Università degli Studi di Trento

Dipartimento di Matematica
Dottorato di Ricerca in Matematica

Ph.D. Thesis

Computational hydraulic techniques for the Saint Venant Equations in arbitrarily shaped geometry

Elisa Aldrighetti

Supervisors

Prof. Vincenzo Casulli and Dott. Paola Zanolli

May 2007

To Mam, Dad and Edoardo

Abstract

A numerical model for the one-dimensional simulation of non-stationary free surface and pressurized flows in open and closed channels with arbitrary cross-section will be derived, discussed and applied.

This technique is an extension of the numerical model proposed by Casulli and Zanolli [10] for open channel flows that uses a semi-implicit discretization in time and a finite volume scheme for the discretization of the Continuity Equation: these choices make the method computationally simple and conservative of the fluid volu-me both locally and globally.

The present work will firstly deal with the elaboration of a semi-implicit numerical scheme for flows in open channels with arbitrary cross-sections that conserves both the volume and the momentum or the energy head of the fluid, in such a way that its numerical solutions present the same characteristics as the physical solutions of the problem considered [3].

The semi-implicit discretization [6] in time leads to a relatively simple and computationally efficient scheme whose stability can be shown to be independent from the wave celerity $\sqrt{gH}$.

The conservation properties allow dealing properly with problems presenting discontinuities in the solution, resulting for example from sharp bottom gradients and hydraulic jumps [46]. The conservation of mass is particularly important when the channel has a non rectangular cross-section.

The numerical method will be therefore extended to the simulation of closed channel flows in case of free-surface, pressurized and transition flows [2].

The accuracy of the proposed method will be controlled by the use of appropriate flux limiting functions in the discretization of the advective terms [52, 35], especially in the case of large gradients of the physical quantities involved in the problem. In the particular case of closed channel flows, a new flux limiter will be defined in order

to better represent the transitions between free-surface and pressurized flows.

The numerical solution, at every time step, will be determined by solving a mildly non-linear system of equations that becomes linear in the particular case that the channel has a rectangular cross-section.

Careful physical and mathematical considerations about the stability of the method and the solvability of the system with respect to the implemented boun-dary conditions will be also provided. The study of the existence and uniqueness of the solution requires the solution of a constrained problem, where the constraint expresses that the feasible solutions are physically meaningful and present a non-negative water depth. From this analysis, it will follow an explicit (dependent only on known quantities) and sufficient condition for the time step to ensure the non-negativity of the water volume. This condition is valid in almost all the physical situations without more restrictive assumptions than those necessary for a correct description of the physical problem.

Two suitable solution procedures, the Newton Method and the conjugate gradient method, will be introduced, adapted and studied for the mildly non linear system arising in the solution of the numerical model.

Several applications will be presented in order to compare the numerical results with those available from the literature or with analytical and experimental solutions. They will illustrate the properties of the present method in terms of stability, accuracy and efficiency.

Acknowledgments

Firstly, I would like to thank Professor Vincenzo Casulli and Paola Zanolli for their help, support and valuable guidance throughout this Ph.D. They always showed me a high-quality way to do research and goad me to do my best.

I would like to express my gratitude to Professor Guus Stelling who made my stay at the Fluid Mechanic Section of the Technical University of Delft possible. His supervision, support and patience were constant throughout the period I spent there. His keen interest and understanding of the topic was inspiring and led to many fruitful discussions. Special thanks also for having had faith in me and in my abilities, it made me feel more confident and positive.

Many thanks to all the Ph.D. students of the Department of Mathematics of the University of Trento. Their reciprocal support not only during the exciting and productive but also the demanding phases of this project has been very important. Amongst the staff members of the Department, I would like to thank particularly Myriam and Amerigo, who were always helpful when I requested something.

I wish to thank the staff and the colleagues of the Fluid Mechanic Section in Delft who always made me feel welcome. Working there afforded the opportunity to meet very good researchers and to find many good friends that made me feel less lonely. We shared very happy moments that I will cherish forever.

I would also like to acknowledge the company Delft Hydraulics that allowed me to use their facilities and laboratories to carry out the experimental part of my research.

I am also grateful to all my friends for their assistance, friendship and encouragements throughout several challenging times during the last years. Especially, I would like to mention Giulia and Betta for their heartfelt affection: I hope we will enjoy each other's company in the years to come.

Many loving thanks to Edoardo for believing in me and for always accepting my choices, whether he liked them or not. He gave me the extra strength, motivation

and love necessary to get things done.

Finally, I would like to thank my family and especially my parents for their infinite love and continuous encouragement. They have been always there as a continual source of support.

Contents

List of Figures

List of Tables

List of main symbols

u	water velocity in the x-direction for the $3D$ model	$[m/s]$
v	water velocity in the y-direction for the $3D$ model	$[m/s]$
w	water velocity in the z-direction for the $3D$ model	$[m/s]$
U_2	water velocity in the x-direction for the $2D_{xz}$ model	$[m/s]$
W_2	water velocity in the z-direction for the $2D_{xz}$ model	$[m/s]$
U	water velocity in the x-direction for the $1D$ model	$[m/s]$
Q	water discharge in the x-direction for the $1D$ model	$[m^3/s]$
η	free surface elevation or pressure head (z in Figures)	$[m]$
$-h$	bottom of the channel	$[m]$
H	total water depth ($H = \eta + h$)	$[m]$
L	lenght of the channel	$[m]$
l_1, L_1	parameters for the curvature of a basin	$[m]$
α	channel's inclination	$[-]$
top	top level value in a closed conduit	$[m]$
ϵ	Preissmann slot width	$[m]$
B	channel width	$[m]$
P	wetted perimeter	$[m]$
A	cross-section area	$[m^2]$
V	water volume	$[m^3]$
R	radius of a pipe	$[m]$
D	diameter of a pipe	$[m]$
c_f	general friction coefficient	[-]
C	Chezy friction coefficient	$[m^{1/2}/s]$
n_M	Manning friction coefficient	[-]
S_0	bed slope	$[-]$

S_f	friction slope	$[-]$
K	conveyance	$[m^3/s]$
R_H	hydraulic radius	$[m]$
k_S	equivalent sand roughness height	$[m]$
Fr	Froude number	$[-]$
x	space	$[m]$
t	time	$[s]$
Δx	increment in the horizontal x-direction	$[m]$
Δt	time step	$[s]$
T	period of time	$[s]$
N	number of nodes of the spatial grid	$[-]$
i, j	space indexes	$[-]$
n	time level superscript	$[-]$
g	gravitational acceleration	$[m/s^2]$
θ	implicitness factor	$[-]$
E	energy head	$[m]$
H_{cr}	critical depth	$[m]$
Ψ	flux limiter function	$[-]$
r^U	regularity of the U data	$[-]$
r^η	regularity of the η data	$[-]$
ρ	fluid density	$[kg/m^3]$
ω	frequency of the motion	$[rad/s]$
Θ	amplitude of the motion	$[m]$

Introduction

The purpose of this doctoral thesis is the study of the numerical techniques for the simulation of free surface and pressurized flows in open and closed channels with arbitrary cross section. The aim of this research is to formulate a new numerical method for hydraulic engineering problems that is capable of predicting subcritical flows, mixed flows (subcritical and supercritical flows) as well as transitions from supercritical to subcritical flows, with particular attention to the robustness and the efficiency of the model and to the conservation of the physical quantities volume, momentum and energy head. This introductive chapter will draw the context where this research has been developed, it will briefly describe the techniques known in the current literature an it will give an idea of the structure of the whole thesis.

Flows in hydrodynamics

The study of free-surface and pressurized water flows in channels has many interesting applications, one of the most important being the modelling of the phenomena in the area of natural water systems (rivers, estuaries) as well as in that of man-made systems (canals, pipes).

For the development of major river engineering projects, such as flood prevention and flood control, there is an increase need to be able to model and predict the consequences of any possible phenomenon on the environment and in particular the new hydraulic characteristics of the system.

Hydraulics has a long tradition of providing a scientific basis for engineering applications [29, 42]. Firstly, conceptual models were designed starting from empirical relations obtained from field observations or model scale experiments.

Lately, mathematics started playing an important role not only to describe the properties of these relations, but also to formulate analytical solutions of particular model situations in order to capture the essential features of those phenomena.

Actually, the research and the applications in the field of computational fluid hydraulics and fluid dynamics evolved with the advent of electronic computers.

The first applications in computational hydraulics concerned programming analytical formulae rather than deriving generic numerical schemes and techniques based on physical principles like conservation laws for mass and momentum. Later developments extended the research and the applications in this field towards simulating complicated flow phenomena in arbitrarily shaped geometries.

The literature

The basic equations expressing hydraulic principles were formulated in the 19th century by Barre de Saint Venant and Valentin-Joseph Bousinnesque.

The original hydraulic model of the Saint Venant Equations [15] is written in the form of a system of two partial differential equations and it is derived under the assumption that the flow in one-dimensional, the cross-sectional velocity is uniform, the streamline curvature is small and the pressure distribution is hydrostatic [60].

One dimensional flows do not actually exists in nature, but the equations remain valid provided the flow is approximately one-dimensional: as pointed out by Steffler and Jin [45], they are inappropriate to analyze free surface flow problems with horizontal length scales close to flow depth.

In the current literature, several numerical techniques for solving the Saint Venant Equations are known. These include the method of characteristics, explicit difference methods, fully implicit methods, Godunov methods [27] and semi-implicit methods [6].

In particular, the method of characteristics is very efficient in the treatment of boundary conditions, but does not guarantee volume and momentum conservation.

The Godunov's type methods (see, e.g., [52, 27, 19]) instead, require the solution of local Riemann problems and, consequently, are very effective on simple channel geometries with flat, horizontal bottom and rectangular cross section. For space varying bottom profiles, however, the bottom slope appears as a source term that may generate artificial flows [53] unless specific treatments of the geometrical source terms are implemented [21]. Moreover, Godunov's type methods [23] are explicit in time and, accordingly, the allowed time step is restricted by a C.F.L. stability condition, which relates the time step to the spatial discretization and the wave speed. These kind of methods are in general based upon non-staggered grids and can achieve higher that first-order accuracy. The Godunov's type methods were originally

developed for gas dynamic and only later extended to hydrodynamic on the basis of the analogy between the equations for isentropic flow of a perfect gas with constant specific heat and the shallow water equations [47, 52].

Alternatively, semi-implicit methods (see, e.g., [6, 7, 33]) can be unconditionally stable and computationally efficient. These methods, however, when do not satisfy momentum conservation, may produce incorrect results if applied to extreme problems having a discontinuous solution. The semi-implicit method presented by Stelling in [46] combines the efficiency of staggered grids with conservation properties and can be applied to problems including rapidly varying flows. A semi-implicit method that conserves the fluid volume when applied to channels with arbitrary cross-sections was presented in [10].

Our contribution

The work presented in this thesis started from the analysis of the numerical model proposed by Casulli and Zanolli [10] for open channel flows that uses a semi-implicit discretization in time and a finite volume scheme for the discretization of the Continuity Equation. These choices make the method computationally simple and conservative of the fluid volume both locally and globally.

This thesis proposes a numerical scheme for flows in open and closed channel with arbitrary cross-sections that conserves both the volume and the momentum or the energy head of the fluid, in such a way that its numerical solutions present the same characteristics as the physical solutions of the problem considered.

It is based upon the classical staggered grids and it combines the computational efficiency of the explicit methods and the unconditional stability of the implicit ones using a semi-implicit time integration.

The high resolution technique called the flux limiter method has been introduced in order to improve the accuracy of the model especially in the case of large gradients of the physical quantities involved in the problem. In the particular case of closed channel flows, a new flux limiter has been defined in order to better represent the transitions between free-surface and pressurized flows.

Different numerical simulations have been performed in order to compare the numerical results with those available from the literature or with the analytical solutions. The results illustrate the applicability of the model to correctly simulate

hydraulic engineering problems such as wetting and drying phenomena [51]. In particular for the case of closed channel flows, some of the numerical results have been also compared with the results obtained in the laboratory. For all the case tested and even for particularly difficult physical situations such as the transitions between free-surface and pressurized flows, the numerical results are definitely satisfying.

A precise theoretical analysis of the stability of the method and of the existence and uniqueness of the numerical solution of the model have also been developed.

An explicit (dependent only on known quantities) and sufficient condition for the time step Δt to ensure the non-negativity of the water volume follows from this analysis and it is valid almost in all the physical situations without more restrictive assumptions than those necessary for a correct description of the physical problem.

A modified version of the Conjugate Gradient Method and one of the Newton Method have been analyzed both from a theoretical and a computational point of view to solve the mildly non linear system arising in the solution of the numerical model.

Several applications included in this work illustrate the potential of the model in simulating real problems and in being an useful engineering tool for the water management.

Structure of the thesis

Chapter 1 of this thesis is devoted to the introduction of the one-dimensional Saint Venant Equations, to their characterization through some of their properties and to their derivation from the Navier-Stokes Equations.

Chapter 2 and 4 describe and formulate the numerical technique that approximates in one dimension water flows in open and closed channels with arbitrary cross-sections, while Chapter 3 presents several open channel flow applications. Chapters 5, 6 and 7 analyze the non-linear system arising from the one-dimensional model from the points of view of existence and uniqueness of its solution, non-negativity of the water volume and solution algorithms.

Below, a description of the contents of each chapter is given.

Chapter 1 introduces the Saint Venant Equations and the main hypotheses used to derive them from the three dimensional Navier Stokes Equations. First of all the three dimensional shallow-water equations are derived under the assumption that the

pressure is hydrostatically distributed and finally they are integrated along the cross section to obtain the Saint Venant Equations.

Chapter 2 describes a new fully conservative semi-implicit finite volume method for the Saint Venant Equations. The mass, the momentum and the energy head conserving equation are discretized on a space staggered grid and are coupled depending on local flow conditions. A high resolution procedure is implemented to deal with steep gradients like the ones that are found in dam break problems or in hydraulic jumps problems. In addition, a new special flux limiter is described and implemented to allow accurate flow simulations near hydraulic structures such as weirs, for both critical and subcritical situations including the transition.

In Chapter 3, the simulation of various test cases illustrates the properties of the proposed method in terms of stability, accuracy and efficiency. The numerical results from the simulation of the unsteady dam break problem over a wet and dry bed in a rectangular channel are given and compared with the analytical solutions. A dam break problem in a triangular channel is also presented to show the applicability of the present algorithm to a problem where precise volume conservation is essential and not easily obtained by traditional linear schemes. Moreover, staedy flows over a discontinuous bed profile are also modelled in order to show the robustness of the proposed method and its ability in dealing with transitions from super to subcritical flows and vice-versa. Finally, two tests describing free fluid oscillations of a planar and of a parabolic surface in an elliptical basin are simulated and prove the correct treatment of the phenomena presenting flooding and drying and the correct computation of the moving wet-dry interface over a sloping topography.

Chapter 4 presents the extension of the numerical model presented in Chapter 2 to simulate pressurized flows in closed channels and pipes with arbitrary cross-section. Flows in closed channels, such as rain storm sewers, often contain transitions from free surface flows to pressurized flows, or vice versa. These phenomena usually require two different sets of equations to model the two different flow regimes. Actually, a few specifications for the geometry of the channel and for the discretization choices can be sufficient to model closed channel flows using only the open channel flow equations. The numerical results obtained solving the pressurization of a hori-zontal pipe are presented and compared with the experimental data known from the literature. Moreover, the numerical scheme is also validated simulating a flow in a horizontal and downwardly inclined pipe and comparing the numerical results with

the experimental data obtained in the laboratory.

Chapter 5 describes a complete analysis of the midly non-linear system arising from the particular discretization of the Saint Venant Equations presented in Chapters 2 and 4. The problem of existence and uniqueness of the solution of this system is investigated with respect to the boundary conditions imposed and it is solved by introducing a few mathematical assumptions that can be justified by physical argumentation.

In Chapter 6, an explicit and an implicit constraint on the time step are derived to ensure the non-negativity of the water volume obtained by the algorithm proposed in Chapters 2 and 4. The advantages of using the explicit constraint are discussed and shown with an interesting numerical example.

Two solution algorithms for solving the midly non-linear system analyzed in Chapter 5 are presented in Chapter 7: the Generalized Newton Method and a particular version of the Conjugate Gradient method. Their convergence is also proved when the requirements for existence and uniqueness of the solution are satisfied and a comparison of these two techniques is presented from the point of view of the computational efficiency.

In the last Chapter, general conclusions on the theoretical results and on the application of the numerical algorithm are formulated. The properties of the proposed numerical model and its potential in dealing with engineering problems are underlined. The chapter closes with recommendations for future research.

1

The Saint Venant Equations (SVE): main assumptions and derivation

The Navier-Stokes Equations are a general model which can be used to model water flows in many applications. However, when considering a specific problem such as shallow-water flows in which the horizontal scale is much larger than the vertical one, the Shallow Water Equations will suffice. The aim of this chapter is to present the one-dimensional Saint Venant Equations and some of their properties starting from their derivation from the Navier-Stokes Equations. First of all, the three dimensional Shallow-Water Equations will be derived under the assumption that the pressure is hydrostatically distributed. Finally, they will be integrated along the cross section to obtain the Saint Venant Equations.

1.1 Basic hypothesis for the SVE

The equations of unsteady channel flow formalize the main concepts and hypotheses used in the mathematical modelling of fluid-flow problems.

These equations consider only the most important flow influences, omitting those which are of secondary importance depending on the purpose of modelling. In this way, they provide a simple model for very complex phenomena.

A general fluid-flow problem involves the prediction of the distribution of different quantities: the fluid pressure, the temperature, the density and the flow velocity.

With this intention, six fundamental equations are considered: the Continuity Equation based on the law of conservation of mass, the Momentum Equations along three orthogonal directions (derived from Newton's second law of motion), the Thermal Energy Equation obtained from the first law of thermodynamics and the

equation of state, which is an empirical relation among fluid pressure, temperature and density.

Channel flow problems do not require the last two equations and therefore can be solved by the Continuity Equation and by the Momentum Equations assuming as constant both density and temperature.

Throughout this thesis, channel flows are assumed to be strictly one-dimensional, although truly one-dimensional flows do not exist in the real life.

The basic one-dimensional equations expressing hydraulic principles are called the Saint Venant Equations [15] and were formulated in the 19th century by two mathematicians, de Saint Venant and Bousinnesque.

These equations can be derived by averaging the three dimensional Reynolds Equations over the cross-section of the channel as it will be presented in the following sections.

The basic assumptions for the analytical derivation of the Saint Venant Equations are the following:

- the flow is one-dimensional, i.e. the velocity is uniform over the cross-section and the water level across the section is represented by a horizontal line
- the streamline curvature is small and the vertical accelerations are negligible, so that the pressure can be taken as hydrostatic
- the effects of boundary friction and turbulence can be accounted for through resistance laws analogous to those used for steady state flow
- the average channel bed slope is small so that the cosine of the angle it makes with the horizontal may be replaced by unity.

These hypotheses do not impose any restriction on the shape of the cross-section of the channel and on its variation along the channel axis, although the latter is limited by the condition of small streamline curvature.

1.2 First step: the 3D Shallow Water Equations

The governing three dimensional primitive variable equations describing constant density, free surface flow of an incompressible fluid are the well known Reynolds-

Averaged Navier-Stokes Equations which express the conservation of mass and momentum. Such equations have the following form

$$u_t + (uu)_x + (uv)_y + (uw)_z = -p_x + (\nu u_x)_x + (\nu u_y)_y + (\nu u_z)_z \quad (1.2.1)$$

$$v_t + (uv)_x + (vv)_y + (vw)_z = -p_y + (\nu v_x)_x + (\nu v_y)_y + (\nu v_z)_z \quad (1.2.2)$$

$$w_t + (uw)_x + (wv)_y + (ww)_z = -p_z + (\nu w_x)_x + (\nu w_y)_y + (\nu w_z)_z - g \quad (1.2.3)$$

$$u_x + v_y + w_z = 0 \quad (1.2.4)$$

where $u(x, y, z, t)$, $v(x, y, z, t)$ and $w(x, y, z, t)$ are the velocity components in the horizontal x, y and in the vertical z-directions. t is the time, p is the normalized pressure, that is the pressure divided by the constant density, g is the gravitational acceleration and ν is an eddy viscosity coefficient which is determined from a specific turbulence model. The discussion about turbulence models is not in the aim of the present work and the eddy viscosity coefficient is a given non-negative function of space and time.

Moreover, assuming that the free surface can be expressed as a single valued function $z = \eta(x, y, t)$, the kinematics condition of the free surface is given by

$$\eta_t + u^s \eta_x + v^s \eta_y = w^s \quad (1.2.5)$$

where $\eta(x, y, t)$ denotes the water surface elevation measured from the undisturbed water surface and u^s, v^s and w^s are the velocity components at the free surface. Under the assumption that the bottom profile can be expressed as a single valued function $z = -h(x, y)$, a similar condition at the bottom boundary is

$$u^b h_x + v^b h_y + w^b = 0 \quad (1.2.6)$$

where $h(x, y)$ is the water depth measured from the undisturbed water surface and u^b, v^b and w^b are the velocity components at the bottom. Condition (1.2.6) states that the velocity component perpendicular to the solid boundaries must vanish.

Integration of the Continuity Equation (1.2.4) over the depth yields

$$\begin{aligned} &\int_{-h}^{\eta} u_x dz + \int_{-h}^{\eta} v_y dz + \int_{-h}^{\eta} w_z dz \\ &= \left[\int_{-h}^{\eta} u dz\right]_x - u^s \eta_x + u^b (-h)_x \\ &+ \left[\int_{-h}^{\eta} v dz\right]_y - v^s \eta_y + v^b (-h)_y + w^s - w^b = 0 \end{aligned} \quad (1.2.7)$$

Thus, by using the conditions (1.2.5)-(1.2.6), the following conservative form of the free surface equation is obtained

$$\eta_t + [\int_{-h}^{\eta} u dz]_x + [\int_{-h}^{\eta} v dz]_y = 0 \tag{1.2.8}$$

Equation (1.2.8) will replace (1.2.5) in the model formulation presented in the following.

In most geophysical flows, the characteristic horizontal length scale is much larger than the characteristic vertical length scale and the characteristic vertical velocity is small in comparison with the characteristic horizontal velocity [45].

These assumptions allow that the terms $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$ are neglected, but more importantly, that the convective and the viscous terms in the third momentum equation can be neglected. Therefore, the following equation for pressure results

$$p_z = -g \tag{1.2.9}$$

This equation yields the following expression for the hydrostatic pressure

$$p(x, y, z, t) = p_a(x, y, t) + g[\eta(x, y, t) - z] \tag{1.2.10}$$

where $p_a(x, y, t)$ is the atmospheric pressure at the free surface which, without loss of generality, will be assumed to be constant.

Substitution of (1.2.10) into the Navier Stokes Equations yields the following three dimensional model equations

$$u_t + (uu)_x + (uv)_y + (uw)_z = -g\eta_x + (\nu u_x)_x + (\nu u_y)_y + (\nu u_z)_z \tag{1.2.11}$$

$$v_t + (uv)_x + (vv)_y + (vw)_z = -g\eta_y + (\nu v_x)_x + (\nu v_y)_y + (\nu v_z)_z \tag{1.2.12}$$

$$u_x + v_y + w_z = 0 \tag{1.2.13}$$

$$\eta_t + [\int_{-h}^{\eta} u dz]_x + [\int_{-h}^{\eta} v dz]_y = 0 \tag{1.2.14}$$

Under the assumption that the free surface is almost flat horizontal, the tangential stress boundary conditions prescribed by

$$\nu(u_z - u_x\eta_x - u_y\eta_y) = \gamma_T(u_a - u^s) \tag{1.2.15}$$

$$\nu(v_z - v_x\eta_x - v_y\eta_y) = \gamma_T(v_a - v^s) \tag{1.2.16}$$

are approximated as follows

$$\nu u_z = \gamma_T(u_a - u^s) \tag{1.2.17}$$

$$\nu v_z = \gamma_T(v_a - v^s) \tag{1.2.18}$$

Similarly, the boundary conditions at the sediment-water interface are given by

$$\nu(u_z + u_x h_x + u_y h_y) = \gamma_B u^b \tag{1.2.19}$$

$$\nu(v_z + v_x h_x + v_y h_y) = \gamma_B v^b \tag{1.2.20}$$

are approximated by

$$\nu u_z = \gamma_B u^b \tag{1.2.21}$$

$$\nu v_z = \gamma_B v^b \tag{1.2.22}$$

With properly specified initial and boundary conditions, Equations (1.2.11)-(1.2.10) form a three dimensional model used in shallow water flow simulations.

1.3 Second step: the laterally averaged Shallow Water Equations

From the fully three dimensional equations, it is possible to derive a simplified $2D$ model for narrow estuaries assuming that the circulation of interest takes place in the vertical $x - z$ plane.

This model is obtained by integrating laterally the Momentum Equations (1.2.11) and (1.2.12). To this purpose, let $y = l(x, z)$ and $y = r(x, z)$ be single-valued functions representing the left and the right walls, respectively, so that $B(x, z) = l(x, z) - r(x, z)$ denotes the width of the estuary. The condition of zero flux through the side walls are derived by requiring that the velocity component perpendicular to the walls must vanish. These conditions are given by

$$u^l l_x + w^l l_z = v^l \tag{1.3.1}$$

$$u^r r_x + w^r r_z = v^r \tag{1.3.2}$$

Similarly, the tangential boundary conditions at the side walls are given by specifying the lateral stresses as

$$\nu(u_x l_x - u_y + u_z l_z) = \gamma_l u \tag{1.3.3}$$

$$-\nu(u_x r_x - u_y + u_z r_z) = \gamma_r u \tag{1.3.4}$$

The laterally averaged momentum can be derived by integrating Equation (1.2.11) from the right $y = r(x, z)$ to the left wall $y = l(x, z)$. Specifically, if $B(x, z)$ denotes

the width of the estuary, $B(x,z) = l(x,z) - r(x,z)$, the laterally averaged velocity U_2 and W_2 and the laterally averaged free surface η_2 are defined as $U_2 = \frac{1}{B}\int_r^l u dy$, $W_2 = \frac{1}{B}\int_r^l w dy$ and $\eta_2 = \frac{1}{B}\int_r^l \eta dy$ respectively.

Thus, by using the boundary conditions (1.3.3)-(1.3.4), the lateral integration of the left hand side of Equation (1.2.11) yields

$$\int_r^l [u_t + (uu)_x + (uv)_y + (uw)_z] dy \tag{1.3.5}$$

$$\begin{aligned} &= (\int_r^l u dy)_t + (\int_r^l uu dy)_x + (\int_r^l uw dy)_z \\ &- u^l[u^l l_x - v^l + w^l l_z] + u^r[u^r r_x - v^r + w^r r_z] \\ &= (BU_2)_t + (BU_2U_2)_x + (BU_2W_2)_z \\ &+ [\int_r^l (u-U_2)^2 dy]_x + [\int_r^l (u-U_2)(w-W_2) dy]_z \end{aligned} \tag{1.3.6}$$

Moreover, the lateral integral of the barotropic pressure gradient term in Equation (1.2.11) yields

$$\int_r^l \eta_x dy = [\int_r^l \eta dy]_x - \eta^l l_x + \eta^r r_x \tag{1.3.7}$$

$$\begin{aligned} &= (B\eta_2)_x - \eta_2 B_x - (\eta^l - \eta_2) l_x + (\eta_r - \eta_2) r_x \\ &= (B\eta_2)_x - (\eta^l - \eta_2) l_x + (\eta_r - \eta_2) r_x \end{aligned} \tag{1.3.8}$$

Finally, by using the boundary conditions (1.3.3)-(1.3.4), the lateral integration of the viscous terms at the right hand side of Equation (1.2.11) yields

$$\int_r^l [(\nu u_x)_x + (\nu u_y)_y + (\nu u_z)_z] dy \tag{1.3.9}$$

$$\begin{aligned} &= (\int_r^l \nu u_x dy)_x + (\int_r^l \nu u_z dy)_z \\ &- \nu(u_x l_x - u_y + u_z l_z)|_{y=l} + \nu(u_x r_x - u_y + u_z r_z)|_{y=r} \\ &= \left[\int_r^l \nu (U_2)_x dy\right]_x + \left[\int_r^l \nu (U_2)_z dy\right]_z \\ &+ [\int_r^l \nu (u-U_2)_x dy]_x + [\int_r^l \nu (u-U_2)_z dy]_z - \gamma_l u^l - \gamma_r u^r \\ &= [\nu_2 B (U_2)_x]_x + [\nu_2 B (U_2)_z]_z - \gamma U_2 \\ &+ [\int_r^l \nu (u-U_2)_x dy]_x + [\int_r^l \nu (u-U_2)_z dy]_z - \gamma_l (u^l - U_2) - \gamma_r (u^r - U_2) \end{aligned} \tag{1.3.10}$$

where $\nu 2 = \frac{1}{B}\int_r^l \nu dy$ is the laterally averaged viscosity coefficient. Thus, after standard approximations on the local velocities with their laterally averaged quantity,

the Momentum Equation (1.2.11) is approximated with

$$\begin{aligned}(BU_2)_t + (BU_2U_2)_x &+ (BU_2W_2)_z = \\ -gB(\eta_2)_x &+ (\nu_2 B(U_2)_x)_x + (\nu_2 B(U_2)_z)_z - \gamma U_2 \end{aligned} \tag{1.3.11}$$

where $\gamma = \gamma_l + \gamma_r$.

Similarly, the laterally integral of the incompressibility condition (1.2.13) yields

$$\int_r^l (u_x + v_y + w_z)dy \tag{1.3.12}$$

$$\begin{aligned} &= (\int_r^l u dy)_x + (\int_r^l w dy)_z \\ &-(ul_x - v + wl_z)|_{y=l} + (ur_x - v + wr_z)|_{y=r} \\ &= (BU_2)_x + (BW_2)_z = 0 \end{aligned} \tag{1.3.13}$$

which represents the exact, laterally averaged incompressibility condition.

Finally, upon integration of the free surface Equation (1.2.13), one gets

$$\int_{r(x,\eta)}^{l(x,\eta)} [\eta_t + (\int_{-h}^{\eta} u dz)_x + (\int_{-h}^{\eta} v dz)_y]dy \tag{1.3.14}$$

$$\begin{aligned} &= (\int_r^l H dy)_t + (\int_r^l dy \int_{-h}^{\eta} u dz)_x \\ &-[Hl_z H_t + (\int_{-h}^{\eta} u dz)(l_x + l_z\eta_x) - \int_{-h}^{\eta} v dz]|_{y=l} \\ &+[Hr_z H_t + (\int_{-h}^{\eta} u dz)(r_x + r_z\eta_x) - \int_{-h}^{\eta} v dz]|_{y=r} \\ &= A_t + [\int_{-h}^{\eta} BU_2 dz]_x = 0 \end{aligned} \tag{1.3.15}$$

where $H = \eta + h$ is the total water depth and $A(x,\eta) = \int_r^l H dy = \int_{-h}^{\eta} dz \int_r^l dy$ is the cross section area.

