THE UNIVERSITY OF READING

DEPARTMENT OF MATHEMATICS

Numerical Methods for Singular Differential Equations Arising from Steady Flows in Channels and Ducts

by A.C.Lemos

Dissertation submitted for the degree of

Doctor of Philosophy

May 2002

Abstract

We study ordinary nonlinear differential equations which arise from steady nonlinear conservation laws with source terms. Two examples of conservation laws which lead to these equations are the Saint-Venant and the Euler equations. In each case there is a reduction to a scalar equation and we use the ideas of upwinding and discretisation of source terms to devise methods for the solution. Numerical results are presented with both the Engquist-Osher and the Roe scheme with different strategies for discretising the source terms based on balance ideas.

Acknowledgements

Firstly, I would like to express my gratitude to Professor Mike Baines. His supervision, support and patience were constant throughout this work and encouraged me to go on. My thanks go also to Professor Nancy Nichols. Her supervision and advice were very helpful. As a team, their supervision complemented each other and I benefited from their knowledge and teaching.

I am grateful to my sponsors in Portugal, *Funda ção para a Ciência e a Tecnologia* (grant PRAXIS XXI/BD/15905/98 from the *Subprograma Ciência e Tecnologia do* 2^{o} *Quadro Comunitário de Apoio*, and the *Escola Superior de Tecnologia e Gestão* from the *Instituto Politécnico de Leiria*, who made this project viable.

I wish to thank the help of staff and colleagues in the Mathematics Department in Reading who always made me feel welcome. Studying in Department of Mathematics of the University of Reading afforded the opportunity to learn with very good teachers and to meet colleagues and fellow researchers.

In Reading, I met new friends and their friendship and support were very important in making me feel less lonely. I would like to thank especially Jessica, Ana Teresa, Hussain and Giovanni. We shared very happy moments that I will cherish forever. I extend my thanks to Helena, who made my stay in the University of Reading possible, and to Fernando, Cacilda, Teresa Mota and Cristine and other members of the Brazilian and Portuguese Speakers Society.

Among the friends I met in in the Mathematics Department, I will remember with gratitude Marek Wlasak and Jo Morgan.

The help of Dr. M. Hubbard and Dr. MacDonald is gratefully acknowledged.

Finally, I would like to thank my family for their love and encouragement. They have been a continual source of support and strength over many years.

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Contents

1 Introduction: overview **1**

2 Reduction of the Saint-Venant Equations and Euler Equations in the steady-state case **5**

2.1 Unsteady Conservation Laws with Source Terms 6

2.2 The Saint-Venant Equations 7

2.2.1 The Saint-Venant Equations for Channels with Variable Breadth . 8

2.2.2 Quasi-linear Representation and Characteristic Speeds 11

2.2.3 The Steady Problem 12

2.2.4 Discontinuous Solutions in the Unsteady and Steady Cases 14

2.2.5 Boundary Conditions 16

2.3 The Euler Equations Modelling One-dimensional Flow in a Nozzle 17

2.3.1 The Euler Equations 18

2.3.2 The Equations of Gas Flow in a Nozzle 22

2.3.3 Matrix Representation and Characteristic Speeds 23

2.3.4 The Steady Problem I: reduction to a singular scalar ODE 24

2.3.5 The Steady Problem II: reduction to a system of two ODEs . . . 28

2.3.6 Discontinuous Solutions 31

2.3.7 Boundary Conditions 33

2.4 Analogy Between Compressible Gas and Water Model 35

3 Vanishing Viscosity Theory Applied to the Steady Problem **37**

3.1 Vanishing Viscosity Theory 37

3.2 The Use of the Vanishing Viscosity Theory in the Steady State Case . . . 42

3.3 Possible Application of Vanishing Viscosity Theory to Steady Problems with Breadth Variation . . . 47

4 Theory on the Upwind Schemes of Engquist-Osher and Roe 49

4.1 Background on Conservative Methods . . . 50

4.2 Flux Function of the Form $\mathbf{F}(\mathbf{U})$ . . . 52

4.2.1 Theory on Conservative Methods . . . 52

4.2.2 The Godunov Method . . . 54

4.2.3 The Scalar Problem . . . 56

4.2.4 The Upwind Scheme of Roe . . . 59

4.2.5 The upwind Scheme of Engquist-Osher (scalar case) . . . 65

4.3 Flux Function of the Form $\mathbf{F}(x, \mathbf{U})$ . . . 66

4.3.1 Theory on Conservative Methods . . . 66

4.3.2 The Godunov, Roe and Engquist-Osher Schemes Extended to Include x-dependence . . . 68

4.4 Nonconservative Form of the Equations . . . 71

4.5 Nonlinear Stability and Higher Order Extensions for the Scalar Equations 73

4.6 Scalar Flux Function Depending on a Discontinuous Coefficient . . . 75

5 Source Terms 77

5.1 The Scalar Unsteady Nonhomogeneous Equation . . . 79

5.2 Conservative Schemes and Numerical Source Terms . . . 80

5.2.1 The Scalar Case . . . 83

5.3 Types of Discretisation of the Source Terms . . . 84

5.4 Roe Scheme Modified to Include Source Terms . . . 86

5.5 Engquist-Osher Scheme Modified to Include Source Terms . . . 88

6 Details of the Numerical Schemes Used 90

6.1 Time-stepping Iteration . . . 91

6.1.1 Convergence of the Time-stepping Iteration . . . 91

6.2 Numerical Schemes and Source Terms Discretisation . . . 93

6.2.1 Direct Approach . . . 94

6.2.2 Indirect Approach 100

6.3 Well-balanced schemes 103

7 Description of the Test Problems 108

7.1 Test Problems for the Steady Saint-Venant Equations 109

7.2 Test problems for the Steady Euler Equations 111

7.2.1 The Analytical Expressions of the Flow Variables and Some Graphs111

7.2.2 The Gas Flow Test Problems 114

8 Results and Discussion 121

9 Conclusions and Further Work 147

Notation

General

$x \in [0, L]$	x is the space variable and L is the length of the channel
t	time
$\mathbf{U}$	vector of conserved variables
$\mathbf{F}$	flux function
$\mathbf{F}^*$	numerical flux function
$\mathbf{D}$	source function
U	conserved variable
F	scalar flux function
F^*	numerical flux function (scalar)
D	source function (scalar)

Saint-Venant equations

$h(x,t)$	depth, i.e. the level of the free surface above the bed level
$u(x,t)$	x-component of the fluid velocity
$A(x,t) = \int_0^h \sigma(x,\eta)\, d\eta$	wetted area
$Q(x,t)$	discharge (total volume of the flux through a given cross-section)
$b(x)$	channel breadth
g	acceleration due to gravity
$\sigma(x,t)$	width of channel as a function of both x and η
$\eta(x,t)$	coordinate which measures height relative to a fixed level
$z_b(x)$	height of the lowest point of the cross-section
$S_0 = -z_b'$	bed slope
S_f	friction slope (Manning formulation)

Euler equations

ρ	density
p	pressure

A	cross-section area
u	velocity (x-component)
T	absolute temperature
c_p	specific heat at constant pressure
c_ν	specific heat at constant volume
γ	ratio of specific heats
E	total energy
e	specific internal energy
H	total specific enthalpy
h	specific enthalpy (i.e. enthalpy per unit mass)
$\mathcal{R}$	gas constant
Q	mass flow
P	flow stress
c	sound speed
K	entropy function
$(.)_*$	denotes sonic flow values
$(.)_-$	denotes values upstream the shock
$(.)_+$	denotes values downstream the shock
$(.)_{in}$	denotes values at inlet
$(.)_e$	denotes values at exit
$(.)_t$	denotes values at the throat

Chapter 1

Introduction: overview

This thesis is concerned with (numerically) solving nonlinear singular differential equations of the form

$$\frac{dF(x,y(x))}{dx} = D(x,y(x)) \tag{1.1}$$

where $F(x,y(x))$ is a flux function that may or may not depend explicitly on x and $D(x,y(x))$ is a driving term. These equations arise in the simulation of steady flows in channels and ducts, becoming singular for critical flow when $\frac{\partial F}{\partial y}$ vanishes within the domain. The singularity of the equations studied occurs when the Froude number or the Mach number are equal to one (see Chapter 2). The solution of the equations by iterative methods is complicated by the fact that the necessary boundary conditions may switch at these critical points depending on the solution y which is not known in advance.

The one-dimensional steady *Saint-Venant Equations* and *Shallow Water Equations* as well as the *Euler Equations* of gas dynamics without heat sources lead to such singular forms. In the former equations, the source term $D(x,y(x))$ arises from bed slope and friction effects while in the latter it comes from friction only. However, the variation of the channel cross-section also leads to a contribution to the source terms. The flux function $F(x,y(x))$ arises from modifying the momentum or energy flux using conservation properties of the steady problem.

The study of steady flow, a flow essentially unchanging in time, is important in Hydraulics and in Gas Dynamics. In fact, the steady form of the Saint-Venant equations and Euler equations has been studied in many classical texts such as [8] and [84], respectively.

The one-dimensional homogeneous unsteady Saint-Venant equations and the Euler equations are hyperbolic systems of conservation laws which admit discontinuous solutions (so-called *weak solutions*). For given well posed boundary conditions there exists at most one smooth solution of these equations but there may be more than one discontinuous solution. In order to pick up the physically relevant solution a *vanishing viscosity method* can be used (see [59]). This method consists of adding higher order derivatives with small (*viscosity*) coefficients to obtain a modified equation with an unique solution and letting the viscosity coefficients tend to zero. The unique solution so obtained is called a *vanishing viscosity solution.* This method has its roots on more realistic models of water and gas flows than the Saint-Venant equations and the Euler equations (such as the Navier Stokes equations) which take into account diffusive or viscous effects. These vanishing viscosity discontinuous solutions are obtained (in the vanishing viscosity limit) from problems (often parabolic in nature) whose solutions are continuous with narrow regions where the solution changes very rapidly (these regions are called *shock layers*). Boundary conditions are imposed at both ends of the region. Since the (vanishing viscosity) limit process is nonuniform (the so-called *singular perturbation problem*) in the limit the procedure allows the appropriate boundary conditions to be chosen from the viscous second order differential equation without the limit problem necessarily having to satisfy either of those boundary conditions (see [59]). Other methods of finding the physical correct solution make use of *entropy conditions* (see, e.g. [49]). In [59] MacDonald (see also [60]) presents theoretical results for the steady flow problem derived from the Saint-Venant equations. The solution of the steady flow problem is constructed as the vanishing viscosity limit of solutions to a singular perturbation problem.

Shock capturing schemes are often used to obtain numerical solutions since these numerical schemes have the ability to deal with all the features of the flow in the entire domain. For example, upwind schemes based in the Roe scheme [75] and the Engquist-Osher [18] scheme are used. These are first-order methods which ensure correct application of the boundary conditions, in particular in the presence of shocks and expansions. An upwind method does not need any numerical boundary condition since these schemes pick up the “wind” direction. Hence, if artificial computational boundaries need to be assigned, (for example, if one can only solve a problem on a bounded domain), an upwind method will not take those values as an incoming signal and they will not affect the solution at interior points.

Higher order shock capturing methods (*high-resolution methods*) have been studied by other authors (see, e.g. [42, 21]) but are not dealt with in this thesis.

As in MacDonald [59] we apply what is called a *scalar approach* to the solution of the steady equations, that is, we apply the numerical schemes to an equation of the form (1.1) obtained in the steady state case by reduction of the full systems corresponding to the Saint-Venant equations and Euler equations to a single scalar equation. The analysis of these scalar problems is much simpler than for systems but still leads to the solution of a nonlinear system of difference equations.

