) which does not match the constant value of discharge. However, when using conservative schemes the intermediate value may be useful to compute a rough estimate of the shock position. The position of the shock inside the cell can be computed imposing conservation of mass as

$$x_S = \frac{h_M - h_R}{h_L - h_R}\,, \tag{111}$$

where $x_S \in [0, 1]$ represents the normalized position of the shock (where 0≡left interface, 0.5≡middle position and 1≡right interface) [42].

As a first approach and before getting into the numerical issues concerning hydraulic jumps, let us compare analytically the solution for the ideal steady hydraulic jump (pure discontinuity) with another solution for the steady hydraulic jump that includes an intermediate state, which resembles the discrete solution provided by Godunov's scheme. Both solutions are weak solutions of the equations and they are both valid. Whereas the former is characterized by two states, namely $\mathbf{U}_L$ and $\mathbf{U}_R$, the latter is given by $\mathbf{U}_L$,

$\mathbf{U}_M$ and $\mathbf{U}_R$. Moreover, the latter does not experience a sudden transition of flow regime, hence it cannot be considered a pure, or ideal, hydraulic jump.

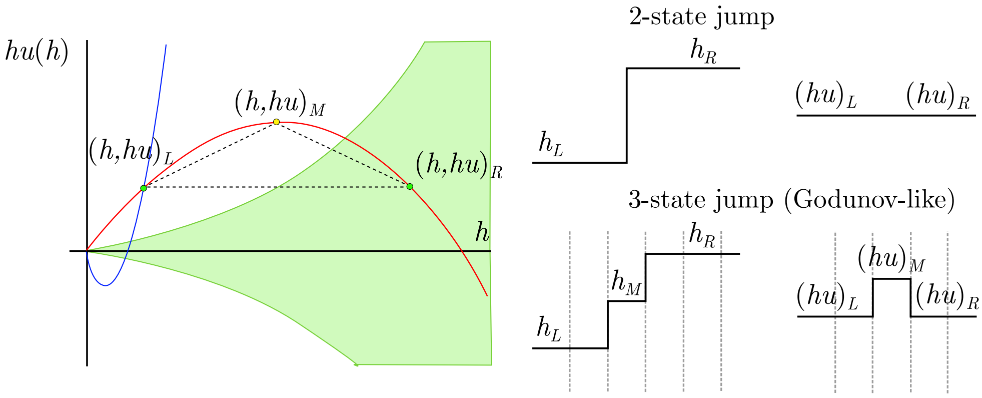

Figure 11: Hugoniot Locus and sketch of the analytical solutions for a 2-state and 3-state hydraulic jumps.

Let us consider first the ideal hydraulic jump composed of two states. This solution consists of a supercritical right-moving steady flow that suddenly decelerates through a pure discontinuity to subcritical conditions, as depicted schematically in Figure 11 (top-right). The Hugoniot locus that connects the left and right states of the jump, Ψ^1, is depicted in Figure 11 (left), showing that such states are located at the intersection of the Hugoniot Locus with the curve of constant discharge $(hu)_L = (hu)_R$, ensuring the steady regime.

On the other hand, when seeking a weak solution of the equations that includes an intermediate state, $\mathbf{U}_M$, as depicted in Figure 11 (bottom-right), we need to look for this additional state on the Hugoniot curve. According to Figure 11 (left), the intermediate state $(h_M, (hu)_M)$ (yellow point) will lie on Hugoniot Locus and is connected to the left and right states (green points) through this curve. From the previous observations, we realize that only a linear Hugoniot Locus would ensure a constant discharge in the intermediate state [42].

If a curve of the family of

$$\breve{\Psi}(\xi) = \begin{pmatrix} h(\xi) \\ (hu)_{steady} \end{pmatrix} \tag{112}$$

was considered in state space, with $(hu)_{steady} \in \mathbb{R}^+$ for a right-moving flow, a constant discharge for the intermediate state would be possible. Only if

Ψ^1 was of the type of $\breve{\Psi}$, constant discharge would be ensured across the intermediate cell. This means that we would have a linear Hugoniot [42]. This concept can be extended to moving hydraulic jumps by examination of Figure 10 (bottom left). Let us redefine the states denoted in the plot by (h', hu') and (h'', hu'') as left state (h_L, hu_L) and right state (h_R, hu_R), respectively. The linear Hugoniot must lie on the line depicted in magenta, with slope $\theta = (h_R - h_L)/(hu_R - hu_R)$ and can be parametrized in terms of x_S in (111). Hence, it can be expressed as

$$\breve{\Psi}(x_S) = \begin{pmatrix} h(x_S) \\ hu(x_S) \end{pmatrix}, \tag{113}$$

where $h(x_S) = x_S(h_R - h_L) + h_L$,

$$hu(x_S) = hu_L + \theta h(x_S) \tag{114}$$

and $x_S \in [0, 1]$. Note that parametrization $\breve{\Psi}(\xi)$ is straighfoward as $\xi = (h_R - h_L)x_S$.

Considering again the steady case described above and depicted in Figure 11, we can observe that the exact Hugoniot is neither linear nor monotone and ψ_2^1 has a global maxima hu_{max} at $h_{max} \in [h_L, h_R] \subset \mathbb{R}^+$ therefore, for any $h_M \in [h_L, h_R] \subset \mathbb{R}^+$, we have that $(hu)_M \geq (hu)_L = (hu)_R \equiv (hu)_{steady}$. This can be observed in Figure 11 (bottom-right), where a spike in the discharge appears.

5.3. Properties of the intermediate state in discrete Godunov-type solutions

Up to this point throughout this section, we have only considered exact solutions to the hydraulic jump. Theoretically, when considering the exact solution, the presence of an intermediate constant state $\mathbf{U}_M = (h_M, (hu)_M)$ is not stable, that is, it cannot be kept under steady conditions. The reason for this is that both jumps (left to middle and middle to right) have non-zero wave velocities of opposite sign, hence both jumps would converge to form a unique jump. This behavior, shown in Figure 12, is only present in the exact solution. On the other hand, when considering a discrete solution in a computational grid, both waves could be kept at a stationary position (at the cell interfaces of the intermediate cell) and the intermediate cell could keep the intermediate value in the steady regime. The reason for this is that the numerical fluxes at the interfaces of such a cell would coincide, that is

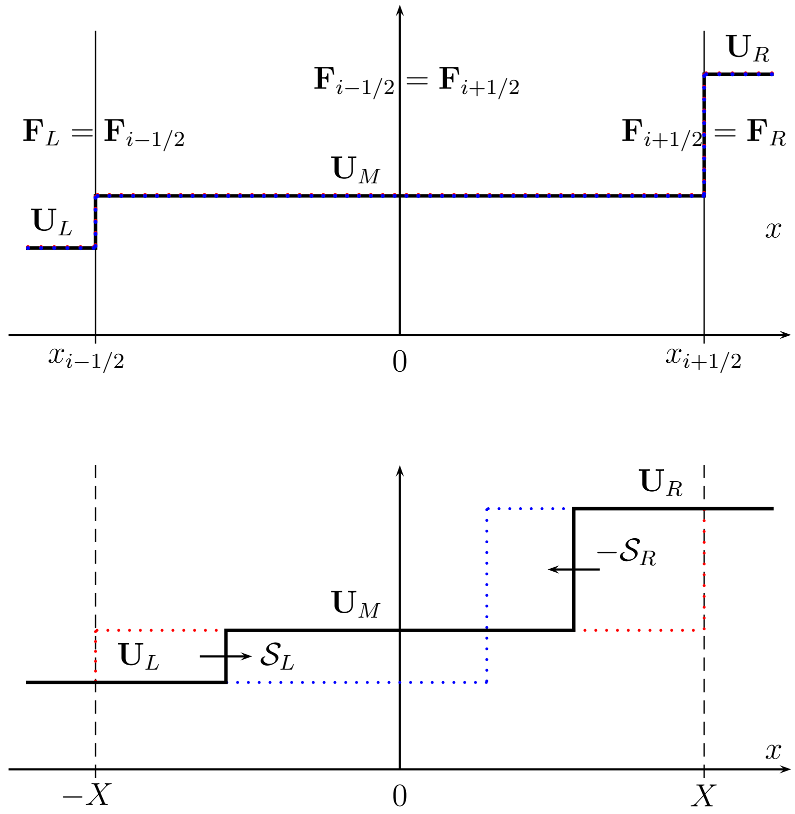

Figure 12: Initial condition considering an intermediate state (red), transient evolution of the discontinuities $\mathbf{U}_L$-$\mathbf{U}_M$ and $\mathbf{U}_M$-$\mathbf{U}_R$ (black) and final steady solution (blue).

$$\mathbf{F}^{-}_{i+1/2} = \mathbf{F}^{+}_{i-1/2}\,, \tag{115}$$

when considering the numerical resolution of the problem by means of FV Godunov's scheme in (44).

Figure 12 depicts the contrasting behavior of the 3-state hydraulic jump when considering the discrete (top) and exact (bottom) solution. The initial condition is represented by red dotted line, the final solution (when steadiness is achieved) is represented by blue dotted line and the solution at an arbitrary time before reaching the steady state is represented by black solid line. It can be observed that the initial condition is maintained in the discrete solution, where the intermediate state, $\mathbf{U}_M$, has been defined inside the cell $[x_{i-1/2}, x_{i+1/2}]$.

There is another important issue worth being mentioned. Only when the intermediate state coincides with the left or right states, the approximate solver would provide the exact solution. Hence, only when the shock position is located exactly at the interface, the approximate solver provides the exact solution [53, 54]. Moreover, it must be borne in mind that the intermediate state, $\mathbf{U}_M$, does depend on the Riemann solver used for the computation of the fluxes, and will only coincide with the value of $\mathbf{U}_M$ provided by the analytical Hugoniot locus when using an exact solver. A exhaustive comparison of the numerical performance in shock-capturing of different flux functions in the framework of Euler equations can be found in [55].

6. Flux fixes for the computation of the hydraulic jump

In this section, some spike-reduction numerical techniques based on flux interpolation are recalled and applied to the Shallow Water Equations (SWE). This idea of flux interpolation was first presented by Zaide and Roe [42], who proposed to find the fluxes in the untrustworthy intermediate cells by extrapolation from trustworthy neighbors and presented two new flux functions. The first one, named by the authors flux function A, was constructed based on the flux-wave approach, by computing the fluctuations in the interpolated fluxes across each wave. The second one, called flux function B, is based on the classical Roe solver and relies on conserved variables to determine the jumps across each wave and the contribution of each wave to the numerical flux. The authors claim that, by enforcing a linear shock structure and unambiguous sub-cell shock position, numerical shockwave anomalies are dramatically reduced.

Zaide and Roe [42] proposed to compute the fluxes in the intermediate cells by extrapolation from neighboring cells, hence a more general idea of a homogeneous flux function of the type $\mathbf{F}^\star_{i+1/2} = \mathbf{F}^\star_{i+1/2}(\mathbf{U}_{i-m}, ..., \mathbf{U}_{i-n})$ was introduced, rather than a Riemann solver that computes the numerical flux as $\mathbf{F}^\star_{i+1/2} = \mathbf{F}^\star_{i+1/2}(\mathbf{U}_i, \mathbf{U}_{i+1})$, with m and n two integer numbers. The authors in [42] outline that the conserved variables must be trusted since this is the only way to ensure conservation, however, the flux values should not be trusted.

Prior to the construction of the novel numerical fluxes $\mathbf{F}^\star_{i+1/2}$, physical fluxes (which are the cell centered fluxes, $\mathbf{F}_i$) are used to construct a novel approximation of the fluxes in every cell. Cell-centered fluxes, $\mathbf{F}_i$, are recomputed by means of extrapolation from neighboring cells. At every cell, the new flux is calculated as

$$\check{\mathbf{F}}_i = \frac{1}{2}(\mathbf{F}_{i+1} + \mathbf{F}_{i-1}) - \frac{1}{2}\tilde{\mathbf{J}}_{i-1,i+1}(\mathbf{U}_{i+1} - 2\mathbf{U}_i + \mathbf{U}_{i-1})\,, \tag{116}$$

with $\widetilde{\mathbf{J}}_{i-1,i+1} = \widetilde{\mathbf{J}}_{i-1,i+1}(\mathbf{U}_{i+1}, \mathbf{U}_{i-1})$ a Jacobian Roe's matrix,

$$\mathbf{F}_{i+1} - \mathbf{F}_{i-1} = \widetilde{\mathbf{J}}_{i-1,i+1}(\mathbf{U}_{i+1} - \mathbf{U}_{i-1}). \tag{117}$$

To construct those more general numerical fluxes, two alternatives, named flux function A and flux function B, are proposed in [42]. Such alternatives, as well as the traditional Roe flux, are detailed below:

- Traditional Roe homogeneous flux:

  The traditional Roe homogeneous flux (B.8) in Appendix B is used. It is constructed using Roe's matrix $\widetilde{\mathbf{J}}_{i+\frac{1}{2}}$,

  $$\mathbf{F}^{\star,Roe}_{i+1/2} = \frac{1}{2}\left(\mathbf{F}_i + \mathbf{F}_{i+1}\right) - \frac{1}{2} \mid \widetilde{\mathbf{J}}_{i+1/2} \mid \delta\mathbf{U}_{i+1/2}\,, \tag{118}$$

  evaluated conventionally as $\widetilde{\mathbf{J}}_{i+\frac{1}{2}} = \widetilde{\mathbf{J}}_{i+\frac{1}{2}}(\mathbf{U}_i, \mathbf{U}_{i+1})$.