In summary, then, the two-dimensional laterally averaged model is given by Equations (1.3.11), (1.3.13) and (1.3.15), that is

$$\begin{aligned}(BU_2)_t + (BU_2U_2)_x + (BU_2W_2)_z &= -gB\eta_2 + [\nu_2 B(U_2)_x]_x \\ &+ [\nu_2 B(U_2)_z]_z - \gamma U_2 \end{aligned} \tag{1.3.16}$$

$$(BU_2)_x + (BW_2)_z = 0 \tag{1.3.17}$$

$$A_t + [\int_{-h}^{\eta} BU_2 dz]_x = 0 \tag{1.3.18}$$

It is interesting to point out that when u, w and η are independent from y, Equations (1.3.16)-(1.3.18) can be derived from the $3D$ model Equations (1.2.11)-(1.2.10)

without any approximation: this is the case when the flow variables coincide with their laterally averaged values.

The boundary conditions at the free surface are specified by the prescribed wind stress as

$$\nu(U_2)_z = \gamma_T(u_a - U_2^s) \tag{1.3.19}$$

and the boundary conditions at the sediment-water interface are given by specifying the bottom stress as

$$\nu(U_2)_z = \gamma_B U_2^b \tag{1.3.20}$$

where γ_B is a non-negative friction coefficient. Typically, γ_B is taken to be $\gamma_B = \frac{g|U_2|}{C^2}$ [11]. With properly specified initial and boundary conditions, Equations (1.3.16)-(1.3.18) form a $2D_{xz}$ model used to simulate shallow water flow in estuarine environment.

1.4 Last step: the 1D Saint Venant Equations

The one dimensional equations for unsteady flow in open channel can be derived by integrating Equations (1.3.16)-(1.3.18) from the sea bed $z = -h$ to the free surface $z = \eta$.

Specifically, defining the cross sectional averaged velocity as $U = \frac{1}{A}\int_{-h}^{\eta} dz \int_r^l u dy$, Equation (1.3.18) becomes

$$A_t + (AU)_x = 0 \tag{1.4.1}$$

Moreover, the vertical integration of Equation (1.3.16) and the application of the boundary conditions (1.3.3)-(1.3.4) yield

$$(AU)_t + (AUU)_x = -gA(\overline{\eta})_x + (\overline{\nu}AU_x)_x + \gamma_T u_a - \gamma U \tag{1.4.2}$$

were $\gamma = \gamma_T + \gamma_B$ and $\overline{\nu} = \frac{1}{A}\int_r^l dy \int_{-h}^{\eta} \nu dz$.

Often, in the current literature [11], Equation (1.4.2) is rewritten as

$$(AU)_t + (AUU + gI_1)_x = gA(S_0 - S_f) + gI_2 + (\overline{\nu}AU_x)_x \tag{1.4.3}$$

where $S_0 = (-h)_x$ is the bed slope, S_f is the friction slope and

$$I_1(x, H) = \int_{-h}^{\eta} (H - z)B dz \tag{1.4.4}$$

In particular, I_2 represents the integral of a reaction force from hydrostatic pressure acting on the boundary and I_1 is a term linked to the hydrostatic forces over the cross-section such that

$$(gI_1)_x = gAH_x + gI_2. \tag{1.4.5}$$

Equation (1.4.1) is called Continuity Equation and expresses the conservation of the fluid volume.

Equation (1.4.2) as well as (1.4.3) is called Momentum Equation. In particular, studying Equation (1.4.3), one can see that it expresses the strict conservation of the momentum $Q = AU$ if and only if its right hand side is equal to zero. When the right hand side is different from zero, momentum is no longer conserved and the free terms act as momentum sources or momentum sinks.

Equations (1.4.1)-(1.4.2) are called the Saint Venant Equations. Regarding the notation, $\overline{\eta}$ will be replaced by η in the following of this work.

1.5 Hyperbolicity and the Saint Venant system

In the first part of this Section, we present some definitions and elementary properties of a particular class of equations, the hyperbolic conservation laws with source terms.

Actually, this kind of equations are particularly interesting in the development of this work, because the Saint Venant Equations reduce to a hyperbolic system of conservations laws in case the effects of the viscosity ν are neglected.

1.5.1 Hyperbolic systems

Conservation laws are systems of PDEs that can be written in the form

$$\mathbf{W}_t + \mathbf{F}(x, \mathbf{W})_x = \mathbf{b}(x, \mathbf{W}) \tag{1.5.1}$$

where $\mathbf{W}$ is the vector of the conserved quantities and $\mathbf{F}$ is a flux function. Assume m the dimension of the system.

If $\mathbf{b} \leq 0$, system (1.5.1) is homogeneous, otherwise it is said to be a system of conservation laws with source terms.

Actually, respect to our purposes, conservation laws of the form (1.5.1) can be rewritten in a more useful way by applying the chain rule to the derivative of the

flux function as follows

$$\frac{d\mathbf{F}}{dx} = \frac{\partial \mathbf{F}}{\partial \mathbf{W}} \frac{\partial \mathbf{W}}{\partial x} + \frac{\partial \mathbf{F}}{\partial x}, \tag{1.5.2}$$

Hence (1.5.1) becomes

$$\mathbf{W}_t + \mathbf{J}\mathbf{W}_x = \mathbf{b}(x, \mathbf{W}) \tag{1.5.3}$$

where $\mathbf{b}(x, \mathbf{W}) = \mathbf{b}(x, \mathbf{W}) - \frac{\partial \mathbf{F}}{\partial x}$ is the new source term and matrix $\mathbf{J} = \mathbf{J}(\mathbf{W}) = \frac{\partial \mathbf{F}}{\partial \mathbf{W}}$ is said the Jacobian of the flux function $\mathbf{F}(\mathbf{W})$.

A system (1.5.3) is said to be hyperbolic at (x, t) if all the eigenvalues λ_i of matrix $\mathbf{J}$ are real and if all its eigenvectors $\mathbf{K}^{(i)}$ are linearly independent. Moreover, this system is said to be strictly hyperbolic if all the eigenvalues λ_i are distinct.

1.5.2 Characteristic curves

The simplest PDE of hyperbolic type is the linear advection equation

$$w_t + a w_x = 0 \tag{1.5.4}$$

where a is a constant wave propagation speed.

From the study of this simple equation one can derive an important characterization extendable also to a more general hyperbolic system of PDE.

For each scalar equation such as (1.5.4) one can introduce the characteristic curve $x = x(t)$ as that curve in the (t, x) plane along which the PDE becomes an ODE.

Consider $x = x(t)$ and regard w as a function of t, that is $w = w(x(t), t)$. The rate of change of w along $x = x(t)$ is

$$\frac{dw}{dt} = \frac{\partial w}{\partial t} + \frac{\partial x}{\partial t} \frac{\partial w}{\partial x} \tag{1.5.5}$$

If the characteristic curve $x = x(t)$ satisfies the ODE

$$\frac{dx}{dt} = a \tag{1.5.6}$$

then the PDE (1.5.4) together with (1.5.5) and (1.5.6) gives

$$\frac{dw}{dt} = \frac{\partial w}{\partial t} + a \frac{\partial w}{\partial x} = 0 \tag{1.5.7}$$

Therefore the rate of change of w along the characteristic curve $x = x(t)$ satisfying (1.5.6) is zero, that is w is constant along the curve $x = x(t)$.

The speed a in (1.5.6) is called characteristic speed and it is the slope of the curve $x = x(t)$ in the (t, x) plane.

In order to extend these properties to a hyperbolic system of PDEs, let consider a hyperbolic system of the form (1.5.1) with a constant Jacobian matrix $\mathbf{J}$.

Given λ_i its real eigenvalues and $\mathbf{K}^{(i)}$ its linearly independent eigenvectors, it is possible to verify that matrix $\mathbf{J}$ is diagonalisable, that means $\mathbf{J}$ can be expressed as

$$\mathbf{J} = \mathbf{K}\Lambda\mathbf{K}^{-1} \tag{1.5.8}$$

in terms of the diagonal matrix

$$\Lambda = diag\{\lambda_1, ...\lambda_m\}$$

and a constant matrix

$$\mathbf{K} = \left[K^{(1)}, ...K^{(m)}\right]$$

.

Therefore, defining new variables $\mathbf{W} = \mathbf{KU}$ and manipulating system (1.5.1), one has

$$\begin{aligned} \mathbf{KU}_t + \mathbf{JKU}_x &= \mathbf{b} \\ \mathbf{K}^{-1}\mathbf{KU}_t + \mathbf{K}^{-1}\mathbf{JKU}_x &= \mathbf{K}^{-1}\mathbf{b} \\ \mathbf{U}_t + \Lambda\mathbf{U}_x &= \widetilde{\mathbf{b}} \end{aligned} \tag{1.5.9}$$

that is called the canonical form of system (1.5.1).

The system is therefore decoupled in m linear advection equations and its characteristic speeds λ_i define m characteristic curves satisfying the m ODEs

$$\frac{dx}{dt} = \lambda_i \qquad i = 1, ...m \tag{1.5.10}$$

1.5.3 Hyperbolic form of the Saint Venant system

The Saint Venant Equations (1.4.1) and (1.4.3) for flows in channels with arbitrary cross-sections take the form (1.5.1).

They constitute a system of two partial differential equations that can be written in the matrix notation (1.5.1) where

$$\mathbf{W} = \begin{pmatrix} A \\ AU \end{pmatrix}, \quad \mathbf{J} = \begin{pmatrix} 0 & 1 \\ c^2 - u^2 & 2U \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ gA(S_0 - S_f) + gI_2 \end{pmatrix} \tag{1.5.11}$$

and c is the wave celerity given by

$$c = \sqrt{\frac{gA}{B}}.$$

Studying the characteristic polynomial of $\mathbf{J}$, one can prove that system (1.5.1) is hyperbolic. In fact, its eigenvalues

$$\lambda_1 = U - c$$

$$\lambda_2 = U + c$$

are real, distinct and correspond to the following right-eigenvectors respectively

$$\mathbf{r}_1 = \begin{pmatrix} 1 \\ u - c \end{pmatrix}, \qquad \mathbf{r}_2 = \begin{pmatrix} 1 \\ u + c \end{pmatrix}. \tag{1.5.12}$$

Therefore, the Saint Venant system can be decomposed in two ODEs that hold along the two characteristic curves given by

$$\frac{dx}{dt} = \lambda_{1,2}.$$

1.5.4 Flow classification and boundary conditions

Given the characteristic speeds, one can classify the flow according to an adimensional parameter called the Froude number and defined as

$$Fr = \frac{|U|}{c}.$$

In the case $Fr < 1$, that means $|U| < c$, the two characteristic speeds have opposite directions. Therefore, the information is transmitted along these curves both upstream and downstream. This kind of flow is known as subcritical flow and occurs when the gravitational forces are dominant over the inertial ones.

In the case $Fr > 1$, that means $|U| > c$, the two characteristic speeds have the same direction of U. Therefore the information is only transmitted downstream. This kind of flow is known as supercritical flow and it occurs when the inertial forces are dominant over the gravitational ones.

Finally, in the case $Fr = 1$, that means $|U| = c$, one characteristic speed is vertical and the other has the same direction of U. This kind of flow is known as critical

flow and occurs when the inertial forces and the gravitational forces are perfectly balanced.

Characteristic theory also suggests the initial and the boundary conditions required in order to have a well-posed problem.

A general rule to consider is the following: "the number of boundary conditions should be equal to the number of characteristic curves entering the domain."

Consider the Saint Venant Equations and assume that $U > 0$. Therefore, $\lambda_2 > 0$ and one variable has to be specified at the inflow for either supercritical or subcritical flows.

Moreover, if the flow is supercritical at the inflow, thus $\lambda_1 > 0$ and another variable has to be specified at the inflow.

On the other hand, if the flow is subcritical at the outflow, thus $\lambda_1 < 0$ and a variable has to be specified at the outflow.

For $t = 0$, since both the characteristics always enter the domain, two independent variables must be always specified. These values are the initial conditions for the problem.

1.6 The resistance laws

The friction slope S_f is used to model the effects due to boundary friction and turbulence and it is usually written in the following form

$$S_f = \frac{Q\,|Q|}{K^2} \tag{1.6.1}$$

where K is a quantity called the conveyance. One of the most widely used form for the conveyance can be expressed by:

$$K = \frac{A^{k_1}}{n_M P^{k_2}}, \tag{1.6.2}$$

where n_M is a positive constant which represents the bed roughness [11],

$$P = B(x,0) + \int_0^H \sqrt{(4 + (B^s)^2)}d\eta \tag{1.6.3}$$

is the wetted perimeter, B^s is the channel width at the free surface and k_1 and k_2 are positive and real constants.

The friction slope S_f can be expressed using the Manning's law with $k_1 = 5/3$, $k_2 = 5/3$ and n_M the Manning friction coefficient [11].

With $k_1 = 3/2$ and $k_2 = 1/2$ one obtains the Chezy formula where $C = 1/n_M$ is the Chezy friction coefficient [11].

These laws are empirical and where originally developed for use with steady state flow [3, 1, 7].

More detailed information about these and other friction laws can be found in [13, 30].

1.7 An energy head formulation for the Momentum Equation

Equations (1.4.1)-(1.4.2) express the conservation of fluid volume and momentum.

Actually, in accordance with the concepts of classical hydraulics [11], in order to provide a complete model for channel flows that deals properly with these phenomena, the Momentum Equation and in particular its advection term should be formulated in such a way to conserve both momentum and energy head.

Strelkoff [48] pointed out that the governing equations developed using the momentum principle is different from those derived based on the energy approach.

Considering the three dimensional flow equations, even though originally both principles are established from Newton's second law of motion, the Momentum Equation is a vectorial relationship in which only the component of the velocity along the direction being considered affects the momentum balance.

On the other hand, the energy equation is a scalar relationship where all the three components of the flow velocity are involved.

Moreover, the energy approach incorporates a term to account for internal losses that it is completely different from the one which is included in the Momentum Equation for external resistance. Chow [13] described that the friction slope in the Momentum Equation stands for the resistance due to external boundary stresses, whereas in the Energy Equation the dissipated energy gradient accounts for the energy dissipation due to internal stresses working over a velocity gradient field.

For the one dimensional flow equations, the energy head conserving Equation is given by

$$U_t + \left(\frac{U^2}{2} + g\eta\right)_x + \gamma U = 0 \tag{1.7.1}$$

that, for steady flows and for frictionless channels, expresses the precise constancy of

the energy head function

$$E = \eta + \frac{U^2}{2g} \tag{1.7.2}$$

In the one dimensional context, the difference between momentum and energy approaches is reflected by the velocity distribution correction factors as well as by certain terms.

Actually, the momentum (1.4.2) and the energy head conserving formulation (1.7.1) are completely equivalent for continuous and sufficiently smooth solutions.

In fact, in case of gradually varied flow situations, the internal energy losses appear to be identical with the losses due to external forces and also the difference between the two velocity correction factors are very small and can be ignored [13]. This indicates that both principles can give an almost identical governing equation for the solution of this type of flow problem.

Moreover, in uniform flows, the rate with which surface forces are doing work is equal to the rate of energy dissipation. In such case, the frictional loss term have identical values.

For the case of rapidly varied flows and at local discontinuities, however, the two principles give flow equations which incorporate different correction factors for the effects of the curvature of the streamlines. Local discontinuities can either be due to discontinuities in the bathymetry or to the effects of bores generated in dam break problems or near hydraulic jumps. Since such flows occur in a short reach of the channel, the frictional losses due to external forces are insignificant.

In general, in order to connect Equation (1.7.1) or (1.4.2) at both sides of the discontinuity, conservation of mass and momentum provide the internal boundary conditions, although, in case of converging flows and steep bottom gradients, conservation of energy head can be applied as well (see, for example, [12]).

According to some authors [52], energy head conservation should be used only if solutions are smooth. For proper shocks speeds and locations, the momentum balance has to be applied.

However, when the discontinuities are not due to shock formation but to the bathymetry and the flow is converging, it is still possible to impose momentum conservation throughout, but energy head conservation is a better assumption [12].

The reason to change the conservation principle depending on the physical conditions can be also explained in terms of energy loss.

In a sudden channel expansion the energy head losses are to be derived from the application of the momentum principle and can be quantified as a function symmetric in the Froude number Fr [11, 46].

This means that if dissipation of energy occurs near expansions then, like wise, increase of energy is obtained near contractions.

This result is totally wrong from a realistic and physical point of view and suggests the use of a combined approach, that is the application of the momentum principle only in expansions and the energy head balance in sudden contractions [46].

Therefore, in the following chapters, both the energy head and the momentum conserving formulation of the Momentum Equation will be modelled and used depending on the local flow conditions.

2

A high resolution scheme for 1D flows in open channels with arbitrary cross-section

The aim of this chapter is to present a numerical scheme to simulate unsteady, one dimensional flows in open channels with arbitrary cross-section. This scheme is fully conservative of volume and switches between momentum and energy head conservation depending on local flow conditions. The derived finite volume method is semi-implicit in time and based on a space staggered grid. A high resolution technique, the flux limiter method, is implemented to control the accuracy of the proposed scheme. Our purpose is to achieve the precision and the stability of the method with respect to the regularity of the data. In addition, a new flux limiter is described and implemented to allow accurate flow simulations near hydraulic structures such as weirs.

2.1 Introduction

The current literature describes several numerical techniques that are suitable for solving Equations (1.4.1), (1.4.2) and (1.7.1). These include the method of characteristics, explicit difference methods, fully implicit methods, Godunov methods [27] and semi-implicit methods [6].

In particular, the method of characteristics is very efficient in the treatment of boundary conditions, but does not guarantee volume and momentum conservation.

The Godunov's type methods (see, e.g., [52]) instead, require the solution of local Riemann problems and, consequently, are very effective on simple channel geometries with flat, horizontal bottom and rectangular cross-section. For space varying bottom

profiles, however, the bottom slope appears as a source term that may generate artificial flows [53] unless specific treatments of the geometrical source terms are implemented [21, 55]. Moreover, Godunov's type methods are explicit in time and, accordingly, the allowed time step is restricted by a C.F.L. stability condition, which relates the time step to the spatial discretization and the wave speed. These kinds of methods are in general based upon non-staggered grids and can achieve higher that first-order accuracy. The Godunov's type methods were originally developed for gas dynamic and only later extended to hydrodynamic on the basis of the analogy between the equations for isentropic flow of a perfect gas with constant specific heat and the shallow water Equations [47, 52].

Alternatively, semi-implicit methods (see, e.g., [6, 7, 10, 3]) can be unconditionally stable and computationally efficient. In particular, a semi-implicit method that conserves the fluid volume when applied to channels with arbitrary cross-sections was firstly introduced and presented in [10]. These methods, however, when do not satisfy the physical conservation property of momentum, may produce incorrect results if applied to extreme problems having a discontinuous solution. Actually, the semi-implicit scheme proposed in [3] as well as that presented by Stelling in [46] combine the efficiency of staggered grids with the conservation of both fluid volume and momentum and can be applied to problems including rapidly varying flows.

In the present chapter a numerical technique to solve Equations (1.4.1), (1.4.2) and (1.7.1) is derived and discussed.

This technique is first order accurate, fully conservative of volume, both locally and globally, and switches between momentum and energy head conservation depending on local flow conditions (see Reference [46] for details), satisfying a correct momentum balance near large gradients.

Moreover, under a suitable constraint on the time interval, it ensures the non-negativity of the water volume, so allowing a correct solution of problems presenting flooding and drying.

A high-resolution method, the flux limiter method, is implemented to control the accuracy of the proposed scheme: our purpose is to achieve the precision and the stability of the method with respect to the regularity of the data.

In addition, a special flux limiter is formulated, described and implemented to allow accurate flow simulations near hydraulic structures such as weirs.

Finally, a proper semi-implicit discretization leads to a scheme that is relatively

simple and highly accurate, even if the C.F.L. condition is violated.

2.2 Time and space discretization

In order to obtain a computationally efficient numerical method that does not suffer from stability problems, the time discretization is chosen to be semi-implicit, that means that only some terms in the governing equations are discretized implicitly.

The determination of the specific form of the semi-implicit discretization follows directly from the analysis of the hyperbolic system (1.4.1)-(1.4.2) and from the study of the Courant, Friedrichs and Lewy (C.F.L.) stability condition [6, 40, 39]

$$\Delta t \leq \frac{\Delta x}{max\left\{\lambda_1, \lambda_1\right\}} = \frac{\Delta x}{|U| + \sqrt{\frac{gA}{B}}} \tag{2.2.1}$$

for explicit numerical methods. This restriction is sufficient, but not necessary and thus it usually requires a much smaller time step than that permitted by accuracy considerations.

On the other hand, a fully implicit discretization of the governing equations leads to methods which are unconditionally stable, but that involve the simultaneous solution of a large number of coupled non-linear equations. Moreover, from the ponit of view of the accuracy, the time step cannot be taken arbitrarily large and therefore these methods often become impractical.

In order to propose a compromise between the explicit and the implicit time discretization, the semi-implicit one seems to be a valid answer.

For simplicity, the derivation of the specific form of the semi-implicit discretization will be carried out assuming that the channel has a rectangular cross-section of constant width B so that the cross-sectional area A is simply given by $A = A(x,t) = BH(x,t)$. Moreover, Equations (1.4.1)-(1.4.2) can be written out in a more extended non conservative form as

$$\eta_t + U\eta_x + HU_x = -Uh_x \tag{2.2.2}$$

$$U_t + UU_x + g\eta_x = \frac{1}{H}(\overline{\nu}HU_x)_x + \frac{(\gamma_T u_a - \gamma U)}{H} \tag{2.2.3}$$

These equations in matrix notation can be written in the form (1.5.3) where

$$\mathbf{W} = \begin{pmatrix} \eta \\ U \end{pmatrix}, \qquad \mathbf{J} = \begin{pmatrix} H & U \\ U & g \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} -Uh_x \\ \frac{1}{H}(\overline{\nu}HU_x)_x + \frac{(\gamma_T u_a - \gamma U)}{H} \end{pmatrix} \tag{2.2.4}$$

When $\nu = 0$ the system of Equations (1.5.3) is strictly hyperbolic and the corresponding characteristic speeds are $\lambda_{1,2} = U \pm \sqrt{gH}$, that clearly depend only on the fluid velocity U and upon the celerity $\sqrt{gH}$. Note that when $|U| << \sqrt{gH}$ the flow is strictly subcritical and the characteristic speeds $\lambda_{1,2}$ have opposite directions. More importantly, the dominant term $\sqrt{gH}$ arises from the off diagonal terms g and H in the matrix $\mathbf{J}$. These are the coefficient of η_x in the Momentum Equation and of U_x in the Continuity Equation. Therefore, these derivatives must be discretized implicitly in order for the stability of the method to be independent of the celerity $\sqrt{gH}$.

Actually, in the schemes presented in the following, the θ-method will be used instead of the implicit one, with θ in $[\frac{1}{2}, 1]$ for stability reasons. The remaining terms will be discretized explicitly.

From the point of view of the space discretization, Equations (1.4.1) and (1.4.2) are discretized in the spatial interval $[0, L]$ on a space staggered grid whose nodes are denoted by x_i and $x_{i+1/2}$. The discrete velocity $U_{i+1/2}$ (or the discharge $Q_{i+1/2} = A_{i+1/2}U_{i+1/2}$) is defined at half integer nodes, while the discrete surface elevation η_i as well af the total water depth H_i, assumed to be constant in the interval $[x_{i-1/2}, x_{i+1/2}]$, are defined at integer nodes. The grid intervals are denoted by $\Delta x_i = x_{i+1/2} - x_{i-1/2}$ and $\Delta x_{i+1/2} = \frac{\Delta x_{i+1} + \Delta x_i}{2}$.

The time interval is taken to be Δt.

2.3 Discretization of the Continuity Equation

The Continuity Equation (1.4.1) expresses the physical law of conservation of volume and it is discretized by a finite volume method in space and by the θ-method in time [10].

Directly from the specifications of the previous section, the discretization of the Continuity Equation follows from the integration in space of (1.4.1)

$$A_t + Q_x = 0 \tag{2.3.1}$$

over the interval $[x_{i-1/2}, x_{i+1/2}]$

$$\int_{x_{i-1/2}}^{x_{i+1/2}} [A_t + Q_x] dx = \frac{\partial}{\partial t} \int_{x_{i-1/2}}^{x_{i+1/2}} A dx + \frac{\partial}{\partial x} \int_{x_{i-1/2}}^{x_{i+1/2}} Q dx = 0 \tag{2.3.2}$$

that leads to

$$\frac{\partial}{\partial t} V_i(\eta_i) + [Q(x_{i+1/2}) - Q(x_{i-1/2})] = 0 \tag{2.3.3}$$

and from the semi-implicit discretization in time

$$V_i(\eta_i^{n+1}) = V_i(\eta_i^n) - \Delta t[Q_{i+1/2}^{n+\theta} - Q_{i-1/2}^{n+\theta}], \tag{2.3.4}$$

where the fluid volume $V_i(\eta_i) = \int_{x_{i-1/2}}^{x_{i+1/2}} A dx$ is, in general, a non linear function of η and $Q^{n+\theta} = \theta Q^{n+1} + (1-\theta)Q^n$.

From the point of view of the time discretization, the discharge is defined as $Q^n = A^n U^n$.

Equation (2.3.4) obviously expresses a discrete conservation of fluid volume.

The particular attention given here to volume conservation is justified by the importance of this conservation when the channel has a non-rectangular cross-section. In this case, traditional numerical methods (and even the Godunov's type methods) apply a linearization technique to the non linear function V in Equation (2.3.4). Specifically,

$$V_i(\eta_i^{n+1}) \approx V_i(\eta_i^n) + \frac{\partial V_i(\eta_i^n)}{\partial \eta}(\eta_i^{n+1} - \eta_i^n), \tag{2.3.5}$$

where $\frac{\partial V_i(\eta_i^n)}{\partial \eta}$ represents the surface area between $x_{i-1/2}$ and $x_{i+1/2}$.

Substitution of (2.3.5) into (2.3.4) yields

$$\frac{\partial V_i(\eta_i^n)}{\partial \eta}(\eta_i^{n+1} - \eta_i^n) + \Delta t[Q_{i+1/2}^{n+\theta} - Q_{i-1/2}^{n+\theta}] = 0, \tag{2.3.6}$$

where the term $\frac{\partial V_i(\eta_i^n)}{\partial \eta}(\eta_i^{n+1} - \eta_i^n)$ is no longer the volume variation unless $\frac{\partial V_i}{\partial \eta}$ is a constant. This is the case, e.g., for channels with rectangular cross-section. In general, however, the linearized Equation (2.3.6) or similar linearizations will not guarantee volume conservation and an artificial loss or creation of mass may result.

In Chapter 3, a wet bed dam break in an open channel with triangular cross section is presented.

2.3.1 Definition of η and $-h$ at $i + 1/2$

From the point of view of the spatial discretization, the discharge is defined as $Q_{i+1/2} = A_{i+1/2} U_{i+1/2}$.

Therefore, remembering the definition of the cross-sectional area A as $A = A(x, \eta(x,t))$, the variable η and the bottom $-h$, are initially defined at integers nodes, it is necessary to define explicitly their value at the half integer node $i + 1/2$.

To do this, the following upwind rule based on the sign of the discharge $Q_{i+1/2}$ is used for the definition of η

$$\eta_{i+1/2} = \begin{cases} \eta_i & \text{if} \quad Q_{i+1/2} \geq 0 \\ \eta_{i+1} & \text{if} \quad Q_{i+1/2} < 0 \end{cases} \tag{2.3.7}$$

while the value of the bottom $-h_{i+1/2}$ is given by

$$-h_{i+1/2} = min(-h_i, -h_{i+1}). \tag{2.3.8}$$

except for the case we can analytically express it as $-h_{i+1/2} = -h(x_{i+1/2})$.

Definition (2.3.8) can be justified as follows.

Assume that in the middle of a channel there is a sill $1m$ height and with vertical walls, that is the slopes of the sill are abrupt within one grid cell (see, e.g., the bottom of the channel in the example presented in Section 3.3). Assume that the bottom is discretized in such a way only one point of the sill's crest is detected, that is $-h_{i-1} = 0m$, $-h_i = 1m$ and $-h_{i+1} = 0m$.

Using Equation (2.3.8), the bottom at $i \pm 1/2$ is given by

$$-h_{i-1/2} = -h_{i+1/2} = 1m$$

and thus a crest has appeared in the bottom profile, giving the correct description of the channel's geometry.

The introduction of a different choice for the definition of the bottom at $i \pm 1/2$ could lead to incorrect results.

For example, applying an average, the bottom at $i \pm 1/2$ is given by the following expressions

$$-h_{i-1/2} = \frac{(-h_i + h_{i-1})}{\Delta x}(x_{i-1/2} - x_{i-1}) - h_{i-1}$$
$$-h_{i+1/2} = \frac{(-h_{i+1} + h_i)}{\Delta x}(x_{i+1/2} - x_{i+1}) - h_i$$

that describe the numerical bottom profile as smooth between $-h_{i-1}$ and $-h_i$ and between $-h_i$ and $-h_{i+1}$.

Alternatively, introducing an upwind rule based on the sign of the discharge

$$-h_{i+1/2} = \begin{cases} -h_i & \text{if} \quad Q_{i+1/2} \geq 0 \\ -h_{i+1} & \text{if} \quad Q_{i+1/2} < 0 \end{cases}, \tag{2.3.9}$$

the bottom profile displays a sill with a crest whose length varies between one point and the space interval Δx_i, depending on the value of the discharge field.

2.4 Discretization of the Momentum Equation

In order to formulate a correct scheme for the Momentum Equation, not only numerical guidelines have to be considered, but also the physical considerations presented in Section 1.7.

In fact, near local discontinuities in the solution, following from, for example, sharp bottom gradients or hydraulic jumps, the order of accuracy concept is meaningless. Conservation properties are more important aspects in such situations.

An energy head and a momentum conservative approximation of the Momentum Equation are presented in the following subsections.

A switch between the two forms is formulated in such a way that energy head can be chosen for converging flows (such as strong contractions) and the momentum for diverging flows.

2.4.1 First formulation: conservation of the momentum

Equation (1.4.2)

$$Q_t + (UQ)_x + gA\eta_x + \gamma U = 0 \tag{2.4.1}$$

is discretized with a conservative method in order to obtain a physically correct solution also under extreme circumstances.

The formulation presented in this section is called Q-formulation, meaning that it will be solved in the variable Q.

Specifically, this scheme is given by centred finite differences for the integration in space of water surface elevation and the semi-implicit method for the time integration (see, e.g., [6, 7, 8, 10]).

Based on the discussion of Section 2.2, the gradient of the free surface elevation will be discretized with the θ-method, while the convective term will be discretized explicitly. For stability, the friction term will be discretized implicitly, but the friction coefficient γ will be evaluated explicitly so that the resulting algebraic system to be solved will be linear.

Finally, the resulting discretization of the Momentum Equation is the following:

$$\begin{aligned} \frac{Q^{n+1}_{i+1/2} - Q^n_{i+1/2}}{\Delta t} + \frac{(UQ)^n_{i+1} - (UQ)^n_i}{\Delta x} + gA^n_{i+1/2}\frac{\eta^{n+\theta}_{i+1} - \eta^{n+\theta}_i}{\Delta x} \quad + \\ \frac{\gamma^n_{i+1/2}}{A^n_{i+1/2}} Q^{n+1}_{i+1/2} \quad = \quad 0 \end{aligned} \tag{2.4.2}$$

that is

$$(1+\frac{\gamma^n_{i+1/2}}{A^n_{i+1/2}}\Delta t)Q^{n+1}_{i+1/2}+gA^n_{i+1/2}\theta\frac{\Delta t}{\Delta x_{i+1/2}}(\eta^{n+1}_{i+1}-\eta^{n+1}_i)=F^n_{i+1/2} \tag{2.4.3}$$

where

$$F^n_{i+1/2}=Q^n_{i+1/2}-\Delta t\frac{[(UQ)^n_{i+1}-(UQ)^n_i]}{\Delta x_{i+1/2}}-gA^n_{i+1/2}(1-\theta)\Delta t\frac{(\eta^n_{i+1}-\eta^n_i)}{\Delta x_{i+1/2}} \tag{2.4.4}$$

is a finite difference operator including the explicit discretizations of the advective and the free surface slope terms.