The possibility of solving this nonlinear system by a *time stepping iteration* or the Newton method was studied in [60, 59]. In this thesis we use a *time stepping iteration*, also called a *pseudo-time iteration*, that models at a discretised level the solving of a time-dependent problem until a steady state is reached.

With a pseudo-time iteration applied to this scalar approach MacDonald [59] showed convergence and uniqueness properties for a form of the Saint-Venant equations (rectangular prismatic channel case).

The test problems used in this thesis for the Saint-Venant problem have been taken from [59] and were developed in previous work (see [60, 61]). These test problems have known analytical solutions and include features like varying channel geometries and discontinuous solutions.

The test problems chosen for the Euler equations, namely a diverging section and a nozzle, have been taken from Wixcey [103]. There exist known exact solutions for these problems although at a practical level the solution values have to be obtained with a certain degree of approximation that we take to be greater than or to the order of convergence of the methods we are using, so that we are able to compare our numerical results with sufficiently accurate values.

We note that equations of the form (1.1) arise from steady conservation laws with source terms. The theory for such nonhomogeneous systems is not so well-developed and the numerical treatment of source terms is also a subject of current research. Some recent work in this area and relevant for this thesis is [3, 99, 42, 20, 43, 50, 21, 33]. Some earlier work with useful discussions is presented in [53, 15, 90].

Numerical difficulties arise if the source terms are stiff and particularly in the steady-case if their magnitude is significant. Much work has been done related to *balancing source terms* [50, 3, 42, 5, 21, 31, 32, 33, 43, 73, 99].

Problems arise when applying *operator splitting methods* to steady problems or near steady problems ([95]). Our choice of upwind methods to deal with steady conservation laws with source terms is supported by work based on using upwind methods combined with *upwinding source terms* (e.g., [23, 3, 99, 20, 42]).

The idea of upwinding the source terms was first put forward by Roe in [78]. Further work was carried out in this direction in e.g. [3, 99]. The idea is to try to build discretisations of the source terms in a way similar to those used to construct the numerical flux functions.

An extra numerical difficulty arises when the flux function $F(x, y)$ depends explicitly on x. Such is the case for water flow in a nonprismatic open channel and is considered in this thesis. Two approaches have been taken. In the *direct* approach the derivative of the flux function $\frac{dF(x,y)}{dx}$ is approximated directly: in the second approach this derivative is split by applying the chain rule and the term $\frac{\partial F}{\partial x}$ is included in the right-hand side source terms. An interesting discussion on the possibility of including derivative terms such as the latter in a new formulation of the numerical flux function is given in [42].

In the next chapter we introduce the Saint-Venant equations and the Euler equations and show how in the steady state these equations can be written as a singular scalar ordinary differential equation of the form (1.1). In Chapter 3 some of the theoretical background is presented. Some general features of the numerical schemes used in the thesis are described in Chapter 4. Chapter 5 is dedicated to a discussion of ways to discretise the source terms. In Chapter 6 the theory is put into practice, that is, we present the test problems used and describe the particular features of the implementation of the numerical schemes (introduced previously in Chapter 4) mainly to water applications although gas applications are also described. Results are then presented in Chapter 7 as well as their discussion. Finally, in Chapter 8 we discuss possible future directions of research.

Chapter 2

Reduction of the Saint-Venant Equations and Euler Equations in the steady-state case

Our aim is to study the singular differential equations which arise from steady one-dimensional systems of hyperbolic conservation laws with source terms. The differential equations are obtained from the unsteady forms of these equations by assuming no time dependence. This leads to a steady system where the mass conservation equation is easily solved, and a system of two partial differential equations (PDEs) can then be reduced to one ordinary differential equation (ODE) (or a system of three PDEs to two ODEs). The differential equations break down at shocks and integral forms (leading to the *Rankine-Hugoniot conditions*) have to be used in this case. The shock situation corresponds to a singularity in the system of PDEs (and hence in the ODE) at an unknown location. It is the treatment of this singularity that is the main study of this thesis. In our approach we analyse an iteration to the solution (using pseudo time) that under steady boundary conditions may eventually reach the correct steady-state.

In the following section, Section 2.1, the general form of unsteady one-dimensional hyperbolic conservation laws with source terms is presented. Practical applications of these systems include the Saint-Venant equations modelling water flow in an open channel and the Euler equations modelling gas flow in a pipe. Properties of the steady problem for the former equations are presented in Section 2.2 and for the latter in Section 2.3.1. The influence of the source terms is discussed in Section 2.2.4. In Section 2.4 we present some analogies between the compressible gas and water models.

2.1 Unsteady Conservation Laws with Source Terms

Our aim is to study the steady state case of one-dimensional systems of hyperbolic conservation laws with source terms of the form

$$\mathbf{U}_t + \mathbf{F}(x, \mathbf{U})_x = \mathbf{D}(x, \mathbf{U}). \tag{2.1}$$

The function $\mathbf{F}$ is a *flux function*, the function $\mathbf{D}$ is a *source term* and $\mathbf{U}$ is the vector of conserved quantities. If $\mathbf{D}(x, \mathbf{U}) = \mathbf{0}$ then equations (2.1) are said to be in *conservative form.*

Particular cases of practical importance are systems where the flux function $\mathbf{F}$ depends only on $\mathbf{U}$

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = \mathbf{D}(x, \mathbf{U}). \tag{2.2}$$

By applying the chain rule to the derivative of the flux function, we can rewrite the systems of conservation laws (2.2) in the form (a quasi-linear form)

$$\mathbf{U}_t + J\mathbf{U}_x = \mathbf{D}(x, \mathbf{U}) \tag{2.3}$$

where J is the *Jacobian* matrix given by

$$J = \frac{d\mathbf{F}}{d\mathbf{U}}. \tag{2.4}$$

For the flux function $\mathbf{F}$ in (2.1) the chain rule yields

$$\frac{\partial \mathbf{F}}{\partial x} = \frac{\partial \mathbf{F}}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial x} + \frac{\partial \mathbf{F}}{\partial x}. \tag{2.5}$$

By including the derivative $\frac{\partial \mathbf{F}}{\partial x}$ in the source terms we can write the system (2.1) in a non-conservative form similar to (2.3), i.e.

$$\mathbf{U}_t + \bar{J}\mathbf{U}_x = \bar{\mathbf{D}}(x, \mathbf{U}). \tag{2.6}$$

where the right-hand side of system (2.6) is

$$\bar{\mathbf{D}}(x, \mathbf{U}) = \mathbf{D}(x, \mathbf{U}) - \frac{\partial \mathbf{F}}{\partial x} \tag{2.7}$$

and the Jacobian matrix $\bar{J}$ is given by

$$\bar{J} = \frac{\partial \mathbf{F}}{\partial \mathbf{U}}. \tag{2.8}$$

The systems (2.1) and (2.2) are of *hyperbolic type* if the Jacobian of the function $\mathbf{F}$ has all eigenvalues real and has a full set of linearly independent eigenvectors.

With x held fixed a practical application of systems of the form (2.2) is in Hydraulics. The flow of water in open prismatic channels can be modelled by the Saint-Venant equations which can be written in the form (2.2). More generally, for water flow in nonprismatic channels the Saint-Venant equations take the form (2.1).

Another practical application of systems of the form (2.2) is in Gas Dynamics. The Euler equations modelling quasi one-dimensional flow of a gas in a pipe with smoothly varying circular cross-section can be written in the form (2.2) (see [15, 22, 2]).

Other applications are possible (see [94] for more details).

Some useful references on the numerical solution of hyperbolic systems of conservation laws are [49, 44, 29, 9, 95, 93, 101]. Other relevant references in the theory of conservation laws are [11, 86, 80].

In the next sections the Saint-Venant and Euler equations are introduced and we show how these equations can be reduced to a singular scalar ODE of the form (1.1) in the steady-state case.

2.2 The Saint-Venant Equations

In this section the Saint-Venant equations are introduced and some of their properties are discussed with the main focus being the steady-state case. As we shall show, under the assumption of steady-state flow, the Saint-Venant equations reduce to a single nonlinear ordinary differential equation describing the variation of the free surface.

In Section 2.2.1 we introduce the Saint-Venant equations for channels with variable breadth function and some notation. The particular case of a channel with constant breadth function is also studied. The characteristic speeds are presented in Section 2.2.2. The steady-state case is the main focus of Section 2.2.3 whereas the occurrence of discontinuous solutions is the main focus of Section 2.2.4. The boundary conditions are discussed in Section 2.2.5.

2.2.1 The Saint-Venant Equations for Channels with Variable Breadth

The one-dimensional free surface water flow in a channel can be modelled by the Saint-Venant equations (see [8, 59]). These equations can be written as a system of equations of the form (2.1) with

$$\mathbf{U} = \begin{pmatrix} A \\ Q \end{pmatrix}, \quad \mathbf{F}(x, \mathbf{U}) = \begin{pmatrix} Q \\ \frac{Q^2}{A} + gI_1 \end{pmatrix}, \quad \mathbf{D}(x, \mathbf{U}) = \begin{pmatrix} 0 \\ gI_2 + gA(S_0 - S_f) \end{pmatrix}, \tag{2.9}$$

where $Q = Au$ is the *discharge*, $A = \int_0^h \sigma \, d\eta$ is the wetted cross-section, g is the acceleration of gravity, S_0 is the bed slope and S_f is the *friction slope* (associated with bed friction). I_1 and I_2 account for pressure forces and are defined by

$$I_1(x, h) = \int_0^h (h - \eta)\sigma \, d\eta \tag{2.10}$$

and

$$I_2(x, h) = \int_0^h (h - \eta)\sigma_x \, d\eta \tag{2.11}$$

where h is the water depth and $\sigma(x, \eta)$ is the channel width at a position η from the bottom ($\sigma(x, h)$ is the *free surface width*). (See Fig. 2.1 and Fig. 2.2). For simplicity, the channel is assumed symmetric about the xz-plane (see Fig. 2.3).

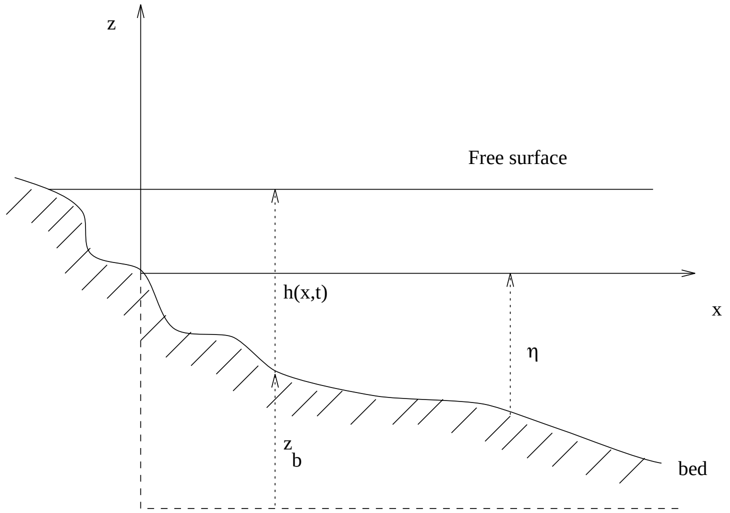

Figure 2.1: y-cross section showing bed and free surface

In this model it is assumed that both σ and S_0 are continuously differentiable functions and that $Q > 0$ everywhere (if $Q < 0$, just reverse the x direction to obtain $Q > 0$).