- Flux function A:

  The extrapolated fluxes, $\check{\mathbf{F}}_i$, computed by (116), can be directly projected onto the Jacobian's eigenvectors basis and upwinded according to the propagation velocities of the Jacobian. The resulting numerical flux is constructed using (B.8), yielding [42]

  $$\mathbf{F}^{\star,A}_{i+1/2} = \frac{1}{2}\left(\check{\mathbf{F}}_i + \check{\mathbf{F}}_{i+1}\right) - \frac{1}{2}\mathrm{sgn}\left(\widetilde{\mathbf{J}}_{i+\frac{1}{2}}\right)\delta\check{\mathbf{F}}_{i+1/2}\,. \tag{119}$$

- Flux function B:

  This new flux function is computed by means of a novel Roe's matrix that spans a wider set of cells, instead of just the two cells at each side of the discontinuity. It reads [42]

  $$\mathbf{F}^{\star,B}_{i+1/2} = \frac{1}{2}\left(\check{\mathbf{F}}_i + \check{\mathbf{F}}_{i+1}\right) - \frac{1}{2} \mid \bar{\mathbf{J}}_{i+1/2} \mid \delta\mathbf{U}_{i+1/2}\,, \tag{120}$$

with $\bar{\mathbf{J}}_{i+1/2} = \bar{\mathbf{J}}_{i+1/2}(\mathbf{U}_{i-1}, \mathbf{U}_{i+2})$ Roe's matrix computed with cells $i-1$ and $i+2$.

6.1. Test case 2: assessment of flux functions A and B for the SWE

In order to test flux functions A and B in the framework of the SWE and compare their performance with the traditional homogeneous Roe flux, the following numerical experiment is proposed. It consists of a RP with initial data $h_L = 0.5$, $(hu)_L = 3$, $h_R = 1.6$ and $(hu)_R = 3.28787832816$, that generates a moving shock wave with speed $\mathcal{S} = 0.26171$. The computational domain is set to $[0, 450]$, with the discontinuity located at $x = 225$. Regarding the numerical discretization, the computational domain is divided in 900 cells of size $\Delta x = 0.5$ and the CFL number is set to 0.8. The simulation time is 25 s.

This test case is computed using the traditional Roe flux in (118) as well as the flux functions A and B in (119) and (120) respectively. The numerical solution for the discharge provided by such methods is plotted in space and time in Figure 13. Complementary results for the study of the spike in the numerical solution are presented in Figure 14, where the evolution in time of cell average values are depicted for the 8 leftmost cells on the right hand side of the RP (e.g. the first cell on the right of the initial discontinuity is depicted in blue, the second one in cyan and so on).

From figures 13 and 14, it is clearly evidenced that whereas the traditional Roe solver leads to a high spike in the discharge, which generates a shedding of spurious waves, when using the novel flux functions the spike is dramatically reduced and hence the shedding of such waves. A closer examination of the numerical results evidences that flux function A provides the best performance concerning the reduction of the spike, on the other hand, flux function B does also reduce this anomalous behavior at the cell where the shock is contained but still leaves a small spike behind it. This particularity of flux function B is clearly noticed in Figure 14 (bottom) where the spikes appear to be shifted to the left, which means that it occurs on the right side of the wavefront, as observed in Figure 13 (bottom).

In Figure 15 (left), the numerical solutions provided by the traditional Roe solver, the solver using flux function A and the solver using flux function B is depicted at $t = 25$ s in purple, green and magenta, respectively. It is observed that both the Roe flux and the flux A capture the exact position of the shock whereas the flux B underestimates the shock speed, hence providing a slightly shifted, though convergent, shock position.

The analysis of the properties of the novel flux functions from [42] can be completed by plotting the numerical results in the phase space. Figure 15 (right) shows the exact and approximate Hugoniot locus for the intermediate states between the left and right states of the RP. The exact Hugoniot locus is represented by a red continuous line, the approximate locus for the traditional Roe flux by purple dots, the approximate locus for flux function A by green dots and that for flux function B by magenta dots. As outlined in [42], the optimal locus that prevents the numerical solution from exhibiting any spike and spurious waves is the straight line between the left and right state. It can be observed in Figure 15 (right) that only flux function A achieves this requirement and therefore it is the preferred technique for the reduction of the spike in the SWE.

6.2. Extension of the flux function A to the SWE with source term

It is evidenced that flux function A is a better choice than B for the resolution of moving hydraulic jumps as it provides a better estimate of the shock speed. Previous numerical experiments do not include the presence of source terms, but most realistic cases are dominated by the action of those sources. In this section, the extension of flux function A to non-homogeneous equations is carried out by means of a suitable correction of the interpolation technique that ensures a virtually exact equilibrium between fluxes and source term. In addition to this, the numerical fluxes at the interfaces must be rewritten to account for the source term.

First, it is time to find out which is the most suitable correction of the flux extrapolation to reduce the spike of discharge in both transient and steady cases. Following a similar procedure than in [42], the idea is to find an approximation of such fluxes that ensures the exact equilibrium between fluxes and source term across cell interfaces under steady conditions, while keeping the idea of having an interpolated flux in the cell contanining the shock in order to prevent the scheme from using the equilibrium flux, which leads to the spike. To this end, it is first required to find the cell where the shock is contained. We propose to use Roe celerities, $\tilde{\lambda}^m$ to unequivocally locate such a cell, since it is known that both celerities at the left interface are positive (supercritical flow entering the cell) while a combination of celerities corresponding to subcritical conditions (one negative and the other one positive) is identified at the right interface.

Let us consider the cells, Ω_i, as single items contained in the domain Ω such that $\Omega = \{\Omega_i \mid i \in [1, ..., N]\}$. Considering the possibility of multiple

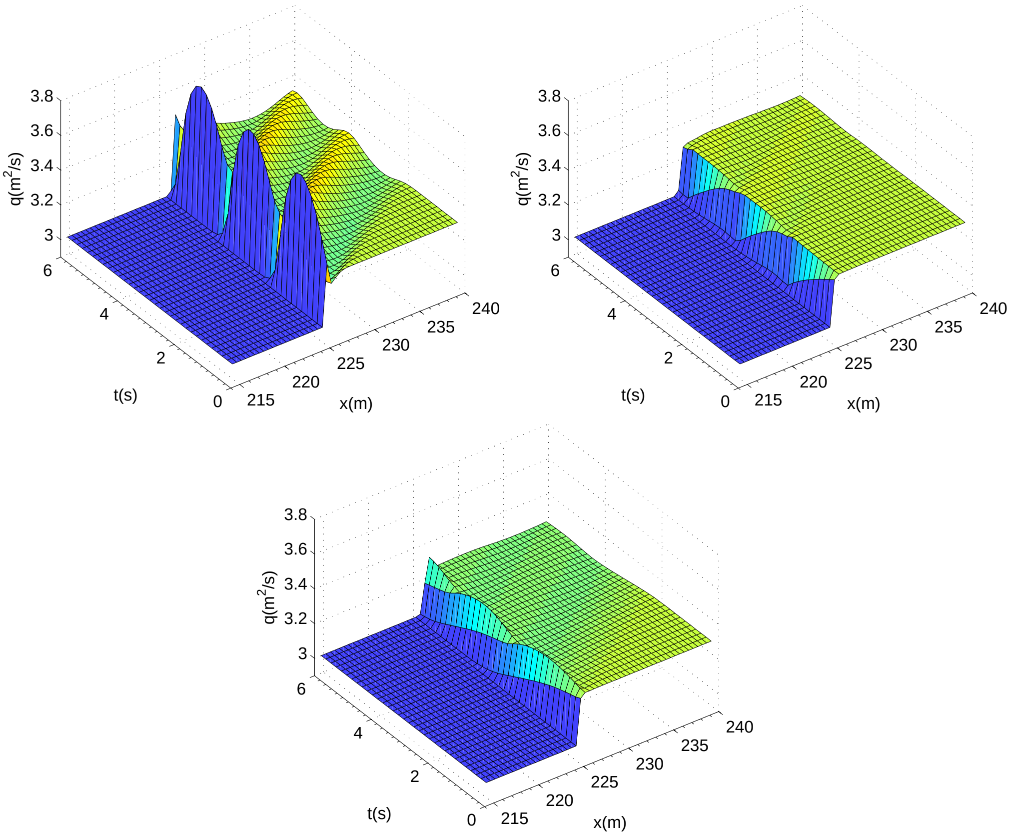

Figure 13: Test case 2. Numerical solution provided by the traditional Roe solver (top-left) as well as the flux functions A (top-right) and B (bottom) proposed in [42] within the time interval $[0, 6]$ s.

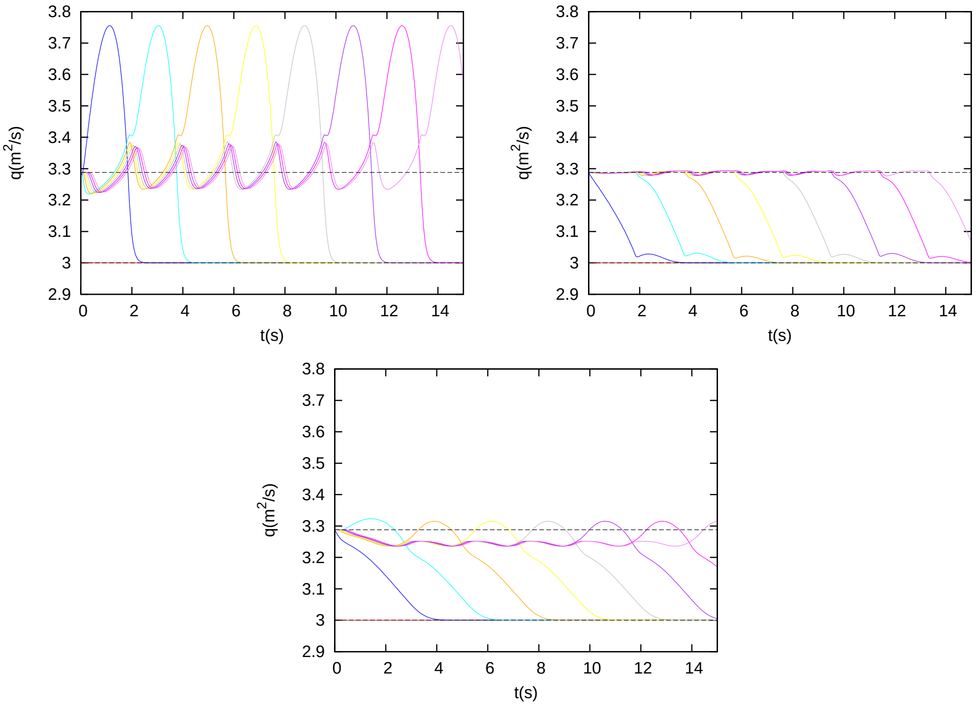

Figure 14: Test case 2. Evolution in time of cell average values for the 8 leftmost cells on the right hand side of the RP using the Roe flux (top-left), flux function A (top-right) and flux function B (bottom).