Regarding the time discretization, one can note that the θ-method has been used for the free surface slope term, the friction has been taken implicitly, while the other terms have been discretized explicitly.

Moreover, the cross-sectional area that multiplies the free surface slope term can be defined at the half integer node $i+1/2$ as $A^n_{i+1/2}=A(x_{i+1/2},\frac{\eta^n_{i+1}+\eta^n_i}{2})$.

Here, it is worth noting that in case of a frictionless channel with rectangular cross-section and flat bottom one has $A(x,\eta)=BH=B(h+\eta)$, where B is the channel width and $-h=constant$ is the channel depth when $\eta=0$.

In this case, Equation (2.4.2) can be regarded as being the semi-implicit time discretization of

$$\frac{dQ_{i+1/2}}{dt}+\frac{(UQ)_{i+1}-(UQ)_i}{\Delta x_{i+1/2}}=-gB\frac{(H_{i+1}+H_i)}{2}\frac{(H_{i+1}-H_i)}{\Delta x_{i+1/2}} \tag{2.4.5}$$

or, equivalently,

$$\frac{dQ_{i+1/2}}{dt}+\frac{(UQ)_{i+1}-(UQ)_i}{\Delta x_{i+1/2}}=-\frac{gB}{2}\frac{(H^2_{i+1}-H^2_i)}{\Delta x_{i+1/2}}. \tag{2.4.6}$$

Interestingly enough, even though the given Momentum Equation (1.4.2) is not written in conservative form, the resulting Equation (2.4.6) represents the precise momentum conservation because it is written in flux form (see, e.g., [46] for further details).

We shall then assume that the more general Equation (2.4.3) is conservative also in the more general case of channels with arbitrary cross-section and with varying bottom slope.

The advective term

The value of UQ at the integer node i, as required by F, may be computed using the following upwind rule based on the sign of the discharge average:

$$(UQ)_i^n = \frac{Q_{i+1/2}^n + Q_{i-1/2}^n}{2} \begin{cases} U_{i-1/2}^n & \text{if} \quad \frac{Q_{i+1/2}^n + Q_{i-1/2}^n}{2} \geq 0 \\ U_{i+1/2}^n & \text{if} \quad \frac{Q_{i+1/2}^n + Q_{i-1/2}^n}{2} < 0 \end{cases} \tag{2.4.7}$$

2.4.2 Second formulation: conservation of the energy head

In order to obtain an energy head conserving scheme expressed in the variable Q, it is convenient to add Equation (1.4.1) multiplied by U to Equation (1.7.1) multiplied by A to obtain

$$Q_t + UQ_x + \frac{1}{2}A(U^2)_x + gA\eta_x + \gamma U = 0. \tag{2.4.8}$$

The discretization in space and time of the reformulated energy head principle (2.4.8) is given by centred finite differences for the integration in space of water surface elevation and by the semi-implicit method presented in the previous Subsection for the time integration:

$$\begin{aligned} \frac{Q_{i+1/2}^{n+1} - Q_{i+1/2}^n}{\Delta t} + U_{i+1/2}^n \frac{Q_{i+1}^n - Q_i^n}{\Delta x_{i+1/2}} + A_{i+1/2}^n \frac{(U^2)_{i+1}^n - (U^2)_i^n}{2\Delta x_{i+1/2}} \quad + \\ gA_{i+1/2}^n \frac{\eta_{i+1}^{n+\theta} - \eta_i^{n+\theta}}{\Delta x_{i+1/2}} + \frac{\gamma_{i+1/2}^n}{A_{i+1/2}^n} Q_{i+1/2}^{n+1} \quad = \quad 0 \end{aligned} \tag{2.4.9}$$

that can be written as Equation (2.4.3) with $F_{i+1/2}^n$ defined as follows

$$\begin{aligned} F_{i+1/2}^n \quad &= \quad Q_{i+1/2}^n - \Delta t U_{i+1/2}^n \frac{Q_{i+1}^n - Q_i^n}{\Delta x_{i+1/2}} - \Delta t A_{i+1/2}^n \frac{(U^2)_{i+1}^n - (U^2)_i^n}{2\Delta x_{i+1/2}} \\ &- \quad gA_{i+1/2}^n (1-\theta)\Delta t \frac{(\eta_{i+1}^n - \eta_i^n)}{\Delta x_{i+1/2}} \end{aligned} \tag{2.4.10}$$

The advective term

The values of U and Q at the integer node i, as required by F, are computed using the following upwind rule based on the sign of the discharge average:

$$U_i, Q_i = \begin{cases} U_{i-1/2}, Q_{i-1/2} & \text{if} \quad \frac{Q_{i+1/2} + Q_{i-1/2}}{2} \geq 0 \\ U_{i+1/2}, Q_{i+1/2} & \text{if} \quad \frac{Q_{i+1/2} + Q_{i-1/2}}{2} < 0 \end{cases} \tag{2.4.11}$$

while the cross-sectional area $A_{i+1/2}$ that multiplies the $\frac{\partial U^2}{\partial x}$ term is defined as explained in Section 2.3.1.

2.5 Switching the conservation

Two possible approaches to the implementation of the switch between momentum and energy head conservation are proposed in this section.

Both of them are such that only a small part of scheme (2.4.2) has to be differently defined to obtain scheme (2.4.9) and therefore the implementation of the switch does not cause problems from the points of view of the computational cost and efficiency of the model.

In the first approach, the switch consists in the choice of the discretization of the advective term and assumes the following form

$$use \begin{cases} U^n_{i+1/2}\frac{Q^n_{i+1}-Q^n_i}{\Delta x_{i+1/2}} + A^n_{i+1/2}\frac{(U^2)^n_{i+1}-(U^2)^n_i}{\Delta x_{i+1/2}} & \text{if } \frac{u_{i+1/2}-u_{i-1/2}}{\Delta x} > \epsilon > 0 \\ \frac{(UQ)^n_{i+1}-(UQ)^n_i}{\Delta x_{i+1/2}} & \text{otherwise} \end{cases} \tag{2.5.1}$$

In the second approach, valid only for steady state flows, formulation (2.4.2) switches to formulation (2.4.9) changing the definition of the cross-sectional area $A_{i+1/2}$ that multiplies the free surface slope term.

In particular, consider the following expression

$$A_{i+1/2} = \frac{2A_{i+1}A_i}{A_{i+1}+A_i} = 2(\frac{1}{A_{i+1}} + \frac{1}{A_i})^{-1} \tag{2.5.2}$$

Dividing Equation (2.4.2) by the factor (2.5.2), one has

$$\frac{1}{2}(\frac{Q_{i+1}}{A_{i+1}} + \frac{Q_i}{A_i})\frac{U_{i+1}-U_i}{\Delta x} + g\frac{\eta_{i+1}-\eta_i}{\Delta x} = 0 \tag{2.5.3}$$

or equivalently, observing that Q is constant everywhere,

$$\frac{1}{2}\frac{U^2_{i+1}-U^2_i}{\Delta x} + g\frac{\eta_{i+1}-\eta_i}{\Delta x} = 0 \tag{2.5.4}$$

that is consistent with the Energy Head conserving Equation (1.7.1).

Thus, the switch of the second approach is the following

$$use\ A_{i+1/2} = \begin{cases} \frac{2A_{i+1}A_i}{A_{i+1}+A_i} & \text{if } \frac{u_{i+1/2}-u_{i-1/2}}{\Delta x} \geq \epsilon > 0 \\ \frac{A_i+A_{i+1}}{2} & \text{otherwise} \end{cases} \tag{2.5.5}$$

For its simplicity, one can decide to use switch (2.5.5) during the whole computation of steady state phenomena and therefore also in the transitions where, although not completely correct, it is still consistent.

2.6 The semi-implicit finite volume method for the SVE

The semi-implicit method obtained in this work for the discretization of the Saint Venant system takes the following form

$$V_i(\eta_i^{n+1}) = V_i(\eta_i^n) - \Delta t[Q_{i+1/2}^{n+\theta} - Q_{i-1/2}^{n+\theta}], \tag{2.6.1}$$

$$(1 + \frac{\gamma_{i+1/2}^n}{A_{i+1/2}^n}\Delta t)Q_{i+1/2}^{n+1} + gA_{i+1/2}^n\theta\frac{\Delta t}{\Delta x_{i+1/2}}(\eta_{i+1}^{n+1} - \eta_i^{n+1}) = F_{i+1/2}^n \tag{2.6.2}$$

Observe that the discretization of the Momentum Equation given by (2.6.2) expresses both the discrete conservation of the energy head and of the momentum depending on the definition of the explicit operator F.

2.7 Order of accuracy and consistency

The numerical method (2.6.1)-(2.6.2) is first order accurate.

Its order of accuracy can be verified through the analysis of the consistency of the method that requires that the original equations can be recovered from the algebraic ones: obviously this is a minimum requirement for any discretization.

Consider negative flow directions ($U < 0$, $Q < 0$) and a Taylor expansion of the individual terms in Equations (2.6.1)-(2.6.2)

$$V_{i+1}^{n+1} = V_i^n + \Delta t\left(\frac{\partial V}{\partial t}\right)_i^n + O(\Delta t^2) \tag{2.7.1}$$

$$Q_{i+1/2}^{n+\theta} = Q_{i+1/2}^n + \theta\Delta t\left(\frac{\partial Q}{\partial t}\right)_i^n + O(\Delta t^2) \tag{2.7.2}$$

$$Q_{i+3/2}^n = Q_{i+1/2}^n + \Delta x\left(\frac{\partial Q}{\partial x}\right)_{i+1/2}^n + O(\Delta x^2) \tag{2.7.3}$$

$$Q_{i+1/2}^n = Q_i^n + \frac{\Delta x}{2}\left(\frac{\partial Q}{\partial x}\right)_i^n + (\frac{\Delta x}{2})^2\frac{1}{2}\left(\frac{\partial Q}{\partial x}\right)_i^n + O(\Delta x^3) \tag{2.7.4}$$

$$Q_{i-1/2}^n = Q_i^n - \frac{\Delta x}{2}\left(\frac{\partial Q}{\partial x}\right)_i^n + (\frac{\Delta x}{2})^2\frac{1}{2}\left(\frac{\partial Q}{\partial x}\right)_i^n + O(\Delta x^3) \tag{2.7.5}$$

$$\eta_{i+1}^{n+\theta} = \eta_{i+1/2}^{n+\theta} + \frac{\Delta x}{2}\left(\frac{\partial \eta}{\partial x}\right)_{i+1/2}^{n+\theta} + (\frac{\Delta x}{2})^2\frac{1}{2}\left(\frac{\partial \eta}{\partial x}\right)_{i+1/2}^{n+\theta} + O(\Delta x^3) \tag{2.7.6}$$

$$\eta_i^{n+\theta} = \eta_{i+1/2}^{n+\theta} - \frac{\Delta x}{2}\left(\frac{\partial \eta}{\partial x}\right)_{i+1/2}^{n+\theta} + (\frac{\Delta x}{2})^2\frac{1}{2}\left(\frac{\partial \eta}{\partial x}\right)_{i+1/2}^{n+\theta} + O(\Delta x^3) \tag{2.7.7}$$

that yields

$$\left(\frac{\partial V}{\partial t}\right)_i^n + O(\Delta t^2) + \Delta x \left(\frac{\partial Q}{\partial x}\right)_{i+1/2}^{n+\theta} + O(\Delta x^3) = 0 \tag{2.7.8}$$

$$\begin{aligned}\left(\frac{\partial Q}{\partial t}\right)_{i+1/2}^n + O(\Delta t^2) \quad &+ \quad \left(\frac{\partial Q}{\partial x}\right)_{i+1/2}^n + O(\Delta x^2) \\ &+ \quad gA\left(\frac{\partial \eta}{\partial x}\right)_{i+1/2}^{n+\theta} + O(\Delta x^3) - \gamma U = 0\end{aligned} \tag{2.7.9}$$

Therefore, the semi-implicit numerical method (2.6.1)-(2.6.2) is first order both in space and in time. From the same expression it follows that this scheme is also consistence with the physical laws that it discretizes.

2.8 Stability of the method

The stability analysis of the semi-implicit method (2.6.1)-(2.6.2) will be carried out by using the von Neumann method under the assumption that our differential equations (1.4.1)-(1.4.2) are linear ($A = BH$), fully implicit in time and defined on an infinite spatial domain, or with periodic boundary conditions on a finite domain.

Consider $\theta = 1$. Hence, the difference Equations (2.6.1)-(2.6.2) reduce to

$$\eta_i^{n+1} = \eta_i^n - \frac{\Delta t}{\Delta x}[Q_{i+1/2}^{n+1} - Q_{i-1/2}^{n+1}], \tag{2.8.1}$$

$$(1 + \frac{\gamma}{H}\Delta t)Q_{i+1/2}^{n+1} + gBH\frac{\Delta t}{\Delta x}(\eta_{i+1}^{n+1} - \eta_i^{n+1}) = F_{i+1/2}^n \tag{2.8.2}$$

where the operator F has been assumed to be linear, all the coefficients H, γ and B have been assumed to be constants and in particular $B = 1$. Now, expressing the two equations in U form, one has

$$\eta_i^{n+1} = \eta_i^n - \frac{\Delta t}{\Delta x}H[U_{i+1/2}^{n+1} - U_{i-1/2}^{n+1}], \tag{2.8.3}$$

$$(H + \gamma\Delta t)U_{i+1/2}^{n+1} + gH\frac{\Delta t}{\Delta x}(\eta_{i+1}^{n+1} - \eta_i^{n+1}) = HF_{i+1/2}^n \tag{2.8.4}$$

Now, by changing variables U and η with $\overline{U} = \sqrt{(H + \gamma\Delta t)}U$ and $\overline{\eta} = \eta\sqrt{g}$, Equations (2.6.1)-(2.6.2) become

$$\overline{\eta}_i^{n+1} = \overline{\eta}_i^n - C^*[\overline{U}_{i+1/2}^{n+1} - \overline{U}_{i-1/2}^{n+1}], \tag{2.8.5}$$

$$\overline{U}_{i+1/2}^{n+1} + C^*(\overline{\eta}_{i+1}^{n+1} - \overline{\eta}_i^{n+1}) = \frac{HF_{i+1/2}^n}{\sqrt{(H+\gamma\Delta t)}} \tag{2.8.6}$$

where $C^* = \frac{\sqrt{g}\Delta t H}{\sqrt{(H+\gamma\Delta t)}\Delta x}$.

In order to analyze the stability of Equations (2.8.5)-(2.8.6) with the von Neumann method, a Fourier mode is introduced for each field variable $\overline{U}$ and $\overline{\eta}$ and the stability analysis is carried out on the corresponding amplitude functions. Specifically, $\overline{U}_{i+1/2}^n$ and $\overline{\eta}_i^n$ are replaced in (2.8.5)-(2.8.6) by $\widehat{U}^n e^{I(i+1/2)\alpha}$ and $\widehat{\eta}^n e^{Ii\alpha}$ respectively, where $\widehat{U}^n$ and $\widehat{\eta}^n$ are the amplitude functions of $\overline{U}$ and $\overline{\eta}$ at the time level t^n, $I = \sqrt{-1}$ and α is the phase angle. Thus, after substituting these expressions and dividing by $e^{I(i+1/2)\alpha}$, Equations (2.8.5)-(2.8.6) become

$$\widehat{\eta}_i^{n+1} = \widehat{\eta}_i^n - C^*\widehat{U}^{n+1}[e^{I\alpha/2} - e^{-I\alpha/2}], \tag{2.8.7}$$

$$\widehat{U}^{n+1} + C^*\widehat{\eta}^{n+1}[e^{I\alpha/2} - e^{-I\alpha/2}] = \frac{Hf}{\sqrt{(H+\gamma\Delta t)}}\widehat{U}^n \tag{2.8.8}$$

where f is the amplification factor of the linearized operator F. Since $e^{I\alpha/2} - e^{-I\alpha/2} = 2Isin(\alpha/2)$, by setting $p = 2C^*sin(\alpha/2)$, Equations (2.8.7)-(2.8.8) in matrix notation become

$$\mathbf{P}\widehat{\mathbf{W}}^{n+1} = \mathbf{Q}\widehat{\mathbf{W}}^n \tag{2.8.9}$$

where

$$\widehat{\mathbf{W}} = \begin{pmatrix} \widehat{U}^n \\ \widehat{\eta}^n \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} 1 & Ip \\ Ip & 1 \end{pmatrix}, \qquad \mathbf{Q} = \begin{pmatrix} \frac{Hf}{\sqrt{(H+\gamma\Delta t)}} & 0 \\ 1 & 0 \end{pmatrix} \tag{2.8.10}$$

Thus, the amplification matrix of the method is $\mathbf{G} = \mathbf{P}^{-1}\mathbf{Q}$ and a necessary and sufficient condition for stability is that $\|\mathbf{G}\|_2 \leq 1$ identically for every α. But, since $\|\mathbf{G}\|_2 \leq \|\mathbf{P}\|_2^{-1} \|\mathbf{Q}\|_2$, we are seeking the conditions for which $\|\mathbf{P}\|_2^{-1} \leq 1$ and $\|\mathbf{Q}\|_2 \leq 1$. Note now, that the two matrices $\mathbf{P}$ and $\mathbf{Q}$ and hence also $\mathbf{P}^{-1}$, are normal matrices; that is, they commute with their respective hermitian conjugate. Thus, the norms of $\mathbf{P}^{-1}$ and of $\mathbf{Q}$ are equal to their respective spectral radius. But, the eigenvalues of $\mathbf{P}$ are

$$\lambda_P = 1 \pm I\,|p|$$

and thus the spectral radius of $\mathbf{P}^{-1}$ is always no greater than unity. Next, the eigenvalues of $\mathbf{Q}$ are

$$\lambda_Q = 1 \qquad \lambda_Q = \frac{Hf}{(h + \gamma \Delta t)}$$

Hence, in order for spectral radius of $\mathbf{Q}$ not to exceed unity, it is sufficient that

$$|f| \leq 1$$

identically for every α. Thus the stability of the semi-implicit method (2.6.1)-(2.6.2) depends only on the choice of the difference operator F used to discretize the convective and the viscous terms.

For example, using an explicit upwind discretization, the stability restriction on the time step is given by

$$\Delta t \leq \frac{\Delta x^2}{|U|\,\Delta x + 2\nu} \tag{2.8.11}$$

If instead, an Eulerian-Lagrangian discretization is used, then the stability restriction on the time step reduces to

$$\Delta t \leq \frac{\Delta x^2}{2\nu} \tag{2.8.12}$$

2.9 Numerical accuracy and high-resolution

The numerical method proposed in this work is only first order accurate. In general, all first order schemes suffer from numerical dissipation and all second order schemes suffer from artificial dispersion, which creates oscillations around discontinuities.

In order to improve the accuracy of the method without running into stability problems but leading it to satisfy the TVD property [53, 59], we introduce a particular class of high-resolution methods.

A high-resolution method can be characterized with the following properties [25]:

- it provides at least second order of accuracy in smooth areas of the flow.
- it produces numerical solutions (relatively) free from spurious oscillations
- in the case of discontinuities, the number of grid points in the transition zone containing the shock wave is smaller in comparison with that of first-order monotone methods

The motivation for the development of high-resolution methods emerges from our effort to circumvent Godunovs theorem [23] that states: *There are no monotone, linear schemes for the linear advection equation of second or higher order of accuracy.*

In other words, second-order accuracy and monotonicity are contradictory requirements. The key to circumvent Godunovs theorem lies on the assumption made in the theorem that the schemes are linear. Therefore, if we want to design a method which provides at least second order of accuracy and at the same time avoids spurious oscillations near large gradients, then we need to develop non-linear methods.

Limiters are the general non-linear mechanism that distinguishes modern methods from classical linear schemes. These are sometimes referred to as flux limiters or slope limiters, but their role is similar: they act as a non-linear switch between more than one linear methods and choice the numerical method to be used depending on the behaviour of the local solution.

Limiters result in non-linear methods even for linear equations in order to achieve second-order accuracy simultaneously with monotonicity.

We present the flux limiter approach [16, 4, 5, 41, 50, 52] in terms of a simple conservation law

$$w_t + f(w)_x = 0 \tag{2.9.1}$$

$$f(w) = aw \tag{2.9.2}$$

as solved by

$$w_i^{n+1} = w_i^n + \frac{\Delta t}{\Delta x}(f_{i+1/2} - f_{i-1/2}) \tag{2.9.3}$$

Given a high order flux $f_{i+1/2}^{HI}$ associated with a scheme of accuracy greater than or equal to two and a low order flux $f_{i+1/2}^{LO}$ associated with a monotone first order scheme, one can define a high order flux $f_{i+1/2}^{*}$ as

$$f_{i+1/2}^{*} = f_{i+1/2}^{LO} + \Psi_{i+1/2}[f_{i+1/2}^{HI} - f_{i+1/2}^{LO}] \tag{2.9.4}$$

where $\Psi_{i+1/2}$ is a flux limiter function that usually depends on a ratio r measuring the regularity of the data at $x_{i+1/2}$: $r \approx 1$ denotes smooth data, while r far from 1 denotes non-regular data.

Definition (2.9.4) produces a high order resolution flux that switches between a second order approximation when the data are sufficiently smooth and a first order approximation near a discontinuity.

There are various choices for the flux limiter function and for example:
Minmod [52] is given by

$$\Psi(r) = \begin{cases} 0 & r \leq 0 \\ r & 0 \leq r \leq 1 \\ 1 & r \geq 1 \end{cases} \tag{2.9.5}$$

Vanleer is given by

$$\Psi(r) = \begin{cases} 0 & r \leq 0 \\ \frac{2r}{1+r} & r > 0 \end{cases} \tag{2.9.6}$$

and Superbee is given by

$$\Psi(r) = \begin{cases} 0 & r \leq 0 \\ 2r & 0 \leq r \leq \frac{1}{2} \\ 1 & \frac{1}{2} \leq r \leq 1 \\ r & 1 \leq r \leq 2 \\ 2 & r \geq 2 \end{cases} \tag{2.9.7}$$

2.9.1 Flux limiters in the present model

The value of $(UQ)_i$, as required by F as well as the value of $\eta_{i+1/2}$ involved in the definition of the momentum $Q_{i+1/2}$ have been chosen with upwind rules, (2.4.7) and (2.3.7) respectively.

With this choice, the resulting numerical scheme is only first order accurate. In order to improve the accuracy without running into stability problems but leading it to satisfy the TVD property [53], the flux limiter method presented in the previous section has been used.

As pointed out, this high order resolution method switches between a second order approximation when the data are sufficiently smooth and a first order appro-ximation near a discontinuity.

In our implementation of the data reconstruction step, this technique adds to the first order numerical flux a correction term limited by a flux limiter function Ψ that depends on the regularity of the data r.

Assuming positive flow direction, the velocity U in the approximation of the advective term (2.4.7) becomes

$$U(x) = U_{i-1/2} + \frac{(x - x_{i-1/2})}{\Delta x}\Psi(r_U(x))(U_{i-1/2} - U_{i-3/2}) \quad x \in [x_{i-1}, x_i] \tag{2.9.8}$$

where

$$r_i^U = \frac{U_{i+1/2} - U_{i-1/2}}{U_{i-1/2} - U_{i-3/2}} \tag{2.9.9}$$

Moreover, the water surface elevation η involved in the definition of $Q_{i+1/2}$ is now given by

$$\eta(x) = \eta_i + \frac{(x - x_i)}{\Delta x}\Psi(r(x))(\eta_i - \eta_{i-1}) \quad x \in [x_{i-1/2}, x_{i+1/2}] \tag{2.9.10}$$

where

$$r_{i+1/2}^{\eta} = \frac{\eta_{i+1} - \eta_i}{\eta_i - \eta_{i-1}} \tag{2.9.11}$$

The flux limiting function Ψ can be chosen in several ways (see, e.g., [35, 34] for details).

2.9.2 A special flux limiter

A special flux limiter function has been defined to be used in the extrapolation of the value $\eta_{i+1/2}$ in the advective term in case of critical flow and it is defined, for positive flow direction, by the following relation

$$\Psi_{i+1/2} = \Psi(x_{i+1/2}) = min(0, max(\frac{-\eta_i/3}{\eta_{i+1} - \eta_i}, 1)) \tag{2.9.12}$$

One can show that

$$0 \leq \Psi_{i+1/2} \leq 1$$

that means that a data reconstruction using the flux limiter function Ψ defined in (2.9.12) is consistent, because it is a Total Variation Non Increasing (TVNI) scheme, as stated in the Harten's Theorem [25].

In particular, the reconstruction of η in the node $i + 1/2$ assumes the following form

$$\eta(x) = \eta_i + \frac{(x - x_i)}{\Delta x}\Psi(x)(\eta_{i+1} - \eta_i) \quad x \in [x_{i-1/2}, x_{i+1/2}] \tag{2.9.13}$$

and can be written in a more compact notation as follows

$$\eta_{i+1/2} = min(\eta_i, max(\frac{2}{3}\eta_i, \eta_{i+1})). \tag{2.9.14}$$

The derivation of this special flux limiter follows from the analysis of the specific energy head function [11] in case of a constant discharge

$$E = H + \frac{U^2}{2g}. \tag{2.9.15}$$

This function assumes its minimum respect to H in the case of critical flows, that is if $Fr = 1$ ($U = \sqrt{gH}$), and its minimum value is

$$E = \frac{3}{2}(\frac{Q^2}{gA^2})^{\frac{1}{3}} \tag{2.9.16}$$

where $H_{cr} = (\frac{U^2}{g})^{\frac{1}{3}}$ is called critical depth.

Thus, in case of critical flow, one has $H = \frac{2}{3}E$ (see, e.g., [11]).

Equation (2.9.14) is finally obtained assuming that the squared velocity is negligible with respect to H and introducing a min-max rule to ensure consistency.

The implementation of this flux limiter improves the accuracy of the method and helps in facing the problems arising in case of low resolution of the grid. An application of this flux limiter can be found in Section 3.3.

3

Numerical results in open channels

The aim of this chapter is to show the properties of the proposed method in terms of stability, accuracy and efficiency in the simulation of various test cases. A few computational examples are given on the classical frictionless dam break problem for rectangular and non-rectangular channels and on wet and dry channel's bed. The numerical results obtained in rectangular channels are compared with the analytical solutions, while those in a triangular channel are presented to show the applicability of the present algorithm to a problem where precise volume conservation is essential and not easily obtained by traditional linear schemes. Steady state problems over a discontinuous bed profile in a rectangular frictionless channel are also modelled. In particular, a steady state problem including a hydraulic jump is exemplified to show the ability of the proposed flux limiter (see Section 2.9.2) in providing a physically correct solution even in case of a low resolution grid. Continuous transitions from subcritical to supercritical flow and vice versa are simulated as an interesting proof of the robustness of the proposed scheme. Two free fluid oscillations in parabolic basins are modelled in order to check the ability of the scheme to compute a moving wet-dry interface over a sloping topography.

3.1 Dam Break problems

The test problems presented in this section belong to the class of the well known frictionless dam break problem introduced by Stoker in 1957 [47].

They consist in the simulation of the phenomenon following the instantaneous removal of a vertical wall separating the water in the middle of a channel.

This kind of events are fortunately rare, but when they occur the consequences are disastrous. The mathematical modelling plays a very important role in understand-

ing the evolution of the breaking process, when it happens, and the whole physical phenomena involved.

Volume and momentum conservation are always applied in order to obtain a physically correct numerical solution. Moreover, the flux limiter method is also implemented to gain in accuracy.

For all the test cases, the domain length is $L = 1m$, the bottom is flat, the vertical wall is situated at $L/2$ and the boundaries are closed.

The water is initially at rest

$$u(x, 0) = 0 \tag{3.1.1}$$

and the water depth is constant on the downstream as well as on the upstream side of the dam

$$\eta(x, 0) = \begin{cases} \eta_l & \text{if} \quad 0 \leq x \leq \frac{L}{2} \\ \eta_r & \text{if} \quad \frac{L}{2} < x \leq L. \end{cases} \tag{3.1.2}$$

For the first problem the channel is rectangular, the upstream depth η_l is $1m$ and the downstream depth η_r is $0m$.

The other physical and computational parameters are the friction coefficient $\gamma = 0$, the gravitational acceleration $g = 1m/s^2$, the grid size $\Delta x = 0.005m$, the parameter $\theta = 1$ and the time step $\Delta t = 10^{-3}s$.

In this example $\eta(x, t)$ also represents the total water depth which is initially zero for $\frac{1}{2} \leq x \leq 1$.

The main task of this test is to correctly simulate the flooding of the second half part of the domain that is dry at time $t = 0$. In particular, the accuracy and reliability in the approximation of the wave arrival time is very important to understand how the process will evolve and to give a contribute to the risk analysis for civil protection [31].

Figures 3.1 and 3.2 show the numerical results and the analytical solution (plotted with a dotted line) at time $T = 0.15s$ for the water surface elevation and for the velocity. These results compare favourably well with those obtained from high-order Godunov's type methods (see, e.g., [53, 18]).

The second test problem is a dam break problem in a rectangular channel: both the velocity and the water surface present a large gradient as in the previous test.

The upstream depth η_l is $1m$, the downstream depth η_r is $0.1m$ and the computational parameters are set as in the first example except for $\theta = 0.5$.