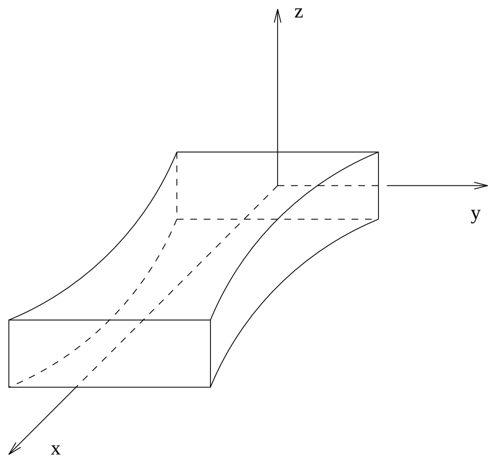

Figure 2.2: Sketch of a channel with rectangular x-cross section and variable breadth function (constant bed slope)

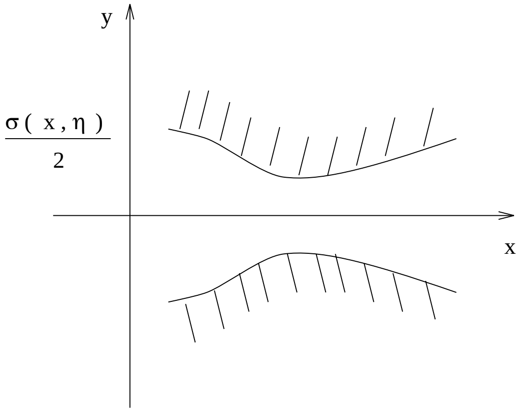

Figure 2.3: Horizontal cross section of a channel at height η

The quantity S_f is usually written (see [59]) in the form

$$S_f = \frac{Q|Q|}{K^2} \tag{2.12}$$

where P is the wetted perimeter given by

$$P = \sigma(x,0) + \int_0^h \sqrt{4+\sigma_\eta^2}\, d\eta, \tag{2.13}$$

K is the *conveyance* given by

$$K = \frac{A^{k_1}}{nP^{k_2}} \tag{2.14}$$

and n is a constant representing the *bed roughness* of the channel. The friction slope S_f can be expressed by using Chezy's or Manning's laws (see [8, 1]). Here the Manning formulation for the friction slope S_f is adopted, i.e. $k_1 = 5/3$, $k_2 = 2/3$ with the *Manning coefficient*, n, taking the value 0.03.

A channel is said to be *prismatic* if its x-cross-section does not change throughout its length. For these channels the breadth function σ is independent of x.

Our aim is to study nonprismatic channels (variable breadth function) with prismatic cross-section. For these channels the function σ can be written as $\sigma = b(x) + 2hZ$ where $b(x)$ is the width at the bottom (see Fig. 2.4).

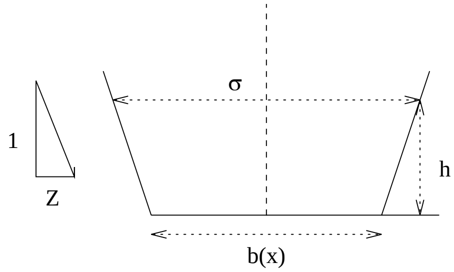

Figure 2.4: x-cross section showing a trapezoidal channel

In the following table, Table 2.1, the particular expressions for channels with rectangular or trapezoidal cross-section are given. Other types of channel cross-sections can be found in [8].

Nonprismatic channel	
Trapezoidal cross-section $b(x) > 0,\ Z > 0$	Rectangular cross-section $b(x) > 0,\ Z = 0$
$\sigma(x, h) = b(x) + 2hZ$	$\sigma(x, h) = b(x)$
$A(x, h) = h(b(x) + hZ)$	$A(x, h) = hb(x)$
$P(x, h) = b(x) + 2h\sqrt{1 + Z^2}$	$P(x, h) = b(x) + 2h$
$I_1 = \frac{1}{2}h^2 b(x) + \frac{1}{3}Zh^3$	$I_1 = \frac{1}{2}h^2 b(x)$
$I_2 = \frac{1}{2}h^2 b'(x)$	$I_2 = \frac{1}{2}h^2 b'(x)$

Table 2.1: Some formulas for the Saint-Venant equations (with variable breadth function) corresponding to rectangular and trapezoidal cross-section channels

In the particular case of a prismatic channel, the flux function in the Saint-Venant equations does not depend explicitly on x. In this case the Saint-Venant equations can be written in the form (2.1) where $\mathbf{U}$, $\mathbf{F}$ and $\mathbf{D}$ are given by (2.9) with the expressions in Table 2.1 simplified as shown in Table 2.2.

Prismatic channel	
Trapezoidal cross-section $b > 0,\ Z > 0$	Rectangular cross-section $b > 0,\ Z = 0$
$\sigma(h) = b + 2hZ$ $A(h) = h(b + hZ)$ $P(h) = b + 2h\sqrt{1+Z^2}$ $I_1 = \frac{1}{2}h^2 b + \frac{1}{3}Zh^3$ $I_2 = 0$	$\sigma(h) = b$ $A(h) = hb$ $P(h) = b + 2h$ $I_1 = \frac{1}{2}h^2 b$ $I_2 = 0$

Table 2.2: Some formulas for the Saint-Venant equations (with constant breadth function) corresponding to rectangular and trapezoidal cross-section channels

Taking into account these simplifications the Saint-Venant equations for a prismatic channel can be written in the form

$$\begin{pmatrix} A \\ Q \end{pmatrix}_t + \begin{pmatrix} Q \\ \frac{Q^2}{A} + gI_1 \end{pmatrix}_x = \begin{pmatrix} 0 \\ gA(S_0 - S_f) \end{pmatrix}. \tag{2.15}$$

The case where the breadth function is variable and the cross section is rectangular ($b(x)$ >0 and $Z = 0$) is the one studied here.

2.2.2 Quasi-linear Representation and Characteristic Speeds

For both the prismatic and nonprismatic cases (with rectangular cross-section) studied in the thesis the expression of the Jacobian matrix is similar. In fact, even if the flux function depends on x, that occurs solely due to the breadth function $b(x)$ depending on x. Hence, the function $\mathbf{F}(x, \mathbf{U})$ in (2.9) can be thought of as a function $\overline{\mathbf{F}}(b(x), \mathbf{U})$ of $b(x)$ and $\mathbf{U}$, and we have

$$\overline{\mathbf{F}}(b(x), \mathbf{U}) = \begin{pmatrix} Q \\ \frac{Q^2}{A} + \frac{1}{2}gAh \end{pmatrix} = \begin{pmatrix} Q \\ \frac{Q^2}{A} + \frac{1}{2}g\frac{A^2}{b(x)} \end{pmatrix}. \tag{2.16}$$

Therefore

$$\frac{\partial \mathbf{F}}{\partial x} = \frac{\partial \overline{\mathbf{F}}}{\partial b}\frac{db}{dx} + \frac{\partial \overline{\mathbf{F}}}{\partial \mathbf{U}}\frac{\partial \mathbf{U}}{\partial x} \tag{2.17}$$

where

$$\frac{\partial \overline{\mathbf{F}}}{\partial b} = \begin{pmatrix} 0 \\ -\frac{1}{2}gh^2 \end{pmatrix} \tag{2.18}$$

and

$$\frac{\partial \overline{\mathbf{F}}}{\partial \mathbf{U}} = \bar{J} = \begin{pmatrix} 0 & 1 \\ gh - u^2 & 2u \end{pmatrix}. \tag{2.19}$$

A similar expression to (2.19) is obtained for the case where the flux function depends only on $\mathbf{U}$ (prismatic channels) except that $\mathbf{F}$ and J do not have overlines (the overline notation means that $\mathbf{F}$ was thought of as a function of $\mathbf{U}$ and $b(x)$ instead of the original $\mathbf{U}$ and x and thus the variables kept constant in the partial differentiation are different). The Jacobian matrix J or $\bar{J}$ has eigenvalues

$$\begin{aligned} \lambda_1 &= u - c \\ \lambda_2 &= u + c \end{aligned} \tag{2.20}$$

where c is the *wave celerity* and is given by

$$c^2 = gA/b \tag{2.21}$$

The expressions (2.20) are the *characteristic speeds* and the corresponding right-eigenvectors of J are given by

$$\mathbf{r}_1 = \begin{pmatrix} 1 \\ u - c \end{pmatrix}, \qquad \mathbf{r}_2 = \begin{pmatrix} 1 \\ u + c \end{pmatrix}. \tag{2.22}$$

2.2.3 The Steady Problem

The steady flow equations can be obtained from the equations (2.1) or (2.2) by assuming no time dependence. In this case equations (2.1) or (2.2) reduce to

$$\frac{dQ}{dx} = 0 \tag{2.23}$$

$$\frac{dF}{dx} = D, \tag{2.24}$$

with F and D being, respectively, the second components of $\mathbf{F}$ and $\mathbf{D}$ (see (2.15)). The first equation (2.23) corresponds to a constant discharge and hence (2.24) can be modified and written as a single nonlinear ordinary differential equation of the form

$$\frac{dF(x,h)}{dx} = D(x,h), \tag{2.25}$$

with

$$F(x,h) = \frac{Q^2}{A} + gI_1 \tag{2.26}$$

and

$$D(x,h) = gI_2 + gA(S_0 - S_f) \tag{2.27}$$

(Q constant).

Equivalently, equation (2.25) (which is of the form (1.1)) can be written as

$$(1 - F_r^2)\frac{dh}{dx} = S_0 - S_f + \frac{Q^2}{gA^3}\int_0^h \sigma_x \, d\eta \tag{2.28}$$

where

$$F_r = \sqrt{\frac{Q^2\sigma(x,h)}{gA^3}} \tag{2.29}$$

is the *Froude number*. (Note that $\sigma(x,h) = b(x)$ in the case of a channel with rectangular cross-section.)

By rewriting equation (2.28) in the form

$$\frac{dh}{dx} = \frac{S_0 - S_f + \frac{Q^2}{gA^3}\int_0^h \sigma_x \, d\eta}{1 - F_r^2} \tag{2.30}$$

we can see that the derivative on the left-hand side of equation (2.30) becomes unbounded when the denominator on the right-hand side of equation is zero ($1 - F_r^2 = 0$). When $F_r = 1$ the differential equation breaks down and the flow smoothness assumption is no longer valid. Hence, the differential equation is called singular and the singularity occurs when

$$F_r = \frac{|u|}{c} = 1, \tag{2.31}$$

which corresponds to *critical* flow. The flow is called *supercritical* if $F_r > 1$ and *subcritical* if $F_r < 1$ and its behaviour differs accordingly.

We assume there is only one *critical depth function* $h_c(x)$ in each x cross-section such that $h = h_c(x)$ solves equation (2.29). (Note that for prismatic channels the critical depth is constant and does not depend on x.)

It is possible to obtain an explicit formula for $h_c(x)$ if the channel cross-section is rectangular. Indeed, in this case, the critical depth takes the form

$$h_c(x) = \sqrt[3]{\frac{Q^2}{gb(x)^2}}. \tag{2.32}$$

For the trapezoidal cross-section, though, the critical depth function must be obtained implicitly as the positive (real) root of a polynomial of the 6th degree. If a value of the critical depth function is needed in a certain cross-section it can be obtained by using the Newton method.

It is also assumed that the width of the channel does not approach zero as the depth becomes large. Then $\frac{\partial F}{\partial h}$ has the following properties at a cross-section x:

- $\frac{\partial F}{\partial h} = 0$ at $h = h_c(x)$ ($F_r = 1$, *critical flow*)
- $\frac{\partial F}{\partial h} > 0$ for $h > h_c(x)$ ($F_r < 1$, *subcritical flow*)
- $\frac{\partial F}{\partial h} < 0$ for $h < h_c(x)$ ($F_r > 1$, *supercritical flow*).

More comments on the solutions of equation (2.25) are given in Section 2.2.4.

We would like to mention that, by reducing the steady Saint-Venant equations, it is also possible to obtain an ODE in the dependent variable A which is singular when the product $(c-u)(c+u)$ is zero.