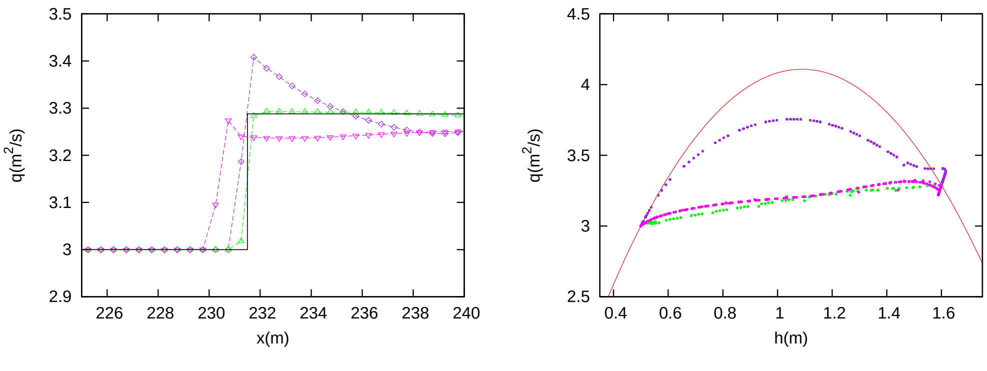

Figure 15: Test case 2. Left: numerical solution using the Roe flux (—◇—), flux function A (—△—) and flux function B (—▽—) at $t = 25$ s. Right: exact Hugoniot locus and approximate locus for the Roe flux, flux function A and flux function B.

hydraulic jumps within the domain, we denote the set of cells containing a positive-flow hydraulic jump as

$$\mathcal{D}^+ = \left\{ \Omega_i \mid \Omega_i \in \Omega \wedge \tilde{\lambda}^1_{i-1/2} \cdot \tilde{\lambda}^1_{i+1/2} < 0 \wedge h_{i-1} < h_{i+1} \right\} \tag{121}$$

and the set of cells containing a negative-flow hydraulic jump as

$$\mathcal{D}^- = \left\{ \Omega_i \mid \Omega_i \in \Omega \wedge \tilde{\lambda}^2_{i-1/2} \cdot \tilde{\lambda}^2_{i+1/2} < 0 \wedge h_{i-1} > h_{i+1} \right\} . \tag{122}$$

Once the hydraulic jumps are found, the following cell-centered fluxes are proposed in order to generate an spike fix

$$\hat{\mathbf{F}}_i = \left\{ \begin{array}{cl} \mathbf{F}_i & \text{if} \quad \Omega_i \notin \mathcal{D}^+ \cup \mathcal{D}^- \\ \check{\mathbf{F}}_i - (1 - x_{\mathcal{S},i}) \bar{\mathbf{S}}_{i-1,i+1} + \bar{\mathbf{S}}_{i-1/2} & \text{if} \quad \Omega_i \in \mathcal{D}^+ \cup \mathcal{D}^- \end{array} \right. \tag{123}$$

with $\check{\mathbf{F}}_i$ the interpolated flux in (116), $\bar{\mathbf{S}}_{i-1,i+1}$ a centered integral of the source term, that can be computed computed as

$$\bar{\mathbf{S}}_{i-1,i+1} = \left(\begin{array}{c} 0 \\ -g \frac{h_{i-1}+h_{i+1}}{2} (z_{i+1} - z_{i-1}) \end{array} \right) , \tag{124}$$

$\bar{\mathbf{S}}_{i-1/2}$ the integral of the source term across the left interface, that can be computed as

$$\bar{\mathbf{S}}_{i-1/2} = \left(\begin{array}{c} 0 \\ -g \frac{h_{i-1}+h_i}{2} (z_i - z_{i-1}) \end{array} \right) . \tag{125}$$

Parameter $x_{\mathcal{S},i}$ accounts for the normalized position of the shock inside the cell, here approximated by

$$x_{\mathcal{S},i} = \frac{h_i - h_{i+1}}{h_{i-1} - h_{i+1}} , \tag{126}$$

if considering that the intermediate state is a linear combination of the left and right states (linear Hugoniot)

$$\mathbf{U}_i = x_{\mathcal{S},i} \mathbf{U}_{i-1} + (1 - x_{\mathcal{S},i}) \mathbf{U}_{i+1} , \tag{127}$$

where $\mathbf{U}_{i-1}$, $\mathbf{U}_i$ and $\mathbf{U}_{i+1}$ are any arbitrary left, middle and right states defining a hydraulic jump as depicted in Figure 12.

It is worth pointing out that the corrected flux in (123) provides an approximation of the cell-centered flux in the shock cell that converges to the exact steady flux, unlike traditional methods, that only converge to an equilibrium flux (different to the exact flux) that allows the steadiness of the solution. The reason why the proposed technique does not always ensure the exact flux with independence of the grid is due to the assumption we make for the definition of (123): the intermediate state (at cell Ω_i where the shock is located) lies on a linear Hugoniot between the left and right states, according to (127), which is not completely true under the presence of a bed step source term. The exact linear Hugoniot would be expressed instead as

$$\mathbf{U}_i = x_{\mathcal{S},i}\mathbf{U}_i^- + (1 - x_{\mathcal{S},i})\mathbf{U}_i^+ \,, \tag{128}$$

where $\mathbf{U}_i^-$ and $\mathbf{U}_i^+$ are the left and right intermediate states at the interfaces of cell Ω_i. In spite of this, the approximation in (127) provides a trustworthy approximation of the shock position when solving for $x_{\mathcal{S},i}$ and what is of most importance, it converges to the exact position as the grid is refined, when dealing with a smooth bed topography.

It is straightforward to show that (123) provides the exact flux under steady conditions by considering the shock located at cell Ω_M and applying steady state conditions to the second equation of (123), as follows

$$\hat{\mathbf{F}}_i = \frac{1}{2}(\mathbf{F}_{i-1}+\mathbf{F}_{i+1})-\frac{1}{2}\tilde{\mathbf{J}}_{i-1,i+1}(\mathbf{U}_{i+1}-2\mathbf{U}_i+\mathbf{U}_{i-1})-(1-x_{\mathcal{S},i})\bar{\mathbf{S}}_{i-1,i+1}+\bar{\mathbf{S}}_{i-1/2} \,, \tag{129}$$

where substitution of $\mathbf{U}_i$ using (127) yields

$$\hat{\mathbf{F}}_i = \frac{1}{2}(\mathbf{F}_{i-1}+\mathbf{F}_{i+1})+\frac{1}{2}(1-2x_{\mathcal{S},i})\tilde{\mathbf{J}}_{i-1,i+1}(\mathbf{U}_{i+1}-\mathbf{U}_{i-1})-(1-x_{\mathcal{S},i})\bar{\mathbf{S}}_{i-1,i+1}+\bar{\mathbf{S}}_{i-1/2} \,. \tag{130}$$

From the definition of Roe's Jacobian matrix, we know that $\tilde{\mathbf{J}}_{i-1,i+1}(\mathbf{U}_{i+1} - \mathbf{U}_{i-1}) = \mathbf{F}_{i+1} - \mathbf{F}_{i-1}$ and under steady conditions $\mathbf{F}_{i+1} - \mathbf{F}_{i-1} = \bar{\mathbf{S}}_{i-1,i+1}$. Substitution of this term into (130) reads

$$\hat{\mathbf{F}}_i = \frac{1}{2}(\mathbf{F}_{i-1}+\mathbf{F}_{i+1})+\frac{1}{2}(1-2x_{\mathcal{S},i})\bar{\mathbf{S}}_{i-1,i+1}-(1-x_{\mathcal{S},i})\bar{\mathbf{S}}_{i-1,i+1}+\bar{\mathbf{S}}_{i-1/2} \,, \tag{131}$$

Now, making use of $\mathbf{F}_{i+1} - \mathbf{F}_{i-1} = \bar{\mathbf{S}}_{i-1,i+1}$ again, it does lead to

$$\hat{\mathbf{F}}_i - \mathbf{F}_{i-1} = \bar{\mathbf{S}}_{i-1/2}\,, \tag{132}$$

the GRH condition.

Finally, the expression for the numerical fluxes at cell interfaces is presented. Using definitions in Section Appendix A, we can write the non-homogeneous version of the numerical flux in (119) to account for the contribution of the source term as

$$\begin{aligned} \mathbf{F}^-_{i+1/2} &= \hat{\mathbf{F}}_i + \sum_{m=1}^{I} [(\hat{\gamma} - \beta)\widetilde{\mathbf{e}}]^m_{i+\frac{1}{2}}\ , \\ \mathbf{F}^+_{i+1/2} &= \hat{\mathbf{F}}_{i+1} - \sum_{m=I+1}^{N_\lambda} [(\hat{\gamma} - \beta)\widetilde{\mathbf{e}}]^m_{i+\frac{1}{2}}\ . \end{aligned} \tag{133}$$

where $\hat{\gamma}$ are the components of $\hat{\mathbf{\Gamma}}_{i+1/2} = \widetilde{\mathbf{P}}^{-1}_{i+1/2}\delta\hat{\mathbf{F}}_{i+1/2}$, the projection of the jump in the extrapolated fluxes across cell interfaces, $\hat{\mathbf{F}}_{i+1/2} = \hat{\mathbf{F}}_{i+1} - \hat{\mathbf{F}}_i$.

6.3. Test case 3: Steady jump over smoothly varying bed profile

In this test case, steady solutions for the flow over the following bed elevation profile

$$z(x) = \begin{cases} 0 & \text{if} \quad x < 8 \\ 0.05(x-8) & \text{if} \quad 8 \leq x \leq 12 \\ 0.2 - 0.05\,(x-12)^2 & \text{if} \quad 12 \leq x \leq 14 \\ 0 & \text{if} \quad x > 12 \end{cases} \tag{134}$$

are computed using the proposed technique. The computational domain is $[0, 20]$ and the solution is computed for $t = 400$ s. CFL number is set to 0.45 for all cases and the computational domain is discretized in 100 cells. The discharge is imposed to 0.6 m^2/s upstream to obtain the sonic point at the cell with the maximum bed elevation, that is $z_{max} = 0.2$. Downstream, the water depth is also imposed in order to generate the hydraulic jump. Different values for h downstream, are chosen to generate the jump at different locations and assess the performance of the proposed scheme. The complete configuration of boundary conditions is presented in Table 2.

Numerical results provided by the novel scheme are presented for test case 1.A in Figure 16 (top) and compared with the results provided by the traditional Roe solver, depicted in Figure 16 (bottom). No differences can be noticed when considering the solution for the water surface elevation, but it is clearly evidenced that the spike in the solution for the discharge at the cell where the shock is located is strongly reduced when using the novel numerical technique.

Case	$q_{BC:left}(\mathrm{m}^2/\mathrm{s})$	$h_{BC:right}(\mathrm{m})$	Shock position (m)	$x_{\mathcal{S}}$
1.A	0.6	0.6185	13.298	0.01
1.B	0.6	0.6200	13.278	0.11
1.C	0.6	0.6220	13.252	0.24
1.D	0.6	0.6256	13.201	0.495
1.E	0.6	0.6280	13.166	0.67
1.F	0.6	0.6300	13.135	0.825
1.G	0.6	0.6320	13.102	0.99

Table 2: Different boundary condition configurations for Test case 3.

To study the behavior of this spike, the solution for the discharge in the shock cell is depicted for tests cases 1.A-1.G in Figure (17) (left). In this plot, the value of discharge against the normalized shock position has been depicted for the results provided by the traditional Roe solver as well as the modified solver using flux interpolation in [42] and the proposed technique. It can be observed that the method in [42] already helps decreasing the spike of discharge but only when including the correction term, as done in the novel method, the spike is virtually reduced to zero.

As outlined before, the proposed scheme does not always provide the exact discharge in the shock cell, however, the numerical estimate of the discharge in this cell converges to the exact value as the grid is refined. This property is of utmost importance, as the novel scheme can be considered L_1, L_2 and L_∞ convergent, while previous schemes were not able to converge when regarding L_∞ error norm. Convergence rate results for L_∞ error norm are presented in Figure 17 (right) for the traditional Roe solver and for the proposed scheme. The convergence rate test has been carried out for case 1.D using four different grids, composed of 100, 200, 400 and 800 cells. It is worth mentioning that the grid is shifted in order to keep a constant distance between the exact position of the jump and the right cell interface. It is clearly evidenced that the proposed technique allows the scheme to converge to the exact solution as the grid is refined, unlike the traditional Roe solver that does not exhibit any convergence with grid refinement because

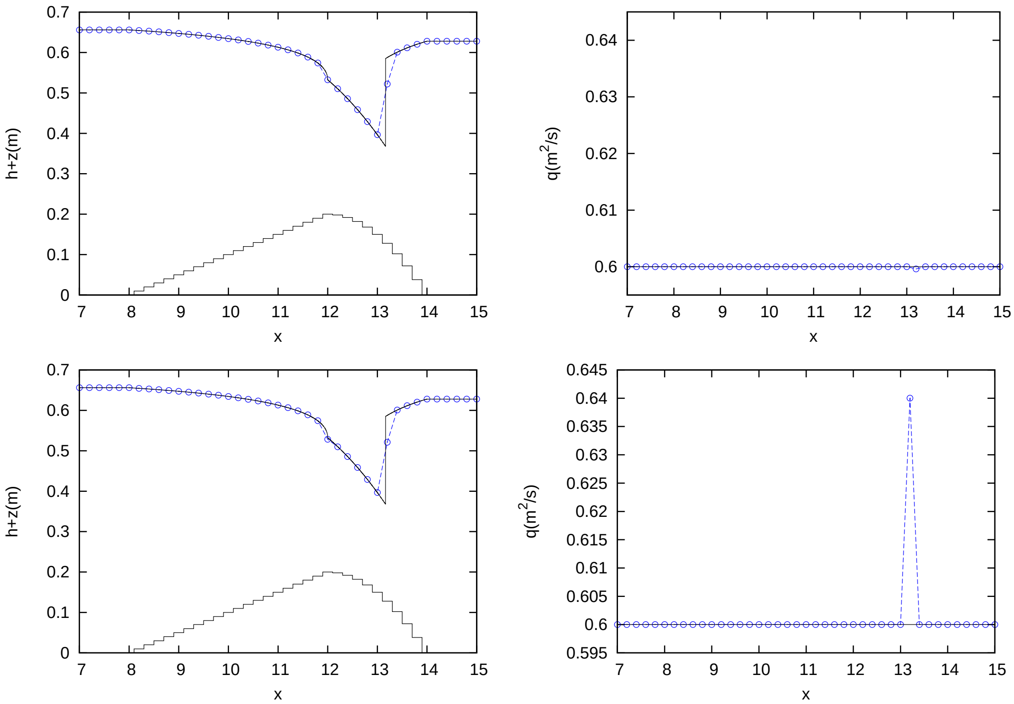

Figure 16: Test case 3. Numerical results for $h + z$ (left) and q (right) provided by the proposed spike-reducing method (top) and by the traditional Roe solver (bottom), compared to the exact solution, using 100 cells and CFL=0.45.

the equilibrium discharge at the shock cell is always different than the exact discharge when the shock is not located at cell interfaces.