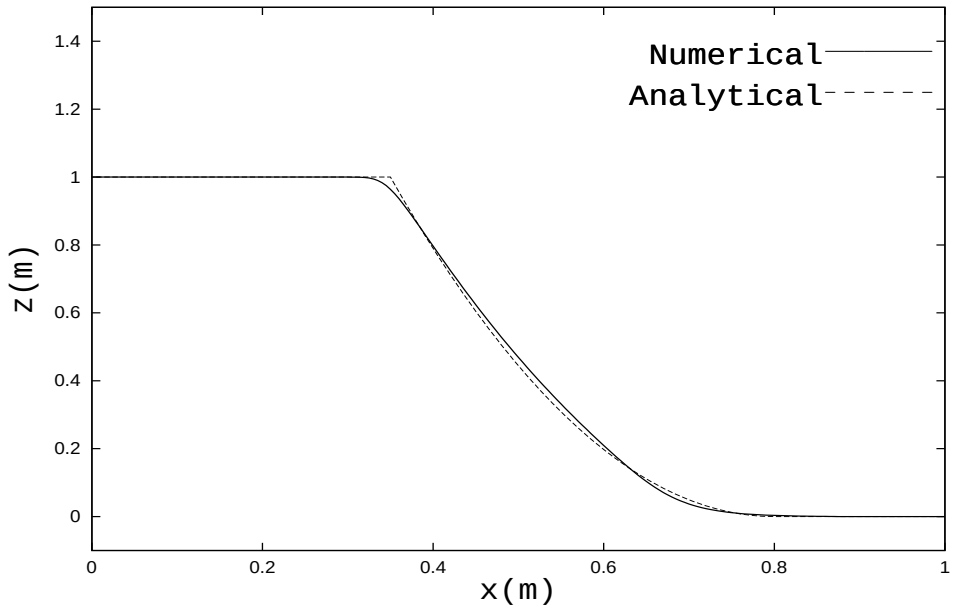

Figure 3.1: Dam break over a dry bed in a rectangular channel: the water elevation

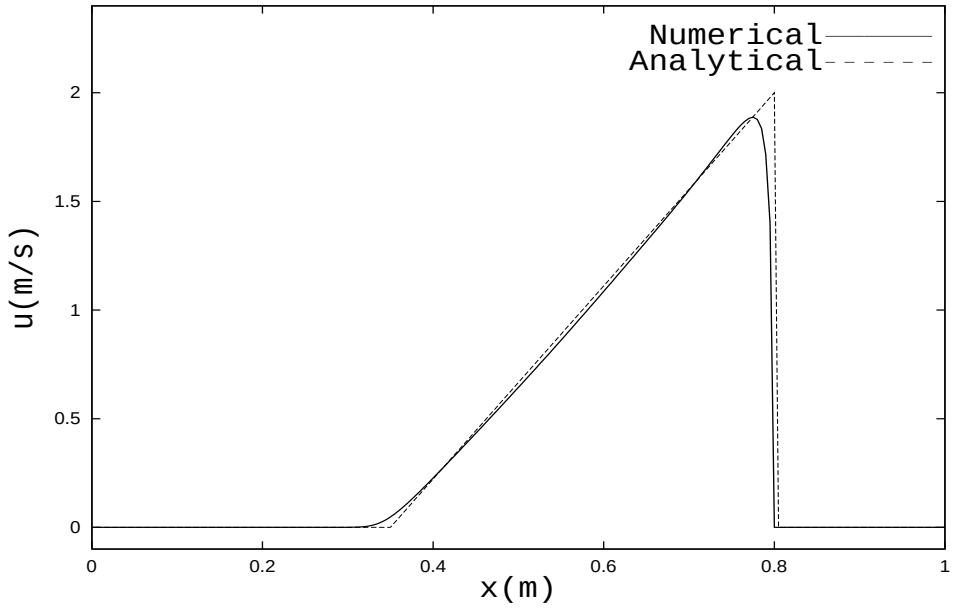

Figure 3.2: Dam break over a dry bed in a rectangular channel: the velocity

Figures 3.3 and 3.4 show the numerical results and the exact solution obtained for water level and velocity profiles after $T = 0.3s$. The comparison is very satisfactory.

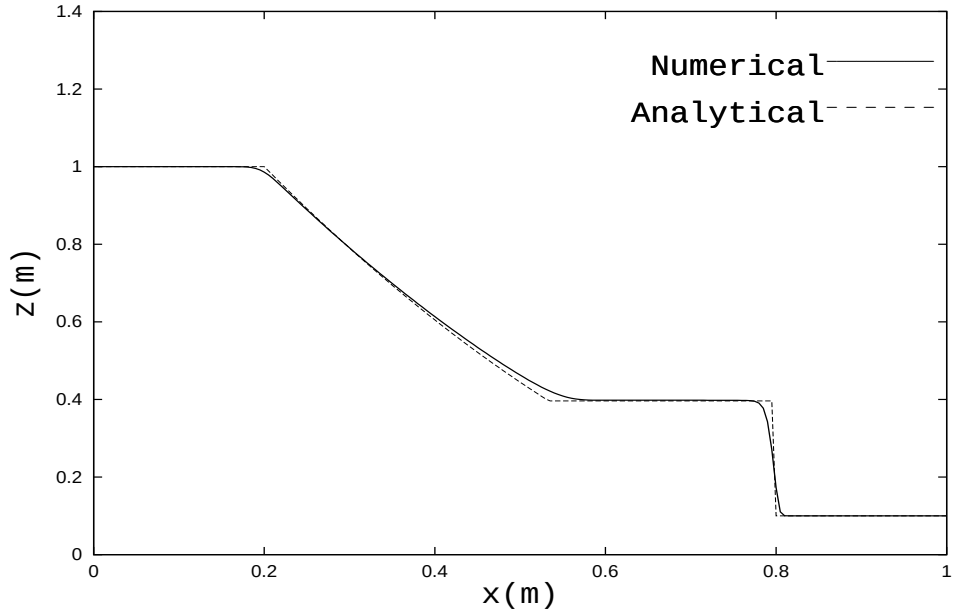

Figure 3.3: Dam break over a wet bed in a rectangular channel: the water elevation

In the third test problem the channel has triangular cross section of area $A = 10\eta^2$.

The initial conditions are the same as in (3.1.1) and (3.1.2) with the downstream depth η_l equal to $1m$ and the upstream depth η_r equal to $0.1m$.

The computational parameters are set as in the first example except for $\theta = 0.5$.

The results obtained at time $T = 0.3s$ are plotted in Figure 3.5.

This example shows the applicability of the present algorithm to a dam break problem where precise volume conservation is essential and not easily obtained by traditional linear schemes.

3.2 Subcritical and transcritical flow over a hump

The tests presented in this section simulate steady state problems over a discontinuous bed profile in a rectangular frictionless channel.

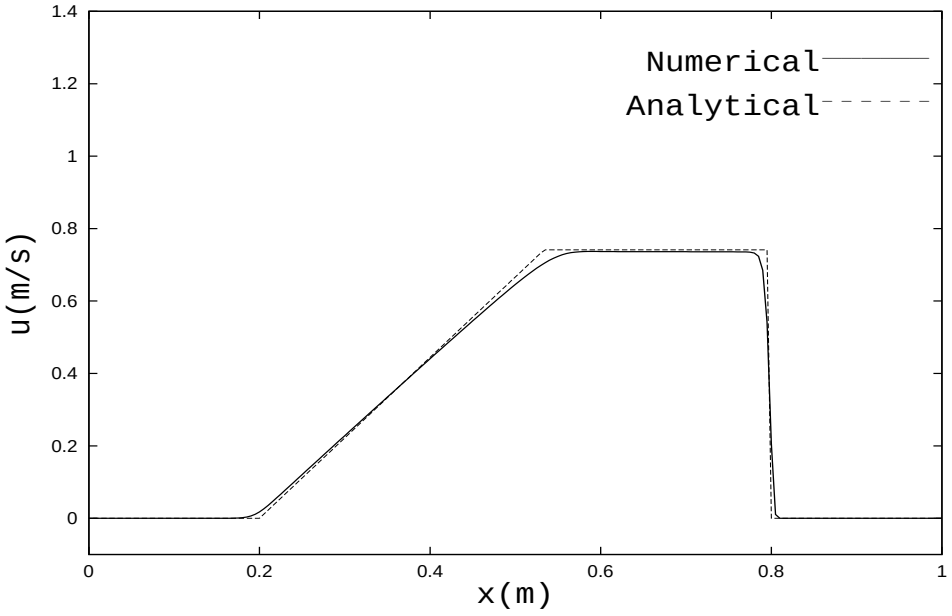

Figure 3.4: Dam break over a wet bed in a rectangular channel: the velocity

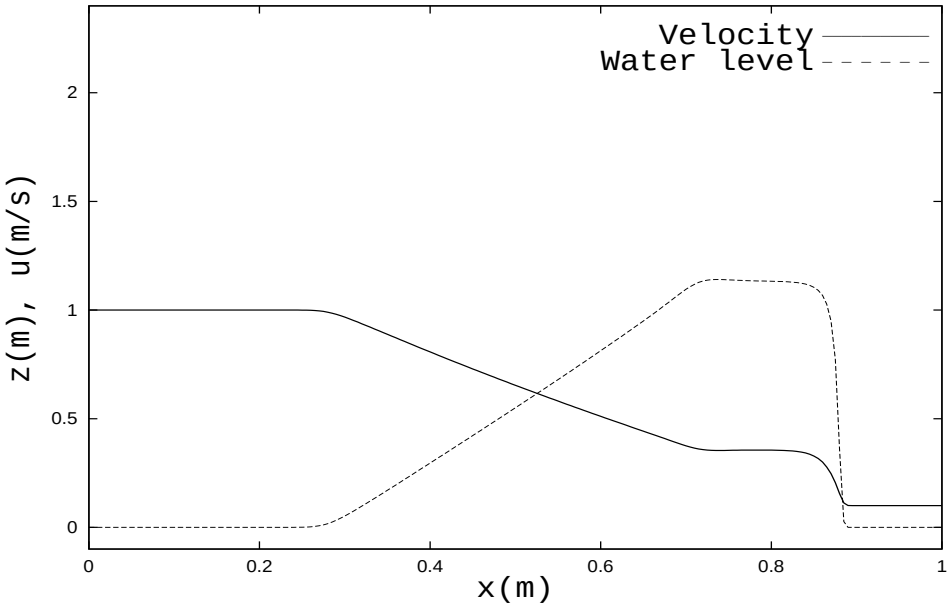

Figure 3.5: Dam break over a wet bed in a triangular channel

The domain length is $L = 50m$ and in the middle of the channel there is a sill with a crest of $0.4m$ height and $8m$ long with vertical walls. Specifically, the bottom profile is given by

$$h(x) = \begin{cases} 0.4m & \text{if} \quad 16m < x < 24m \\ 0m & \text{if} \quad otherwise \end{cases} \tag{3.2.1}$$

According to the boundary and initial conditions, the flow may be subcritical, transcritical with a steady shock, or supercritical.

As boundary conditions, the discharge and the water depth are imposed, respectively, at the inflow and at the outflow.

A constant water level equal to the level imposed downstream and a discharge equal to zero are chosen as initial conditions.

$\gamma = 0$ and $g = 9.81m/s^2$, while the discretization parameters are set to $\Delta x = 0.25m$, $\theta = 1$ and $\Delta t = 10^{-2}s$. In particular, the numerical representation of the bottom profile is such that it changes between zero and 0.4 within one grid cell just next to the sill in the middle of the channel.

In the first problem a discharge of $2.42m^3/s$ and a water depth of $2m$ are imposed at the two open boundaries.

The flow is subcritical and the results obtained for the water level, the velocity and the discharge are plotted in Figures 3.6 and 3.7.

In the second problem the upstream discharge is $1.53m^3/s$ and the downstream water depth is $0.5m$.

The discretization parameters are the same as for the subcritical problem.

The flow is transcritical without shock: Figures 3.8 and 3.9 show the numerical results for the water level, the velocity and the discharge.

3.3 Transitions from super to subcritical flows

The first test presented in this section simulates a steady state problem including a hydraulic jump over a non-flat bed profile in a rectangular frictionless channel.

A hydraulic jump consists in the transition from a supercritical flow to a subcritical flow, it is extremely turbulent, it is characterized by strong energy dissipation and it necessitates proper local conservation properties to be correctly represented.

In the analysis of supercritical flows, the main aspect to be investigated is the location of the hydraulic jump.

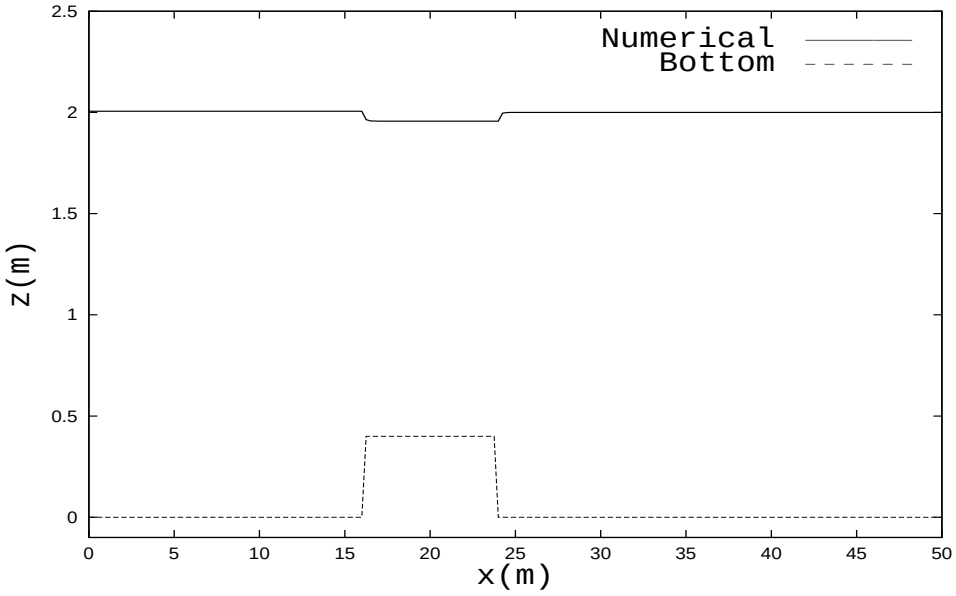

Figure 3.6: Subcritical flow over a sill: water elevation

On the other hand, in case of subcritical flows, a precise estimation of the energy head loss due to the hydraulic jump is essential to have the correct upwind water level and the correct discharge over the sill once the downstream water level is fixed.

The numerical test presented in this section shows the ability of the numerical method and of the flux limiter function provided by (2.9.14) in fulfilling these requirements, even in the case of a low resolution grid.

Moreover, Energy head Conservation is used in contractions and Momentum Conservation elsewhere.

The physical properties of the channel are the following: its length is $L = 100m$ and a sill with a crest of $1m$ height and $10m$ long and vertical walls is situated in the middle.

Moreover, there are two open boundaries, the inflow and the outflow, where a discharge of $1m^3/s$ and a water depth of $1m$, respectively, are imposed [46].

The discretization parameters are $\theta = 1$ and $\Delta t = 10^{-3}s$, while $\gamma = 0$ and $g = 9.81m/s^2$.

Figure 3.10 shows a comparison between the numerical solutions obtained for 100

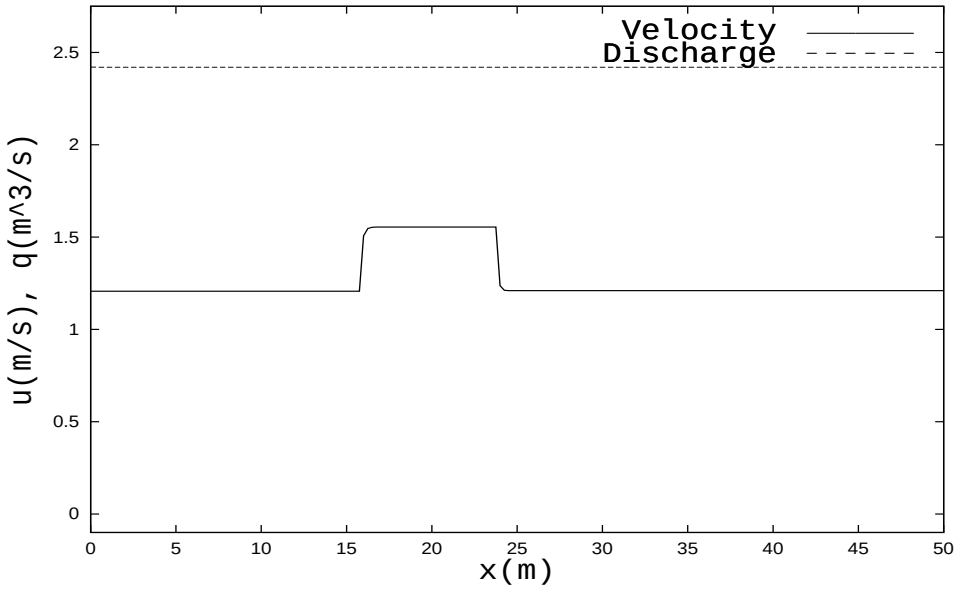

Figure 3.7: Subcritical flow over a sill: velocity and discharge

grid points ($\Delta x = 1m$) using the flux limiter (2.9.12) only over the sill (*Solution 1*) and the numerical solutions obtained for 20 grid points ($\Delta x = 5m$) with (*Solution 2*) and without (*Solution 3*) the help of the flux limiter.

The numerical *Solutions* 1 and 2 are coincident in almost all the nodes in common (and in particular at the upstream end) although the second grid is five times coarser than the first.

Moreover, on equal grid size, the numerical solution obtained using the limiter (*Solution 2*) shows an upstream water level that is consistent with that of *Solution 1* and higher than that obtained without the limiter (*Solution 3*): the reduction of the resolution of the grid causes the upstream water level to decrease in the numerical solution of the first order model.

The quality of the results can also be appreciated from the approximation of the energy line plotted in Figure 3.10: as one can see, it is constant everywhere, except near the hydraulic jump where the energy head drops as is to be expected by considerations based on open channel hydraulics [11].

The second test presented in this section is an interesting proof of the robust-

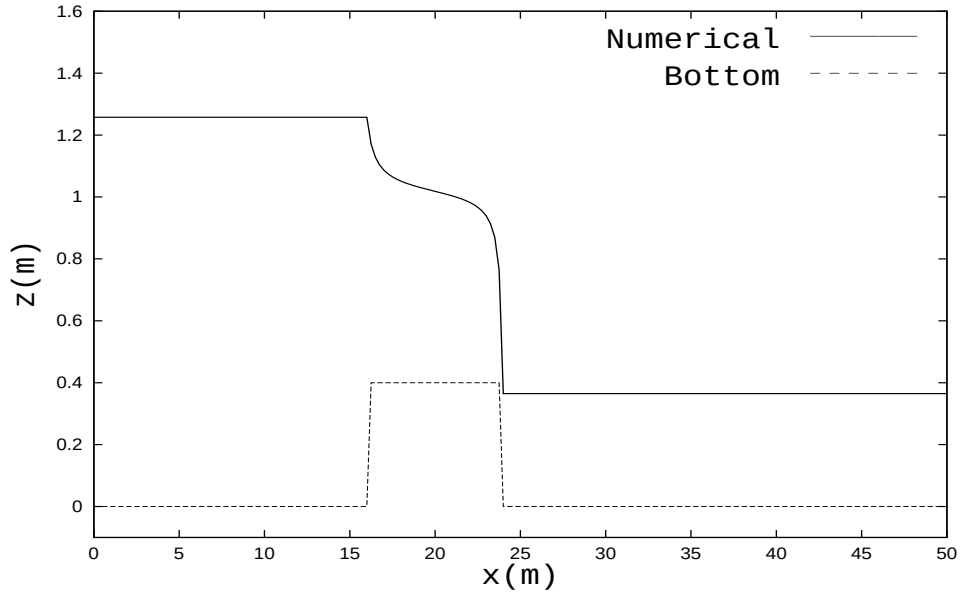

Figure 3.8: Transcritical flow over a sill: water elevation

ness of the proposed scheme in simulating continuous transitions from subcritical to supercritical flow and vice versa.

These transitions are obtained imposing as downstream boundary condition a water level following the hydrograph depicted in Figure 3.11 and described by the equation

$$\eta(L,t) = 0.8sin(0.01t) + 1 \tag{3.3.1}$$

The physical domain considered is the same as that of the previous test.

Figure 3.12 shows the numerical results obtained for the upstream water level during two complete oscillations of the downstream boundary condition (3.3.1).

As expected, in the range for $\eta(L,\cdot)$ corresponding to imperfect weirs, any small change of its value affects the upstream flow condition, because the wave celerity is larger than the flow velocity. Note that in Figure 3.12 the oscillations of η at the border of each smooth peak are qualitatively correct and not due to numerical reasons,

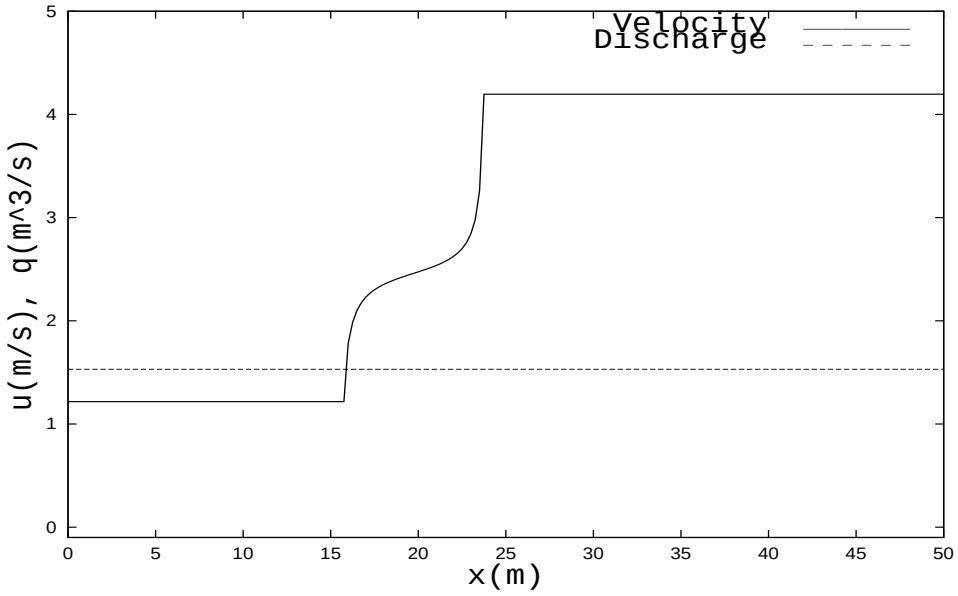

Figure 3.9: Transcritical flow over a sill: velocity and discharge

because they represent the settlement of the η-values caused by the perturbation of the downstream water level.

On the other hand, in the range for $\eta(L, \cdot)$ corresponding to perfect weirs, a downstream disturbance does not travel upstream and identical upstream depth estimations are produced.

3.4 Wetting, drying and moving boundaries

The non-linear Shallow Water Equations with topography cannot in general be solved exactly. Therefore, it is not possible to validate a numerical method in all cases, and the problems where an exact solution is known are important test cases.

In 1981 Thacker [51] described analytically the solution to the Shallow Water Equations for two particular test cases of two dimensional motion: the oscillations of a planar surface and of a parabolic surface in an elliptical basin for a frictionless fluid.

These are important and severe test cases because they present a moving wet-dry

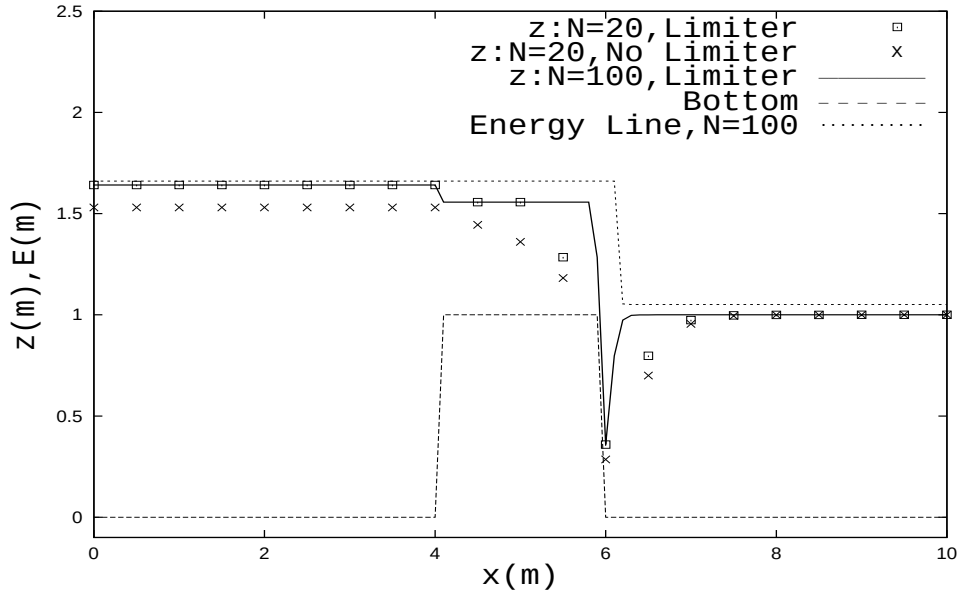

Figure 3.10: High and low resolution grids: effect of the flux limiter

front in the domain and because the absence of bottom friction can cause a loss of stability of the numerical solution.

3.5 Oscillations with planar surface

Consider the shallow basin given by the elliptical paraboloid of equation

$$h(x,y) = h_0(1 - \frac{x^2}{l^2} - \frac{y^2}{L^2}), \tag{3.5.1}$$

where h_0 represents the maximal depth and l and L are parameters for the curvature of the basin.

For this problem, assume that the initial water surface elevation is planar, that the velocities in the x and y directions are constant in space and that the earth's rotation is neglected.

Therefore, if the basin (3.5.1) is a canal with parabolic cross-section ($l >> L$), Thacker [51] provides the following solution for the two dimensional Shallow Water

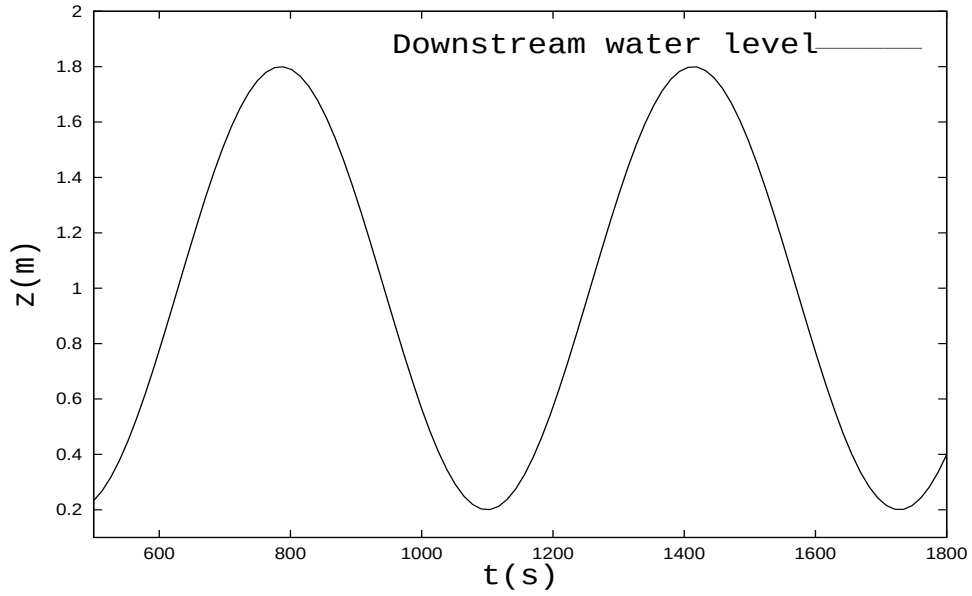

Figure 3.11: Downstream boundary condition on the water level

Equations:

$$\begin{cases} u(x,t) = -\Theta\omega sin\omega t \\ v(x,t) = 0 \\ \eta(x,t) = 2\Theta\frac{h_0}{l}cos\omega t(\frac{x}{l} - \frac{\Theta}{2l}cos\omega t) \end{cases} \tag{3.5.2}$$

where

$$\omega = \sqrt{\frac{2gh_0}{l^2}} \tag{3.5.3}$$

is the frequency and Θ is the amplitude of the motion.

The shorelines for this solution are determined by the condition $H = 0$ and are given by

$$x = \Theta cos\omega t \pm l. \tag{3.5.4}$$

As one can note from Equations (3.5.2), the water surface elevation remains planar and its inclination varies during the evolution of the phenomenon.

To test the one-dimensional numerical model presented in [3] on this problem,

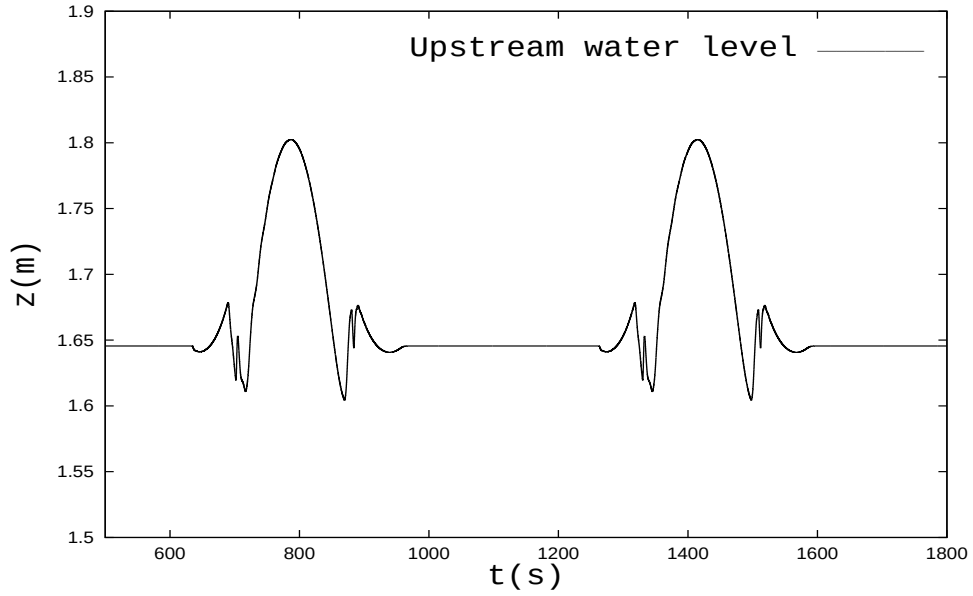

Figure 3.12: Varying downstream boundary condition: Upstream water level

the bathymetry of the canal is assumed to be

$$h(x) = h_0(1 - \frac{x^2}{l^2}), \tag{3.5.5}$$

while its cross-section at the point x is described by the following function

$$A(x, \eta) = h(x) - h_0\frac{\eta^2}{L^2}. \tag{3.5.6}$$

The initial conditions consist in an initial zero velocity

$$u(x, 0) = 0 \tag{3.5.7}$$

and in a planar water surface elevation presenting an inclination related to Θ

$$\eta(x, 0) = 2\Theta\frac{h_0}{l}(\frac{x}{l} - \frac{\Theta}{2l}). \tag{3.5.8}$$

With this configuration, the numerical solution approximates the analytical one (3.5.2) favourably well.

$g = 9.81m/s^2$, the physical parameters are $h_0 = 1m$, $l = 50m$ and $L = 4m$, and the computational ones are $N = 625$, $\Delta x = 0.2m$, $\Delta t = 2.e-02s$, $\theta = 1$ and $\Theta = 2.m$.

From these data it follows that the frequency of the motion is $\omega = 0.0886rad/s$ and the period of the motion is $T = \frac{2\pi}{\omega} = 70.925s$. The percentage of wetting and drying is around 5%.

The numerical simulation covers two complete oscillations.

In Figure 3.13 one can observe a good agreement between the numerical and the analytical velocity u at the center of the basin, plotted in function of time for the first two periods.