2.2.4 Discontinuous Solutions in the Unsteady and Steady Cases

The differential equations in the Saint-Venant model (2.1),(2.9) break down when a shock occurs. This shock is known in Hydraulics as *hydraulic bore* or simply *bore* and represents a discontinuity in $\mathbf{U}$. From the integral form of the equations it is possible to obtain the *jump conditions* (also called *Rankine-Hugoniot conditions*, especially in Gas Dynamics) characterizing the shock (see, e.g. [11],[49]and [80]). These conditions are independent of the system being homogeneous or nonhomogeneous (by the inclusion of source terms). The Saint-Venant equations together with the Rankine-Hugoniot conditions ensure that we get a weak solution but they do not guarantee uniqueness. An extra condition on the shock is needed and is called an *entropy condition* by analogy with Gas Dynamics (see Sections 2.3.6 and 2.4 for more details). For the Saint-Venant problem this extra condition is motivated by the fact that a hydraulic bore is a dissipative phenomena with no mechanism to create energy. Hence it is a constraint on the jump in energy across a shock (see [88], [59] and [102]). Instead of including more "physics" (an extra condition) to get the unique physically relevant solution it is possible to use a *vanishing viscosity solution* argument (see, e.g. Smoller [86] and Thomas [93]). In this way the unique solution is obtained as the limit solution of a viscous problem when the viscosity vanishes. Physically this approach can be thought of as obtaining a solution approximating a model that includes some small amount of dissipation (see Chapter 3 for more details).

For a shock propagating with speed s the *Rankine-Hugoniot (jump) condition*

$$[\mathbf{F}(\mathbf{U})] = s[\mathbf{U}] \tag{2.33}$$

must be satisfied (see, e.g. [49]), that is from (2.15),

$$[Q] = s[A] \tag{2.34}$$

$$[\frac{Q^2}{A} + gI_1] = s[Q] \tag{2.35}$$

where $[.] = (.)_R - (.)_L$. The subscripts "L" and "R" refer to computed values at either side of the shock, respectively, on the left and on the right of the shock.

It is possible to derive an "entropy" condition on the energy for the Saint-Venant equations in a manner similar to that in Stoker [88] for frictionless flow. In fact, the jump conditions (2.34)-(2.35) imply the following "entropy" condition (see [59])

$$m(E_R - E_L - s(u_R - u_L)) \leq 0 \tag{2.36}$$

where $u = \frac{Q}{A}$ is the fluid velocity, m is given by

$$m = Q_R - sA_R = Q_L - sA_L \tag{2.37}$$

and the energy E is given by

$$E = \frac{u^2}{2} + gh. \tag{2.38}$$

As we can see from equation (2.23), at steady state the discharge is constant and hence any bore must be stationary, i.e. have zero velocity. This is known as a *hydraulic jump*. Therefore for steady flow $s = 0$ and the jump conditions (2.34)-(2.35) simplify to

$$[Q] \quad = \quad 0 \tag{2.39}$$

$$[\frac{Q^2}{A} + gI_1] \quad = \quad 0, \tag{2.40}$$

or

$$\begin{aligned} Q_L &= Q_R & (2.41) \\ \frac{Q_L^2}{A_L} + g(I_1)_L &= \frac{Q_R^2}{A_R} + g(I_1)_R. & (2.42) \end{aligned}$$

For steady flow the "entropy" condition (2.36) simplifies to

$$E_R \leq E_L. \tag{2.43}$$

We have assumed, without loss of generality, that the discharge Q is positive since the x-direction can be reversed if $Q < 0$. The case of $Q = 0$ is trivial and corresponds to a horizontal free surface solution.

It can be proved that a hydraulic jump can only occur if and only if

$$h_L < h_c < h_R. \tag{2.44}$$

More details on using the vanishing viscosity theory with the steady problem are given in Chapter 3.

2.2.5 Boundary Conditions

The boundary condition requirements to the steady Saint-Venant problem are thought having in mind that the steady problem is a particular example of an unsteady problem. Hence, the boundary condition requirements should be the same but obeying the rule of being constant in time.

From the theory of characteristics it is known that in order to have a well-posed problem, the initial and boundary conditions to impose, must have in consideration the geometry of the characteristics. Furthermore, those boundary and initial conditions determine uniquely, either explicitly or implicitly, the Riemann invariants (see [101], Chapter 8). For one-dimensional homogeneous systems of conservation laws the Riemann invariants are constant along characteristics whose slope is given by the eigenvalues (2.20) (also called *characteristics directions*). For more details see, e.g. [1, 101].

A general rule to take in consideration is:

the number of boundary conditions should be equal to the number of characteristics entering the domain (usually the direction along the characteristic is that of time increasing).

Furthermore, even if the analytical problem does not require boundary conditions to be specified, the use of a numerical method may require extra boundary conditions to be specified for computation purposes. These numerical boundary conditions can be computed extrapolating from values of the domain interior. If the aim is to maintain a balance (like for finite volume schemes), numerical boundary conditions may need to be prescribed (e.g. through a fictitious cell). The subject of numerical boundary conditions will be addressed in Chapter 6.

For the Saint-Venant equations if $Q > 0$, we have also $u > 0$ and consequently, $\lambda_2 > 0$. Hence one variable has to be specified at inflow (for either supercritical or subcritical flow) and that variable should be Q which we know remains constant in the steady-sate case.

If the flow is supercritical at inflow ($\lambda_1 > 0$) we have to specify another variable at inflow, and we choose the depth h. No variables have to be specified for supercritical outflow.

If the flow is subcritical at outflow ($\lambda_1 < 0$) we have to specify another variable at outflow, and we choose the depth h. No additional variable has to be specified for subcritical inflow.

More has to be said on the particular boundary values that, along with other conditions, will guarantee the existence of a steady solution of our problem. We address this subject in Chapter 3.

2.3 The Euler Equations Modelling One-dimensional Flow in a Nozzle

In this section the Euler equations are introduced and some of their properties are discussed. A special form of these equations for ducts with axi-symmetric geometries is presented as well. The main focus is the steady-state case where the solution of the conservation of mass (ordinary) differential equation allows its substitution in the momentum and energy equations. In the absence of source terms in the the conservation of energy equation, the Euler equations can be reduced to a single nonlinear ordinary differential equation describing the variation of the density throughout the pipe. This reduction is similar to the one described in Section 2.2. If there is a source term in the energy equation, it is possible to reduce the Euler equations to a system of two ODEs.

For the so-called *quasi one-dimensional* gas flow in a nozzle that we aim to study, the gas flow can be thought of as homentropic (isentropic throughout the whole flow field). In this case the resulting ODE depends explicitly, both through the flux function and the source term, on the (constant) entropy function which is discontinuous across a shock. This raises some extra problems (lower continuity assumptions) that were not found when reducing the steady Saint-Venant equations to an ODE through a similar reduction process (see Section 2.2.3).

In Section 2.3.1 the Euler equations are introduced as well as the assumptions on the gas. In the following section, Section 2.3.2, the equations modelling quasi-one dimensional flow in a nozzle are introduced and it is shown how to transform them, if needed, into the Euler form presented in Section 2.3.1. The characteristic speeds and the steady-state case are the main focus of Section 2.3.3. In Sections 2.3.4 and 2.3.5 it is shown how to reduce the Euler system of equations in the steady case to, respectively, one ODE and two ODEs. The occurrence of discontinuous solutions is the main focus of Section 2.3.6 whereas the boundary conditions are the main focus of Section 2.3.7.

2.3.1 The Euler Equations

A form of the Euler equations modelling gas flow which includes some source terms is

$$\begin{array}{llcl} \text{mass} & \dfrac{\partial \rho}{\partial t} + \dfrac{\partial(\rho u)}{\partial x} & = & 0 \qquad (2.45) \\ \text{momentum} & \dfrac{\partial(\rho u)}{\partial t} + \dfrac{\partial(\rho u^2 + p)}{\partial x} & = & b(x, \mathbf{U}) \qquad (2.46) \\ \text{energy} & \dfrac{\partial E}{\partial t} + \dfrac{\partial(u(E+p))}{\partial x} & = & \Omega. \qquad (2.47) \end{array}$$

where p is the *pressure*, ρ is the *density*, u is the *velocity*, $b(x, \mathbf{U})$ is a friction term (to be described later), Ω is the heat input and E is the *total energy* defined by

$$E = \rho(e + \frac{1}{2}u^2), \qquad (2.48)$$

e being the *specific internal energy*. The specific internal energy is related to the *specific* *enthalpy* h through the formula

$$h = e + \frac{p}{\rho}. \tag{2.49}$$

As we shall see when studying the steady-state case, it is useful to define the *total specific enthalpy*, H, which is given by the formula

$$H = \frac{E + p}{\rho}. \tag{2.50}$$

By using equations (2.48) and (2.49), H is also given by

$$H = h + \frac{1}{2}u^2. \tag{2.51}$$

The equations (2.45)-(2.47) are of the form (2.2) with

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad \mathbf{F}(x, \mathbf{U}) = \begin{pmatrix} \rho u \\ p + \rho u^2 \\ u(E + p) \end{pmatrix}, \quad \mathbf{D}(x, \mathbf{U}) = \begin{pmatrix} 0 \\ b \\ \Omega \end{pmatrix}. \tag{2.52}$$

The three Euler equations (2.45)-(2.47) have four unknowns: ρ, u, p and E. In order to be able to solve this system a further equation relating the unknowns is needed and that will be an *equation of state*. If we assume that the internal energy is a known function of pressure and density, the equation of state is of the form

$$e = e(p, \rho). \tag{2.53}$$

This equation of state depends on the gas under consideration.

Even if viscous effects are taken in account it can be assumed that the gas is *perfect* (see [40]). For a perfect gas, the internal energy (per unit mass) e is a function of the temperature alone, $e = e(T)$ (see, e.g. [94, 49]), and we have a perfect gas law relating pressure, density and temperature which can be written in the form

$$p = \mathcal{R}\rho T \tag{2.54}$$

where $\mathcal{R}$ is the gas constant per unit of mass (equal to the universal gas constant divided by the molecular mass of the gas) and we see that e is a function of p/ρ.

For air, using the S.I. units, we have

$$\mathcal{R} = 287 m.N/Kg.K \tag{2.55}$$

(see, e.g. [65]). Henceforth the quantities are defined in S.I. units unless stated otherwise.

In the particular case of a *polytropic* (or *calorically perfect*) gas, the specific heat capacity c_ν is constant and we have

$$e = c_\nu T \tag{2.56}$$

$$h = c_p T \tag{2.57}$$

and also (by the perfect gas law)

$$c_p - c_\nu = R. \tag{2.58}$$

Hence, assuming the gas is *polytropic*, we have an equation of state of the form $e = e(p, \rho)$ which can be written in the form

$$e = \frac{p}{(\gamma - 1)\rho} \tag{2.59}$$

where the *ratio of specific heats* γ (or *adiabatic exponent*) is defined as

$$\gamma = \frac{c_p}{c_\nu}. \tag{2.60}$$

For air in standard conditions

$$\gamma = 1.4 \tag{2.61}$$

and this is the value used in the thesis.

Using equation (2.59) in equation (2.48) we obtain another form of the equation of state of a polytropic gas:

$$E = \frac{p}{(\gamma - 1)} + \frac{1}{2}\rho u^2. \tag{2.62}$$

The equation of state (2.59) or (2.62) can be used in conjunction with the Euler equations (2.45)-(2.47) to solve for any of the unknowns ρ, u, p and e (or E).