6.4. Test case 4: Traveling jump over different bed profiles

In this test case, traveling shock waves over different bed elevation profiles $z(x)$ are computed. For all bed profiles, the maximum bed elevation is $z_{max} = 0.2$ m and the bed elevation at the boundaries is zero. To construct a solution consisting of a single jump traveling across the domain, we first compute a steady transcritical solution over the bed profile by imposing a constant discharge upstream of $q = 0.6$ m^2/s. When the steady regime is reached, the boundary condition upstream is redefined, imposing now $q = 0.556749458405104$ m^2/s and $h = 0.12$ m, which generates a supercritical state that is connected with the original subcritical state by means of a traveling hydraulic jump, according to the Hugoniot locus. The computational domain is $[0, 560]$ and the solution is computed at $t = 610$ s. The CFL

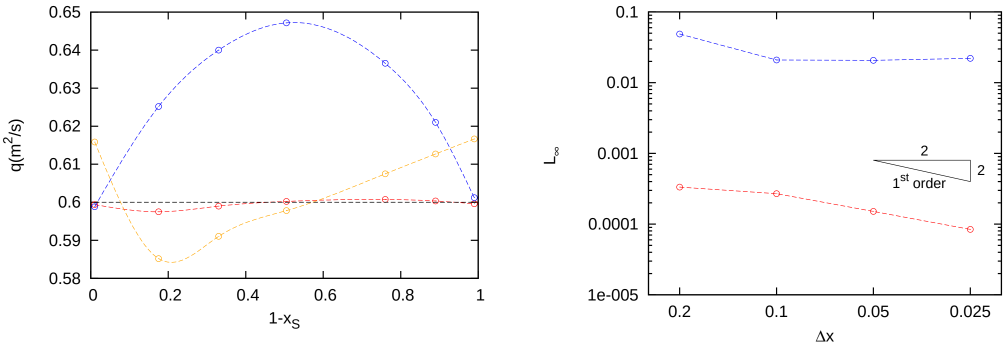

Figure 17: Test case 3. Left: representation of the spike of discharge against the position of the shock within the cell for the traditional Roe flux (− ○ −), for the method using the interpolated flux in [42] (− ○ −) and for the proposed spike-reducing method (− ○ −), using 100 cells and CFL=0.45. Right: convergence rate test for the traditional Roe method (− ○ −) and for the proposed method (− ○ −), using CFL=0.45.

number is set to 0.45 and the domain is discretized in 140 computational cells.

The bed profile will be constructed as

$$z(x) = \begin{cases} \frac{0.2}{276}(x-4)+g(x) & \text{if} \quad 4 \leq x < 280 \\ 0.2 - \frac{0.2}{276}(x-280) & \text{if} \quad 280 \leq x \leq 556 \\ 0 & \qquad \text{otherwise} \end{cases} \tag{135}$$

where $g(x)$ is an additional geometric function that allows to make variations in the basic constant slope profile (when $g(x) = 0$). Three different bed slopes are defined:

- *Constant slope (Test 4.1):* The first test is carried out over a constant slope profile, setting $g(x) = 0$ in (135).

- *Sinusoidal variations in a constant slope (Test 4.2):* Now, a sinusoidal variation is added to (135) by means of

$$g(x) = \begin{cases} 0.02\sin(0.04\pi(x-12)) & \text{if} \quad 12 \leq x < 212 \\ 0 & \qquad \text{otherwise} \end{cases} \tag{136}$$

- *Discontinuities in the constant slope (Test 4.3):* Here, some discontinuities are added to (135) by means of

$$g(x) = \begin{cases} 0.02 & \text{if} \quad 12 \leq x < 32 \\ -0.02 & \text{if} \quad 32 \leq x < 52 \\ 0.04 & \text{if} \quad 52 \leq x < 72 \\ -0.04 & \text{if} \quad 72 \leq x < 92 \\ 0 & \quad\quad \text{otherwise} \end{cases} \tag{137}$$

Numerical results for tests 4.1, 4.2 and 4.3 are presented in Figures 18, 19, 20 and 21. Figure 18 shows the numerical solution at $t = 610$ s for the water surface elevation and discharge provided by the ARoe scheme and by the proposed spike-reducing method in Section 6.2. For all the test, the SEBF discretization of the source term is chosen. In the figures mentioned above, major differences are observed in the solution of the discharge, which is much more oscillatory when computed by the ARoe method. On the other hand, differences on the water surface elevation are less sensitive to the spike. A space-time representation of the numerical discharge is presented in Figure 19, where the elimination of post-shock oscillations can be observed. In Figure 20, the numerical solution for the water surface elevation and discharge inside the cell with maximum bed elevation (cell 71) is plotted in time, showing that the proposed spike-reducing scheme performs adequately with independence of the bed profile, as it prevents the solution from generating oscillations. On the other hand, the numerical solution computed by means of the traditional ARoe scheme shows the oscillations produced by the spike, which travel downwards at a higher speed than the hydraulic jump. In order to carry out an exhaustive analysis on the spike reducing effect of the proposed method, the evolution in time of the numerical solution for the discharge in cells 2 to 11, computed by means of the aforementioned schemes, is plotted in Figure 21. It is evidenced that the numerical solution provided by the proposed scheme completely reduces the spike and only leaves very small peaks that are virtually bounded by the values of the discharge at each side of the shock, hence they are not of any relevance.

6.5. Test case 5: Interaction of two jumps over a smooth bed profile

In this case, two hydraulic jumps moving in opposite directions are introduced in a steady transcritical flow over the bed profile in (134), inside the domain $[0, 20]$. The initial condition corresponds to the steady solution generated when setting $q = 0.6$ m^2/s upstream in most part of the domain, and also includes the two jumps as

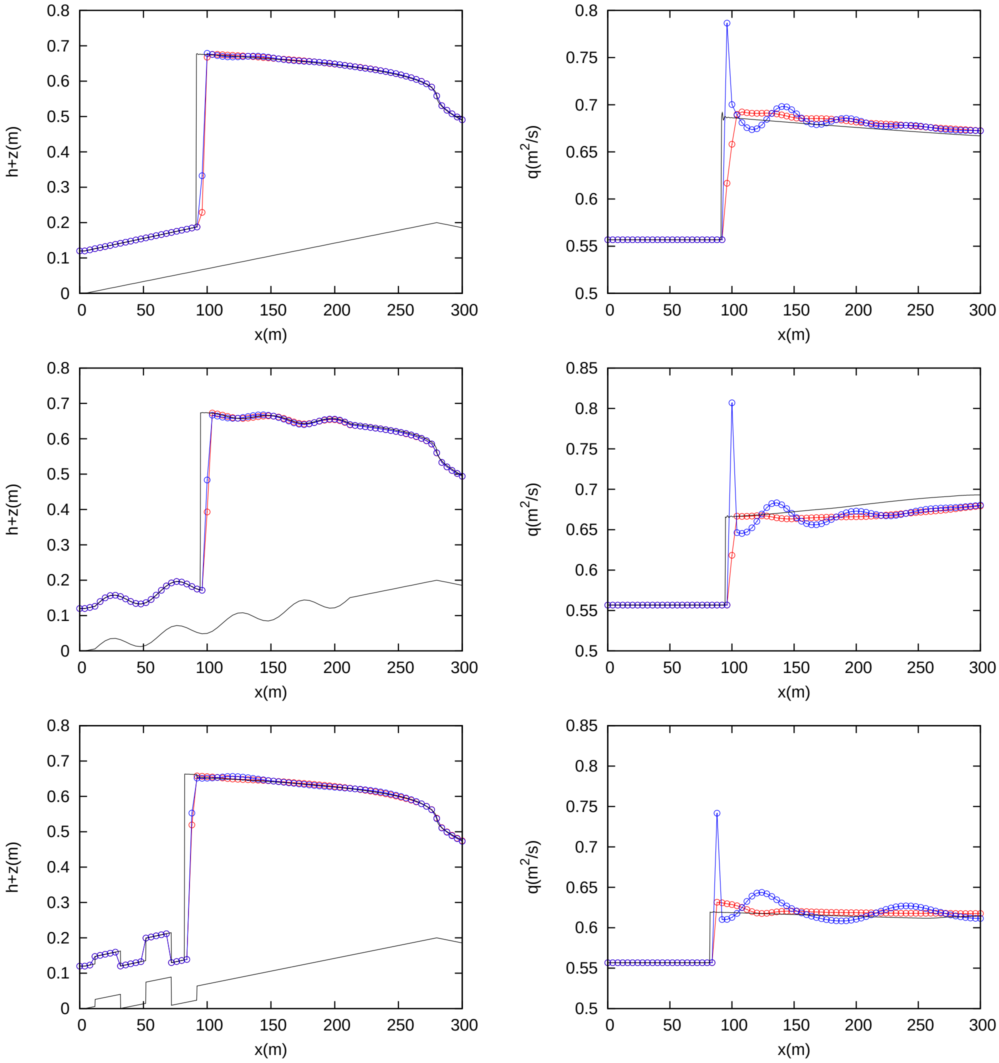

Figure 18: Test case 4. Numerical solution at $t = 610$ s for the water surface elevation (left) and discharge (right) provided by the traditional Roe flux ($-\circ-$) and by the proposed spike-reducing method ($-\circ-$), using 140 cells and CFL=0.45.

$$\mathbf{U}(x) = \begin{cases} \mathbf{U}_{in} & \text{if} \quad 0 \leq x \leq 1 \\ \mathbf{U}_{s} & \text{if} \quad 1 < x < 17 \\ \mathbf{U}_{out} & \text{if} \quad 17 \leq x \leq 20 \end{cases} \tag{138}$$

1289 where $\mathbf{U}_s$ is the steady energy-conservative solution with $q = 0.6 \text{ m}^2/\text{s}$, $\mathbf{U}_{in} =$

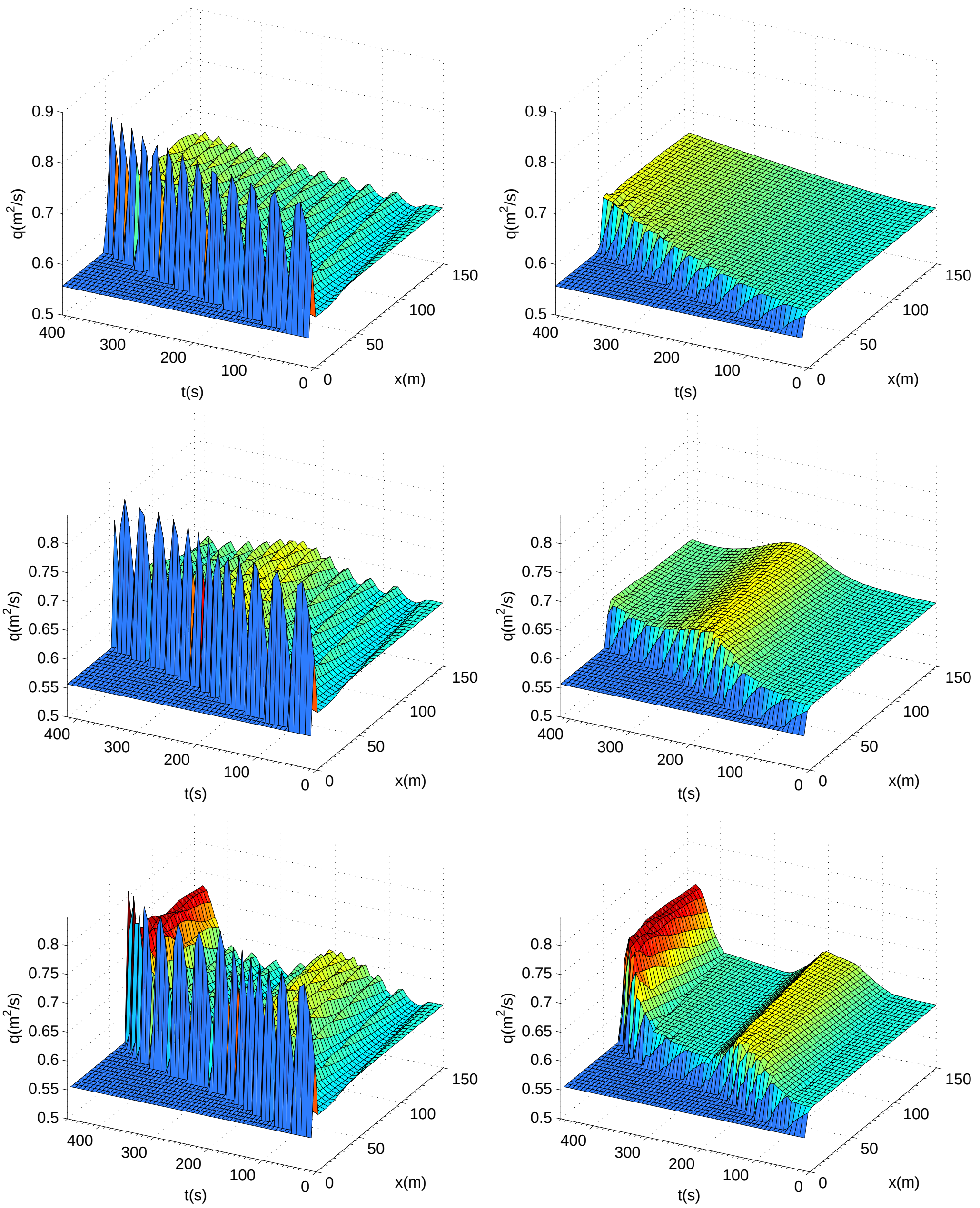

Figure 19: Test case 4. Space-time representation of the numerical discharge provided by the traditional Roe flux (left) and by the proposed spike-reducing method (right), using 140 cells and CFL=0.45.