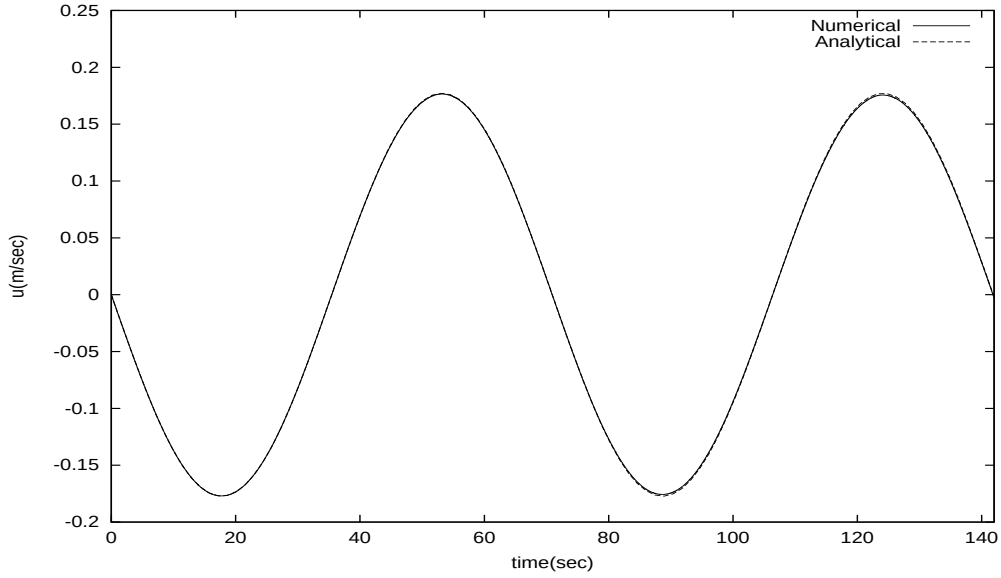

Figure 3.13: Numerical and analytical velocities at the center of the basin for the oscillations of a planar surface

Moreover, the numerical velocity is constant in space and, after two periods, the difference from the theoretical value is of the order of a millimeter per second, as the velocity is $-0.2m/s^{-1}$.

The numerical water surface elevation remains planar during the evolution of the oscillations and compares favourably well with the analytical value given by (3.5.2),

even for frequency and amplitude.

The results depicted in Figure 3.14 represent the water surface drawn every $10.13s$, while the dashed line shows its initial position.

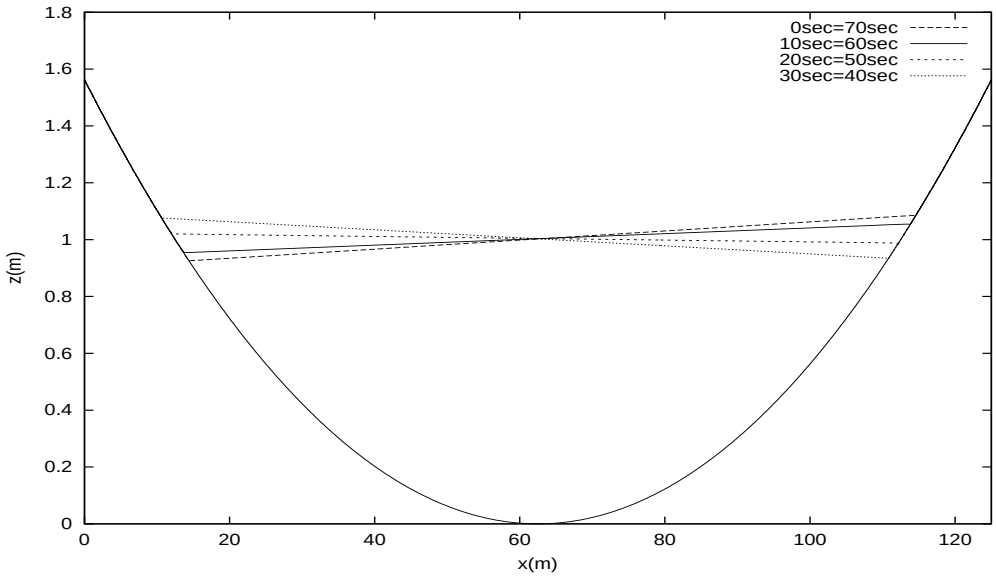

Figure 3.14: Oscillations of a planar surface in a parabolic basin

Finally, the numerical shorelines appear to differ slightly from those given by (3.5.4), being situated from zero to tree spatial intervals far from the theoretical values with a maximum error of $0.52m$.

Figure 3.15 shows the comparison between the numerical and the analytical left shoreline for the first two periods.

3.6 Oscillations with parabolic surface

For this problem, consider a canal with a parabolic cross-section of bathymetry given by (3.5.5).

Assume that the initial water surface elevation is parabolic, that the initial velocities in the x and y directions are zero and that the earth's rotation is neglected.

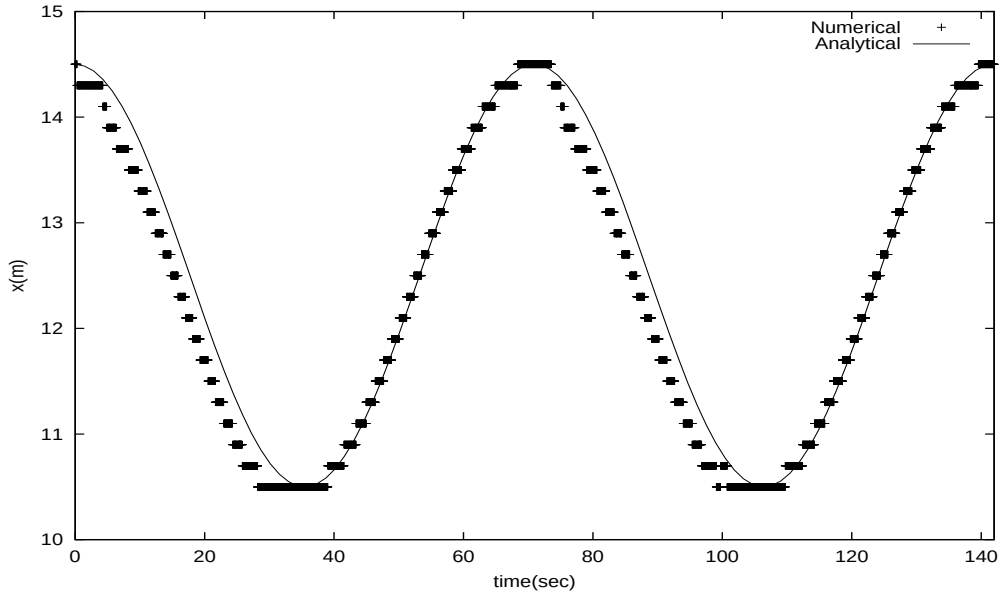

Figure 3.15: Numerical and analytical left shoreline

Although it is not possible to derive the analytical solution for this problem [51], the behaviour of the water surface elevation during the evolution of the oscillations is known: it remains parabolic and its extreme states are the initial water surface elevation and a parabolic surface of opposite concavity. Moreover, the frequency of the oscillations seems to depend on the amplitude of the motion [51].

Therefore, testing the one-dimensional numerical model presented in [3] on this problem can be evaluated only qualitatively.

The initial conditions consist in an initial zero velocity

$$u(x,0) = 0 \tag{3.6.1}$$

and in the following parabolic water surface elevation

$$\eta(x,0) = \Theta(1 - \frac{2x^2}{l^2}). \tag{3.6.2}$$

The physical and computational parameters are $g = 9.81m/s^2$, $N = 1500$, $\Delta x = 0.2m$, $\Delta t = 1.e-01s$, $\theta = 1$, $h_0 = 0.01m$, $l = 65m$, $L = 2m$ and $\Theta = 0.1m$.

The numerical simulation covers two complete oscillations.

Figure 3.16 shows the numerical results for the water surface elevation every $50s$, while the dashed line shows its initial position.

The results depicted in Figure 3.17 represent the sinusoidal behaviour of the numerical left shoreline in function of time.

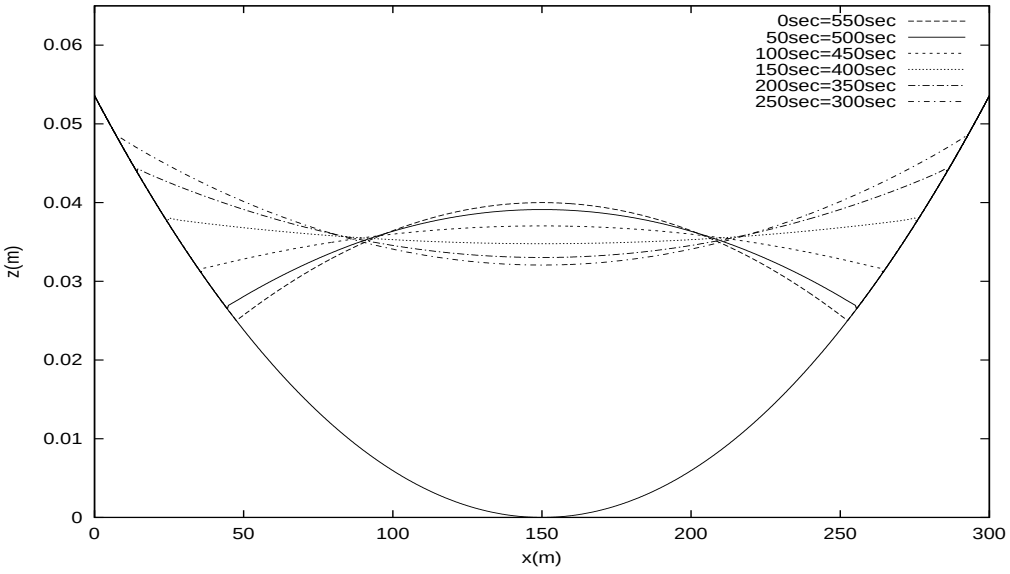

Figure 3.16: Oscillations of a parabolic surface in a parabolic basin

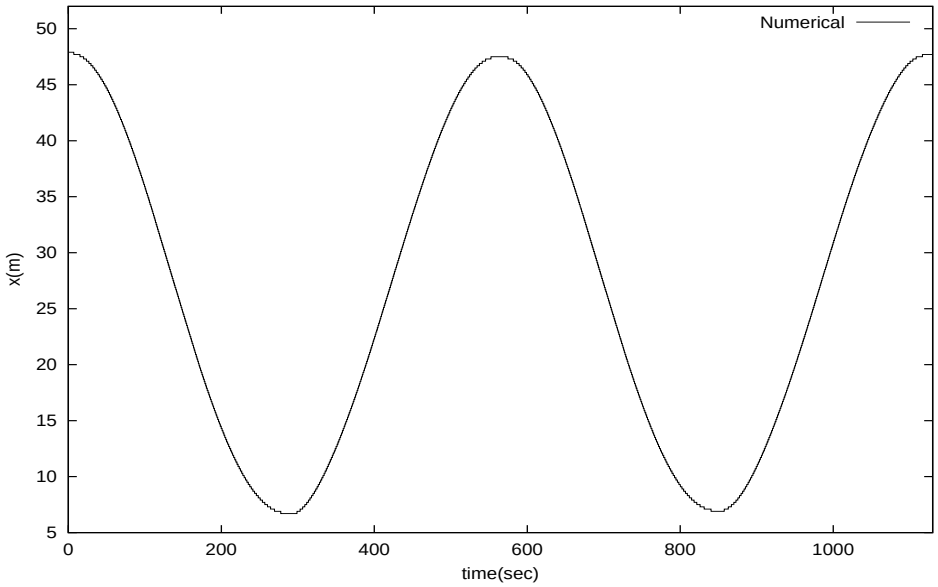

Figure 3.17: Numerical left shoreline

4

Extension to closed channel flows

The aim of this chapter is to present the extension of the numerical scheme for one dimensional open channel flows described in Chapter 2, to one dimensional closed channel flows. Flows in closed channels, such as rain storm sewers, often contain transitions from free surface flows to pressurized flows, or vice versa. These phenomena usually require two different sets of equations to model the two different flow regimes. Actually, a few specifications for the geometry of the channel and for the discretization choices can be sufficient to model closed channel flows using only the open channel flow equations. The numerical results obtained solving the pressurization of a horizontal pipe are presented and compared with the experimental data known from the literature. Moreover, the numerical scheme is also validated simulating a flow in a horizontal and downwardly inclined pipe and comparing the numerical results with the experimental data obtained in the laboratory.

4.1 Flows in closed channels

The transition from free surface to pressurized flow or vice versa is a phenomenon often occurring in closed channels.

This situation may happen for example in storm sewers systems during heavy storm events or even in a closed channel with initially free surface flow as a result of the start-up of machinery (turbines, pumps, gates).

Because of the wide range of practical problems involving closed channel flows, numerical methods are needed to predict the water profile, pressure and discharge during pipes pressurization and depressurization.

The one-dimensional equations for free surface as well as pressurized flows in closed channels are essentially the Saint Venant Equations and two types of algo-

rithms broadly used in the literature to solve them numerically are the Saint Venant Equations (1.4.1)-(1.4.3).

Explicit algorithms are such that the time step is limited to the Courant condition. This limitation cannot be fulfilled for pressurized flows due to the infinite propagation velocities. In fact, assuming the incompressibility of water, the wave celerity is infinite in pressurized sections and the same explicit algorithm used for the free surface flow part of the domain cannot be used to solve the pressurized parts.

To avoid this inconvenience, almost all existing models use the Preissmann slot technique [30, 20, 44], that is an approximation of the real, closed section with an open section displaying a very small top width, called Preissmann slot.

In case of free surface flows the slot has no effects and the open channel flow equations apply as usual.

Moreover, in case of pressurized flows, the small slot allows a finite value of the wave celerity and the use of the free surface flow model everywhere in the computational domain.

A delicate issue is the choice of the slot width ϵ. In fact, if ϵ is too small, the use of the Preissmann approximation can produce a large wave celerity and a corresponding strict time step limitation, while, if ϵ is too large, inaccuracies may results [43].

On the other hand, unconditionally stable methods like fully implicit methods [7, 54] are able to simulate the transition from free surface to pressurized flow in channels with closed sections without any approximation of the section geometry. In fact, assuming the incompressibility of water, they can manage instantaneous transmission of pressure and velocity changes arising in the pressurized part of the channel.

Therefore, using a fully implicit discretization in time, the numerical scheme presented in Chapter 2 can be used to simulate free surface as well as pressurized flows [2].

4.2 Geometrical and physical specifications

The water depth H and the cross-sectional area A are related with the variable η.

In case of free surface flows in a closed channel as well as for open channel flows, the quantities η, H and A have the usual definitions.

In case of pressurized flows, η plays the role of the pressure head, the water height

H is the maximum height reachable $H_{top} = \eta_{top} + h$ and the wetted area A is the area of the whole cross section A_{top}.

Therefore, the total water depth H in a closed channel can be expressed as follows

$$H = \begin{cases} \eta + h & \text{if} \quad \eta \leq \eta_{top} \\ H_{top} & \text{if} \quad \eta > \eta_{top} \end{cases} \tag{4.2.1}$$

Moreover, the cross-sectional area A in a closed channel is a piecewise derivable non decreasing functions of η and it is defined depending on the channel geometry.

For a rectangular closed channel with constant width B one has $A = BH$, while for the special case of a circular channel with diameter D it holds

$$A = \begin{cases} \frac{D^2}{4}\left[arccos(1-2\frac{H}{D}) - (1-2\frac{H}{D})\sqrt{1-(1-2\frac{H}{D})^2}\right] & \text{if} \quad \eta \leq \eta_{top} \\ \pi(D/2)^2 & \text{if} \quad \eta > \eta_{top} \end{cases} \tag{4.2.2}$$

4.3 Numerical results in closed channels

The numerical results obtained solving the pressurization of a horizontal pipe are presented and compared with the experimental data known from the literature. Moreover, the numerical scheme is also validated simulating a flow in a horizontal and downwardly inclined pipe and comparing the numerical results with the experimental data obtained in the laboratory.

4.3.1 Pressurization in a horizontal pipe

This test [20, 36] reproduces a free surface and pressurized flow in a horizontal, rough, rectangular, closed channel of length $L = 10m$, width $B = 0.51m$, height $H_{top} = 0.148m$ and $c_f = g\frac{n_M^2}{R_H^{1/3}}$, where $n_M = 0.12$ is the Manning's roughness coefficient [11].

The upstream boundary condition is the hydrograph for the pressure head described in Figure 4.1, while the downstream boundary condition is a fixed water level, $H_{N+1} = 0.128m$.

Initially the following free surface flow conditions with still water are present:

$$U(x,0) = 0m/s, \qquad \eta(x,0) = top(x) = 0.128 \tag{4.3.1}$$

Then a wave, coming from the outside left side, causes the closed channel to pressurize starting from upstream. The interface separating pressurized from free surface flow moves from upstream to downstream as a front wave.

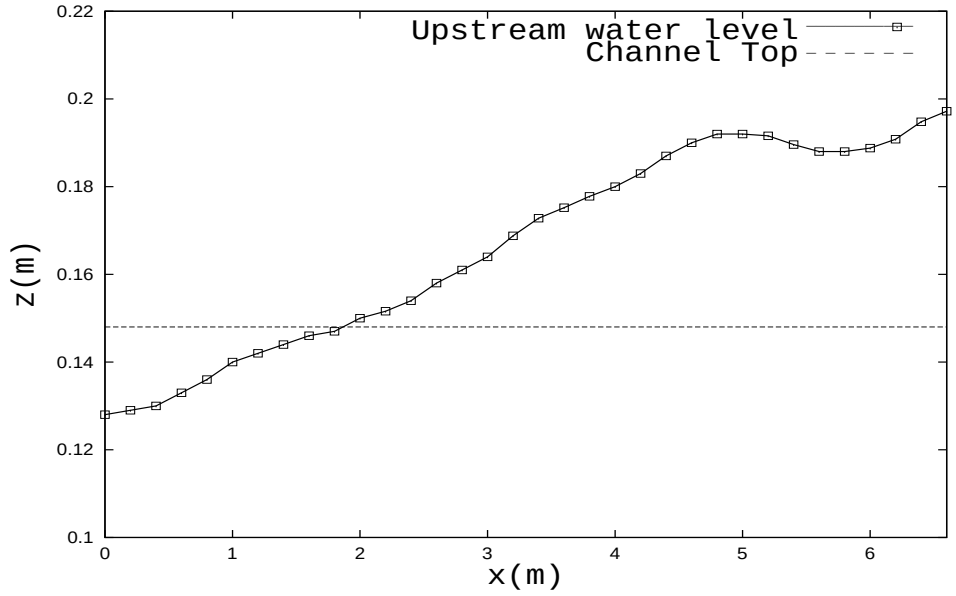

Figure 4.1: Water height at the upstream boundary against time.

The physical and computational parameters are $g = 9.81m/s^2$, $\Delta x = 0.1m$, $\theta = 1.$ and $\Delta t = 5.\ 10^{-3}s$.

Figure 4.2 shows the behaviour of the numerical instantaneous pressure head η against time at $x = 3.5m$ compared with the experimental data obtained by Wiggert [56, 57]. As one can see from the Figure below, the experimental and the numerical data agree fairly well.

4.3.2 Hydraulic jump in a circular pipe

These experiments have been carried out by the University of Delft and Delft Hydraulics in collaboration with the majority water boards in the Netherlands [14].

The aim of these experiments is the investigation about the air-water phenomena in wastewater pressure mains with respect to transportation and dynamic hydraulic behaviour. Free gas in pressurized pipelines can in fact significantly reduce the flow capacity and may cause undesirable efficiency loss.

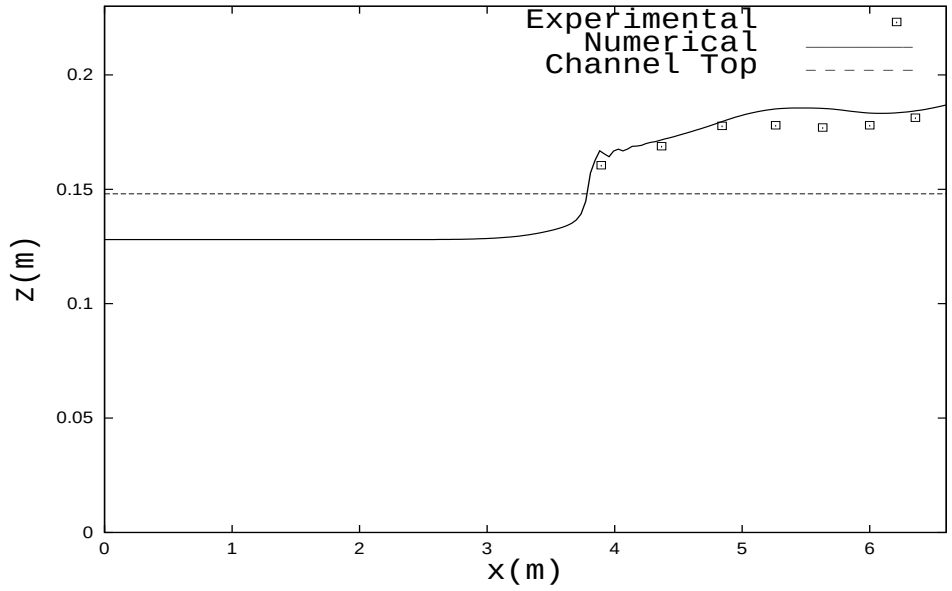

Figure 4.2: η at $x = 3.5$ against the time.

These experiments have been conducted in a dedicated facility for research on gas pockets that are located at the transition from horizontal to inclined pipes.

The test section of the pipe consists of three parts: a horizontal pipe of length $L_1 = 2m$, a downward inclined pipe ($\alpha = 10°$) of length $L_2 = 4m$ and a horizontal pipe of length $L_3 = 2m$. The pipes have an inner diameter of $220mm$ and are made of transparent material (Perspex with equivalent sand roughness height of $k_s = 0$).

Injecting air into the water and preserving a constant water discharge at the inlet of the pipe and a constant pressure head downstream, an air pocket appears in the inclined part of the pipe and the obtained configuration presents similarities with hydraulic jumps in open channels.

The numerical results of the present model for the pressure head at the steady state of the phenomenon are compared with the experimental data. They are given as measurements of the water depth at specific nodes located along the air pocket at a distance of about $30cm$ one to the other. The hydraulic jump is located after at most $30cm$ from the last measurement. In the fully pressurized part of the pipe, the pressure head is constant and its value corresponds to that of the boundary condition imposed downstream.

Table 4.1 summarizes the boundary conditions imposed on the scheme in performing different tests.

Test	1	2	3	4
water flow rate upstream (l/s)	30	36	40	45
pressure head downstream (m.w.c.)	0.554	0.583	0.634	0.69

Table 4.1: Boundary Conditions

The physical and computational parameters are $g = 9.81m/s^2$, $\Delta x = 0.06m$, $\theta = 1.$ and $\Delta t = 10^{-2}s$.

Figures 4.3, 4.4, 4.5, 4.6 show a good agreement between the measured and the predicted data. Moreover, the pressure head η is constant everywhere in the pressurized part of the pipe and its value corresponds to that of the downstream boundary condition.

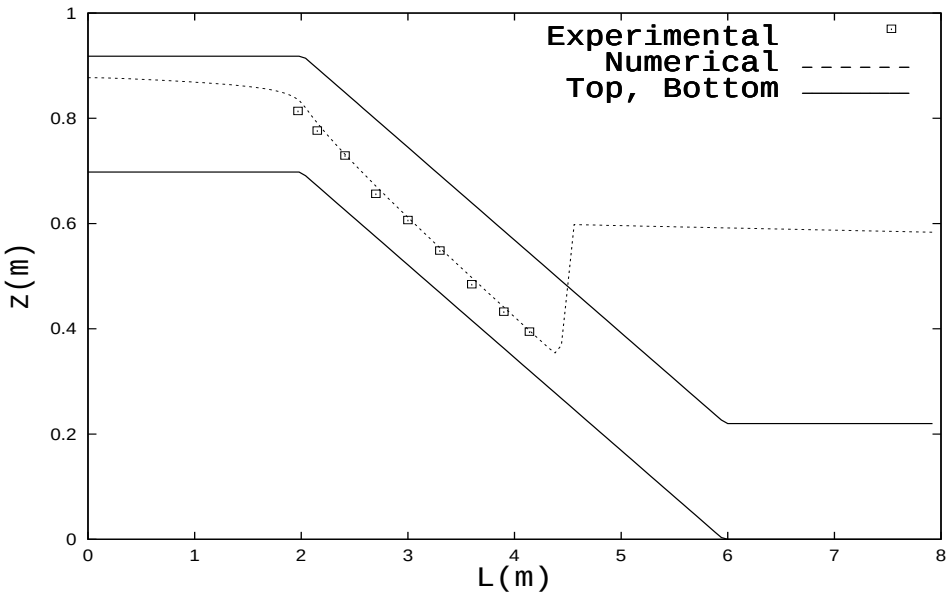

Figure 4.3: Hydraulic Jump in a circular pipe: Test 1.

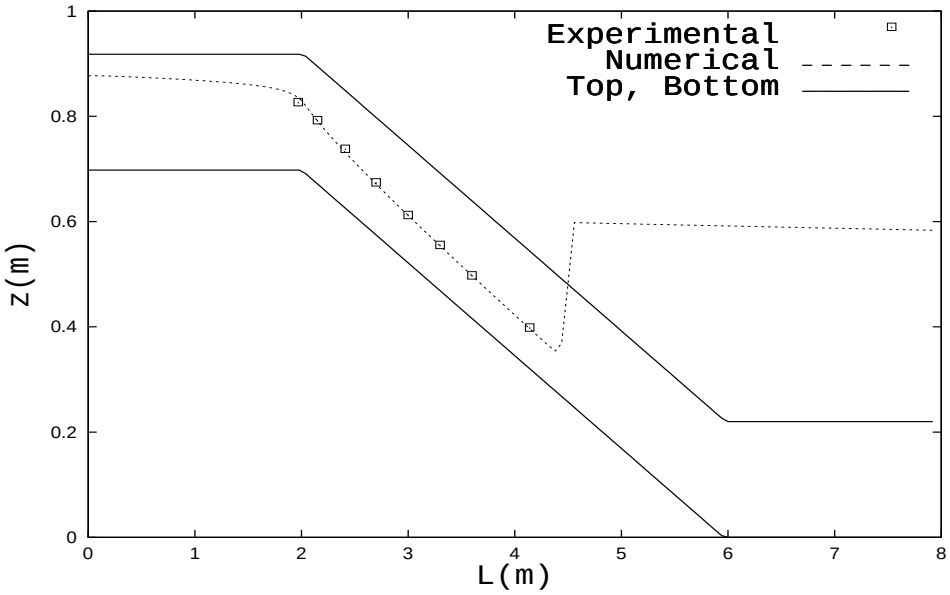

Figure 4.4: Hydraulic Jump in a circular pipe: Test 2.

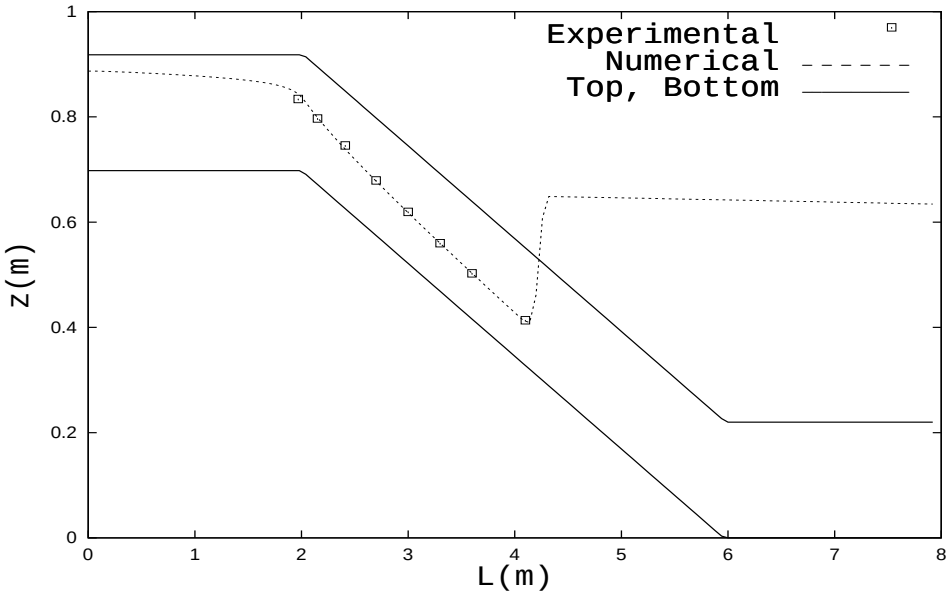

Figure 4.5: Hydraulic Jump in a circular pipe: Test 3.

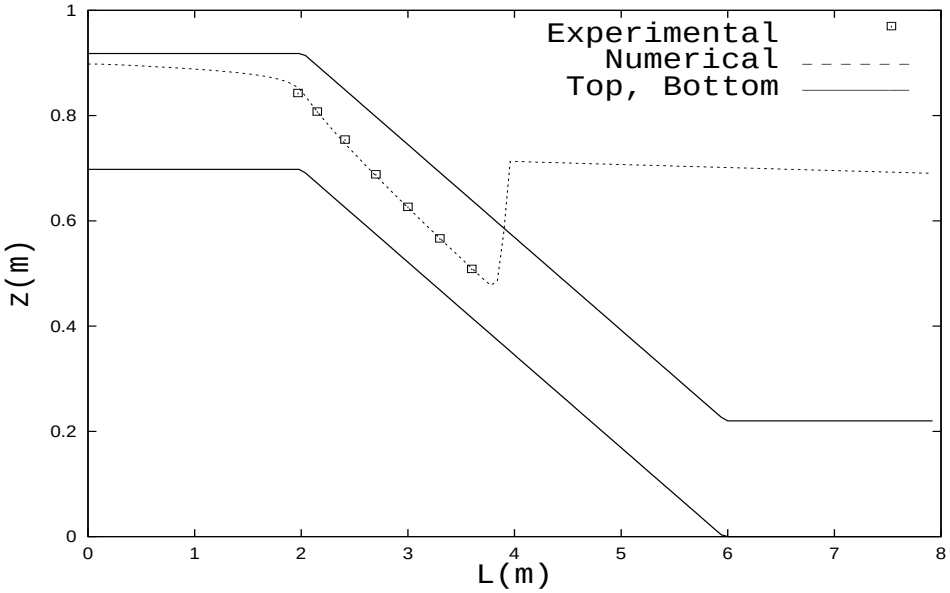

Figure 4.6: Hydraulic Jump in a circular pipe: Test 4.

5

Existence and uniqueness of the numerical solution

The aim of this chapter is to prove the existence and uniqueness of the numerical solution of the scheme presented in Chapter 2 and 4 by introducing a few mathematical assumptions that can be justified by physical argumentations.

5.1 The solution algorithm

At each time step Equations (2.3.4) and (2.4.3) for $i = 1, ...N$ form a system of non-linear equations with unknowns $Q_{i+1/2}^{n+1}$ and η_i^{n+1} over the entire computational mesh.

This system can be reduced for computational convenience to a smaller one in which η_i^{n+1} $i = 1, ...N$ are the only unknowns.

Specifically, the expressions for $Q_{i\pm1/2}^{n+1}$ can be substituted from (2.4.3) into (2.3.4) to obtain

$$V_i(\eta_i^{n+1}) + p_{i-1/2}^n \eta_{i-1}^{n+1} + d_i^n \eta_i^{n+1} + p_{i+1/2}^n \eta_{i+1}^{n+1} = f_i^n \tag{5.1.1}$$

that, for $i = 1, ...N$, constitute the solution system.