Another thermodynamic quantity noticeably important for gases is the *entropy* since the pressure depends on it. The entropy S can be defined up to an additive constant by (see, e.g. [11, 49, 94])

$$S = c_\nu \ln\left(\frac{p}{\rho^\gamma}\right) + C_0 \tag{2.63}$$

where C_0 is a constant. Solving this equation for p gives

$$p = K(S)\rho^\gamma \tag{2.64}$$

where $K(S) = Ce^{\frac{S}{c_\nu}}$ and C is a constant. The entropy S satisfies the equation

$$S_t + uS_x = 0 \tag{2.65}$$

that is, the entropy is constant along particle paths of smooth flow. On those paths, $K(S)$ is a function of the initial entropy S_0, $K = K(S_0)$, but this value changes from path to path. Hence in the streamline equation (2.65) can be written in the form

$$p = K\rho^\gamma \tag{2.66}$$

which is called the *entropic equation of state*.

For discontinuous flow there is a jump in the value of the entropy. In fact, if the particle crosses a shock then, by the second law of thermodynamics, the entropy must increase and therefore there is a jump in the entropy across the shock.

A quantity that plays an important role is the local *sound speed* which, for a polytropic gas, is given by

$$c = \sqrt{\left(\frac{\partial p}{\partial \rho}\right)_{S=\text{constant}}} = \sqrt{\gamma K(S)\rho^{\gamma-1}} = \sqrt{\frac{\gamma p}{\rho}}. \tag{2.67}$$

If the entropy S is constant in the whole fluid domain we talk about *isentropic flow* or *homentropic flow* (see [4]). In particular, in a homentropic flow with no shock waves, the entropy function K is constant on every particle path since $S = S_0$ and hence equation (2.66) is valid in all the fluid domain. Therefore the equation (2.66) is called an *isentropic equation of state*. Moreover equation (2.59) is only a function of ρ and u and can be written in the form

$$e = \frac{K}{\gamma - 1}\rho^{\gamma-1}. \tag{2.68}$$

Furthermore the energy equation (2.47) becomes redundant with the values of E, e and p being computed through algebraic relations. Thus the Euler equations reduce to the mass and momentum equations, in the variables ρ and u. These equations modelling isentropic flow are one of the objects of study in the thesis or more specifically, their steady form that can be reduced to a single singular ODE.

Another quantity worth defining is the *Mach number*, M, which is defined by

$$M = \frac{|u|}{c} \tag{2.69}$$

where c is the sound speed.

2.3.2 The Equations of Gas Flow in a Nozzle

We are interested in a model for the one-dimensional or quasi one-dimensional flow of a compressible gas in a pipe with axi-symmetric geometry. The pipe is assumed to have 'slowly varying' circular cross-section so that to a first approximation the flow is in the (positive) x-direction only. This asserts that the fluid velocity (or fluid speed) is positive and thus the mass flow is also positive. The gas is also assumed to be inviscid. The equations modelling this flow may be derived by expressing the physical conservation of mass, momentum (in the x direction) and energy.

The quasi one-dimensional approximation of gas flow in a duct or nozzle with variable cross-section (neglecting friction and heat conduction) is given by the system of equations

$$\begin{aligned}
&\text{mass} & \frac{\partial(\rho A)}{\partial t}+\frac{\partial(\rho A u)}{\partial x} &= 0 & (2.70)\\
&\text{momentum} & \frac{\partial(\rho A u)}{\partial t}+\frac{\partial(\rho A u^2+Ap)}{\partial x} &= p\frac{dA}{dx} & (2.71)\\
&\text{energy} & \frac{\partial(AE)}{\partial t}+\frac{\partial(Au(E+p))}{\partial x} &= 0. & (2.72)
\end{aligned}$$

The system (2.70)-(2.72) consists of three equations in four variables: ρ, e, p and u. A fourth equation relating the variables is given by the equation of state of a polytropic gas (2.59).

These equations are of the form (2.2) with

$$\mathbf{U}=\begin{pmatrix} A\rho \\ A\rho u \\ AE \end{pmatrix}, \qquad \mathbf{F}=\begin{pmatrix} A\rho u \\ A(p+\rho u^2) \\ Au(E+p) \end{pmatrix}, \qquad \mathbf{D}=\begin{pmatrix} 0 \\ pA'(x) \\ 0 \end{pmatrix} \qquad (2.73)$$

where E is given by definition (2.48). A is the variable cross-section area and depends only on x. The sound speed c is given by (2.67).

The equations (2.70)-(2.72) can be written in a form similar to (2.45)-(2.47) if we change to new variables (see [22]). In fact, if we introduce the new variables

$$\begin{aligned}
\bar{p} &= pA\\
\bar{u} &= u\\
\bar{\rho} &= \rho A\\
\bar{E} &= AE = A\rho(e+\frac{u^2}{2}) \qquad (2.74)
\end{aligned}$$

we can rewrite the system (2.70)-(2.72) as

$$\begin{pmatrix} \bar{\rho} \\ \bar{\rho}\bar{u} \\ \bar{E} \end{pmatrix}_t + \begin{pmatrix} \bar{\rho}\bar{u} \\ \bar{p} + \bar{\rho}\bar{u}^2 \\ \bar{u}(\bar{E} + \bar{p}) \end{pmatrix}_x = \begin{pmatrix} 0 \\ \bar{p}\frac{A'(x)}{A(x)} \\ 0 \end{pmatrix}. \tag{2.75}$$

2.3.3 Matrix Representation and Characteristic Speeds

The system (2.70)-(2.72) can be written in the form (2.3) by using the chain rule, assuming that the flux function **F** is differentiable. The Jacobian is given by

$$J = \begin{pmatrix} 0 & 1 & 0 \\ \frac{\gamma-3}{2}u^2 & (3-\gamma)u & \gamma - 1 \\ \frac{\gamma-2}{2}u^3 - \frac{c^2 u}{\gamma-1} & \frac{3-2\gamma}{2}u^2 + \frac{c^2}{\gamma-1} & \gamma u \end{pmatrix} \tag{2.76}$$

where c is the sound speed.

It can be shown (see, e.g. [94]) that the eigenvalues of J are

$$\lambda_1 = u - c \tag{2.77}$$

$$\lambda_2 = u \tag{2.78}$$

$$\lambda_3 = u + c \tag{2.79}$$

and the corresponding right eigenvectors are given by

$$\mathbf{r}_1 = \begin{pmatrix} 1 \\ u - c \\ H - uc \end{pmatrix} \tag{2.80}$$

$$\mathbf{r}_2 = \begin{pmatrix} 1 \\ u \\ \frac{1}{2}u^2 \end{pmatrix} \tag{2.81}$$

$$\mathbf{r}_3 = \begin{pmatrix} 1 \\ u + c \\ H + uc \end{pmatrix} \tag{2.82}$$

where H is given by equation (2.50) or, equivalently, by equation (2.51). Since the eigenvalues are real and the eigenvectors form a complete set of linearly independent eigenvectors the one-dimensional Euler equations for ideal gases are *hyperbolic* (*strictly hyperbolic* if the eigenvalues are distinct, i.e. if the sound speed $c \neq 0$).

2.3.4 The Steady Problem I: reduction to a singular scalar ODE

In this section we show how to reduce the steady system of Euler equations to one ordinary differential equation in a way similar to that taken to study the Saint-Venant equations. Furthermore, we show how to obtain some relations between the flow variables that allow us to characterize the gas flow in more detail.

The Reduced ODE

By assuming no time dependence, the velocity, pressure, density and entropy are unchanged in time and the flow field can be described by streamlines invariant in time. Hence on each streamline the entropy is constant and the entropy function $K(S)$ is constant as well (see, e.g. [103] for more details).

Furthermore, in quasi one-dimensional flow, if the streamlines are assumed indistinguishable then any of them may be thought of as a representative streamline (see [103]). The flow may, therefore, be assumed to be isentropic with equation of state given by (2.66). The representative streamline (representing the full flow) will be chosen along the duct axis having a prescribed value for the total specific enthalpy and constant entropy function. These values are given in Chapter 7, where the test problems are described.

When a stationary shock, which is perpendicular to a streamline in the flow field, occurs the flow properties may be considered approximately one-dimensional but the flow is not isentropic anymore. We can distinguish three regions: before the shock, across the shock and after the shock. In the first and the last regions the flow can be assumed to be isentropic (with different equations of state) and across the shock the flow variables must satisfy jump conditions. These jump conditions, which a discontinuous solution must satisfy, are studied in more detail in Section 2.3.6.

The system of equations (2.45)-(2.47) can be written in the form

$$\frac{d(\rho A u)}{dx} = 0 \tag{2.83}$$

$$\frac{d(\rho A u^2 + Ap)}{dx} = p\frac{dA}{dx} \tag{2.84}$$

$$\frac{d(Au(E+p))}{dx} = 0. \tag{2.85}$$

By solving the mass equation (2.83) we get

$$A\rho u = AQ = m = \text{constant} \tag{2.86}$$

where the *mass flow* Q is given by

$$Q = \rho u. \tag{2.87}$$

If we use (2.86) to solve the energy equation (2.85) we obtain

$$\frac{p}{(\gamma-1)\rho} + \frac{p}{\rho} + \frac{u^2}{2} = H = \text{constant} \tag{2.88}$$

which is the steady form of *Bernoulli's equation.* (See also equations (2.50) and (2.51).) Since the flow is assumed to be isentropic, equation (2.88) can be written in the form

$$\frac{\gamma}{\gamma-1}K\rho^{\gamma-1} + \frac{u^2}{2} = H = \text{constant} \tag{2.89}$$

where K is the constant entropy function in (2.66). As we have seen, the total specific enthalpy is constant and thus if its value is known the system of three ODEs (2.83)-(2.85) can be reduced to both the mass and momentum equation on the variables ρ, u and p plus the isentropic equation of state (2.66).

We can also substitute equation (2.86) into equation (2.84) to get

$$\frac{d}{dx}\left(Ap + \frac{m^2}{A\rho}\right) = p\frac{dA}{dx}, \tag{2.90}$$

or, if we use the isentropic equation of state (2.66),

$$\frac{d}{dx}\left(AK\rho^\gamma + \frac{m^2}{A\rho}\right) = K\rho^\gamma\frac{dA}{dx}. \tag{2.91}$$

Note that this equation is very similar in form to the one obtained by reducing the Saint-Venant equations in the steady-state case (cf. equation (2.25)). Actually, equation (2.91) is of the form (1.1) since it can be writen as

$$\frac{d}{dx}F(x,\rho) = D(x,\rho) \tag{2.92}$$

with the functions F and D given , respectively, by

$$F(x,\rho) = AK\rho^\gamma + \frac{m^2}{A\rho} \tag{2.93}$$

and

$$D(x,\rho) = K\rho^\gamma\frac{dA}{dx}. \tag{2.94}$$

Alternatively, an ODE in the dependent variable p (instead of ρ) is obtained if isentropic flow is assumed and $\rho = \left(\frac{p}{K}\right)^{\frac{1}{\gamma}}$ is used. Other reductions are possible (see [87]).

Equivalently, equation (2.91) can be written as

$$(c^2 - u^2)\frac{d\rho}{dx} = \frac{m^2}{A^3\rho}\frac{dA}{dx}, \tag{2.95}$$

where the sound speed c is given by equation (2.67) and m is given by equation (2.86).

It is also possible to write equation (2.95) in a way to bring out the role of the Mach number, i.e.

$$(1 - M^2)\frac{d\rho}{dx} = \frac{m^2}{\gamma K A^3 \rho^\gamma}\frac{dA}{dx}$$

or, equivalently,

$$\frac{d\rho}{dx} = \frac{\frac{m^2}{\gamma K A^3 \rho^\gamma}\frac{dA}{dx}}{1 - M^2} \tag{2.96}$$

where M is the Mach number defined by equation (2.69).

Comparing (2.96) with equation (2.30) in Section 2.2.3 we see that the Mach number plays a role similar to the Froude number. When $M = 1$ the derivative $\frac{d\rho}{dx}$ becomes unbounded and the differential equation breaks down. Hence a singularity in equation (2.92) occurs when $M = 1$ and the flow is said to be *sonic* in this case. The flow is called *supersonic* if $M > 1$ and *subsonic* if $M < 1$.