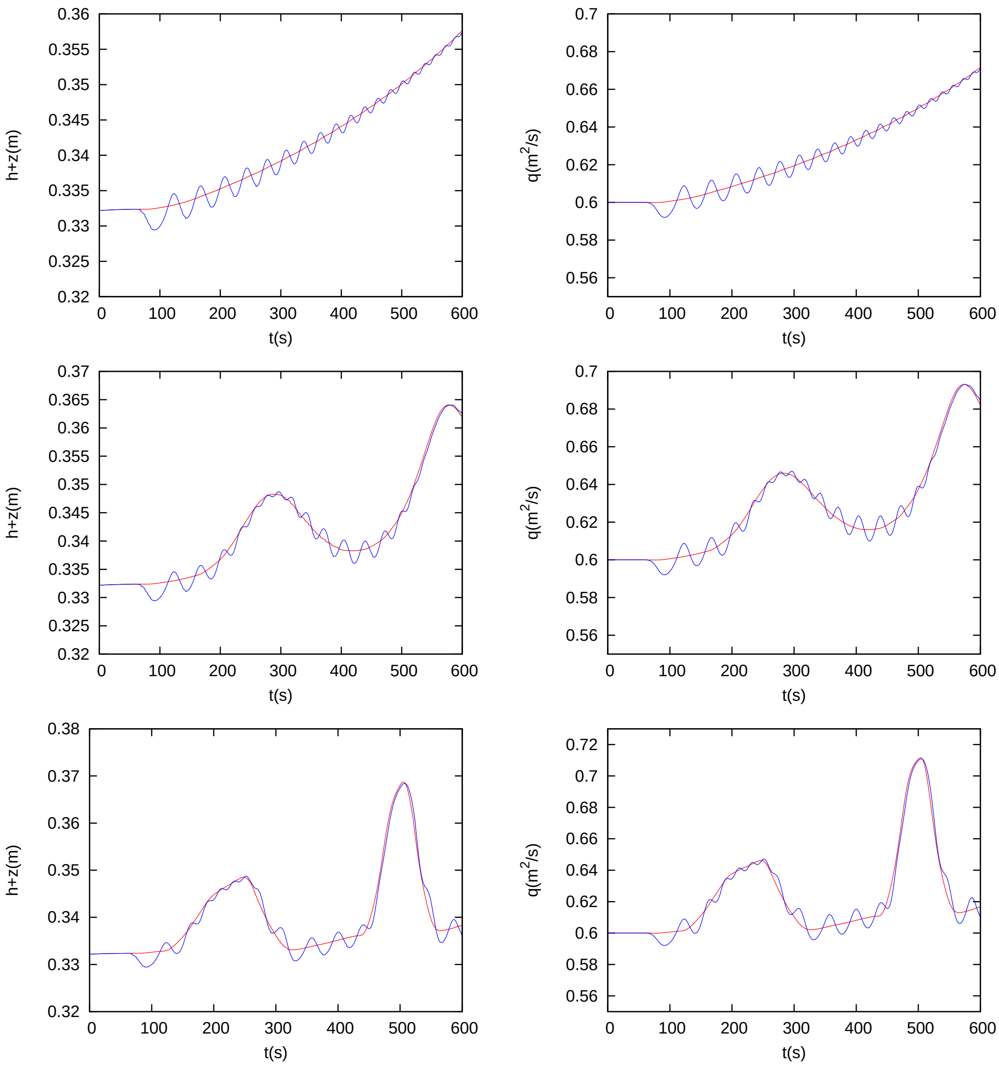

Figure 20: Test case 4. Evolution in time of the numerical solution for the water surface elevation (left) and discharge (right) in the cell with initial $Fr = 1$ (cell 71) provided by the traditional Roe flux (–) and by the proposed spike-reducing method (–), using 140 cells and CFL=0.45.

(h_{in}, q_{in}) and $\mathbf{U}_{out} = (h_{out}, q_{out})$, with $h_{in} = 0.12$ m, $q_{in} = 0.556749458405104$ m^2/s, $h_{out} = 0.62$ m and $q_{out} = 0.410276289759429$ m^2/s

In order to maintain the hydraulic jumps, the boundary conditions are set supercritical upstream and subcritical downstream, hence we impose $h = h_{in}$

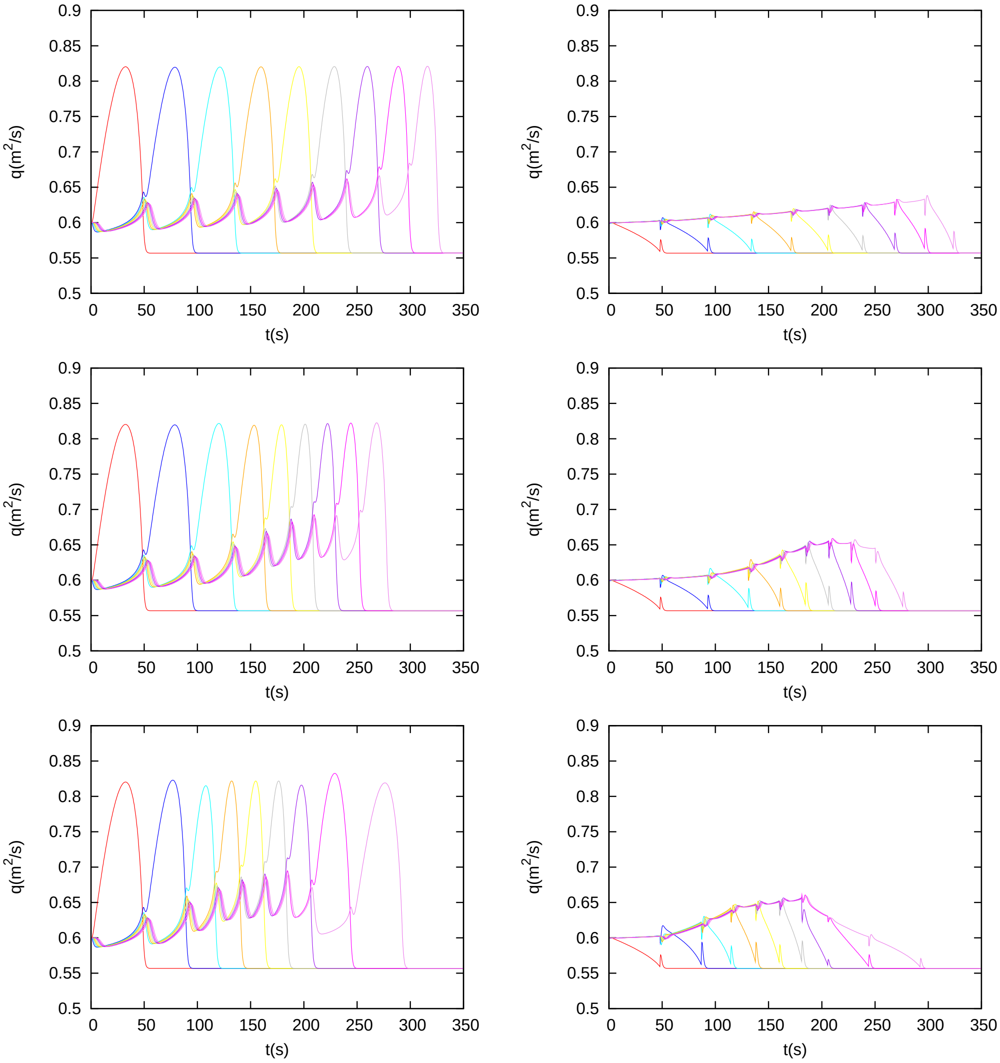

Figure 21: Test case 4. Evolution in time of the numerical solution for the discharge inside cells 2 to 11 provided by the traditional Roe flux (left plot) and by the proposed spike-reducing method (right plot), using 140 cells and CFL=0.45.

and $q = q_{in}$ upstream and $h = h_{out}$ downstream. For this test case, we set $\Delta x = 0.2$ and $\Delta x = 0.1$ m and CFL=0.45. As time goes forward, the left-moving shock on the right decelerates and eventually stops, as the thrust exerted by the bed slope is sufficiently large for it. On the other hand, the right-moving shock on the left does not stop and continuously moves along the domain. In most part of this simulation, the aforementioned shock moves over a flat bottom.

The numerical solution computed by the ARoe scheme and the proposed spike-reducing method are presented in Figures 22 and 23, for grid sizes $\Delta x = 0.2$ and $\Delta x = 0.1$ m respectively. The top plots show the solution for the water surface elevation and discharge at $t = 70$ s and the bottom plots show the evolution in time of such quantities inside the cell where the right jump stops and remains steady. It is observed that the spike-reducing method provides a numerical solution much closer to the reference solution as no shedding of spurious oscillation occurs, unlike the traditional Roe scheme that is unable to avoid those oscillations. It is also observed that oscillations are barely reduced with mesh refinement. This is because the spike is still present, as the approximate Hugoniot locus of the Roe solver does not depend on the discretization (the hydraulic jump is still produced between the same left and right states). This means that only the spike-reducing method can ensure convergence with mesh refinement.

7. Conclusions

This work focuses on the study and design of efficient and robust numerical schemes for the computation of hyperbolic conservation laws with source terms, with application to the SWE. The goal of the methods proposed here is to overcome some present difficulties that have been well documented in previous literature, such as the exact conservation of the discrete energy (when necessary), the accurate positioning of steady shockwaves and the reduction of the numerical shockwave anomalies arising from slowly-moving shocks, among others.

Regarding the conservation of energy in the numerical solution of the Shallow Water Equations (SWE), we carry out a theoretical study on the relations among variables across the bed step contact wave, showing that the conservation of energy can be ensured by imposing conservation of the Riemann invariants associated to this wave, or in other words, making the Generalized Hugoniot locus (GHL) and the Integral Curve (IC) coincide. We consider then the design of a suitable source term discretization (STD) that ensures the conservation of energy, showing that the WEBF [25] can be derived from these assumptions under the conditions of steady state. The WEBF has proven a good performance in a variety of situations, however,

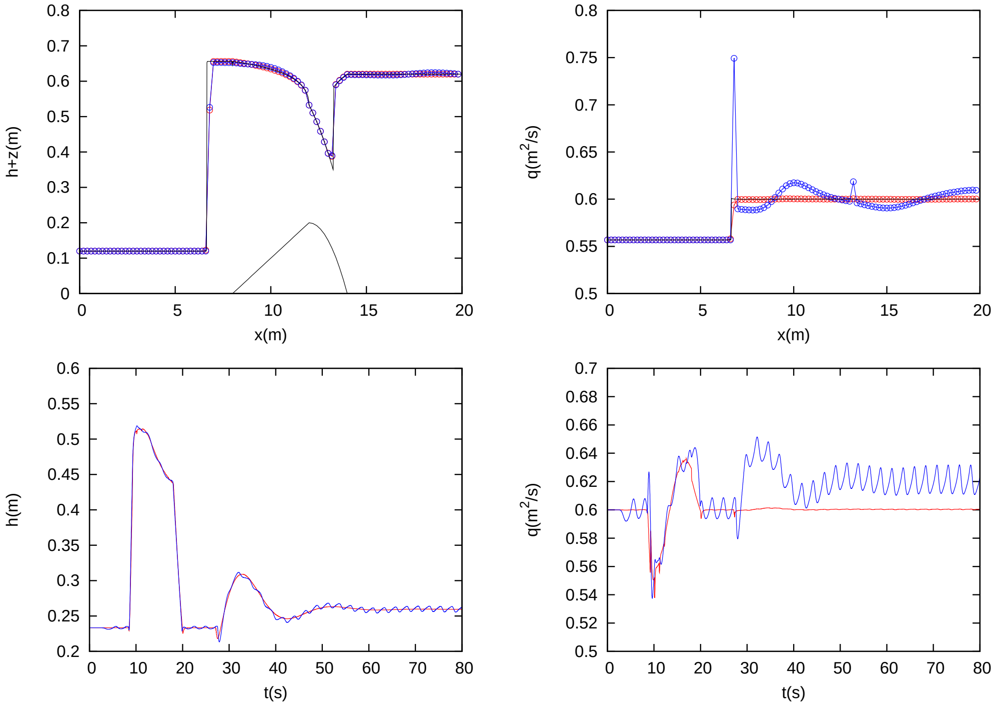

Figure 22: Test case 5. Top: Numerical solution at $t = 70$ s for the water surface elevation (left) and discharge (right) provided by the traditional Roe flux (—○—) and by the proposed spike-reducing method (—○—). Bottom: Numerical solution inside cell containing the right jump for the water depth (left) and discharge (right), provided by the traditional Roe flux (—) and by the proposed spike-reducing method (—). Grid size is set to $\Delta x = 0.2$.

when using it for the computation of hydraulic jumps, it is not able to provide an accurate positioning of the discontinuity.