The coefficients $p_{i\pm1/2}^n$ on the sub- and superdiagonal of system (5.1.1) are given by

$$p_{i\pm1/2}^n = -\frac{g(\theta\Delta t)^2 A_{i\pm1/2}^n}{\Delta x_{i\pm1/2}(1 + \frac{\gamma_{i\pm1/2}^n}{A_{i\pm1/2}^n}\Delta t)} \quad i = 1, ...N$$

while the coefficients d_i^n on the main diagonal and the known terms f_i^n are defined as

$$d_i^n = -p_{i+1/2}^n - p_{i-1/2}^n$$

and

$$\begin{aligned} f_i^n &= V_i(\eta_i^n) - (1-\theta)\Delta t[Q_{i+1/2}^n - Q_{i-1/2}^n] \\ &- \theta\Delta t[\frac{F_{i+1/2}^n}{(1+\frac{\gamma_{i+1/2}^n}{A_{i+1/2}^n}\Delta t)} - \frac{F_{i-1/2}^n}{(1+\frac{\gamma_{i-1/2}^n}{A_{i-1/2}^n}\Delta t)}] \end{aligned} \tag{5.1.2}$$

for $i = 2, ...N-1$.

The applied boundary conditions complete the definition of the solution system, specifying the elements of the main diagonal and of the known terms on the first and on the N-th rows.

For every time step n, system (5.1.1) can be written in a more compact matrix notation as follows

$$\mathbf{V}(\eta) + \mathbf{M}\eta = \mathbf{f}, \tag{5.1.3}$$

where η=$(\eta_1, \eta_2, ..., \eta_N)^T$ is the vector of the unknowns representing the water level for free surface flows and the pressure head for pressurized flows,

$$\mathbf{V}(\eta) = \begin{pmatrix} V_1(\eta_1) \\ V_2(\eta_2) \\ ... \\ V_N(\eta_N) \end{pmatrix}, \qquad \mathbf{M} = \begin{pmatrix} d_1 & p_{\frac{3}{2}} & \dots & 0 \\ p_{\frac{3}{2}} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_{N-\frac{1}{2}} \\ 0 & ... & p_{N-\frac{1}{2}} & d_N \end{pmatrix}, \qquad \mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ ... \\ f_N \end{pmatrix}. \tag{5.1.4}$$

Once system (5.1.3) has been solved and the solution for η^{n+1} has been determined, Q^{n+1} can be easily computed by substituting η^{n+1} in (2.4.3).

System (5.1.3) is mildly non linear.

The coefficient matrix $\mathbf{M}$ is symmetric and tridiagonal. Moreover, one can assume, without loss of generality, that the elements on the main diagonal are positive and those on the sub- and superdiagonal are negative.

In fact, in the case it exists an $\bar{i}$ such that $p_{\bar{i}+1/2} = 0$, it follows that $A_{\bar{i}+1/2} = \frac{A_{\bar{i}}+A_{\bar{i}+1}}{2} = 0$ and therefore both the $\bar{i}$-th and the $(\bar{i}+1)$-th cell of the spatial domain are empty at time t_n.

Moreover, writing Equation (5.1.1) for $i = \bar{i}$ and for $i = \bar{i}+1$

$$V_{\bar{i}}(\eta_{\bar{i}}^{n+1}) + p_{\bar{i}-1/2}^n \eta_{\bar{i}-1}^{n+1} + d_{\bar{i}}^n \eta_{\bar{i}}^{n+1} = f_{\bar{i}}^n \tag{5.1.5}$$

$$V_{\bar{i}+1}(\eta_{\bar{i}+1}^{n+1}) + d_{\bar{i}+1}^n \eta_{\bar{i}+1}^{n+1} + p_{\bar{i}+3/2}^n \eta_{\bar{i}+2}^{n+1} = f_{\bar{i}+1}^n \tag{5.1.6}$$

one can observe that they are no longer related to each other and therefore system (5.1.3) breaks into two independent systems, specifically Equation (5.1.1) for $i = 1, ...\bar{i}$ and Equation (5.1.1) for $i = \bar{i} + 1, ...N$.

The same procedure can be repeated for every $\bar{i}$ such that $A_{\bar{i}+1/2} = 0$ and a set of independent systems can be obtained.

These new systems are such that the coefficients $p_{i+1/2}$ on their diagonals are all negative and all of them can be linked to one of the couples of boundary conditions that will be introduced in the following sections.

Regarding the non-linear part, $\mathbf{V}$ is a diagonal function and, representing water volumes, it is also non-decreasing.

About its regularity, one can assume that $\mathbf{V}$ is Lipschitz continuous and thus, for every r and s in $\Re$, it holds

$$\mid V_i(r) - V_i(s) \mid \leq L_i \mid r - s \mid \qquad i = 1, ...N$$

where L_i is the Lipschitz constant of V_i. Observe that L_i is positive because the case $L_i = 0$ corresponds to $V_i \equiv constant$.

The diagonal matrix $\mathbf{L}$ such that its main diagonal contains the Lipschitz constants of the components of $\mathbf{V}$, that is $\mathbf{L} = \mathbf{diag}(L_1, L_2, ...L_N)$, will be useful in the following.

The hypothesis of Lipschitz continuity on $\mathbf{V}$ is realistic and consistent with the applications, because, representing V_i the water volume in the cell i, it means that the surface area is always bounded for every η and thus the flow is assumed to be confined within the channel banks.

In the following sections, each component V_i $i = 1, ...N$ of function $\mathbf{V}$ will be properly defined on $\Re$ for open and closed channels.

Actually, observe that a function volume does not have sense for a negative water depth and thus, from the physical point of view, any definition for V_i corresponding to η_i in the range $[-\infty, -h_i]$ will be allowed and meaningless at the same time.

Moreover, the physics of the problem is only interested in $\eta_i \geq -h_i$, but the mathematics involved in the proofs of existence and uniqueness of the solution of system (5.1.3) and in the construction of the constraint on Δt for the non-negativity of the water volume, requires the definition of each function V_i on $\Re$ with particular properties.

5.2 Boundary conditions

The Saint Venant Equations are a hyperbolic system of two partial differential equations such that the existence and uniqueness of their solution is guaranteed if the boundary data satisfy proper conditions.

From the theory of characteristics (see, e.g., [47]) it is known that, in order to have a well-posed problem, boundary conditions should be imposed. Moreover, since the object of our interest is the study of subcritical flows, the boundary conditions have to be assigned one for each boundary of the domain.

From the numerical point of view, one can observe that this choice closes system (5.1.3), in the sense that its number of the equations becomes equal to its number of the unknowns.

One can explicitly show that, studying Equations (2.3.4)-(2.4.3) for $i = 1$

$$V_1(\eta_1^{n+1}) = V_1(\eta_1^n) - \Delta t[Q_{3/2}^{n+\theta} - Q_{1/2}^{n+\theta}] \tag{5.2.1}$$

$$(1 + \frac{\gamma_{3/2}^n}{A_{3/2}^n}\Delta t)Q_{3/2}^{n+1} + gA_{3/2}^n\theta\Delta t\frac{(\eta_2^{n+1} - \eta_1^{n+1})}{\Delta x_{3/2}} = F_{3/2}^n \tag{5.2.2}$$

and for $i = N$

$$V_N(\eta_N^{n+1}) = V_N(\eta_N^n) - \Delta t[Q_{N+1/2}^{n+\theta} - Q_{N-1/2}^{n+\theta}] \tag{5.2.3}$$

$$(1 + \frac{\gamma_{N+1/2}^n}{A_{N+1/2}^n}\Delta t)Q_{N+1/2}^{n+1} + gA_{N+1/2}^n\theta\Delta t\frac{(\eta_{N+1}^{n+1} - \eta_N^{n+1})}{\Delta x_{N+1/2}} = F_{N+1/2}^n, \tag{5.2.4}$$

both the two couples of Equations (5.2.1)-(5.2.2) and (5.2.3)-(5.2.4) require Q or η as boundary condition and, specifically, $Q_{1/2}$ or η_0 and $Q_{N+1/2}$ or η_{N+1} respectively.

In general, in the following, we will talk about Q-type boundary conditions and η-type boundary conditions.

Depending on the chosen type of boundary conditions, the location of the first and of the last node of the spatial grid can change together with the form and the properties of the non-linear system that at each time step Equations (2.3.4)-(2.4.3) form.

In particular, the next two subsections will present the form of the first and of the last row of system (5.1.3) after the application of the boundary conditions.

5.2.1 Q-type boundary conditions

The application of a Q-type boundary condition at the inflow leads the first node being considered to be $x_{1/2}$ and the first row of system (5.1.1) to assume the following form:

$$V_1(\eta_1^{n+1}) - p_{3/2}^n \eta_1^{n+1} + p_{3/2}^n \eta_2^{n+1} = f_1^n, \tag{5.2.5}$$

where

$$f_1^n = V_1(\eta_1^n) - \Delta t\theta \frac{F_{3/2}^n}{(1 + \frac{\gamma_{3/2}^n}{A_{3/2}^n}\Delta t)} + \Delta t Q_{1/2}^{n+\theta} - \Delta t(1-\theta)Q_{3/2}^n.$$

Regarding the outflow, using a Q-type boundary condition leads the last node being considered to be $x_{n+1/2}$ and the last row of system (5.1.1) to become

$$V_N(\eta_N^{n+1}) + p_{N-1/2}^n \eta_{N-1}^{n+1} - p_{N-1/2}^n \eta_N^{n+1} = f_N^n, \tag{5.2.6}$$

where

$$f_N^n = V_N(\eta_N^n) + \Delta t\theta \frac{F_{N-1/2}^n}{(1 + \frac{\gamma_{N-1/2}^n}{A_{N-1/2}^n}\Delta t)} - \Delta t Q_{N+1/2}^{n+\theta} + \Delta t(1-\theta)Q_{N-1/2}^n$$

One can observe that the main diagonal coefficients of Equations (5.2.5) and (5.2.6) are equal to the opposite of the super- and subdiagonal coefficient of the same equation respectively.

5.2.2 η-type boundary conditions

Applying a η-type boundary condition at the inflow, x_1 is the first node of the spatial grid and the first equation of system (5.1.1) assumes the following form

$$V_1(\eta_1^{n+1}) - (p_{1/2}^n + p_{3/2}^n)\eta_1^{n+1} + p_{3/2}^n \eta_2^{n+1} = f_1^n, \tag{5.2.7}$$

where

$$\begin{aligned} f_1^n &= V_1(\eta_1^n) - \Delta t\theta\left[\frac{F_{3/2}^n}{(1 + \frac{\gamma_{3/2}^n}{A_{3/2}^n}\Delta t)} - \frac{F_{1/2}^n}{(1 + \frac{\gamma_{1/2}^n}{A_{1/2}^n}\Delta t)}\right] - \Delta t(1-\theta)[Q_{3/2}^n - Q_{1/2}^n] \\ &+ \frac{g(\theta\Delta t)^2 A_{1/2}^n}{\Delta x_{1/2}(1 + \frac{\gamma_{1/2}^n}{A_{1/2}^n}\Delta t)}\eta_0^{n+1} \end{aligned} \tag{5.2.8}$$

On the other hand, using a η-type boundary condition at the outflow, x_{N+1} is the last node of the spatial grid and the N-th equation of system (5.1.1) becomes

$$V_N(\eta_N^{n+1}) + p_{N-1/2}^n \eta_{N-1}^{n+1} - (p_{N-1/2}^n + p_{N+1/2}^n)\eta_N^{n+1} \quad = f_N^n \tag{5.2.9}$$

where, extending notation (5.1.2) to the node N,

$$\begin{aligned} f_N^n &= V_N(\eta_N^n) - \Delta t\theta[\frac{F_{N+1/2}^n}{(1+\frac{\gamma_{N+1/2}^n}{A_{N+1/2}^n}\Delta t)} - \frac{F_{N-1/2}^n}{(1+\frac{\gamma_{N-1/2}^n}{A_{N-1/2}^n}\Delta t)}] \\ &- \Delta t(1-\theta)[Q_{N+1/2}^n - Q_{N-1/2}^n] + \frac{g(\theta\Delta t)^2 A_{N+1/2}^n}{\Delta x_{N+1/2}(1+\frac{\gamma_{N+1/2}^n}{A_{N+1/2}^n}\Delta t)}\eta_{N+1}^{n+1} \end{aligned} \tag{5.2.10}$$

One can observe that the main diagonal coefficients of Equations (5.2.5) and (5.2.6) are greater than the opposite of the super- and subdiagonal coefficient of the same equation respectively.

5.3 Existence and uniqueness of the solution of system (5.1.3) with at least a η-type boundary condition

The aim of this section is to prove the existence and uniqueness of the solution of system (5.1.3), assuming that at least one of the boundary conditions is of the η-type.

Under this hypothesis, let characterize system (5.1.3) by setting the assumptions for the proof of the final result.

As previously mentioned, matrix $\mathbf{M}$ is tridiagonal, symmetric, with positive elements on the main diagonal and negative ones on the sub- and superdiagonal. Therefore, it is said to be irreducible, because

Definition 5.3.1 *A tridiagonal matrix $\boldsymbol{M} \in L(\Re^N)$ is irreducible whenever the entries of the super- and subdiagonal are non-zero.*

Moreover, $\mathbf{M}$ is also diagonally dominant, in the sense that

Definition 5.3.2 *A matrix $\boldsymbol{M} = (m_{i,j})$ in $L(\Re^N)$ is diagonally dominant if and only if it holds*

$$|m_{ii}| \geq \sum_{j=1, j\neq i}^{n} |m_{ij}|, \qquad i = 1, ...N \tag{5.3.1}$$

with strict inequality valid for at least one value of i.

The previous definition is in fact satisfied, because the application of at least one η-type boundary condition at the boundaries assures inequality (5.3.1) to be strict for at least one row (where the η-type boundary condition is applied).

Therefore, by the following theorem [32], the linear part of system (5.1.1) is also positive definite and thus non-singular.

Theorem 5.3.3 *If matrix $\boldsymbol{M} \in L(\Re^N)$ is symmetric, irreducible, diagonally dominant and has positive diagonal elements, then $\boldsymbol{M}$ is positive definite. The determinant of a positive definite matrix is always positive, so a positive definite matrix is always non-singular.*

Regarding the non-linear part of system (5.1.1), function $\mathbf{V}$ represents the water volume in the cells of the channel and therefore, for its physical meaning, it is an isotone function, where

Definition 5.3.4 *A mapping $\boldsymbol{P} : \Re^N \to \Re^N$ is said to be isotone (non-decreasing) if*

$$\boldsymbol{P}(\boldsymbol{x}) \leq \boldsymbol{P}(\boldsymbol{y}) \tag{5.3.2}$$

whenever $\boldsymbol{x} \leq \boldsymbol{y}$, $\boldsymbol{x}, \boldsymbol{y} \in \Re^N$. $\boldsymbol{P}$ is strictly isotone (or increasing) if strict inequality holds in (5.3.2) whenever $\boldsymbol{x} \neq \boldsymbol{y}$.

In Definition 5.3.4 and in the following of this work, the comparison of two vectors of $\Re^N$ will be done element by element. This one may do by means of the *natural or component-wise partial ordering* on $\Re^N$ defined by

$$\mathbf{x}, \mathbf{y} \in \Re^N, \quad \mathbf{x} \leq \mathbf{y} \quad \text{if and only if} \quad x_i \leq y_i, i = 1, ...N$$

No stronger assumptions are required on $\mathbf{V}$ and thus one of the possible ways to define its components V_i $i = 1, ...N$ is the following

$$V_i(\eta_i) = \begin{cases} 0 & \text{if } \eta_i \leq -h_i \\ V_i(\eta_i) & \text{if } -h_i \leq \eta_i \leq top_i \\ V_i(top_i) & \text{if } \eta_i \geq top_i \end{cases} \tag{5.3.3}$$

where top_i is the maximum value allowed for η_i in the cell i and corresponds to $+\infty$ only in the case of an open channel.

Observe that the definition of each function volume V_i is univocal only for $\eta_i \geq -h_i$. In this interval, V_i is isotone for closed channels and strictly isotone for open ones.

Moreover, for η_i in the range $[-\infty, -h_i]$, any expression is mathematically admissible, but, as already said, physically meaningless at the same time.

In particular for a closed channel, the function volume V_i is isotone on $\Re$ regardless its expression in $[-\infty, -h_i]$.

On the other hand, when the channel is open, the monotonic behaviour of V_i on $\Re$ depends on the properties of its definition in this interval and V_i results to be strictly isotone if and only if it is strictly isotone also in $[-\infty, -h_i]$ (see, e.g., Equation (5.4.1)).

Finally, collecting all these hypotheses, let introduce the following theorem [32] that helps in proving the final result.

Theorem 5.3.5 *Let $\boldsymbol{M} \in L(\Re^N)$ be symmetric, positive definite and suppose that $\boldsymbol{V}$ is continuous, diagonal and isotone on $\Re^n$.*

Then mapping $\boldsymbol{P} : \Re^N \to \Re^N$ defined by $\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{V}(\boldsymbol{x})$ is a homeomorphism of $\Re^N$ onto $\Re^N$.

Here, by homeomorphism we mean that

Definition 5.3.6 *A mapping $\boldsymbol{P} : \boldsymbol{D} \subset \Re^N \to \Re^N$ is a homeomorphism of $\boldsymbol{D}$ onto $\boldsymbol{P}(\boldsymbol{D})$ if $\boldsymbol{P}$ is one-to-one on $\boldsymbol{D}$ and $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$ are continuous on $\boldsymbol{D}$ and $\boldsymbol{P}(\boldsymbol{D})$ respectively.*

and by one-to-one the following definition holds

Definition 5.3.7 *A mapping $\boldsymbol{P} : \boldsymbol{D} \subset \Re^N \to \Re^N$ is one-to-one on $\boldsymbol{U} \subset \boldsymbol{D}$ if $\boldsymbol{P}(\boldsymbol{x}) \neq \boldsymbol{P}(\boldsymbol{y})$ whenever $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{U}$, $\boldsymbol{x} \neq \boldsymbol{y}$.*

Observe that the mapping $\mathbf{P}$ of Theorem 5.3.5 is a homeomorphism of $\Re^N$ onto itself and therefore its domain $\mathbf{D}$ and codomain $\mathbf{P}(\mathbf{D})$ are both $\Re^N$.

Finally, let remark that, when at least one boundary condition is of the η-type and the channel is either open or closed, system (5.1.3) satisfies all the assumptions on the domain and on the properties of the mapping $\mathbf{P}$ of Theorem 5.3.5.

Therefore, the following corollary can be applied to prove the existence and uniqueness of its numerical solution.

Corollary 5.3.8 *Under the same hypotheses of Theorem 5.3.5 and for any* $\boldsymbol{f} \in \Re^N$, *system (5.1.3) given by* $\boldsymbol{V}(\eta) + \boldsymbol{M}\eta = \boldsymbol{f}$ *has a unique solution.*

Actually, the existence and uniqueness of the numerical solution do not ensure the physical meaning and therefore the computed η could result less that the channel bottom in some of the cells of the spatial domain.

Chapter 6 will provide a constraint on the time step Δt in order to ensure the physicality of the solution and therefore the non-negativity of the water volume.

5.4 Existence and uniqueness of the solution of system (5.1.3) with two Q-type boundary conditions for open channel flows

The aim of this section is to prove, when possible, the existence and uniqueness of the solution of system (5.1.3), assuming that both the boundary conditions are of the Q-type.

Let first suppose that function $\mathbf{V}$ is isotone and therefore consider the case of a closed channel, because the volume of any open channel can be defined as strictly isotone.

Under this set of hypotheses, the existence and uniqueness of the solution of system (5.1.3) cannot be usually proved.

Actually, this is physically correct, because the solution of a flow in a closed and fully pressurized channel is not unique. In fact, given η a numerical solution of (2.3.4)-(2.4.3), it can be proved directly from these two Equations that infinitely many other solutions can be obtained adding any constant $\mathbf{K} \in \Re^N$ to η.

Therefore, the existence and uniqueness of the solution of (5.1.3) will be studied here assuming that the channel is open. Morever, from the mathematical point of view, such a system could also be impossible to solve. Actually, in the following we will assume that it exists at least one solution.

Let first of all characterize system (5.1.3) by setting the hypotheses for the proof of the final result.

Matrix $\mathbf{M}$ is tridiagonal, irreducible and symmetric, with positive elements on the main diagonal and negative ones on the sub- and superdiagonal.

It is not diagonally dominant, because inequality (5.3.1) is actually an equality for every $i = 1, ...N$ and therefore $\mathbf{M}$ is singular and positive semi-definite.

On the other hand, the non-linear part $\mathbf{V}$ of system (5.1.3) is required to be a strictly isotone function, that can be realized only in the case of open channels defining its components V_i $i = 1, ...N$ in the following way

$$V_i(\eta_i) = \begin{cases} -V_i(-\eta_i - 2h_i) & \text{if } \eta_i \leq -h_i \\ V_i(\eta_i) & \text{if } \eta_i \geq -h_i \end{cases} \tag{5.4.1}$$

Actually, this requirement on function $\mathbf{V}$ is not strong enough to prove, together with the other assumptions, the final result.

The following property is therefore introduced. Let be $\mathbf{x}$ and $\mathbf{y} \in \Re^N$. Thus, there exist a positive constant c in $\Re$ independent on $\mathbf{x}$ and $\mathbf{y}$ such that it holds

$$\mid V_i(x_i) - V_i(y_i) \mid \geq c \mid x_i - y_i \mid \qquad i = 1, ...N \tag{5.4.2}$$

Property (5.4.2) states that the absolute value of the V_i's incremental ratio has c as lower bound.

Moreover, assuming that the derivative of V_i exists on $\Re$, the previous condition consists in requiring that V_i does not have horizontal asymptotes or, in other words, that the surface area $\overline{A}_i = \frac{\partial V_i}{\partial \eta_i}$ is such that $\overline{A}_i \geq c$ for every $\eta_i > -h_i$.

In fact, applying the Mean-Value Theorem [32] to V_i on $[x_i, y_i]$, there exists $\xi_i \in (x_i, y_i)$ such that

$$V_i(x_i) - V_i(y_i) = \overline{A}_i(\xi_i)(x_i - y_i). \tag{5.4.3}$$

$\overline{A}_i(\xi_i) \geq c > 0$ follows directly from the comparison between Equations (5.4.2) and (5.4.3).

On the other hand, for a Lipschitz and not differentiable function V_i, it is known that there exist a constant S_i dependent on x_i and y_i, $0 \leq S_i(x_i, y_i) \leq L_i$, such that

$$V_i(x_i) - V_i(y_i) = S_i(x_i - y_i). \tag{5.4.4}$$

By the strict isotonicity of V_i it results that $S_i(x_i, y_i) > 0$ for every $x_i \neq y_i$.

Therefore, the previous considerations on the surface area $\overline{A}_i$ can be referred to the constant S_i, requiring that the latter has a lower bound $c \in \Re$, $c > 0$ such that $S_i \geq c > 0$.

Let now introduce the following theorem [32], that will be used in the proof of the final result.

Theorem 5.4.1 *Assume that* $\Phi : \Re^N \to \Re$ *is strongly convex and continuously differentiable on* $\Re^N$. *Then the mapping* $\boldsymbol{P} : \Re^N \to \Re^N$ *defined by* $\boldsymbol{P}(\boldsymbol{x}) = \nabla\Phi(\boldsymbol{x}), \boldsymbol{x} \in \Re^N$, *is a homeomorphism from* $\Re^N$ *onto* $\Re^N$.

Here, by *strongly convex* we mean that there exist $c \in \Re, c > 0$ such that

$$[\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{y})]^T(\mathbf{x} - \mathbf{y}) \geq c\|\mathbf{x} - \mathbf{y}\|^2 \forall \mathbf{x}, \mathbf{y} \in R^N.$$

A consequence of the above theorem is the following variation of Theorem 5.3.5.

Theorem 5.4.2 *Let* $\boldsymbol{M} \in L(\Re^N)$ *be symmetric, positive semi-definite. Suppose that the function* $\boldsymbol{V}$ *is a Lipschitz continuous function, diagonal, strictly isotone and such that it satisfies (5.4.2).*

Then mapping $\boldsymbol{P} : \Re^N \to \Re^N$ *defined by* $\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{V}(\boldsymbol{x})$ *is a homeomorphism of* $\Re^N$ *onto* $\Re^N$.

Proof. Define

$$\Phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{M}\mathbf{x} + \sum_{i=1}^{N}\int_0^{x_i} V_i(\xi)d\xi. \tag{5.4.5}$$

Φ is continuously differentiable on $\Re^N$ by definition and $\nabla\Phi(\mathbf{x}) = \mathbf{P}(\mathbf{x})^T$.

Thus, given $\mathbf{x}, \mathbf{y} \in \Re^N$, it holds

$$[\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{y})]^T(\mathbf{x} - \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T\mathbf{M}(\mathbf{x} - \mathbf{y}) + (\mathbf{V}(\mathbf{x}) - \mathbf{V}(\mathbf{y}))^T(\mathbf{x} - \mathbf{y})$$

and because matrix $\mathbf{M}$ is positive semi-definite

$$[\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{y})]^T(\mathbf{x} - \mathbf{y}) \geq (\mathbf{V}(\mathbf{x}) - \mathbf{V}(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}).$$

Now, introducing the property (5.4.4) for Lipschitz and strictly isotone functions, one has

$$\begin{aligned}
(\mathbf{V}(\mathbf{x}) - \mathbf{V}(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}) &= \sum_{i=1}^{N}(V_i(x_i) - V_i(y_i))(x_i - y_i) \\
&= \sum_{i=1}^{N} S_i(x_i, y_i)(x_i - y_i)^2 \\
&= \sum_{i=1, x_i \neq y_i}^{N} S_i(x_i, y_i)(x_i - y_i)^2 \\
&\geq min_{i=1,\ldots N, x_i \neq y_i} S_i(x_i, y_i)\|\mathbf{x} - \mathbf{y}\|^2. \qquad (5.4.6)
\end{aligned}$$

Therefore

$$(\mathbf{V}(\mathbf{x}) - \mathbf{V}(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}) \geq c\|\mathbf{x} - \mathbf{y}\|^2$$

where c is the constant defined in Equation (5.4.2) such that $S_i(s, r) \geq c > 0 \; \forall s, r \in \Re, s \neq r$.

The relation

$$[\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{y})]^T(\mathbf{x} - \mathbf{y}) \geq (\mathbf{V}(\mathbf{x}) - \mathbf{V}(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}) \geq c\|\mathbf{x} - \mathbf{y}\|^2$$

means that Φ is strongly convex.

Therefore, the application of Theorem 5.4.1 proves the theorem. ■

Finally, let remark that, when the two boundary conditions are of the Q-type and the channel is open with the characteristic that the surface area has a lower bound greater than zero for every non zero water depth, system (5.1.3) satisfies all the assumptions on the domain and on the properties of the mapping $\mathbf{P}$ of Theorem 5.4.2.

Therefore, the following corollary can be applied to prove the existence and uniqueness of its numerical solution.

Corollary 5.4.3 *Under the same hypotheses of Theorem 5.4.2 and for any* $\boldsymbol{f} \in \Re^N$, *system (5.1.3) given by* $\boldsymbol{V}(\eta) + \boldsymbol{M}\eta = \boldsymbol{f}$ *has a unique solution.*

Observe that the existence and uniqueness of the numerical solution do not ensure the physical meaning and therefore the computed η could result less that the channel bottom in some of the cells of the spatial domain.

Chapter 6 will provide a constraint on the time step Δt in order to ensure the physicality of the solution and therefore the non-negativity of the water volume.

6

Non-negativity of the water volume

The aim of this chapter is to formulate an explicit and an implicit constraint on the time step Δt to ensures the non-negativity of the numerical water volume obtained by the algorithm proposed in Chapter 2 and and 4. The advantages of using the explicit constraint are discussed and shown with an interesting numerical example.

6.1 Introduction

Existence and uniqueness do not ensure that the numerical solution is physically meaningful.

It could happen in fact, that somewhere the computed numerical water surface or pressure head results less than the bottom of the channel and thus the water volume in those cells is negative.

Non-negativity is a very important physical property that the solution of a numerical scheme for Equations (1.4.1)-(1.4.3) should have, first of all because it ensures a correct treatment of the phenomena of flooding and drying and a physical meaningful solution.

Consider everything is known at the time t_n, $\eta^n \geq -\mathbf{h}$ and assume we want to compute the new numerical solution η^{n+1} solving system (5.1.3) under the assumptions that assure existence and uniqueness of its solutions.

6.2 An implicit constraint on Δt

From Equation (2.3.4) one can easily derive a condition on the time step Δt to ensure non-negativity of the water volume, that is:

$$[Q^{n+\theta}_{i+1/2} - Q^{n+\theta}_{i-1/2}]\Delta t \leq V_i(\eta^n_i) \quad \forall i. \tag{6.2.1}$$

This constraint is algebraically very easy to be calculated, both in the case of a rectangular and of a non-rectangular channel, but it is only useful as *a posteriori* check, because it is implicit in time in the sense that it involves quantities not yet computed.

Only at the steady state of a phenomenon, inequality (6.2.1) could be considered *almost explicit* and sufficiently correct substituting the time level $n + \theta$ with n.

6.3 An explicit constraint on Δt

The analysis of the solution system (5.1.3) from a different point of view can lead to an explicit condition on the time step Δt to ensure the non-negativity of the water volume when the existence and uniqueness of its solution can be proved.

In this section, a few mathematical properties of system (5.1.3) will be pointed out in order to introduce this *a-priori* check on Δt.

First of all, let recall the following definition [32].

Definition 6.3.1 *A mapping $\boldsymbol{P} : \Re^N \to \Re^N$ is inverse isotone if $\boldsymbol{P}(\boldsymbol{x}) \leq \boldsymbol{P}(\boldsymbol{y})$ for any $\boldsymbol{x}, \boldsymbol{y} \in \Re^N$ implies that $\boldsymbol{x} \leq \boldsymbol{y}$.*

In particular, it is possible to prove that function $\mathbf{P}(\mathbf{x}) = \mathbf{V}(\mathbf{x}) + \mathbf{M}\mathbf{x}$ is an inverse isotone function both in the case that $\mathbf{M}$ is a positive-definite matrix and $\mathbf{V}$ is a diagonal, continuous and isotone function (hypotheses of Section 5.3) and in the case that $\mathbf{M}$ is a semi-positive definite matrix and $\mathbf{V}$ is a diagonal, continuous and strictly isotone function (hypotheses of Section 5.4).

Once these results are proved, we can conclude that if the check $F(\eta^{n+1}) \geq F(-\mathbf{h})$ is satisfied, the solution we will get at the new time t_{n+1} will be greater than $-\mathbf{h}$, and therefore will be physically meaningful.

Let now proceed with the proof of the inverse isotonicity of function $\mathbf{P}(\mathbf{x}) = \mathbf{V}(\mathbf{x}) + \mathbf{M}\mathbf{x}$ in the two cases mentioned above.

Under the assumptions of Theorem 5.3.5 in Section 5.3, matrix $\mathbf{M}$ is an M-matrix, that is

Definition 6.3.2 *A matrix $\boldsymbol{M} \in L(\Re^N)$ is an M-matrix if $\boldsymbol{M}$ is invertible, $\boldsymbol{M}^{-1} \geq \boldsymbol{0}$ and $m_{i,j} \leq 0$ for all $i, j = 1, ...N$, $i \neq j$.*

and that can be proved by the following result [32].