Furthermore, the flux F given by (2.93) is convex in the variable ρ since

$$\frac{\partial^2 F}{\partial \rho^2} = AK\gamma(\gamma - 1)\rho^{\gamma-2} + \frac{2m^2}{A\rho^3} > 0. \tag{2.97}$$

Note that if equation (2.85) were nonhomogeneous ($\Omega \neq 0$ in equation (2.47)), the steady system of Euler equations

$$\frac{d(\rho A u)}{dx} = 0 \tag{2.98}$$

$$\frac{d(\rho A u^2 + Ap)}{dx} = p\frac{dA}{dx} \tag{2.99}$$

$$\frac{d(Au(E + p))}{dx} = \Omega \tag{2.100}$$

(obtained from the unsteady system (2.45)-(2.47)) cannot be reduced to one ODE in the same way as described above (see equation (2.91)). Instead, it is possible to reduce system (2.98)-(2.100) to a two-equation nonhomogeneous system by using the conservation of mass equation (2.86). This will be discussed further in Chapter 9.

As in Section 2.2.3 it is possible to define a *critical density*, $\rho_c(x)$. We assume that there is only one critical density $\rho_c(x)$ in each x cross-section such that $\rho = \rho_c(x)$ solves the equation $M = \frac{u}{c} = 1$ (equivalently, $\rho_c(x)$ is the solution of $\frac{\partial F}{\partial \rho} = 0$). This critical density is given explicitly by

$$\rho_c(x) = \left(\frac{m^2}{\gamma K A^2} \right)^{\frac{1}{\gamma+1}}. \tag{2.101}$$

This possibility of expressing explicitly the critical density is very similar to the rectangular cross-section (nonprismatic) channel in the Saint-Venant equations studied in Section 2.2.3.

The derivative $\frac{\partial F}{\partial \rho}$ has the following properties at a particular cross-section x

- $\frac{\partial F}{\partial \rho} = 0$ at $\rho = \rho_c(x)$ ($M = 1$, *sonic flow*)
- $\frac{\partial F}{\partial \rho} > 0$ for $\rho > \rho_c(x)$ ($M < 1$, *subsonic flow*)
- $\frac{\partial F}{\partial \rho} < 0$ for $\rho < \rho_c(x)$ ($M > 1$, *supersonic flow*).

Some Properties of Steady Gas Flow

It is possible to obtain more relations between the flow variables that allow us to characterize the flow in more detail. One of them is the *maximum speed* allowed. This quantity comes out from equation (2.89) when we solve it for the variable ρ to get

$$\rho = K^{\frac{1}{1-\gamma}} \left(\frac{\gamma - 1}{\gamma} \left(H - \frac{u^2}{2} \right) \right)^{\frac{1}{\gamma-1}}. \tag{2.102}$$

Note that equation (2.102) expresses ρ as a function of the flow speed u (for a particle moving on a streamline H and K are constant and in steady flow m is constant as well).

As we can see from equation (2.102) we must have

$$H - \frac{1}{2}u^2 = H - \frac{1}{2}\frac{m^2}{A^2\rho^2} \geq 0 \tag{2.103}$$

since $p, \rho \geq 0$ and $\gamma = 1.4$ or, equivalently,

$$0 \leq u \leq \sqrt{2H}. \tag{2.104}$$

Hence the *maximum speed* allowed, $u_{\max}$, is given by

$$u_{\max} = \sqrt{2H}. \tag{2.105}$$

By using equation (2.105) it is possible to rewrite the steady Bernoulli equation (2.88) in the form

$$(1-\beta^2)c^2 + \beta^2 u^2 = 2H\beta^2 \tag{2.106}$$

or, using equation (2.105), as

$$(1-\beta^2)c^2 + \beta^2 u^2 = \beta^2 u_{\max}^2 \equiv C_*^2 \tag{2.107}$$

where the constant β is given by

$$\beta^2 = \frac{\gamma-1}{\gamma+1}. \tag{2.108}$$

The constant C_* is called the *critical speed* and is attained when the (scalar) velocity coincides with the local speed of sound c (note that c is not constant). Note that the critical speed is independent of the entropy.

The significance of the critical speed in characterizing the type of flow can be seen through another form of equation (2.106), i.e.

$$u^2 - C_*^2 = (1-\beta^2)(u^2 - c^2). \tag{2.109}$$

We can see that;

if $|u| < C_*$ then $|u| < c$ $(M < 1)$ and the flow is *subsonic*

if $|u| = C_*$ then $|u| = C_* = c$ $(M = 1)$ and the flow is *sonic* or *critical*

if $|u| > C_*$ then $|u| > c$ $(M > 1)$ and the flow is *supersonic*.

Note that the "modulus" sign can be removed since we assumed $u > 0$.

2.3.5 The Steady Problem II: reduction to a system of two ODEs

In Section 2.3.4 it was shown how to reduce, in the steady case, the Euler equations (2.70)-(2.72) to a scalar singular ODE. That reduction was possible since the form of the Euler equations studied did not include source terms in the energy equation or mass equation. If there is a source term in the energy equation it is still possible to reduce, in the steady case, the three-equation system to a system of two ODEs. The application of a similar reduction process to a mass equation with a term on the right-hand side of equation (2.70) is not so straightforward and is not studied in this thesis.

The reduction of the Euler system of equations (2.45)-(2.47), which has a source term in the energy equation ($\Omega \neq 0$), is described in this section. By assuming no time dependence, the conservation of mass differential equation is easily solvable and its solution is used in the remaining differential equations yielding a system of two ODEs in the variables ρ and E.

The quasi-linear form of the Euler equations (2.45)-(2.47) can be written as (see (2.76)),

$$\begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}_t + \begin{pmatrix} 0 & 1 & 0 \\ \frac{\gamma-3}{2}u^2 & (3-\gamma)u & \gamma-1 \\ \frac{\gamma-2}{2}u^3 - \frac{c^2 u}{\gamma-1} & \frac{c^2}{\gamma-1}\frac{3-\gamma}{2}u^2 & \gamma u \end{pmatrix} \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}_x = \begin{pmatrix} 0 \\ b \\ \Omega \end{pmatrix}. \tag{2.110}$$

In the steady state, the conservation of mass equation (2.86) can be used to reduce the system (2.110) to a two-equation system of the form

$$\begin{pmatrix} \frac{\gamma-3}{2}u^2 & \gamma-1 \\ \frac{\gamma-2}{2}u^3 - \frac{c^2 u}{\gamma-1} & \gamma u \end{pmatrix} \begin{pmatrix} \rho \\ E \end{pmatrix}_x = \begin{pmatrix} b \\ \Omega \end{pmatrix}. \tag{2.111}$$

By using $u = m/\rho$, the system (2.111) can be written as

$$\begin{pmatrix} \frac{\gamma-3}{2}\frac{m^2}{\rho^2} & \gamma-1 \\ (\gamma-1)\frac{m^3}{\rho^3} - \gamma m \frac{E}{\rho^2} & \gamma \frac{m}{\rho} \end{pmatrix} \begin{pmatrix} \rho \\ E \end{pmatrix}_x = \begin{pmatrix} b \\ \Omega \end{pmatrix} \tag{2.112}$$

where the vector of variables is

$$\mathbf{w} = \begin{pmatrix} \rho \\ E \end{pmatrix}. \tag{2.113}$$

Another approach is to start from the conservation form of the Euler equations with source terms followed by the use of the algebraic conservation of mass equation (2.86) in the remaining ODEs. This yields the system

$$\begin{pmatrix} \frac{3-\gamma}{2}\frac{m^2}{\rho} + (\gamma-1)E \\ \gamma m \frac{E}{\rho} + \frac{1-\gamma}{2}\frac{m^3}{\rho^2} \end{pmatrix}_x = \begin{pmatrix} b \\ \Omega \end{pmatrix} \tag{2.114}$$

in the conserved variables ρ and E.

For the nozzle system of equations (2.70)-(2.72), with a energy source term Ω included, it is possible to obtain a similar (reduced) system of equations in the variables $A\rho$ and AE.

The Jacobian matrix in equation (2.111) is

$$J = \begin{pmatrix} \frac{\gamma-3}{2}u^2 & \gamma-1 \\ \frac{\gamma-2}{2}u^3 - \frac{c^2 u}{\gamma-1} & \gamma u \end{pmatrix} \tag{2.115}$$

and its eigenvalues μ are the roots of the polynomial equation

$$\mu^2 - \left(\frac{\gamma-3}{2}u^2 + \gamma u\right)\mu - \left(u^3 - c^2 u\right) = 0. \tag{2.116}$$

Hence the eigenvalues are

$$\mu_1 = \frac{\phi - \sqrt{\phi^2 + 4\lambda_1\lambda_2\lambda_3}}{2} \tag{2.117}$$

$$\mu_2 = \frac{\phi + \sqrt{\phi^2 + 4\lambda_1\lambda_2\lambda_3}}{2} \tag{2.118}$$

where λ_1, λ_2 and λ_3 are the eigenvalues of the original steady Euler system and are given by equations (2.77)-(2.79) and ϕ is defined by

$$\phi = \frac{\gamma-3}{2}u^2 + \gamma u. \tag{2.119}$$

Is also worth mention that the eigenvalues μ_1 and μ_2 of this reduced system are related to the eigenvalues λ_1, λ_2 and λ_3 of the original (3 equations) system through

$$\mu_1\mu_2 = -\lambda_1\lambda_2\lambda_3. \tag{2.120}$$

If $m > 0$ then $u > 0$ and $c > 0$. Thus, as we have seen, $\lambda_2, \lambda_3 > 0$ but λ_1 may change sign. For strict hyperbolicity (real and distinct eigenvalues) we should have

$$\phi^2 + 4\lambda_1\lambda_2\lambda_3 > 0. \tag{2.121}$$

When $\lambda_1 = u - c > 0$ strict hyperbolicity is automatically verified and $\mu_1 < 0$ and $\mu_2 > 0$ whereas if $\lambda_1 = u - c < 0$ the inequality (2.121) is verified when $\phi < -\sqrt{-4\lambda_1\lambda_2\lambda_3}$ or $\phi > \sqrt{-4\lambda_1\lambda_2\lambda_3}$, with both μ_1 and μ_2 having the same sign, which is that of ϕ. The sign of ϕ changes accordingly to the sign of

$$\frac{\gamma-3}{2}u + \gamma.$$

The corresponding right eigenvectors are

$$\mathbf{r}_1 = \begin{pmatrix} \gamma-1 \\ \mu_1 - \frac{\gamma-3}{2}u^2 \end{pmatrix} \tag{2.122}$$

and

$$\mathbf{r}_2 = \begin{pmatrix} \gamma-1 \\ \mu_2 - \frac{\gamma-3}{2}u^2 \end{pmatrix}. \tag{2.123}$$

2.3.6 Discontinuous Solutions

Here we study the effect of irreversible thermodynamic processes that can occur in a nozzle having the particular form of a stationary shock front. It is known (see, e.g. [11],[45],[84]) that under certain conditions (mass flow and pressure) a *normal shock* (shock which is locally perpendicular to a streamline) can occur in the divergent part of the nozzle. This discontinuous flow can be modelled mathematically (assuming the fluid to be inviscid) by a jump discontinuity and will have an effect of increasing the entropy (see Section 2.3.2). Adequate boundary conditions have to be considered as well.