To address the aforementioned issues of shock positioning, a novel discretization of the source term that ensures the exact conservation of the discrete energy while capturing the exact position of the hydraulic jump is proposed. This technique allows to unequivocally identify the position of hydraulic jumps and dissipate the exact amount of energy across them. It is referred to as selective energy balanced formulation (SEBF) of the integral of the source term and can be applied to the ARoe and HLLS solvers, and their high order versions.

Numerical shockwave anomalies in the framework of the SWE, particularly the so-called slowly-moving shock anomalies, are also considered in this work. Following the approach in [42], we propose a novel spike-reducing

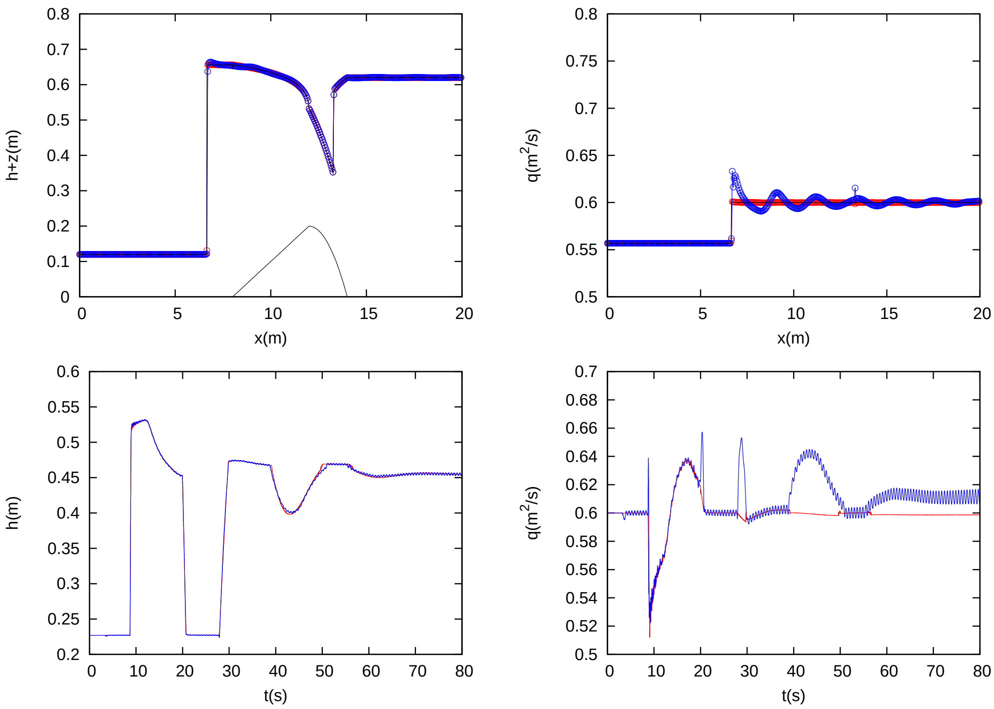

Figure 23: Test case 5. Top: Numerical solution at $t = 70$ s for the water surface elevation (left) and discharge (right) provided by the traditional Roe flux (—○—) and by the proposed spike-reducing method (—○—). Bottom: Numerical solution inside cell containing the right jump for the water depth (left) and discharge (right), provided by the traditional Roe flux (—) and by the proposed spike-reducing method (—). Grid size is set to $\Delta x = 0.1$.

flux function for the SWE with varying bed. To this end, we first study the problem of slowly-moving shocks in the SWE and notice that they are only produced when dealing with hydraulic jumps. A complete description of such kind of waves is provided and a thorough study on the shock structure, comparing exact and Godunov type solutions, is carried out by using the phase space representation. Moreover, prior to the presentation of the proposed technique, flux functions A and B in [42] are assessed for the computation of moving hydraulic jumps over flat bed, evidencing a strong reduction of the spike when using such methods.

The novel spike-reducing flux proposed in this work is computed in the same way than function A [42], but with two main differences. First, a modified flux interpolation technique is carried out in order to account for the contribution of the source. Second, the novel flux function includes the source strengths across each wave as done in the ARoe solver in [25]. Here we propose to modify the interpolation in [42] by means of a correction term that leads to the exact balance between sources and fluxes in the steady state. This spike fix is based on the hypothesis that the intermediate state should lie on a linear Hugoniot that connects the left and right states, which is not completely general, specially for large discontinuities in the bed elevation, but still leads to satisfactory numerical results for any practical purpose.

The proposed technique is assessed in a variety of situations, including steady and transient cases, over continuous and discontinuous bed. Numerical results evidence that the spike is dramatically reduced to a point where the shedding of spurious waves is virtually not noticeable and also that the proposed scheme leads to a convergent numerical solution because the size of the spike can now be reduced with mesh refinement. For the numerical tests presented in this work, the new scheme does not impose additional stability restrictions and the numerical solution is stable for any CFL number below the traditional bound of 1.0. Numerical results for steady cases with hydraulic jumps are presented, proving that the proposed scheme leads to a convergent solution, even when measured with L_∞ error norm.

Appendix A. The ARoe solver for systems of N_λ waves

Depending on the nature of the source term, a centered integration of this term may prevent the numerical scheme from preserving the exact balance between fluxes and sources under steady state. This is the case of the so-called geometric source terms, described in (3). In this case, the so-called augmented Riemann solvers are of application for the resolution of the RP, providing an approximation of the numerical fluxes that includes the contribution of the source term. Numerical fluxes can be generally expressed as $\mathbf{F}^-_{i+\frac{1}{2}} = \mathbf{F}^-_{i+\frac{1}{2}}(\mathbf{U}^n_i, \mathbf{U}^n_{i+1}; \bar{\mathbf{S}}_{i+1/2})$, $\mathbf{F}^+_{i-\frac{1}{2}} = \mathbf{F}^+_{i-\frac{1}{2}}(\mathbf{U}^n_{i-1}, \mathbf{U}^n_i; \bar{\mathbf{S}}_{i-1/2})$, where $\bar{\mathbf{S}}_{i+1/2}$ is a suitable approximation of the integral of the source term across the cell edge.

Riemann Problems are defined at each interface, as depicted in Figure A.24, as

$$
\mathrm{RP}(\mathbf{U}_i, \mathbf{U}_{i+1}) : \begin{cases} \dfrac{\partial \mathbf{U}}{\partial t} + \dfrac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = \mathbf{S} \\ \\ \mathbf{U}(x,0) = \begin{cases} \mathbf{U}_i & x < 0 \\ \mathbf{U}_{i+1} & x > 0 \end{cases} \end{cases} \tag{A.1}
$$

It is worth mentioning that, for each RP, spatial and temporal variables are redefined setting the reference for the spatial coordinate at $x_{i+\frac{1}{2}}$ to $x = 0$ and for the time t^n to $t = 0$. Superscript n is also dropped. As mentioned before, the contribution of the source term is included in the solution of the Riemann Problems as a pointwise quantity at the interface.

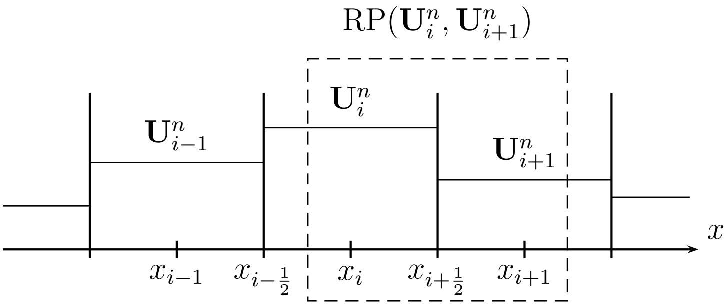

Figure A.24: Neighbouring region of cell Ω_i and representation of piecewise defined data, showing RP at $x_{i+\frac{1}{2}}$ that will be referred to as $\mathrm{RP}(\mathbf{U}_i^n, \mathbf{U}_{i+1}^n)$.

RP in (A.1) can be approximated by exactly solving the following constant coefficient linear RP [13]

$$
\begin{cases} \dfrac{\partial \hat{\mathbf{U}}}{\partial t} + \widetilde{\mathbf{J}}_{i+\frac{1}{2}} \dfrac{\partial \hat{\mathbf{U}}}{\partial x} = \mathbf{S} \\ \\ \hat{\mathbf{U}}(x,0) = \begin{cases} \mathbf{U}_i & x < 0 \\ \mathbf{U}_{i+1} & x > 0 \end{cases} \end{cases} \tag{A.2}
$$

where $\hat{\mathbf{U}}(x,t)$ is the approximate solution of (A.1) and $\widetilde{\mathbf{J}}_{i+\frac{1}{2}} = \widetilde{\mathbf{J}}_{i+\frac{1}{2}}(\mathbf{U}_i, \mathbf{U}_{i+1})$ is a constant matrix defined as a function of left and right states that represents an approximation of the Jacobian at $x_{i+\frac{1}{2}}$. This matrix is chosen so that

$$
\delta \mathbf{F}_{i+\frac{1}{2}} = \widetilde{\mathbf{J}}_{i+\frac{1}{2}} \delta \mathbf{U}_{i+\frac{1}{2}} \tag{A.3}
$$

holds [8]. Matrix $\widetilde{\mathbf{J}}_{i+\frac{1}{2}}$ is considered to be diagonalizable with N_λ approximate real eigenvalues

$$\widetilde{\lambda}^1_{i+\frac{1}{2}} < \ldots < \widetilde{\lambda}^I_{i+\frac{1}{2}} < 0 < \widetilde{\lambda}^{I+1}_{i+\frac{1}{2}} < ... < \widetilde{\lambda}^{N_\lambda}_{i+\frac{1}{2}} \tag{A.4}$$

and N_λ eigenvectors $\widetilde{\mathbf{e}}^1, ..., \widetilde{\mathbf{e}}^{N_\lambda}$. With them, two approximate matrices, $\widetilde{\mathbf{P}}_{i+\frac{1}{2}} = (\widetilde{\mathbf{e}}^1, ..., \widetilde{\mathbf{e}}^{N_\lambda})_{i+\frac{1}{2}}$ and $\widetilde{\mathbf{P}}^{-1}_{i+\frac{1}{2}}$ are built with the following property

$$\widetilde{\mathbf{J}}_{i+\frac{1}{2}} = (\widetilde{\mathbf{P}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{P}}^{-1})_{i+\frac{1}{2}}, \qquad \widetilde{\mathbf{\Lambda}}_{i+\frac{1}{2}} = \begin{pmatrix} \widetilde{\lambda}^1 & & 0 \\ & \ddots & \\ 0 & & \widetilde{\lambda}^{N_\lambda} \end{pmatrix}_{i+\frac{1}{2}} \tag{A.5}$$

where $\widetilde{\mathbf{\Lambda}}_{i+\frac{1}{2}}$ is a diagonal matrix with approximate eigenvalues in the main diagonal. System in (A.2) can be transformed using $\widetilde{\mathbf{P}}^{-1}$ matrix as follows

$$\frac{\partial \hat{\mathbf{W}}}{\partial t} + \widetilde{\mathbf{\Lambda}}_{i+\frac{1}{2}} \frac{\partial \hat{\mathbf{W}}}{\partial x} = \mathbf{B}_{i+\frac{1}{2}} \tag{A.6}$$

expressing (A.2) in terms of the characteristic variables $\hat{\mathbf{W}} = \widetilde{\mathbf{P}}^{-1}_{i+\frac{1}{2}}\hat{\mathbf{U}}$, with $\hat{\mathbf{W}} = (\hat{w}^1, ..., \hat{w}^{N_\lambda})$ and $\mathbf{B}_{i+\frac{1}{2}} = \left(\widetilde{\mathbf{P}}^{-1}\mathbf{S}\right)_{i+\frac{1}{2}}$