Theorem 6.3.3 *Let $\boldsymbol{M} \in L(\Re^N)$ be irreducible and diagonally dominant and assume $m_{i,j} \leq 0$, $i \neq j$, and that $m_{i,i} > 0$, $i = 1, ...N$. Then $\boldsymbol{M}$ is an M-matrix.*

Now, the inverse isotonicity of $\mathbf{P}$ is given by the following theorem [32].

Theorem 6.3.4 *Let $\boldsymbol{M} \in L(\Re^N)$ be an M-matrix and suppose that $\boldsymbol{V} : \Re^N \to \Re^N$ is continuous, diagonal and isotone. Then mapping $\boldsymbol{P} : \Re^N \to \Re^N$ defined by $\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{V}(\boldsymbol{x})$ is inverse isotone.*

On the other hand, under the same hypotheses of Theorem 5.4.2 in Section 5.4, the strict isotonicity of $\mathbf{P}$ can be proved.

To do this, let first recall the following result [32].

Theorem 6.3.5 *Let $\boldsymbol{A}_1 \in L(\Re^N)$ be an M-matrix with diagonal part $\boldsymbol{D}_1$ and off-diagonal part $-\boldsymbol{B}_1 = \boldsymbol{A}_1 - \boldsymbol{D}_1$. If $\boldsymbol{D}_2 \in L(\Re^N)$ is any non-negative diagonal matrix and $\boldsymbol{B}_2 \in L(\Re^N)$ any non-negative matrix with zero diagonal satisfying $\boldsymbol{B}_2 \leq \boldsymbol{B}_1$, then $\boldsymbol{A} = \boldsymbol{D}_1 + \boldsymbol{D}_2 - (\boldsymbol{B}_1 - \boldsymbol{B}_2)$ is an M-matrix and $\boldsymbol{A}^{-1} \leq \boldsymbol{A}_1^{-1}$.*

Finally, the strict isotonicity of $\mathbf{P}$ can be proved by the following theorem.

Theorem 6.3.6 *Let $\boldsymbol{M} \in L(\Re^N)$ be a tridiagonal, irreducible matrix such that*

$$\begin{aligned} & m_{1,1} = -m_{1,2} \\ & m_{i,i} = -m_{i,i-1} - m_{i,i+1} \qquad i = 2, ...N-1 \\ & m_{N,N} = -m_{N,N-1} \end{aligned} \tag{6.3.1}$$

Suppose that $\boldsymbol{V} : \Re^N \to \Re^N$ is Lipschitz continuous, diagonal, strictly isotone and it satisfies (5.4.2). Then mapping $\boldsymbol{P} : \Re^N \to \Re^N$ defined by $\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{V}(\boldsymbol{x})$ is inverse isotone.

Proof. Suppose that $\mathbf{P}(\mathbf{x}) \leq \mathbf{P}(\mathbf{y})$ for some $\mathbf{x}, \mathbf{y} \in \Re^N$ for which $\mathbf{x} \leq \mathbf{y}$ does not hold. Set $\mathbf{N} = \{1 \leq j \leq n \mid x_j > y_j\}$. Consider $j \in \mathbf{N}$. Then

$$0 \leq P_j(\mathbf{y}) - P_j(\mathbf{x}) = V_j(y_j) - V_j(x_j) + \sum_{k=j-1}^{j+1} m_{j,k}(y_k - x_k) \tag{6.3.2}$$

and by strict isotonicity and Lipschitz continuity of V_j one obtains

$$0 \leq P_j(\mathbf{y}) - P_j(\mathbf{x}) \;\; = \;\; S_j(x_j, y_j)(y_j - x_j) + \sum_{k=j-1}^{j+1} m_{j,k}(y_k - x_k)$$

$$= S_j(x_j, y_j)(y_j - x_j) + \sum_{k \in \mathbf{N}, k=j-1}^{j+1} m_{j,k}(y_k - x_k)$$

$$+ \sum_{k \notin \mathbf{N}, k=j-1}^{j+1} m_{j,k}(y_k - x_k) \tag{6.3.3}$$

where $S_j(x_j, y_j)$ is the same constant is the same constant introduced in (5.4.4).

Regarding Equation (6.3.3), one can observe that, for every $k \notin \mathbf{N}$, $y_k - x_k > 0$ and $m_{j,k} < 0$. Thus

$$0 \leq P_j(\mathbf{y}) - P_j(\mathbf{x}) \leq (m_{j,j} + S_j(x_j, y_j))(y_j - x_j) + \sum_{k \in \mathbf{N}, k=j\pm1} m_{j,k}(y_k - x_k) \tag{6.3.4}$$

Let now observe that matrix $\mathbf{G} = \mathbf{M} + \mathbf{S}$ is an M-matrix, because it satisfies the assumptions o Theorem 6.3.3.

Therefore, by Theorem 6.3.5 it follows that the submatrix $\mathbf{A} = (g_{j,k} \mid j, k \in \mathbf{N})$ is also an M-matrix and therefore has the property that, given $\mathbf{x} \in \Re^N$, $\mathbf{Ax} \geq \mathbf{0}$ implies that $\mathbf{x} \geq \mathbf{0}$.

Finally, rewriting Equation (6.3.4) as follows

$$0 \leq P_j(\mathbf{y}) - P_j(\mathbf{x}) \leq (\mathbf{A}(\mathbf{y} - \mathbf{x}))_j \tag{6.3.5}$$

one can show that $y_j \geq x_j$ for all $j \in \mathbf{N}$, that is a contradiction. This proves that $\mathbf{P}$ is inverse isotone. ■

Restarting from Definition 6.3.1, we will explicitly show that it establishes itself a criterion for the non-negativity of the water volume.

In fact, setting $\mathbf{x}$ equal to the channel bottom $-\mathbf{h}$ and $\mathbf{y}$ equal to the solution η we will get at time t_{n+1}, the check for the non-negativity of the water volume assumes the following form

$$\mathbf{P}(\eta) \geq \mathbf{P}(-\mathbf{h}) \Rightarrow \eta \geq -\mathbf{h} \tag{6.3.6}$$

In case the comparison between $\mathbf{P}(-\mathbf{h})$ and $\mathbf{P}(\eta)$ could be done explicitly at time t_n and would be expressed in function of the time step Δt, we would know *a priori* the range for Δt that ensures $\eta \geq -\mathbf{h}$.

Observe that, being $-\mathbf{h}$ a known quantity and not a part of the solution, $\mathbf{P}(-\mathbf{h})$ can be explicitly written as

$$P_j(-\mathbf{h}) = (\mathbf{V}(-\mathbf{h}) + \mathbf{M}(-\mathbf{h}))_j$$

$$
\begin{aligned}
&= p^n_{j-1/2}(-h_{j-1}) + d^n_j(-h_j) + p^n_{j+1/2}(-h_{j+1}) \\
&= \frac{g(\theta\Delta t)^2 A^n_{j-1/2}}{\Delta x_{j-1/2}(1+\frac{\gamma^n_{j-1/2}}{A^n_{j-1/2}}\Delta t)} h_{j-1} \\
&- \left(\frac{g(\theta\Delta t)^2 A^n_{j-1/2}}{\Delta x_{j-1/2}(1+\frac{\gamma^n_{j-1/2}}{A^n_{j-1/2}}\Delta t)} + \frac{g(\theta\Delta t)^2 A^n_{j+1/2}}{\Delta x_{j+1/2}(1+\frac{\gamma^n_{j+1/2}}{A^n_{j+1/2}}\Delta t)}\right) h_j \\
&+ \frac{g(\theta\Delta t)^2 A^n_{j+1/2}}{\Delta x_{j+1/2}(1+\frac{\gamma^n_{j+1/2}}{A^n_{j+1/2}}\Delta t)} h_{j+1}
\end{aligned} \tag{6.3.7}
$$

for $j = 2, \ldots N-1$, while for $j = 1$ and $j = N$ its definition depends on the boundary condition applied on the first and on last cell respectively.

In particular, assuming that a Q-type boundary condition is imposed at $j = 1$,

$$
\begin{aligned}
P_1(-\mathbf{h}) &= (\mathbf{V}(-\mathbf{h}) + \mathbf{M}(-\mathbf{h}))_1 = -p^n_{3/2}(-h_1) + p^n_{3/2}(-h_2) \\
&= -\frac{g(\theta\Delta t)^2 A^n_{3/2}}{\Delta x_{3/2}(1+\frac{\gamma^n_{3/2}}{A^n_{3/2}}\Delta t)} h_1 + \frac{g(\theta\Delta t)^2 A^n_{3/2}}{\Delta x_{3/2}(1+\frac{\gamma^n_{3/2}}{A^n_{3/2}}\Delta t)} h_2
\end{aligned} \tag{6.3.8}
$$

while, in the case a η-type boundary condition is chosen for the first cell, one has

$$
\begin{aligned}
P_1(-\mathbf{h}) &= -(p^n_{1/2} + p^n_{3/2})(-h_1) + p^n_{3/2}(-h_2) \\
&= -\left(\frac{g(\theta\Delta t)^2 A^n_{1/2}}{\Delta x_{1/2}(1+\frac{\gamma^n_{1/2}}{A^n_{1/2}}\Delta t)} + \frac{g(\theta\Delta t)^2 A^n_{3/2}}{\Delta x_{3/2}(1+\frac{\gamma^n_{3/2}}{A^n_{3/2}}\Delta t)}\right) h_1 \\
&+ \frac{g(\theta\Delta t)^2 A^n_{3/2}}{\Delta x_{3/2}(1+\frac{\gamma^n_{3/2}}{A^n_{3/2}}\Delta t)} h_2
\end{aligned} \tag{6.3.9}
$$

With the same procedure one can find the expression for $P_N(-\mathbf{h})$ depending on the boundary condition given on the cell N.

On the other hand, let consider η as the solution of system (5.1.3).

Therefore, $\mathbf{P}(\eta)$ is equal to the known term $\mathbf{f}$ of the system and can be expressed by Equation (5.1.2) in terms of known quantities. Specifically

$$
\begin{aligned}
P_j(\eta) &= V_j(\eta^n_j) - (1-\theta)\Delta t[Q^n_{j+1/2} - Q^n_{j-1/2}] \\
&- \frac{\theta\Delta t}{(1+\frac{\gamma^n_{j+1/2}}{A^n_{j+1/2}}\Delta t)}\Big[Q^n_{j+1/2} - \Delta t\frac{[(UQ)^n_{j+1} - (UQ)^n_j]}{\Delta x_{j+1/2}} \\
&\qquad - gA^n_{j+1/2}(1-\theta)\Delta t\frac{(\eta^n_{j+1} - \eta^n_j)}{\Delta x_{j+1/2}}\Big]
\end{aligned}
$$

$$+ \quad \frac{\theta\Delta t}{(1+\frac{\gamma^n_{j-1/2}}{A^n_{j-1/2}}\Delta t)}[Q^n_{j-1/2} - \Delta t\frac{[(UQ)^n_j - (UQ)^n_{j-1}]}{\Delta x_{j-1/2}} - gA^n_{j-1/2}(1-\theta)\Delta t\frac{(\eta^n_j - \eta^n_{j-1})}{\Delta x_{j-1/2}}] \quad (6.3.10)$$

for $j = 2, ...N-1$, while for $j = 1$ and $j = N$ its definition depends on the boundary condition applied on the first and on last cell respectively.

In particular, assuming that a Q-type boundary condition is imposed at $j = 1$,

$$\begin{aligned} P_1(\eta) &= V_1(\eta^n_1) - \Delta t(1-\theta)Q^n_{3/2} + \Delta t Q^{n+\theta}_{1/2} \\ &- \frac{\theta\Delta t}{(1+\frac{\gamma^n_{3/2}}{A^n_{3/2}}\Delta t)}[Q^n_{3/2} - \Delta t\frac{[(UQ)^n_2 - (UQ)^n_1]}{\Delta x_{3/2}} - gA^n_{3/2}(1-\theta)\Delta t\frac{(\eta^n_2 - \eta^n_1)}{\Delta x_{3/2}}] \end{aligned} \quad (6.3.11)$$

while, in the case a η-type boundary condition is chosen for the first cell, one has

$$\begin{aligned} P_1(\eta) &= V_1(\eta^n_1) - (1-\theta)\Delta t[Q^n_{3/2} - Q^n_{1/2}] \\ &- \frac{\theta\Delta t}{(1+\frac{\gamma^n_{3/2}}{A^n_{3/2}}\Delta t)}[Q^n_{3/2} - \Delta t\frac{[(UQ)^n_2 - (UQ)^n_1]}{\Delta x_{3/2}} - gA^n_{3/2}(1-\theta)\Delta t\frac{(\eta^n_2 - \eta^n_1)}{\Delta x_{3/2}}] \\ &+ \frac{\theta\Delta t}{(1+\frac{\gamma^n_{1/2}}{A^n_{1/2}}\Delta t)}[Q^n_{1/2} - \Delta t\frac{[(UQ)^n_1 - (UQ)^n_0]}{\Delta x_{1/2}} - gA^n_{1/2}(1-\theta)\Delta t\frac{(\eta^n_1 - \eta^n_0)}{\Delta x_{1/2}}] \\ &+ \frac{(gA^n_{1/2}\theta\Delta t)^2}{(1+\frac{\gamma^n_{1/2}}{A^n_{1/2}}\Delta t)\Delta x_{1/2}}\eta^{n+1}_0 \end{aligned} \quad (6.3.12)$$

With the same procedure one can find the expression for $P_N(\eta)$ depending on the boundary condition given on the cell N.

Finally, we are able to explicitly express the relation $\mathbf{P}(\eta) \geq \mathbf{P}(-\mathbf{h})$ and to translate it into a constraint on the time step Δt as follows.

To simplify the calculations, consider a frictionless fluid ($\gamma = 0$).

Therefore, from Equations (6.3.7) and (6.3.10), $P_j(\eta) \geq P_j(-\mathbf{h})$ can be written as

$$a_1(\Delta t)^2 + a_2\Delta t + a_3 \geq 0 \quad (6.3.13)$$

where the coefficients a_1, a_2 and a_3 are defined as follows

$$a_1 = \theta\frac{[(UQ)^n_{j+1} - (UQ)^n_j]}{\Delta x_{j+1/2}} - \theta\frac{[(UQ)^n_j - (UQ)^n_{j-1}]}{\Delta x_{j-1/2}}$$

$$
\begin{aligned}
&+ \quad g\theta(1-\theta)[A^n_{j+1/2}\frac{(\eta^n_{j+1}-\eta^n_j)}{\Delta x_{j+1/2}} - A^n_{j-1/2}\frac{(\eta^n_j-\eta^n_{j-1})}{\Delta x_{j-1/2}}] \\
&- \quad g\theta^2[A^n_{j+1/2}\frac{(h_{j+1}-h_j)}{\Delta x_{j+1/2}} - A^n_{j-1/2}\frac{(h_j-h_{j-1})}{\Delta x_{j-1/2}}] \\
a_2 &= Q^n_{j-1/2} - Q^n_{j+1/2} \\
a_3 &= V_j(\eta^n_j)
\end{aligned} \tag{6.3.14}
$$

for $j = 2, ...N-1$.

Observe that two of the coefficients have a clear physical meaning.

Specifically, a_3 is the water volume in the cell j and a_2 is the difference between the water discharge going into and out of the cell j at time t_n.

Moreover, the coefficient a_1 contains the discretization of the advective terms and the variation of the variable η and of the bottom $-\mathbf{h}$ in the cells $j-1$, j and $j+1$.

For $j = 1$ and $j = N$, the coefficients of the constraint (6.3.13) depend on the boundary conditions.

If $Q_{1/2}$ is chosen as boundary condition for the first cell, the coefficients a_1, a_2 and a_3 assume the following form

$$
\begin{aligned}
a_1 &= \theta\frac{[(UQ)^n_2 - (UQ)^n_1]}{\Delta x_{3/2}} + g\theta(1-\theta)A^n_{3/2}\frac{(\eta^n_2-\eta^n_1)}{\Delta x_{3/2}} - g\theta^2 A^n_{3/2}\frac{(h_2-h_1)}{\Delta x_{3/2}} \\
a_2 &= Q^{n+\theta}_{1/2} - Q^n_{3/2} \\
a_3 &= V_1(\eta^n_1)
\end{aligned} \tag{6.3.15}
$$

On the other hand, if η_0 is imposed, one has

$$
\begin{aligned}
a_1 &= \theta\frac{[(UQ)^n_2 - (UQ)^n_1]}{\Delta x_{3/2}} - \theta\frac{[(UQ)^n_1 - (UQ)^n_0]}{\Delta x_{1/2}} \\
&+ \quad g\theta(1-\theta)[A^n_{3/2}\frac{(\eta^n_2-\eta^n_1)}{\Delta x_{3/2}} - A^n_{1/2}\frac{(\eta^n_1-\eta^n_0)}{\Delta x_{1/2}}] \\
&- \quad g\theta^2 A^n_{3/2}\frac{(h_2-h_1)}{\Delta x_{3/2}} + g\theta^2 A^n_{1/2}\frac{h_1}{\Delta x_{1/2}} + g\theta^2 A^n_{1/2}\frac{\eta^{n+1}_0}{\Delta x_{1/2}} \\
a_2 &= Q^n_{1/2} - Q^n_{3/2} \\
a_3 &= V_1(\eta^n_1)
\end{aligned} \tag{6.3.16}
$$

With the same procedure, one can determine the precise form of the constraint (6.3.13) for $j = N$, depending on the boundary condition chosen for the last cell.

Depending on the sign of the coefficients a_1, a_2 and a_3 we can determine the solution of inequality (6.3.13), that is the values of the time step Δt such that the water volume at the new time t_{n+1} is non-negative.

The results of this study are summarized in Table 6.1, with the rule that

$$sign(a) = \begin{cases} +1 & \text{if } a > 0 \\ 0 & \text{if } a = 0 \\ -1 & \text{if } a < 0 \end{cases} \tag{6.3.17}$$

$sign(a_1)$	$sign(a_2)$	$sign(a_3)$	Results
0	0	0	$\forall \Delta t$
0	+1	0	$\forall \Delta t$
0	-1	0	$\Delta t = 0$
0	0	+1	$\forall \Delta t$
0	+1	+1	$\forall \Delta t$
0	-1	+1	$\Delta t \in [0, a_3/\left\|a_2\right\|]$
+1	0	0	$\forall \Delta t$
+1	+1	0	$\forall \Delta t$
+1	-1	0	$\Delta t = 0$
+1	0	+1	$\forall \Delta t$
+1	+1	+1	$\forall \Delta t$
+1	-1	+1	$\Delta t \in [0, \frac{\left\|a_2\right\| - \sqrt{a_2^2 - 4a_1a_3}}{2a_1}]$
-1	0	0	$\Delta t = 0$
-1	+1	0	$\Delta t \in [0, a_2/\left\|a_1\right\|]$
-1	-1	0	$\Delta t = 0$
-1	0	+1	$\Delta t \in [0, \sqrt{a_3/\left\|a_1\right\|}]$
-1	+1	+1	$\Delta t \in [0, \frac{a_2 + \sqrt{a_2^2 + 4\left\|a_1\right\|a_3}}{2\left\|a_1\right\|}]$
-1	-1	+1	$\Delta t \in [0, \frac{-\left\|a_2\right\| + \sqrt{a_2^2 + 4\left\|a_1\right\|a_3}}{2\left\|a_1\right\|}]$

Table 6.1: Range for Δt

One can note that, for every possible combination of the signs of a_1, a_2 and a_3 and for every cell I_j, there exists a local range $[0, \Delta t_j]$ for Δt, such that the corresponding water volume in that cell and at the new time t_{n+1} is non-negative.

Moreover, in the case at time t_n the cell j is empty ($a_3 = 0$) and the water discharge going out of it is bigger than that going into it ($a_2 < 0$), it is physical that the only Δt allowed is $\Delta t = 0$, that means that the computation cannot go further.

Finally, in order to ensure at the same time the non-negativity of the water volume in each cell $j = 1, ...N$, the time step Δt has to be such that

$$0 \leq \Delta t \leq min\Delta t_i \tag{6.3.18}$$

$\Delta t = min\Delta t_i$ will be called Δt_{min}.

Observe that Δt_{min} is positive both in the case the channel is completely wet and in the case for every dry cell I_j at time t_n ($a_3 = 0$) the water discharge going out of it is less than that going into it ($a_2 > 0$).

On the other hand, Δt_{min} is zero only in the particular case at least one of the Δt_j is zero, that corresponds to the draining ($a_2 < 0$) of an empty cell I_j ($a_3 = 0$).

6.4 A test on the non-negativity of the water volume

The aim of this section is to prove the advantages of satisfying the constraint (6.3.13) in the solution algorithm (1.4.1)-(1.4.3) in order ensure the non-negativity of the water volume.

The proposed example is a hydraulic jump test problem in a $10m$ long rectangular channel. In the middle of the channel, there is a sill with a crest of $1m$ height and with vertical walls, that is the slopes of the sill are abrupt within one grid cell.

There are two open boundaries, the inflow and the outflow, where a discharge of $1m^3/s$ and a water depth of $1m$, respectively, are imposed.

The discretization parameters are $\gamma = 0$, $g = 9.81m/s^2$, $\Delta x = 0.08m$ and $\theta = 1$. The time step $\Delta t = 10^{-1}s$ is given as a data of the problem and the duration of the test is $T = 1s$.

Checking the constraints (6.2.1) and (6.3.13), one can observe that whenever the former fails, it fails also the latter, but not vice versa.

Moreover, if there are M cells, $M < N$, such that Δt does not satisfy the constraint (6.2.1) or (6.3.13), one cannot conclude that the resulting water volume is negative, because the conditions (6.2.1) and (6.3.13) are sufficient, but not necessary to the non-negativity of the water volume.

Furthermore, if the water volume is negative in the cell j, Δt does not satisfy the constraints (6.2.1) and (6.3.13) corresponding to the j-th cell.

Comparing the numerical results obtained using a constant $\Delta t = 10^{-1}s$ along the computation (method 1) and computing Δt with the a priori check (6.3.13) (method 3), one can appreciate the behaviour of the second choice.

In fact, the first method leads to a numerical solution that presents a negative water volume in several cells. This let understand that the time step chosen is not appropriate for this test and should be smaller than $10^{-1}s$.

This problem could be faced using the a posteriori constraint (6.2.1), that at each time t_n checks if the Δt proposed is valid or not.

Actually, when Δt turns out to be too large, the control (6.2.1) is not able to give indications regarding the correct time step to be used.

Therefore, the only information available is that the new time step has to be less than the old one, but it may happen that the new Δt, again, does not satisfy the constraint (6.2.1) or that it is far from the optimal value it should have.

On the other hand, the a priori constraint (6.3.13) allows to find the maximum time step Δt_{min} that guarantees the non-negativity of the water volume at time t_{n+1} and to decide whether or no the Δt proposed for the test problem is acceptable.

Moreover, in case $\Delta t < \Delta t_{min}$ and Δt_{min} has a reasonable value (less than $+\infty$), Δt can be replaced by the Δt_{min} in order to optimize the performance of the algorithm.

Figure 6.1 shows the comparison between the numerical solutions obtained at $T = 3s$ satisfying (Solution 3) or no (Solution 1) the constraint (6.3.13) on Δt.

Moreover, the solution algorithm without any check on the time step (method 1) causes an overflow of the numerical solution in case the computation goes further in time than $T = 3s$.

Figure 6.2 shows the water surface elevation with respect to the time in one of the nodes presenting a negative water depth and before the overflow appears.

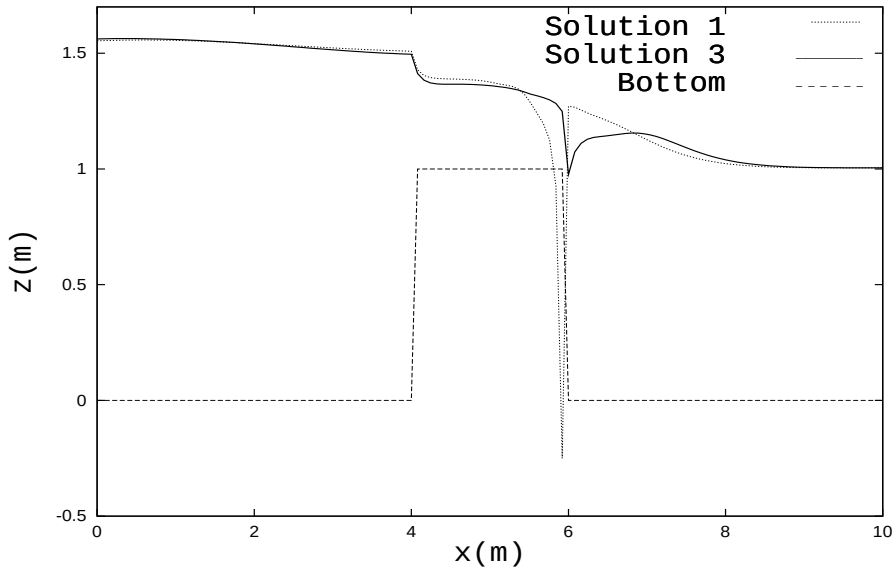

Figure 6.1: Numerical η obtained satisfying or no the explicit constraint on Δt

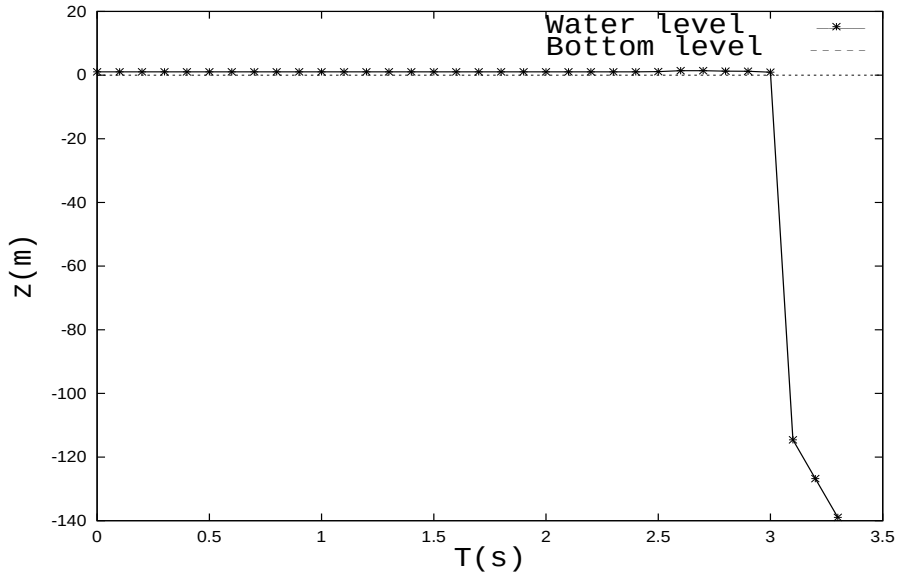

Figure 6.2: The water surface elevation at $x = 5.84m$ with respect to the time

7

Two Solution Algorithms

The aim of this chapter is to provide two solution algorithms [24] to solve system (5.1.1) and to prove their convergence in case of existence and uniqueness of the solution. A comparison of these two techniques is presented from the point of view of the computational efficiency.

7.1 Generalized Newton method (GNM)

The generalized Newton method (see, e.g., [9, 10]) is an iterative method applicable both to linear and to non-linear systems.

It is not direct, it needs an initial guess and the number of steps it takes varies with the accuracy one requests for the answer.

In particular, given a starting vector η^0, the k-th iteration of the generalized Newton method for system (5.1.3) is defined by

$$\eta^{k+1} = \eta^k - \omega[\mathbf{M} + \mathbf{V}'(\eta^k)]^{-1}[\mathbf{M}\eta^k + \mathbf{V}(\eta^k) - \mathbf{f}] \tag{7.1.1}$$

where

$$\mathbf{V}'(\eta) = diag(\, V_1{}'(\eta_1), \ldots V_N{}'(\eta_N)). \tag{7.1.2}$$

It is important to point out that the applicability of the Newton method is ensured if $0 < \omega < 2$ and $V_i'(\eta_i)$ is defined and continuous for each i, that is a very strong condition in practical applications. Moreover, the convergence of (7.1.1) can only be assured if η^0 is sufficiently close to the solution of (5.1.1).

For these reasons, it is possible to modify the generalized Newton method in order to obtain a method which works also when $\mathbf{V}(\eta)$ is only Lipschitz continuous and not differentiable.

Under this hypothesis, let modify iteration (7.1.1) substituting $\mathbf{V}'$ with $\mathbf{L}$ to obtain

$$\eta^{k+1} = \eta^k - \omega[\mathbf{M} + \mathbf{L}]^{-1}[\mathbf{M}\eta^k + \mathbf{V}(\eta^k) - \mathbf{f}] \tag{7.1.3}$$

7.1.1 Convergence of the modified GNM

The aim of this section is to prove the convergence of the modified version of the generalized Newton method in solving system (5.1.3), assuming that existence and uniqueness of its solution can be proved.

Let consider the first case analyzed in Section 5.3, that is function $\mathbf{V}$ is isotone and at least a η-type boundary condition is imposed on system (5.1.3).

Therefore, the result of convergence follows directly from the following theorem [10].

Theorem 7.1.1 *Let $\boldsymbol{M}$ be a symmetric, tridiagonal and positive definite M-matrix. Let $\boldsymbol{V}(\eta)$ be a vector function whose components V_i depend only on the variable η_i and are isotone and Lipschitz continuous. Let $0 < \omega < 2$. Then, the vector function*

$$\boldsymbol{G}(\eta) = \eta - \omega[\boldsymbol{M} + \boldsymbol{L}]^{-1}[\boldsymbol{M}\eta^k + \boldsymbol{V}(\eta^k) - \boldsymbol{f}] \tag{7.1.4}$$

is a contraction, i.e. for every two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, one has

$$\| \boldsymbol{G}(\boldsymbol{x}) - \boldsymbol{G}(\boldsymbol{y}) \| \leq C \| \boldsymbol{x} - \boldsymbol{y} \| \tag{7.1.5}$$

where $0 \leq C < 1$ is a constant independent from $\boldsymbol{x}$ and $\boldsymbol{y}$.

In case two Q-type boundary conditions are imposed and function $\mathbf{V}$ is strictly isotone (see Section 5.4), the convergence of the modified version of the generalized Newton method can be proved by the following theorem.

Theorem 7.1.2 *Let $\boldsymbol{M}$ be a symmetric, tridiagonal and positive semi-definite matrix. Let $\boldsymbol{V}(\eta)$ be a vector function whose components V_i depend only on the variable η_i and are strictly isotone and Lipschitz continuous. Let $0 < \omega < 2$. Then, the vector function*

$$\boldsymbol{G}(\eta) = \eta - \omega[\boldsymbol{M} + \boldsymbol{L}]^{-1}[\boldsymbol{M}\eta^k + \boldsymbol{V}(\eta^k) - \boldsymbol{f}] \tag{7.1.6}$$

is a contraction, i.e. for every two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, one has

$$\| \boldsymbol{G}(\boldsymbol{x}) - \boldsymbol{G}(\boldsymbol{y}) \| \leq C \| \boldsymbol{x} - \boldsymbol{y} \| \tag{7.1.7}$$

where $0 \leq C < 1$ is a constant independent from $\boldsymbol{x}$ and $\boldsymbol{y}$.