For a shock propagating with speed s the *Rankine-Hugoniot (jump) condition*[1] must be satisfied (see, e.g. [49]), that is

$$[\mathbf{F}(\mathbf{U})] = s[\mathbf{U}] \tag{2.124}$$

where $[.] = (.)_R - (.)_L$, the subscripts "L" and "R" refer to computed values on either side of the shock, respectively, behind and after the shock. For steady nozzle flow (see equations (2.70)-(2.72)), a stationary shock ($s = 0$) must satisfy the jump conditions given by

$$[A\rho u] = 0 \tag{2.125}$$

$$[A(p + \rho u^2)] = 0 \tag{2.126}$$

$$[Au(E + p)] = 0. \tag{2.127}$$

Hence, since at the shock there is no change in area, $A_L = A_R$, we obtain from equation (2.125)

$$Q_L = \rho_L u_L = \rho_R u_R = Q_R = Q \tag{2.128}$$

i.e. the mass flow Q is conserved across the shock.

Furthermore, from equation (2.126) we get

$$P_L = p_l + \rho_L u_L^2 = p_R + \rho_R u_R^2 = P_R = P. \tag{2.129}$$

Thus the total momentum flux P is conserved across the shock.

[1] The jump conditions in Gas Dynamics were stated correctly by Rankine and by Hugoniot although they had been stated incorrectly before by Riemann (he conserved entropy instead of energy) [47].

The jump condition for the energy, (2.127), can be written in the form

$$\left(A\rho u\left(e+\frac{p}{\rho}+\frac{u^2}{2}\right)\right)_L = \left(A\rho u\left(e+\frac{p}{\rho}+\frac{u^2}{2}\right)\right)_R. \tag{2.130}$$

which, by using the fact that mass is conserved $A\rho u = m =$ constant, yields

$$H_L = h_L + \frac{u_L^2}{2} = h_R + \frac{u_R^2}{2} = H_R = H. \tag{2.131}$$

Thus the total enthalpy H is also conserved across the (normal) shock.

Note that equations (2.128) and (2.129) are valid for any fluid, irrespective of its equation of state. The equation (2.131), though, incorporates the thermodynamics of the flow and reflects the fact that the steady Bernoulli equation holds across a shock, although now the fluid speed u and enthalpy h will be discontinuous. Likewise, the entropy is discontinuous across a shock, increasing its value (irreversible process). This translates in terms of the entropy function as

$$K_L \leq K_R. \tag{2.132}$$

Several useful properties and relations can be derived from these conditions (see [11] or [84] for more details). One such relation is the *Prandtl-Meyer relation*, i.e.

$$u_L u_R = C_*^2. \tag{2.133}$$

This relation shows that either $|u_L| > C_*$ and $|u_R| < C_*$, or vice-versa; so the flow at one side of the shock is always supersonic and at the other side subsonic (see section 2.3.4). By using the entropy condition it can be proved (see, e.g. [101]) that a shock can only occur from supersonic flow to subsonic flow, i.e.

$$u_L > C_* > u_R \tag{2.134}$$

or

$$u_L > c > u_R. \tag{2.135}$$

In Table 2.3, we summarise some of the variations in the flow variables across a shock (occurring in the diverging section of a nozzle). For more details on the formulae yielding these conclusions, see [11] or [84].

Flow variable	Type of variation
density ρ	increases: $\rho_L < \rho_R$
speed u	decreases: $u_L > u_R$
pressure p	increases: $p_L < p_R$
temperature T	increases: $T_L < T_R$
entropy K	increases: $K_L < K_R$

Table 2.3: The variation of some flow variables across a shock in a nozzle

2.3.7 Boundary Conditions

Like in Section 2.2.5 we look at the boundary conditions for the steady scalar equation obtained from reducing the steady (area dependent) Euler equations by considering them as particular cases of unsteady boundary conditions but remaining constant in time.

It is known that the original (unsteady) nonlinear system of differential equations can be diagonalized and the Jacobian matrix of resulting system (in characteristic variables) has the same eigenvalues. The theory of characteristics leads us to study the signs of the eigenvalues (2.77)-(2.79) yielding the slope of characteristics where the Riemann invariants (see, e.g. [7]) are constant. (Note that the Riemann invariants are constant for the homogeneous case.)

Furthermore, it is also know that boundary and initial conditions that lead to a well-posed problem for the Euler equations are those that determine explicitly or implicitly the Riemann invariants (see, e.g. [101, 44]). The numerical implementation of these conditions is not straightforward (for more comments see [101], Chapter 10). Usually values of the primitive variables are specified at boundaries when needed. The general rule is the same, i.e., that the number of boundary conditions in a point on the boundary must be equal to the number of characteristics entering the domain at that point.

In steady isentropic flow both the total specific enthalpy H ($H > 0$) and $m = \rho A u$ are constant. If we consider $m > 0$ then we have also $Q > 0$ and $u > 0$ (since A and ρ are positive). Consequently, $\lambda_2 > 0$ and $\lambda_3 > 0$ (see Section 2.3.3). Hence, at inflow two conditions must be imposed. Furthermore, the sign of λ_1 is determined by the type of flow, being positive if supersonic flow occurs and negative otherwise. Hence at a subsonic inflow boundary or at a supersonic outflow boundary no extra condition is needed. Quite the contrary occurs at a supersonic inflow boundary or at a subsonic outflow boundary. Our choice relatively at this change in the sign of the eigenvalue is to specify values for the density ρ although in most books the value specified is the pressure p. Note that knowing K, which is constant for isentropic flow, and ρ, we can compute the value of the pressure p through the isentropic equation of state. So it is equivalent to give one or the other but ρ is induced by our choice of reduced equation (see (2.91)).

Furthermore, test problems with convergent-divergent nozzles impose other restrictions on the flow with the position of throat having a determinant role (see Chapter 7 for more details).

We discuss subsonic inflow boundaries in more detail.

Isentropic flow in a divergent nozzle (K is given)

At a subsonic inflow boundary two conditions are needed and we choose to specify the values of the variables H and m (or Q) (which are constant in isentropic flow) and thus allowing ρ to float. A subsonic outflow boundary corresponds to have $\lambda_1 < 0$ so we must specify ρ at outflow. There is no need to specify ρ at a supersonic outflow boundary.

Isentropic flow in a de Laval nozzle (K is given)

Different types of flow can occur in this case and they are determined by the shape of the nozzle (area variation) and by the boundary conditions at inflow and outflow. The flow can be

(i). entirely subsonic

(ii). subsonic (but sonic at the throat)

(iii). subsonic-supersonic (sonic at the throat)

In case (iii), at the inflow subsonic boundary we specify H and and m (or Q). At the supersonic outflow boundary there is no need to specify any variable.

There are a infinite number of possible subsonic isentropic solutions (case (i)). Those subsonic solutions are determined from different values of the density ρ (our variable of choice) specified at outflow. The values of the density vary in the range of the value of density that corresponds to a stagnation pressure and the value of density at outflow that corresponds to have local sonic conditions at the minimum area section (case (ii)).

Normal shock occurring in the divergent section of a de Laval nozzle (K_L, K_R are given)

A normal shock occurs in the diverging section of the nozzle (limit case is the shock at the outlet section) when a subsonic boundary (we use density again) is prescribed that causes the flow to become *choked* at the throat (i.e. the flow remains sonic at the throat). The flow goes supersonic after the throat (still isentropic) until a shock occurs somewhere in the diverging section. After the normal shock the flow becomes subsonic and will remain subsonic (isentropic flow again) verifying the outlet boundary condition, which will be a prescribed value of ρ that gives also the shock location (see, e.g. [2]).

The particular values of the boundary conditions used in the test problems originating different types of flow profiles are described in Chapter 7.

2.4 Analogy Between Compressible Gas and Water Model

It is possible to write the Saint-Venant (without friction) or Shallow water equations in a form analogous to the Euler equations of Gas Dynamics (compressible gas) by choosing a different set of dependent variables. The equations obtained are analogous to the ones for a polytropic gas obeying an adiabatic law (2.64) with $\gamma = 2$ (e.g. see [88], [11] or [49]).

For gases the *entropy* is an important thermodynamic quantity since the pressure depends on it. Not so for water where the influence of changes in entropy is negligible and p may be considered a function of the density alone (see [11]).

Furthermore, regardless of the *"entropy" condition* (an inequality) being called "entropy" it is in fact a condition on the energy for the Saint-Venant equations, not a condition on the entropy as in the Gas Dynamics case. In fact, as discussed in Whitman [102] and in Stoker [88], although the conservation of mass and momentum differential equations imply a conservation of energy differential equation this conservation of energy does not hold across a shock in the Shallow Water Theory. (In the discussion presented there source terms were not considered but the jump conditions are independent of the system being homogeneous or nonhomogeneous.)

For the Saint-Venant equations there is not a third jump condition for the conservation of energy across a shock as in the case in Gas Dynamics theory which allows mechanical energy to be converted in heat. Instead, in Shallow Water Theory, the energy plays a role similar to the entropy in Gas Dynamics. The losses of mechanical energy across a shock in water correspond to an increase in entropy across a shock in Gas Dynamics.

Hence, for the Saint-Venant equations we have a two equation model with two jump conditions holding across a shock and a third extra condition on the energy, an inequality that is called *"entropy"* condition in analogy with what happens in Gas Dynamics. For the Euler equations of Gas Dynamics we have a three differential equations model with three jump conditions holding across a shock and an extra inequality, a condition on the entropy.

In the next chapter the numerical schemes used to approximate (2.25) are described. The work of MacDonald [59] is central to this thesis and to the next chapter in particular. We have followed some of his ideas and extended others and used some of the test problems in [59] to test our algorithms.

Chapter 3

Vanishing Viscosity Theory Applied to the Steady Problem

In this chapter we present some of the existing vanishing viscosity theory that can be applied to the scalar flow problems arising from either the Saint-Venant equations or the Euler equations in the steady state case. The theory guarantees the existence of (weak) discontinuous solutions that are physically relevant (the so-called *entropy solutions*) and for which the differential equation breaks down (smooth assumptions are no longer valid). These entropy solutions can be looked at as solutions of viscous differential equations (parabolic) obtained when considering the limit when a viscous coefficient tends to zero. These limit solutions are unique.

We present some of the work done by different authors (e.g. [71, 59, 54, 55]), that is relevant to this thesis since it points to a possible way of studying, for example, the singular ODE arising from the steady Saint-Venant equations with a nonprismatic channel. A particular attention is given to the work of MacDonald [59] concerning the Saint-Venant equations with a prismatic channel.

3.1 Vanishing Viscosity Theory

Hyperbolic systems of conservation laws, such as the homogeneous Saint-Venant equations or the Euler equations, arise from models of physical processes that do not include viscous or dispersive mechanisms. More accurate models are obtained if we take these mechanisms in account. Such is the case when the differential equations are modified by the inclusion of *viscous terms*, that is, terms with higher order derivatives multiplied by coefficients that are small. For consistency we would like these more general problems to have solutions that, in the limit when the *viscous coefficients* vanish, are solutions of the original hyperbolic problem. Note that the original first order system can have more solutions besides this limit solution, also called *vanishing viscosity solutions*.

By using this *vanishing viscosity method* it is possible to obtain some results concerning the existence and uniqueness of physical solutions of the original hyperbolic system. In fact, the vanishing viscosity theory can provide a mechanism to discriminate against unphysical solutions.

In general the higher order system is parabolic, thus having smooth solutions. The allowable discontinuous solutions of the first-order (hyperbolic) system can be thought of as vanishing viscosity limit solutions of smooth solutions with narrow regions where the solution changes rapidly. These regions are called *shock layers*.

It is possible in certain cases to obtain conditions independent of the viscous limit that allow us to choose the physically relevant solution.