Approximate fluxes on the left and right side of the t axis, $\mathbf{F}^-_i$ and $\mathbf{F}^+_{i+1}$, can be derived using the results for the scalar equation. Combination of the solutions for the characteristic variables, $\hat{w}^m(x,t)$, allows to construct the numerical fluxes at the interface as [13]

$$\begin{aligned} \mathbf{F}^-_i &= \mathbf{F}_i + \sum_{m=1}^{I} \left[\left(\widetilde{\lambda}\alpha - \bar{\beta} \right) \widetilde{\mathbf{e}} \right]^m_{i+\frac{1}{2}} , \\ \mathbf{F}^+_{i+1} &= \mathbf{F}_{i+1} - \sum_{m=I+1}^{N_\lambda} \left[\left(\widetilde{\lambda}\alpha - \bar{\beta} \right) \widetilde{\mathbf{e}} \right]^m_{i+\frac{1}{2}} , \end{aligned} \tag{A.7}$$

where the set of wave strengths is defined as

$$\mathbf{A}_{i+\frac{1}{2}} = \left(\alpha^1, ..., \alpha^{N_\lambda}\right)^T_{i+\frac{1}{2}} = \left(\widetilde{\mathbf{P}}^{-1}\delta\mathbf{U}\right)_{i+\frac{1}{2}} , \tag{A.8}$$

and the set of source strengths as

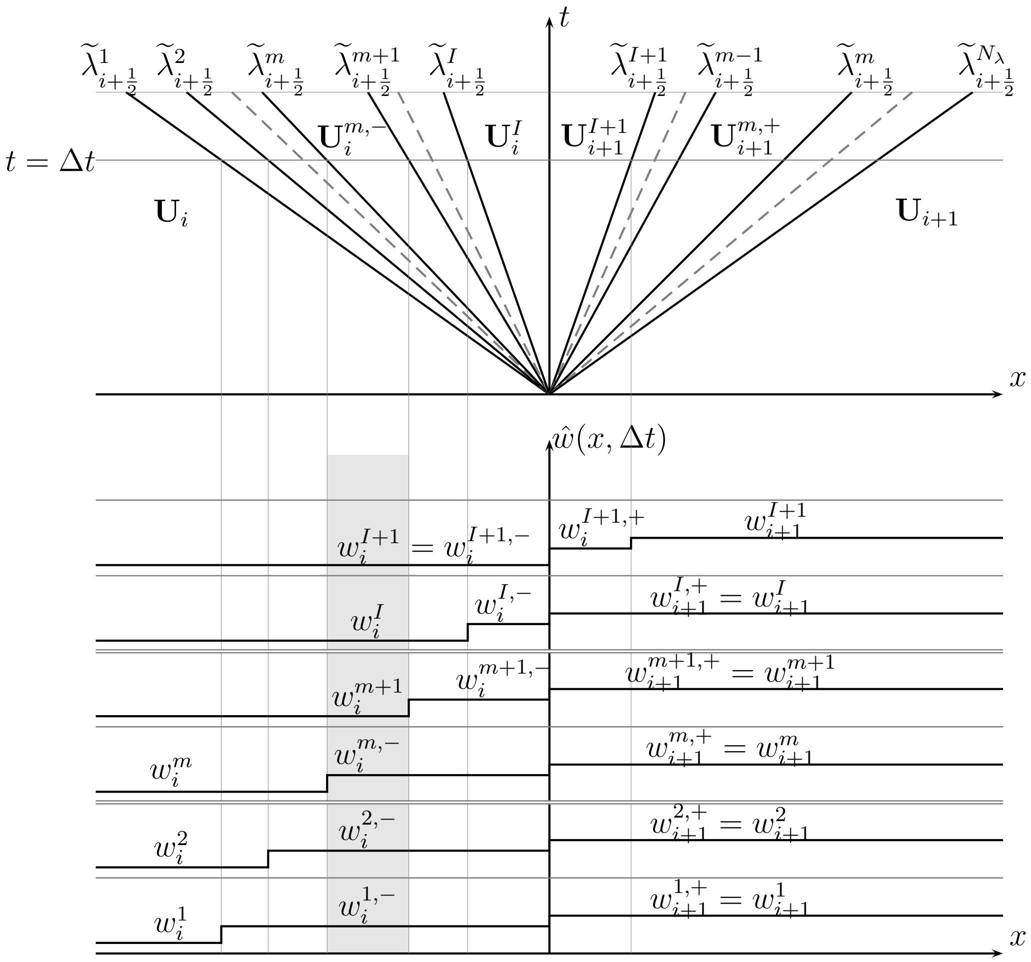

Figure A.25: Upper: Approximate solution $\hat{\mathbf{U}}(x,t)$. The solution consist of N_λ inner constant states separated by a stationary contact discontinuity, with celerity $S = 0$ at $x = 0$. Lower: The solution for characteristic variables $\hat{w}^m(x,t)$ for $m = 1, ..., I+1$ is depicted at $t = \Delta t$.

$$\bar{\mathbf{B}}_{i+\frac{1}{2}} = \left(\bar{\beta}^1, ..., \bar{\beta}^{N_\lambda}\right)^T_{i+\frac{1}{2}} = \left(\widetilde{\mathbf{P}}^{-1}\bar{\mathbf{S}}\right)_{i+\frac{1}{2}} . \tag{A.9}$$

1416 It is worth recalling that $\delta w^m_{i+\frac{1}{2}} = \alpha^m_{i+\frac{1}{2}}$. Analogously, if defining $\delta\mathbf{F}_{i+1/2} =$

$\widetilde{\mathbf{P}}_{i+1/2}\mathbf{\Gamma}_{i+1/2}$, it is straightforward to obtain the following relation

$$\mathbf{\Gamma}_{i+1/2} = \widetilde{\mathbf{\Lambda}}_{i+1/2}\widetilde{\mathbf{A}}_{i+1/2} \tag{A.10}$$

with $\mathbf{\Gamma}_{i+1/2} = (\gamma^1, ..., \gamma^{N_\lambda})_{i+1/2}$, that allows to rewrite (A.7) as

$$\begin{aligned} \mathbf{F}^-_{i+1/2} &= \hat{\mathbf{F}}_i + \sum_{m=1}^{I} \left[(\gamma - \bar{\beta})\widetilde{\mathbf{e}}\right]^m_{i+\frac{1}{2}} \,, \\ \mathbf{F}^+_{i+1/2} &= \hat{\mathbf{F}}_{i+1} - \sum_{m=I+1}^{N_\lambda} \left[(\gamma - \bar{\beta})\widetilde{\mathbf{e}}\right]^m_{i+\frac{1}{2}} \,. \end{aligned} \tag{A.11}$$

For the sake of simplicity, the term $(\gamma - \bar{\beta})^m_{i+\frac{1}{2}}$, or $\left(\widetilde{\lambda}\alpha - \bar{\beta}\right)^m_{i+\frac{1}{2}}$ analogously, can be expressed as $(\widetilde{\lambda}\theta\alpha)^m_{i+\frac{1}{2}}$, where $\theta^m_{i+\frac{1}{2}} = 1 - \bar{\beta}/\widetilde{\lambda}\alpha$. Using this compact form, the difference between left and right states across the interface can be expressed as

$$\mathbf{U}^+_{i+1} - \mathbf{U}^-_i = \mathbf{U}_{i+1} - \mathbf{U}_i - \sum_{m_1=1}^{N_\lambda} (\theta\alpha\widetilde{\mathbf{e}})^{m_1}_{i+\frac{1}{2}} \tag{A.12}$$

where wave contributions can be written in their matrix form as

$$\sum_{m_1=1}^{N_\lambda} (\theta\alpha\widetilde{\mathbf{e}})^{m_1}_{i+\frac{1}{2}} = \left(\widetilde{\mathbf{P}}\Theta\mathbf{A}\right)_{i+\frac{1}{2}} \tag{A.13}$$

with $\Theta_{i+\frac{1}{2}} = diag(\theta^1_{i+\frac{1}{2}}, \theta^2_{i+\frac{1}{2}}, ..., \theta^{N_\lambda}_{i+\frac{1}{2}})$ a diagonal matrix that allows to rewrite $\widetilde{\mathbf{P}}\Theta\mathbf{A} = \widetilde{\mathbf{P}}\mathbf{A} - \widetilde{\mathbf{P}}\widetilde{\Lambda}^{-1}\bar{\mathbf{B}}$. Substituting the previous results in (A.12) and noticing that $\widetilde{\mathbf{P}}\mathbf{A}_{i+\frac{1}{2}} = \mathbf{U}_{i+1} - \mathbf{U}_i$, it becomes

$$\mathbf{U}^+_{i+1} - \mathbf{U}^-_i = \left(\widetilde{\mathbf{P}}\widetilde{\Lambda}^{-1}\bar{\mathbf{B}}\right)_{i+\frac{1}{2}} \tag{A.14}$$

from which it can be observed that the difference between left and right states is only due to the presence of the source term. Expressing $\bar{\mathbf{B}}_{i+\frac{1}{2}} = \left(\tilde{\mathbf{P}}^{-1}\bar{\mathbf{S}}\right)_{i+\frac{1}{2}}$, the following relation is noticed

$$\bar{\mathbf{S}}_{i+\frac{1}{2}} = \left(\widetilde{\mathbf{J}}^{-1}\right)_{i+\frac{1}{2}} \left(\mathbf{U}^+_{i+1} - \mathbf{U}^-_i\right) \,. \tag{A.15}$$

This relation is worth keeping in mind, as it will come along with other derivations within the text.

When using the ARoe numerical fluxes, the first order Godunov scheme in (44) reads

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x}[\mathbf{F}_i^- - \mathbf{F}_i^+] \,. \tag{A.16}$$

Appendix B. The traditional Roe solver

When considering a homogeneous RP, that is, the contribution of the source term is nil, RH condition across the interface yields $\mathbf{F}_i^- = \mathbf{F}_{i+1}^+$, according to the notation used in this work. Such fluxes are now a unique value and are denoted by $\mathbf{F}^\star_{i+1/2}$, which can be expressed in terms of the left or right contributions according to (A.7) as follows

$$\begin{aligned} \mathbf{F}^\star_{i+1/2} &= \mathbf{F}_i + \sum_{m_1=1}^{I} \left(\widetilde{\lambda}\alpha\widetilde{\mathbf{e}}\right)^{m_1}_{i+\frac{1}{2}} \\ \mathbf{F}^\star_{i+1/2} &= \mathbf{F}_{i+1} - \sum_{m_1=I+1}^{N_\lambda} \left(\widetilde{\lambda}\alpha\widetilde{\mathbf{e}}\right)^{m_1}_{i+\frac{1}{2}} \,. \end{aligned} \tag{B.1}$$

Combination of the expressions in (B.1) leads to

$$\mathbf{F}^\star_{i+1/2} = \frac{\mathbf{F}_i + \mathbf{F}_{i+1}}{2} - \frac{1}{2}\sum_{m_1=1}^{N_\lambda} \left(\left|\widetilde{\lambda}\right| \alpha\widetilde{\mathbf{e}}\right)^{m_1}_{i+\frac{1}{2}} \tag{B.2}$$

that can be rewritten in matrix form as

$$\mathbf{F}^\star_{i+1/2} = \frac{\mathbf{F}_i + \mathbf{F}_{i+1}}{2} - \frac{1}{2}\left(\widetilde{\mathbf{P}} \mid \widetilde{\mathbf{\Lambda}} \mid \widetilde{\mathbf{A}}\right)_{i+\frac{1}{2}} \tag{B.3}$$

where

$$\mid \widetilde{\mathbf{\Lambda}} \mid_{i+\frac{1}{2}} = \begin{pmatrix} \mid \widetilde{\lambda}^1 \mid & & 0 \\ & \ddots & \\ 0 & & \mid \widetilde{\lambda}^{N_\lambda} \mid \end{pmatrix}_{i+\frac{1}{2}} \tag{B.4}$$

If defining $\mid \widetilde{\mathbf{J}} \mid_{i+\frac{1}{2}} = \left(\widetilde{\mathbf{P}}\left|\widetilde{\mathbf{\Lambda}}\right|\widetilde{\mathbf{P}}^{-1}\right)_{i+\frac{1}{2}}$, the last term in Equation (B.3) can be rewritten as

$$\left(\widetilde{\mathbf{P}} \mid \widetilde{\mathbf{\Lambda}} \mid \widetilde{\mathbf{A}}\right)_{i+\frac{1}{2}} = \left(\widetilde{\mathbf{P}} \mid \widetilde{\mathbf{\Lambda}} \mid \widetilde{\mathbf{P}}^{-1}\delta\mathbf{U}\right)_{i+\frac{1}{2}} = \left(\mid \widetilde{\mathbf{J}} \mid \delta\mathbf{U}\right)_{i+\frac{1}{2}} \tag{B.5}$$

leading to the following intercell homogeneous flux

$$\mathbf{F}^{\star}_{i+1/2} = \frac{\mathbf{F}_i + \mathbf{F}_{i+1}}{2} - \frac{1}{2}\left(\mid \widetilde{\mathbf{J}} \mid \delta\mathbf{U}\right)_{i+\frac{1}{2}} \tag{B.6}$$