Proof. The hypotheses of $\mathbf{V}$ diagonal and Lipschitz continuous assure that, for each $i = 1, ...N$ and for any given x_i and y_i in $\Re$, there exist a constant S_i dependent on x_i and y_i, such that $0 \leq S_i(x_i, y_i) \leq L_i$ and

$$V_i(x_i) - V_i(y_i) = S_i(x_i - y_i). \tag{7.1.8}$$

Moreover, by the strict isotonicity of $\mathbf{V}$, one can observe that $S_i(x_i, y_i) > 0$ for every x_i, $y_i \in \Re$, $x_i \neq y_i$.

Thus, if $\mathbf{S}$ denotes the diagonal matrix whose elements are S_i, the difference $\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{y})$ can be written as:

$$\begin{aligned} \mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{y}) &= (\mathbf{x} - \mathbf{y}) - (\mathbf{M} + \mathbf{L})^{-1}[\mathbf{M}(\mathbf{x} - \mathbf{y}) + \mathbf{V}(\mathbf{x}) - \mathbf{V}(\mathbf{y})] \\ &= [\mathbf{I} - (\mathbf{M} + \mathbf{L})^{-1}(\mathbf{M} + \mathbf{S})](\mathbf{x} - \mathbf{y}) \end{aligned} \tag{7.1.9}$$

Now, in order to prove that $\mathbf{G}(\mathbf{x})$ is a contraction, it is sufficient to prove that the spectral radius of the matrix $\mathbf{N} = [\mathbf{I} - (\mathbf{M} + \mathbf{L})^{-1}(\mathbf{M} + \mathbf{S})]$ is smaller than a constant which is less than one. For this purpose, since

$$\begin{aligned} det(\mathbf{N} - \lambda\mathbf{I}) &= det[(1 - \lambda)\mathbf{I} - (\mathbf{M} + \mathbf{L})^{-1}(\mathbf{M} + \mathbf{S})] \\ &= det[(\mathbf{M} + \mathbf{L})^{-1}]det[(1 - \lambda)(\mathbf{M} + \mathbf{L}) - (\mathbf{M} + \mathbf{S})] \end{aligned} \tag{7.1.10}$$

and the matrix $(\mathbf{M} + \mathbf{L})^{-1}$ is non-singular for Theorem 5.3.3, the eigenvalues of $\mathbf{N}$ are the solutions of the following equation

$$det[(1 - \lambda)(\mathbf{M} + \mathbf{L}) - (\mathbf{M} + \mathbf{S})] = det[-\lambda\mathbf{M} - \mathbf{S} + (1 - \lambda)\mathbf{L}] = 0.$$

Note first that since $\mathbf{N}$ is a symmetric matrix, it only has real eigenvalues [32]. Additionally, for $\lambda < 0$, the matrix $(1 - \lambda)(\mathbf{M} + \mathbf{L}) - (\mathbf{M} + \mathbf{S})$ is non-singular, because it is tridiagonal, diagonally dominant, with positive elements on the main diagonal and negative elsewhere. Thus, the eigenvalues of $\mathbf{N}$ are all non-negative.

Furthermore, by denoting with σ the minimum eigenvalue of $\mathbf{M}$, one has $\sigma \geq 0$. Moreover, by definition of eigenvalues, $det(\mathbf{M} - \sigma\mathbf{I}) = det(\lambda\mathbf{M} - \lambda\sigma\mathbf{I}) = 0$ and, if $\lambda > 0$, for any matrix $\epsilon = diag(\epsilon_1, ..., \epsilon_N)$ with $\epsilon_i > 0$ for $i = 1, ...N$, one has $det(\lambda\mathbf{M} - \lambda\sigma\mathbf{I} + \epsilon) \neq 0$.

Thus, if the matrix $-\lambda\mathbf{M} - \mathbf{S} + (1 - \lambda)\mathbf{L}$ is of the same sort of the matrix $-\lambda\mathbf{M} + \lambda\sigma\mathbf{I} - \epsilon$, that is if

$$S_i - (1 - \lambda)L_i + \lambda\sigma = \epsilon_i > 0, \tag{7.1.11}$$

then it is non-singular.

Then, the eigenvalue λ of $\mathbf{N}$ must satisfy the following inequality

$$S_i - (1-\lambda)L_i + \lambda\sigma \leq 0, \tag{7.1.12}$$

that is

$$0 \leq \lambda \leq max_{i=1,\dots N}(\frac{L_i - S_i}{L_i + \sigma}). \tag{7.1.13}$$

Because $\mathbf{M}$ is singular, that is $\sigma = 0$, then

$$0 \leq \lambda \leq max_{i=1,\dots N}(\frac{L_i - S_i}{L_i + \sigma}) = max_{i=1,\dots N}(1 - \frac{S_i}{L_i}) = 1 - min_{i=1,\dots N}(\frac{S_i}{L_i}) < 1$$

that proves the theorem. ■

From the above theorems 7.1.2 and 7.1.1 we have immediately the following corollary that proves the final result.

Corollary 7.1.3 *Under the same hypotheses of Theorem 7.1.1 or Theorem 7.1.2, the iterative scheme (7.1.3) converges to the solution of the mildly non-linear system (5.1.3).*

7.2 Conjugate gradient method (CGM)

The conjugate gradient method is a solution procedure widely used in the literature in finding an unconstrained minimum of a function Φ in N variables.

The general form of the conjugate gradient method is the following

$$\mathbf{d}_k := \begin{cases} -\mathbf{g}_k & \text{for} \quad k = 1 \\ -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1} & \text{for} \quad k > 1 \end{cases} \tag{7.2.1}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \tag{7.2.2}$$

where $\mathbf{g}_k$ denotes the gradient $\nabla\Phi(\mathbf{x}_k)$, α_k is a step-length obtained by means of a one-dimensional search and β_k is a scalar chosen so that d_k becomes the k-th conjugate direction when the function Φ is quadratic and the search of α_k is exact.

Some well-known formulas for β_k are called the Fletcher-Reeves (FR), Polak-Ribiere (PR), Hestenes-Stiefel (HS) and Conjugate Descent Method formulas (see, e.g., [17, 37]) and are given by

$$\beta_k^{FR} = \|\mathbf{g}_k\|^2 / \left\|\mathbf{g}_{k-1}\right\|^2, \tag{7.2.3}$$

$$\beta_k^{PR} = \mathbf{g}_k^T(\mathbf{g}_k - \mathbf{g}_{k-1}) / \left\|\mathbf{g}_{k-1}\right\|^2, \tag{7.2.4}$$

$$\beta_k^{HS} = \mathbf{g}_k^T(\mathbf{g}_k - \mathbf{g}_{k-1}) / [\mathbf{d}_{k-1}^T(\mathbf{g}_k - \mathbf{g}_{k-1})], \tag{7.2.5}$$

$$\beta_k^{CD} = -\|\mathbf{g}_k\|^2 / (\mathbf{d}_{k-1}^T \mathbf{g}_{k-1}), \tag{7.2.6}$$

where $\|\cdot\|$ is the Euclidean norm.

In order to use this method in solving system (5.1.3), the latter has to be reformulated as an unconstrained minimization problem of the form

$$min_{\mathbf{X}\in\Re^N} \Phi(\mathbf{x}) \tag{7.2.7}$$

and the equivalence between the minimization problem (7.2.7) and the non-linear system (5.1.3) has to be shown.

To do this, let first of all define the function Φ corresponding to system (5.1.3) as in Equation (5.4.5) and let specify its mildly non-linear part as

$$\mathbf{P}(\eta) = \sum_{j=1}^{N} P_j(\eta_j) = \sum_{j=1}^{N} (\int_0^{\eta_j} V_j(\xi) d\xi) \tag{7.2.8}$$

such that $\bigtriangledown \mathbf{P}(\eta) = V(\eta)$.

The following result proves the equivalence between the minimization problem (7.2.7) and the non-linear system (5.1.3) by showing the identity between the sets of the solution of these two problems.

Theorem 7.2.1 *Each solution η of the minimization problem (7.2.7) satisfies system (5.1.3) and vice versa*

Proof. Let consider η a solution of the minimization problem (7.2.7). Therefore

$$\bigtriangledown \Phi(\eta) = \mathbf{V}(\eta) + \mathbf{M}\eta - \mathbf{f} = 0.$$

that proves that η is also a solution of system (5.1.3).

Let be η is a solution of system (5.1.3). The aim of the second part of this theorem is to prove that inequality

$$\Phi(\eta) \leq \Phi(\eta + e) \tag{7.2.9}$$

holds, where $\mathbf{e} \in \Re^N$ is an error term from the exact solution η.

One can rewrite the second member of the previous inequality as

$$\Phi(\eta + e) = P(\eta + e) + \frac{1}{2}(\eta + e)^T \mathbf{M}(\eta + e) - (\eta + e)^T \mathbf{f}$$

that is, being $\mathbf{M}$ a symmetric matrix,

$$\Phi(\eta + e) = P(\eta + e) + [\frac{1}{2}\eta^T \mathbf{M}\eta - \eta^T \mathbf{f}] + e^T \mathbf{M}\eta + [\frac{1}{2}e^T \mathbf{M}e - e^T \mathbf{f}]$$

or equivalently

$$\Phi(\eta + e) = P(\eta + e) + [\Phi(\eta) - P(\eta)] + e^T \mathbf{M}\eta + [\frac{1}{2}e^T \mathbf{M}e - e^T \mathbf{f}].$$

Because η is the solution of system (5.1.3), it follows

$$\Phi(\eta + e) - \Phi(\eta) = \mathbf{P}(\eta + e) - \mathbf{P}(\eta) - e^T \mathbf{V}(\eta) + \frac{1}{2}e^T \mathbf{M}e.$$

One can observe that each

$$P_j : \mathbf{R} \rightarrow \mathbf{R}$$

is a convex function for every $j = 1, ...N$, because it is a differentiable function on $\mathbf{R}$ and its first derivative is monotone non decreasing. Therefore, function $\mathbf{P}$ is a convex function too and the following inequality holds

$$\mathbf{P}(\eta + e) - \mathbf{P}(\eta) \geq \bigtriangledown \mathbf{P}(\eta)^T (\eta + e - \eta) = \mathbf{V}(\eta)^T e \tag{7.2.10}$$

and therefore

$$\Phi(\eta + e) - \Phi(\eta) \geq \frac{1}{2}e^T \mathbf{M}e \tag{7.2.11}$$

Observing that $\mathbf{M}$ is positive semi-definite, one can conclude that $e^T \mathbf{M}e$ is non-negative for every error-term e. This finishes our proof. ■

7.2.1 Convergence of the CGM

The aim of this section is to prove the convergence of the conjugate gradient method in solving system (5.1.3), assuming that existence and uniqueness of its solution can be proved.

The global convergence of the conjugate gradient method with β_k defined by Equations (7.2.3), (7.2.4), (7.2.5) and (7.2.6) or others has been investigated by many authors (see, e.g., [22, 26, 28, 38, 61]) and the choice of the step-length α_k has always been addressed as a fundamental for the global convergence.

Different techniques have been proposed for the computation of α_k (see, e.g., [22, 38, 58]), among which one can find the classical exact line search defined by

$$\alpha_k := argmin_{\alpha \geq 0} \Phi(\mathbf{x}_k + \alpha \mathbf{d}_k).$$

Now consider the following assumptions

Assumption 7.2.2 *The function Φ is LC^1 in a neighbourhood $\boldsymbol{N}$ of the level set $\boldsymbol{D} := \{\boldsymbol{x} \in \Re^n \,|\Phi(\boldsymbol{x}) \leq \Phi(\boldsymbol{x}_1)\}$ and $\boldsymbol{D}$ is bounded. Here, by LC^1 we mean that the gradient $\nabla\Phi(\boldsymbol{x})$ is Lipschitz continuous with modulus μ, i.e., there exists $\mu > 0$ such that $\|\Phi(\boldsymbol{x}_{k+1}) - \Phi(\boldsymbol{x}_k)\| \leq \mu \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|$ for any $\boldsymbol{x}_{k+1}, \boldsymbol{x}_k \in \boldsymbol{N}$.*

Assumption 7.2.3 *The function Φ is LC^1 and strongly convex on $\boldsymbol{N}$.*

and note that Assumption 7.2.2 implies Assumption 7.2.3, since a strongly convex function has bounded level sets [32].

Let now consider the first case analyzed in Section 5.3, that is function $\mathbf{V}$ is isotone and at least a η-type boundary condition is imposed on system (5.1.3).

Under these assumptions Φ defined by (5.4.5) is clearly strongly convex, because

$$[\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{y})]^T(\mathbf{x} - \mathbf{y}) \geq (\mathbf{x} - \mathbf{y})^T\mathbf{M}(\mathbf{x} - \mathbf{y}) \geq c\|\mathbf{x} - \mathbf{y}\|^2 \qquad (7.2.12)$$

where $c > 0$ is the minimum of the eigenvalues of $\mathbf{M}$.

Moreover, also in the case two Q-type boundary conditions are imposed on system (5.1.3) and function $\mathbf{V}$ is strictly isotone (see Section 5.4), function Φ defined by (5.4.5) is strongly convex, as already proved in the proof of Theorem 5.4.2 .

Let remember the result presented in [49].

Theorem 7.2.4 *Let $\{\boldsymbol{A}_k\}_k$ be a sequence of positive definite matrices and assume that there exist $\nu_{min} > 0$ and $\nu_{max} > 0$ such that $\forall d \in \Re^N$*

$$\nu_{min}\boldsymbol{d}^T\boldsymbol{d} \leq \boldsymbol{d}^T\boldsymbol{A}_k\boldsymbol{d} \leq \nu_{max}\boldsymbol{d}^T\boldsymbol{d}. \tag{7.2.13}$$

Define the step-length formula as follows

$$\alpha_k = \frac{-\delta\boldsymbol{g}_k^T\boldsymbol{d}_k}{\boldsymbol{d}_k^T\boldsymbol{A}_k\boldsymbol{d}_k} \tag{7.2.14}$$

where $\delta \in (0, \frac{\nu_{min}}{\mu})$.
A unified formula for α_k like (7.2.14) can ensure global convergence for many cases, which include: 1. The FR method and the HS method applied to a strongly convex LC^1 objective function (Assumption 7.2.3); 2. The PR method and the CD method applied to a general LC^1 objective function (Assumption 7.2.2).

Observe that, in order to apply Theorem (7.2.4) to our system (5.1.3), one has to define the sequence $\{\mathbf{A}_k\}_k$ of positive definite matrices involved in the computation (7.2.14) of the step-length α_k.

Our choice is $\mathbf{A}_k = \mathbf{A}\ \forall k$, where $\mathbf{A} = \mathbf{M} + \mathbf{L}$.

From the above results and considerations we have the following Corollary.

Corollary 7.2.5 *Consider the minimization problem (7.2.7) with objective function Φ defined by (5.4.5) and assume existence and uniqueness of its solution. Thus, the solution algorithm (7.2.1)-(7.2.2) with the step-length formula defined by (7.2.14) with $\boldsymbol{A}_k = \boldsymbol{M} + \boldsymbol{L}\ \forall k$ converges globally.*

7.3 Computational efficiency

This section proposes a comparison of the two algorithms previously presented, in terms of their computational efficiency in solving system (5.1.1).

The operations that mainly contribute to the computational cost of the modified version of the Generalized Newton Method (GNM) are the matrix-vector product $\mathbf{M}\eta$ and the evaluation of the non-linear function $\mathbf{V}(\eta)$. Therefore, the complexity of the algorithm is of order $O(N) + \sum_{i=1}^{N} O(V_i)$, that becomes $O(N)$ in the particular case of a linear function $\mathbf{V}$.

Regarding the Conjugate Gradient Method (CGM), the order of complexity depends on the following operations: the computation of the step-length α_k and the computation of the search direction $\mathbf{d}_k$.

The computation of the step-length α_k in case $\mathbf{V}$ is non-linear and Φ is non-quadratic can be done by formula (7.2.14) with $\mathbf{A}_k = \mathbf{M}+\mathbf{L}$ for all k. The complexity of this formula depends on two factors: the evaluation of $\nabla\Phi(\mathbf{x}_k) = \mathbf{g}_k = \mathbf{V}(\mathbf{x}_k) - \mathbf{M}\mathbf{x}_k - \mathbf{f}$, that costs $O(N) + \sum_{i=1}^{N} O(V_i)$ and the matrix-vector product $\mathbf{A}d_k$, that has complexity of order the number of non-zero entries of matrix $\mathbf{A}$, that is $O(N)$.

The computation of the search direction $\mathbf{d}_k$ is given by (7.2.1), where β_k can be computed following different formulas. For example, the FR formula has complexity $O(N) + \sum_{i=1}^{N} O(V_i)$, because depends on the evaluation of $\nabla\Phi(\mathbf{x}_k)$.

Therefore, the complexity of the Conjugate Gradient Method is given by $O(N) + \sum_{i=1}^{N} O(V_i)$, that becomes $O(N)$ in the particular case of a linear function $\mathbf{V}$.

From the point of view of the convergence rate, it is known that in case (5.1.1) is linear, the Newton Method converges with order 2, while the Conjugate Gradient Method converges in at least N steps.

Table 7.1 illustrates the performance of the two methods solving the system arising from the Hydraulic Jump test in a rectangular channel presented in [3]. In this test $\Delta t = 10^{-2}$ and $\theta = 1$, while the duration of the simulation is $T_f = 2s$. In this period of time the solution does not reach its steady state.

		Conjugate Gradient Method		Generalized Newton Method	
Δx	N	CPU time(sec)	No.It	CPU time(sec)	No.It
0,5	200	0,04	2	0,07	14
0,1	1000	0,16	2	0,5	17
0,05	2000	0,44	3	1,3	26
0,02	5000	3,13	10	9,9	81
0,01	10000	21,7	38	59,9	245

Table 7.1: Performance of the CGM and the GNM for the Hydraulic Jump Test

Table 7.2 illustrates the performance of the two methods solving the system arising from the Dam Break Test test over a wet bed in a semicircular channel. In this test $\Delta t = 10^{-3}$ and $\theta = 0.5$, while $T_f = 0.3s$.

Fixed the time step, for each grid size the measures of performance are given by

		Conjugate Gradient Method		Generalized Newton Method	
Δx	N	CPU time(sec)	No.It	CPU time(sec)	No.It
0,02	50	0,06	13	0,1	24
0,01	100	0,12	13	0,15	25
0,005	200	0,25	13	0,33	28
0,002	500	0,57	13	0,9	29
0,00167	600	0,7	13	1,1	29

Table 7.2: Performance of the CGM and the GNM for the Dam Break Test (Semicircular channel)

the mean number of iterations (rounded to the nearest integer) for each time step and the total CPU taken by the algorithms.

The tolerance used to test the convergence is $tol = 10^{-7}$.

Analysing Tables 7.1, one can observe that the Conjugate Gradient Method is faster than the Generalized Newton Method solving this linear problem.

In the Hydraulic Jump test in fact, the reduction of the size of the spatial grid causes an increase of the number of iterations and of the CPU time that is more conspicuous for the Generalized Newton Method than for the Conjugate Gradient Method. This behaviour can be brought back to the rate of convergence of the two algorithms.

On the other hand, one can note that the gap between the two algorithms becomes thinner for a non-linear problem.

In fact, considering the Dam Break Test in a Semicircular channel, the results listed in Table 7.2 show that the Conjugate Gradient Method is still preferable to the Generalized Newton Method both for the CPU time and for the number of iterations, but the differences between the data of the two methods are smaller than those obtained in Table 7.1 for a linear problem.

Conclusions and recommendations

The aim of this final chapter is to formulate general conclusions on the numerical scheme presented in this thesis emphasising its specific properties and its potential for dealing with hydraulic engineering problems. The chapter closes with recommendations for future work.

Conclusions

In the present thesis, a semi-implicit numerical model for the one-dimensional simulation of non-stationary free surface in open channels with arbitrary cross-section has been derived, discussed and applied.

The semi-implicit discretization (see, e.g., [6]) leads to a relatively simple (explicit part) and computationally efficient (fully implicit part) scheme whose stability can be shown to be independent from the wave celerity $\sqrt{gH}$.

The conservation properties allow dealing properly with problems presenting discontinuities in the solution, resulting for example from sharp bottom gradients and hydraulic jumps. The conservation of mass is particularly important when the channel has a non rectangular cross-section. The possibility to switch between momentum and energy head conservation depending on local flow conditions leads the numerical solutions to present the same characteristics as the physical ones.

The accuracy of the proposed method is controlled by the use of appropriate flux limiting functions in the discretization of the advective terms [35], especially in the case of large gradients of the physical quantities involved in the problem (i.e. the water level).

The fully implicit version of the method has been easily extended to solve the closed channel flow equations: assuming the incompressibility of water, implicit schemes are able to manage instantaneous transmission of pressure and velocity changes arising in the pressurized part of the channel. Therefore they can simulate the transition from free surface to pressurized flow in channels with arbitrarily shaped closed sections without any approximation of the section geometry and thus

preserving precise volume conservation.

The method allows the simulation of hydraulic engineering situations such as subcritical flows, mixed flows (subcritical and supercritical) as well as transitions from supercritical to subcritical flows such as hydraulic jumps. Wetting and drying phenomena are correctly treated without the use of specific procedures.

Careful physical and mathematical considerations about the stability of the method and the solvability of the related mildly non-linear system with respect to the implemented boundary conditions have been also provided together with suitable solution procedures. An explicit and sufficient condition on the time step for the non-negativity of the water volume has been formulated and it is valid under not more restrictive assumptions than those necessary for a correct description of the physical problem.

Recommendations for further research

Future work on the topic could address the extension of the model to sewer networks and to 2 and 3 dimensions on structured and unstructured grids.

Also the analytical results of existence and uniqueness of the numerical solutions should be extended to those cases.

Conservation properties deserve intense studying at the junctions between more channels of the same network and in case the grid is unstructured.

It could be also possible to extend the method in order to include the air-phenomena occurring in closed pipes as described in the tests presented in Chapter 4: this could be done, for example, by designing a isopycnal type method or a multiphase gas-liquid flow method.

Further research could be devoted to a more detailed study of the explicit constraint for the non-negativity of the water volume.

Much effort must still be put into research about solution algorithms and in particular about the Conjugate Gradient Method for mildly non-linear systems: interesting results could consider computational efficiency estimation and convergence properties.

Finally, more experimental tests are also recommended.

Bibliography

[1] MB. Abbott. *Computational Hydraulics*. Worcester: Ashgate, 1992.

[2] E. Aldrighetti and G. Stelling. A robust scheme for free surface and pressurized flows in channels with arbitrary cross-sections. *Technical Report , Matematica, University of Trento*, UTM 694, 2006.

[3] E. Aldrighetti and P. Zanolli. A high resolution scheme for flows in open channels with arbitrary cross-section. *International Journal for Numerical Methods in Fluids*, 47:817–824, 2005.

[4] DL. Book and JP. Boris. Flux corrected transport. i. shasta, afluid transport algorithm that works. *Journal of Computational Physics*, 11:38–69, 1973.

[5] DL. Book and JP. Boris. Flux corrected transport iii: Minimal error fct algorithms. *Journal of Computational Physics*, 20:397–431, 1976.

[6] V. Casulli. Semi-implicit finite difference methods for the two-dimensional shallow water equations. *Journal of Computational Physics*, 86:56–74, 1990.

[7] V. Casulli and E. Cattani. Stability, accuracy and efficiency of a semi-implicit method for three dimensional *Computers & Mathematics with Application*, 15:629–648, 1994.

[8] V. Casulli and RT. Cheng. Stability analysis of eulerian-lagrangian method for the one-dimensional shallow-water equations. *Applied Mathematical Modelling*, 14:122–131, 1990.

[9] V. Casulli and D. Greenspan. *Numerical analysis for applied mathematics, science and engineering*. Addison-Wesley Publishing Company, 1988.

[10] V. Casulli and P. Zanolli. A conservative semi-implicit scheme for open channel flows. *International Journal of Applied Science & Computations*, 5:1–10, 1998.

[11] H. Chanson. *The Hydraulics of Open Channel Flow*. John Wiley & Sons Ltd, 1999.

[12] M. H. Chaudry. *Open-Channel Flow*. Prentice Hall, 1993.

[13] VT. Chow. *Open-Channel Hydraulics*. McGraw-Hill, Inc., 1959.

[14] F.H.L.R. Clemens C.L. Lubbers. Capacity reduction caused by air intake at wastewater pumping stations. (experiments of transport of gas in pressurized wastewater mains. *Technical Report, Section of Sanitary Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, The Netherlands*, 2005.

[15] B. de Saint Venant. Théorie du mouvement non-permanent des eaux avec application aux crues des rivières at à l'introdution des marées dans leur lit. *Acad. Sci. Comptes Rendus Paris*, 73:147–154, 1871.

[16] JP. Boris DL. Book and K. Hain. Flux corrected transport ii: Generalization of the method. *Journal of Computational Physics*, 18:248–283, 1975.

[17] R. Fletcher. *Practical methods of optimization*. John Wiley, New York, 1980.

[18] P. Garcia-navarro. *Dam Break Flow Simulation*. University of Zaragoza, Spain, 1999.

[19] P. Garcia-Navarro and F. Alcrudo. A high-resolution godunov-type scheme in finite volumes for the 2-d shallow water equations. *Int. J. Num. Meth. in Fluids*, 16:489–505, 1991.

[20] P. Garcia-Navarro and A. Priestley. An implicit method for water flow mo-delling in channels and pipes. *Journal of Hydraulic Research*, 32:721–742, 1994.

[21] P. Garcia-Navarro and ME. Vazquez-Cendon. On numerical treatment of the source terms in the shallow water equations. *Computers & Fluids*, 29:951–979, 2000.

[22] LC. Gilbert and J. Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on optimization*, 2:21–42, 1992.

[23] SK. Godunov. A difference scheme for numerical computation of discontinuous solution of hydro dynamic equations. *Math. Sbornik*, 47:271–306, 1959.

[24] GH. Golub and CF. Van Loan. *Matrix Computations*. Baltimore: John Hopkins, 1989.

[25] A. Harten. High resolution schemes for hyperbolic conservation laws. *J. Comput. Phys*, 49:357393, 1983.

[26] MR. Hestenes and E. Stiefel. Method of conjugate gradient for solving linear systems. *J. Res. Nat. Bur. Stand.*, 49:409–436, 1952.

[27] C. Hirsh. *Numerical Computation of Internal and External Flows. Volume 2: Computational Methods for Inviscid and Viscous Flows*. John Wiley & Sons Ltd, 1990.

[28] Y. Hu and C. Storey. Global convergence result for conjugate gradient methods. *Journal of Optimization Theory and Applications*, 71:399–405, 1991.

[29] AT. Ippen. *Estuary and Coastline Hydrodynamics*. McGraw-Hill, 1966.

[30] F.M. Holly J.A. Cunge and A. Verwey. *Practical Aspects of computational river hydraulics*. Pitman Publishing, 1980.

[31] MH. Chaudhry JA. Roberson, JJ. Cassidy. *Hydraulic Engineering*. Houghton Mifflin Company, Boston, 1988.

[32] WC. Rheimboldt JM. Ortega. *Iterative Solution of Non-Linear Equation in several variables*. Academics Press: New York, 1970.

[33] JJ. Leendertse. *Aspects of a computational method for long period water-wave propagation*. Memorandum RM-5294-PR., The RAND Corp., Santa Monica, California, 1967.

[34] RJ. LeVeque. *Numerical methods for conservation laws*. Birkhaeuser, Basel/Boston/Berlin, 1992.

[35] RJ. LeVeque. *Finite Volume Methods for hyperbolic Problems*. Cambridge University Press, 2002.

[36] A. Maffio MJ. Baines and A. Di Filippo. Unsteady 1d flows with steep waves in plant channels: The use of roe's upwind tvd difference scheme. *Advances in water resources*, 15:89–94, 1992.

[37] W. Murray PE. Gill and MH. Wright. *Practical optimization*. Academic Press, New York, 1981.

[38] MJD. Powell. Convergence properties of algorithms for nonlinear optimization. *SIAM Review*, 28:487–500, 1986.

[39] KO. Friedrichs nd H. Lewy R. Courant. Ueber die partiellen differenzengleichungen der matematisches physik. *Math.Ann.*, 100:32–74, 1928.

[40] KO. Friedrichs nd H. Lewy R. Courant. On the partial differential equations of mathematical physics. *IBM Journal*, 11:215–234, 1967.

[41] PL. Roe. *Some contributions to the modelling of discontinuous flows*. Proceedings of the SIAM/AMS Seminar, San Diego, 1997.

[42] H. Rouse. *Engineering Hydraulics*. John Wiley, New York, 1950.

[43] PG. Samuels and CP. Skeels. Stability limits for preissmann's scheme. *Journal of Hydraulic Engineering*, 116:997–1012, 1990.

[44] A. Sjoberg. *Calculation of Unsteady Flows in Regulated Rivers and Storm Sewer Systems*. Chalmers Univ. of Technology, Goteborg, Sweden, 1976.

[45] PM. Steffler and YC. Jin. Depth averaged and moment equations for moderately shallow free surface flow. *Journal of Hydraulic Research, IAHR*, 31:5–18, 1993.

[46] GS. Stelling and SPA. Duinmeijer. A staggered conservative scheme for every froude number in rapidly varied shallow water flows. *International Journal for Numerical Methods in Fluids*, 43:1329–1354, 2003.

[47] J.J. Stoker. *Water waves*. The mathematical theory with applications. Interscience publisher, 1957.

[48] T. Strelkoff. One-dimensional equations for open-channel flow. *J. Hydr. Div., ASCE*, 953:861–876, 1969.

[49] J. Sun and J. Zhang. Global convergence of conjugate gradient methods without line search. *Annals of operations research*, 103:161–173, 2001.

[50] PK. Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM Journal Numer. Anal.*, 21:995–1011, 1984.

[51] C. Thacker. Some exact solutions to the nonlinear shallow-water wave equations. *Journal of Fluid Mechanics*, 107:499–508, 1981.

[52] EF. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics.* Springer, 1997.

[53] EF. Toro. *Shock capturing Methods for free surface shallow flows.* John Wiley & Sons Ltd, 2001.

[54] T. Tucciarelli. A new algorithm for a robust solution of the fully dynamic saint venant equations. *Journal of Hydraulic Research, IAHR*, 3:239–243, 2003.

[55] ME. Vasquez-Cendon. Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry. *Journal of Computational Physics*, 148:497–526, 1999.

[56] D.C. Wiggert. Transient flow in free-surface, pressurized systems. *Journal of hydraulic division*, 98:11–27, 1972.

[57] D.C. Wiggert and M.J. Sundquist. Fixed grid characteristics for pipeline transients. *Journal of hydraulic division*, 103:1403–1415, 1978.

[58] P. Wolfe. Convergence conditions for ascent methods. *SIAM Review*, 11:226–235, 1969.

[59] HC. Yee. Construction of explicit and implicit symmetric tvd schemes and their applications. *Journal of Computational Physics*, 68:151–179, 1987.

[60] BC. Yen. Open channel flow equations revisited. *Journal of Engineering Mechanics Division, ASCEE*, 99:979–1009, 1973.

[61] G. Zoutendijk. Nonlinear programming, computational methods. *Integer and Nonlinear Programming*, pages 37–87, 1970.