Analogously, if defining $\delta\mathbf{F}_{i+1/2} = \widetilde{\mathbf{P}}_{i+1/2}\mathbf{\Gamma}_{i+1/2}$, it is straightforward to obtain the following relation

$$\mathbf{\Gamma}_{i+1/2} = \widetilde{\mathbf{\Lambda}}_{i+1/2}\widetilde{\mathbf{A}}_{i+1/2} \tag{B.7}$$

with $\mathbf{\Gamma}_{i+1/2} = (\gamma^1, ..., \gamma^{N_\lambda})_{i+1/2}$, that can be introduced in (B.3) to obtain

$$\mathbf{F}^{\star}_{i+1/2} = \frac{\mathbf{F}_i + \mathbf{F}_{i+1}}{2} - \frac{1}{2}\mathrm{sgn}(\widetilde{\mathbf{J}}_{i+\frac{1}{2}})\delta\mathbf{F}_{i+1/2} \tag{B.8}$$

where $\mathrm{sgn}(\widetilde{\mathbf{J}}_{i+\frac{1}{2}}) = \left(\widetilde{\mathbf{P}} \mid \widetilde{\mathbf{\Lambda}} \mid \widetilde{\mathbf{\Lambda}}^{-1}\widetilde{\mathbf{P}}^{-1}\right)_{i+\frac{1}{2}}$ is the upwinding matrix. The previous equation can be rewritten as follows

$$\mathbf{F}^{\star}_{i+1/2} = \frac{\mathbf{F}_i + \mathbf{F}_{i+1}}{2} - \frac{1}{2}\sum_{m_1=1}^{N_\lambda}\left(\mathrm{sgn}(\widetilde{\lambda})\gamma\widetilde{\mathbf{e}}\right)^{m_1}_{i+\frac{1}{2}} \tag{B.9}$$

or, analogously to equation (B.1)

$$\begin{aligned} \mathbf{F}^{\star}_{i+1/2} &= \mathbf{F}_i + \sum_{m_1=1}^{I}(\gamma\widetilde{\mathbf{e}})^{m_1}_{i+\frac{1}{2}} \\ \mathbf{F}^{\star}_{i+1/2} &= \mathbf{F}_{i+1} - \sum_{m_1=I+1}^{N_\lambda}(\gamma\widetilde{\mathbf{e}})^{m_1}_{i+\frac{1}{2}}\,. \end{aligned} \tag{B.10}$$

When using the homogeneous Roe fluxes, the first order Godunov scheme in (44) reads

$$\mathbf{U}^{n+1}_i = \mathbf{U}^n_i - \frac{\Delta t}{\Delta x}[\mathbf{F}^{\star}_{i+1/2} - \mathbf{F}^{\star}_{i-1/2}] \tag{B.11}$$

and can be used to solve a homogeneous PDE.

Acknowledgment

This work has been funded by the Spanish Ministerio de Economía y Competitividad under research project CGL2015–66114–R.

[1] S. Sahmima, F. Benkhaldounb, F. Alcrudo, A sign matrix based scheme for non-homogeneous PDE's with an analysis of the convergence stagnation phenomenon, J. Comput. Phys. 148 (2007) 1753-1783.

[2] L. O. Mller and E. F. Toro, A global multiscale mathematical model for the human circulation with emphasis on the venous system, Int. J. Numer. Meth. Biomed. Engng. 30 (2014)) 681-725 .

[3] S.K. Godunov, Finite difference methods for the computation of discontinuous solutions of the equations of fluid dynamics, Mat. Sb. 47 (1959) 271–306.

[4] E. Godlewski, P.-A. Raviart *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Springer Science and Business Media, Berlin, 2013.

[5] D.L. George. Augmented Riemann solvers for the shallow water equations over variable topography with steady states and inundation, J. Comput. Phys. 227 (2008) 3089–3113.

[6] J. Murillo, J. Burguete, P. Brufau, P. García-Navarro. The influence of source terms on stability, accuracy and conservation in two-dimensional shallow flow simulation using triangular finite volumes, Int. J. Numer. Meth. Fluids (2007) 54 543–590.

[7] J. Murillo, P. García-Navarro, Augmented Roe's approaches for Riemann problems including source terms: definition of stability region with application to the shallow water equations with rigid and deformable bed. In M. E. Vázquez-Cendón and A. Hidalgo and P. García-Navarro and L. Cea, eds., Numerical Methods for Hyperbolic Equations. Theory and Applications, pages 149–154. Taylor-Francis Group, 2013.

[8] Roe, Approximate Riemann solvers, parameter vectors, and difference schemes, J. Comput. Phys. 43 (1981) 357–372.

[9] A. Harten, P. Lax, B. van Leer, On upstream differencing and Godunov type methods for hyperbolic conservation laws, SIAM review. 25 (1983) 35–61.

[10] E.F. Toro, M. Spruce, W. Spears, Restoration of the contact surface in the HLL Riemann solver, Shock Waves. 4 (1994) 25–34.

[11] J. Murillo, P. García-Navarro, Weak solutions for partial differential equations with source terms: application to the shallow water equations, J. Comput. Phys. 229 (2010) 4327–4368.

[12] J. Murillo, P. García-Navarro, Augmented versions of the HLL and HLLC Riemann Solvers including source terms in one and two dimensions for shallow flow applications, J. Comput. Phys. 231 (2012) 6861–6906.

[13] J. Murillo and A. Navas-Montilla, A comprehensive explanation and exercise of the source terms in hyperbolic systems using Roe type solutions. Application to the 1D-2D shallow water equations, Advances in Water Resources 98 (2016) 70–96.

[14] G. Rosatti, L. Begnudelli, The Riemann Problem for the one-dimensional, free-surface Shallow Water Equations with a bed step: theoretical analysis and numerical simulations, J. Comput. Phys. 229 (2010) 760-787.

[15] A. Bermudez and M.E. Vázquez-Cendón, Upwind methods for hyperbolic conservation laws with source terms, Comput. Fluids. 23 (1994) 1049–1071.

[16] M.E. Vázquez-Cendón. Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry, J. Comput. Phys. 148 (1999) 497–498.

[17] 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.

[18] P. García-Navarro, M.E. Vázquez-Cendón. On numerical treatment of the source terms in the shallow water equations, Comput. and Fluids. 29 (2000) 951–979.

[19] A. Chinnayya, A.-Y. LeRoux, N. Seguin, A well-balanced numerical scheme for the approximation of the shallow water equations with topography: the resonance phenomenon, Int. J. Finite Vol. 1 (2004) 1–33.

[20] M. E. Hubbard, P. García-Navarro, Flux difference splitting and the balancing of source terms and flux gradients. J. Comp. Phys. 165 (2000) 89–125.

[21] Noelle, S., Xing, Y., Shu, C., High-order well-balanced finite volume WENO schemes for shallow water equation with moving water, J. Comput. Phys. 226 (2007) 29–58.

[22] U.S. Fjordholm, S. Mishra, E. Tadmor, Well-balanced and energy stable schemes for the shallow water equations with discontinuous topography, J. Comput. Phys. 230 (2011) 5587–5609.

[23] M.J. Castro Díaz , J.A. López-García, Carlos Parés, High order exactly well-balanced numerical methods for shallow water systems, J. Comput. Phys. 246 (2013) 242–264.

[24] Y. Xing,Exactly well-balanced discontinuous Galerkin methods for the shallow water equations with moving water equilibrium, J. Comput. Phys. 257 (2014) 536–553.

[25] J. Murillo, P. García-Navarro, Energy balance numerical schemes for shallow water equations with discontinuous topography, J. Comput. Phys. 236 (2012) 119–142.

[26] J. Murillo, P. García-Navarro, Accurate numerical modeling of 1D flow in channels with arbitrary shape. Application of the energy balanced property, J. Comput. Phys. 260 (2014) 222–248.

[27] A. Navas-Montilla, J. Murillo, Energy balanced numerical schemes with very high order. The Augmented Roe Flux ADER scheme. Application to the shallow water equations, J. Comput. Phys. 290 (2015) 188–218.

[28] A. Navas-Montilla, J. Murillo, Asymptotically and exactly energy balanced augmented flux-ADER schemes with application to hyperbolic conservation laws with geometric source terms, J. Comput. Phys. 317 (2016) 108–147.

[29] A. Chinnayya, A. Y. LeRoux, N. Seguin, A well-balanced numerical scheme for the approximation of the shallow-water equations with topography: the resonance phenomenon, Int. J. Finite Volumes 1 (2004) 1-33.

[30] T. Galloet, J.M. Herard, N. Seguin, Some approximate Godunov schemes to compute shallow-water equations with topography, Computers and Fluids 32 (2003) 479-513.

[31] F. Alcrudo, F. Benkhaldoun, Exact solutions to the Riemann problem of the shallow water equations with a bottom step, Comput. Fluids 30 (2001) 643–671.

[32] R. Bernetti, V.A. Titarev, E.F. Toro, Exact solution of the Riemann problem for the shallow water equations with discontinuous bottom geometry, J. Comput. Phys. 227 (2008) 3212–3243.

[33] D. S. Balsara, T. Rumpf, M. Dumbser, C.-D. Munz, Efficient, high accuracy ADER-WENO schemes for hydrodynamics and divergence-free magnetohydrodynamics, J. Comput. Phys., 228 (2009) 2480-2516.

[34] F. Franzini, S. Soares-Frazão Efficiency and accuracy of Lateralized HLL, HLLS and Augmented Roes scheme with energy balance for river flows in irregular channels, Appl. Math. Model. 40 (2016) 7427–7446.

[35] K.M. Peery and S.T. Imlay, Blunt-body flow simulations, AIAA paper, (1988) 88-2924.

[36] K. Kitamura, E Shima, and PL Roe, Three-dimensional carbuncles and euler fluxes, Proceedings of the 48th AIAA Aerospace Sciences Meeting (2010).

[37] T. W. Roberts, The behavior of flux difference splitting schemes near slowly moving shock waves, J. Comput. Phys., 90 (1990) 141–160.

[38] D. W. Zaide, Numerical Shockwave Anomalies, PhD thesis, Aerospace Engineering and Scientific Computing, University of Michigan, 2012.

[39] R. S. Myong, and P. L. Roe, Shock waves and rarefaction waves in magnetohydrodynamics. part 2. the mhd system. J. Plasma Ph., 58 (1997) 21–552.

[40] M. Arora, P. L. Roe, On postshock oscillations due to shock capturing schemes in unsteady flows, J. Comput. Phys., 130 (1997) 25–40.

[41] W.F. Noh, Errors for calculations of strong shocks using an artificial viscosity and an artificial heat flux, J. Comput. Phys., 72 (1987) 78-120.

[42] D. W. Zaide, P. L. Roe, Flux functions for reducing numerical shockwave anomalies. ICCFD7, Big Island, Hawaii, (2012) 9–13.

[43] G. Cameron, An analysis of the errors caused by using artificial viscosity terms to represent steady-state shock waves. J. Comput. Phys. 1 (1966) 1–20.

[44] A. Emery, An evaluation of several differencing methods for inviscid fluid flow problems, J. Comput. Phys., 2 (1968) 306–331.

[45] S. Karni, S. Canic, Computations of slowly moving shocks, J. Comput. Phys., 136 (1997) 132–139.

[46] S. Jin, J. G. Liu, The Effects of Numerical Viscosities, J. Comput. Phys., 126 (1996) 373–389.

[47] M. H. Carpenter, J. H. Casper, Accuracy of Shock Capturing in Two Spatial Dimensions, AIAA Journal, 37 (1999) 1072–1079.

[48] N. K. Yamaleev, M. H. Carpenter, On accuracy of adaptive grid methods for captured shocks, J. Comput. Phys., 181 (2002) 280–316.

[49] Y. Stiriba, R. Donat, A numerical study of postshock oscillations in slowly moving shock waves, Comput. Math. with Appl., 46 (2003) 719–739.

[50] E. Johnsen, S. K. Lele, Numerical errors generated in simulations of slowly moving shocks, Center for Turbulence Research, Annual Research Briefs, (2008) 1–12.

[51] R. J. LeVeque, Finite volume methods for hyperbolic problems (Vol. 31). Cambridge university press, (2002).

[52] E.F. Toro, Riemann solvers and numerical methods for fluid dynamics: a practical introduction, third ed., Springer-Verlag, Berlin, Heidelberg, 2009.

[53] T.J. Barth, "Some Notes on Shock-Resolving Flux Functions Part 1: Stationary Characteristics," NASA TM-101087 (1989)

[54] P.L. Roe, Fluctuations and Signals - A Framework for Numerical Evolution Problems, Numerical Methods for Fluid Dynamics, edited by K. W. Morton, and M. J. Baines, Academic Press, New York, (1982) 219–257.

[55] K. Kitamura, P.L. Roe, F. Ismail, Evaluation of Euler fluxes for hypersonic flow computations, AIAA Journal, 47 (2009) 44–53