As we have seen in Chapter 2, the steady Saint-Venant equations and the steady Euler equations can be reduced to one differential equation of the form

$$\frac{dF}{dx} = D(x, w), \qquad 0 \leq x \leq L, \qquad t > 0 \tag{3.1}$$

where the flux function F depends always on w ($w > 0$) and may depend on x as well (as is the case for the Saint-Venant equations with a nonprismatic channel). The variable w is the depth h in the case of the Saint-Venant equations and the density ρ or the pressure p for the Euler equations. Independently of where this equation (3.1) came from (a steady hyperbolic system), we can think of it as the steady limit of a new unsteady differential equation of the form

$$\frac{\partial w}{\partial t} + \alpha \frac{\partial}{\partial x} F(x, w) = \alpha D(x, w) \tag{3.2}$$

with appropriate boundary conditions and $\alpha = \pm 1$. In turn, this last equation arises from the integral conservation law

$$\int_{x_1}^{x_2} [w]_{t_1}^{t_2} dx + \alpha \int_{t_1}^{t_2} [F(x, w)]_{x_1}^{x_2} dt = \alpha \int_{t_1}^{t_2} \int_{x_1}^{x_2} D(x, w) dx dt, \tag{3.3}$$

where $t_2 \geq t_1 \geq 0$ and $0 \leq x_1 \leq x_2 \leq L$ are arbitrary. At steady state this integral form yields

$$[F(x,w)]_{x_1}^{x_2} dt = \int_{x_1}^{x_2} D(x,w)dx, \tag{3.4}$$

which has the same form as the steady integral form of the Saint-Venant equation and Euler equations. Therefore, at steady state, equation (3.3) and the Saint-Venant equations or Euler equations have the same weak solutions, even though the transient behaviour in unrelated.

We recall that for a scalar homogeneous problem of the form

$$\begin{aligned} \frac{\partial w}{\partial t} \quad &+ \quad \frac{\partial}{\partial x} f(w) = 0, \\ t > 0, \qquad -\infty < \quad & x \quad < \infty, \qquad w(x,0) = w_0(x), \end{aligned} \tag{3.5}$$

the only physically relevant solution can be defined as the vanishing viscosity solution, i.e. the limit solution as the viscosity coefficient $\epsilon \downarrow 0$ of the parabolic equation

$$\frac{\partial w}{\partial t} + \frac{\partial}{\partial x} f(w) = \epsilon \frac{\partial^2 w}{\partial x^2}. \tag{3.6}$$

Oleinik [68] demonstrates the existence and uniqueness of a vanishing viscosity solution for given initial data $w_0(x)$ which satisfies the so-called *Oleinik (entropy) condition*

$$\frac{f(w) - f(w_L)}{w - w_L} \geq s \geq \frac{f(w) - f(w_R)}{w - w_R} \tag{3.7}$$

for all w between w_L and w_R where s is the speed of the shock given by

$$s = \frac{f(w_R) - f(w_L)}{w_R - w_L}. \tag{3.8}$$

The entropy condition identifies the physically allowed discontinuities, and the weak solutions satisfying the entropy condition (3.7) are called *entropy (satisfying) solutions.*

If we assume that the Oleinik [68] entropy condition still discriminates towards the physically relevant solution of the nonhomogeneous equation (3.2), then at steady state ($s = 0$) the condition corresponding to (3.7) reduces to

$$\alpha \left(\frac{F(x,w) - F(x,w_L)}{w - w_L} \right) \geq 0 \tag{3.9}$$

for all w between w_L and w_R, and this implies that the physical entropy condition (2.43) holds if we take $\alpha = -1$ (the converse is not necessarily true). That is, although the steady solutions of both (3.2) and the steady Saint-Venant or Euler systems are the same, if $\alpha = +1$ is considered, the solutions will be entropy violating solutions. MacDonald [59] shows this for the Saint-Venant equations and it can be shown also for the Euler equations. Indeed, a particular case of Oleinik's entropy condition holding for a convex scalar flux on the variable w is

$$f_w(x, w_L) > s > f_w(x, w_R) \tag{3.10}$$

and this holds if $w_L > w_R$. Since, for the Euler equations our original flux (2.93) is convex (see equation (2.97)) and we have $\rho_L < \rho_R$ across a shock (see Table 2.3), we should start from a concave flux, that is, we should consider $f = -F$.

Hence, we can compute steady (entropy) solutions to the Saint-Venant or Euler equations via computing steady solutions of the differential equation

$$\frac{\partial w}{\partial t} - \frac{\partial}{\partial x} F(x, w) = -D(x, w) \tag{3.11}$$

where w is the depth h for the Saint-Venant equations and w is the density ρ for the Euler equations, with corresponding F and D given Chapter 2.

MacDonald [59], studying the Saint-Venant problem, shows that for prismatic channels with cross-section having a single critical depth, a solution satisfying the physical entropy condition (2.43) is also a solution satisfying Oleinik's (steady) entropy condition.

Summarising, instead of trying to apply the vanishing viscosity theory to a steady system of equations (usually by adding a viscous term to the momentum equation), which will be much harder also because of the source terms (see, e.g. [86]), we can reduce the steady system to one differential equation and apply the vanishing viscosity theory to the scalar case.

Hence, to obtain results concerning the existence and uniqueness of entropy solutions of problem (3.11) we could study the viscous problem

$$\frac{\partial w}{\partial t} - \frac{\partial}{\partial x} F(x, w) = -D(x, w) + \epsilon \frac{\partial^2 w}{\partial x^2}, \tag{3.12}$$

where $\epsilon > 0$.

Moreover, since our interest is steady solutions, we will look at the steady viscous problem (see Fig. 3.1)

$$\begin{aligned} \epsilon \frac{d^2 w_\epsilon}{dx^2} + \frac{d}{dx} F(x, w_\epsilon) = D(x, w_\epsilon) \quad , &\qquad w_\epsilon > 0, \quad 0 \leq x \leq L \\ w_\epsilon(0) = \gamma_0 \quad , &\qquad w_\epsilon(L) = \gamma_1 \end{aligned} \tag{3.13}$$

$$\begin{tikzcd}
& \boxed{\text{Unsteady}} & \boxed{\text{Steady}} & \boxed{\text{Unsteady}} \\
\boxed{\text{Integral form}} & \int_{x_1}^{x_2}[w]_{t_1}^{t_2}dx - \int_{t_1}^{t_2}[F(x,w)]_{x_1}^{x_2}dt = -\int_{t_1}^{t_2}\int_{x_1}^{x_2} D(x,w)dxdt \arrow[r] & {[F(x,w)]_{x_1}^{x_2} = \int_{x_1}^{x_2} D(x,w)dx} & \int_{x_1}^{x_2}[\mathbf{w}]_{t_1}^{t_2}dx - \int_{t_1}^{t_2}[\mathbf{F}(x,\mathbf{w})]_{x_1}^{x_2}dt = -\int_{t_1}^{t_2}\int_{x_1}^{x_2} \mathbf{D}(x,\mathbf{w})dxdt \arrow[l] \\
\boxed{\text{Differential form}} & \frac{\partial w}{\partial t} - \frac{\partial}{\partial x}F(x,w) = -D(x,w) \arrow[r] & \boxed{\frac{d}{dx}F = D(x,w)} & \frac{\partial \mathbf{w}}{\partial t} - \frac{\partial}{\partial x}\mathbf{F}(x,\mathbf{w}) = -\mathbf{D}(x,\mathbf{w}) \arrow[l] \\
\boxed{\text{Viscous form}} & \frac{\partial w}{\partial t} - \frac{\partial}{\partial x}F(x,w) = -D(x,w) + \varepsilon\frac{\partial^2 w}{\partial x^2} \arrow[r] & \varepsilon\frac{d^2 w}{dx^2} + \frac{d}{dx}F = D(x,w) & \frac{\partial \mathbf{w}}{\partial t} - \frac{\partial}{\partial x}\mathbf{F}(x,\mathbf{w}) = -\mathbf{D}(x,\mathbf{w}) + \varepsilon M\frac{\partial^2 \mathbf{w}}{\partial x^2} \arrow[l]
\end{tikzcd}$$

Figure 3.1: Ways of looking at the steady problem (in integral, differential and viscous form) starting from an unsteady problem

where ϵ, γ_0, $\gamma_1 > 0$.

Since the differential equation of the viscous problem (3.13) is second order, two boundary conditions are needed, the simplest being Dirichlet boundary conditions. Problem (3.13) is a *singular perturbation problem* (see, e.g. [69]) since the order of the differential equation reduces from second order to first order as ϵ vanishes. Furthermore, we cannot expect the solution of the limit problem to satisfy both boundary conditions of the viscous problem since we have a nonuniform convergence. Hence, although the choice of two boundary conditions may appear to be against the actual choice of boundary conditions for the steady problem (which may have to be specified at both ends, either or neither end of the channel as discussed in Chapter 2), the nonuniform nature of the limiting process allows the choice of boundary conditions that the limit solution does not need to satisfy. Nevertheless, with the choice of the boundary values γ_0 and γ_1 for the viscous problem we would like to control the behaviour of the limiting solution. For example, we would like the limit solution to satisfy the jump (or Rankine-Hugoniot) condition.

In Section 3.2 we discuss useful work done by different authors which uses the vanishing viscosity theory to prove convergence to steady state solutions of monotone finite difference schemes. Then, in Section 3.3 we discuss possible applications of the vanishing viscosity theory to more general problems and to the particular problems studied in this thesis.

3.2 The Use of the Vanishing Viscosity Theory in the Steady State Case

In [71], Osher used the vanishing viscosity theory and artificial time stepping to prove convergence to a unique steady state solution of a nonlinear singular perturbation problem using the Engquist-Osher scheme [17, 16, 18]. Osher [71] recognised that the conservative approximation to the spatial derivative can be used in approximating singular perturbation problems of the form (3.13). The problem discussed in [71] is of the form

$$\begin{aligned} \epsilon y'' - a(y)y' - b(x,y) &= H(x), \quad -1 \le x \le 1 \\ y(-1) = A \quad , &\qquad y(1) = B \end{aligned} \tag{3.14}$$

where $0 < \epsilon \ll 1$ and A and B are arbitrary constants, $a(y)$, $b(x, y)$ are C^2 functions verifying

$$b(x, 0), \quad b_y(x, y) \geq 0. \tag{3.15}$$

The problem is studied for both $H(x) = 0$ and $H(x) \neq 0$ by using numerical monotone schemes devised in [17, 16, 18] to approximate unsteady homogeneous scalar conservation laws. By discretising the unsteady problem corresponding to (3.14), Osher is able to prove convergence of the solutions of this unsteady problem to solutions of the analogue discrete version of (3.14) as $t \to \infty$ and independently of the initial guess and independently of ϵ.

The idea in Osher's work of using the unsteady discrete problem to study the discrete related steady problem obtained in the limit (as $t \to \infty$) is that of using a pseudo-time iteration to solve the discrete nonlinear system of equations obtained by using finite differences discretisation to numerically solve the steady problem and studying the convergence of this iteration. Proofs of convergence of this pseudo time iteration (which is a Picard iteration) rely on the contraction mapping theorem (see, e.g. [70]).

Viscous steady problems of the form (3.13) with F of the form $F(w)$ have been studied by several authors (e.g., [59, 54, 55, 56]). Less work has been done in the case $F(x, w)$.

The theory applied in MacDonald [59] to the steady Saint-Venant problem for prismatic channels makes use of the class of functions which have bounded total variation and is based on the theory used in Lorenz [54, 55]. A function $w \in BV[c, d]$, i.e. w has total variation bounded if it is bounded and all points of discontinuity are simple ($w(x-)$ and $w(x+)$ exist) and the set of discontinuities is countable. By grouping all the elements in the space $BV[c, d]$ into equivalence classes of almost everywhere equal functions, a normalised space $NBV[c, d]$ can be constructed (see, e.g. [59]).

MacDonald [59, 60], studying the steady Saint-Venant problem, only considers (positive) solutions that are bounded below away from zero (otherwise, physical quantities such as energy would become unbounded). Those solutions are in the set $NBV_+[c, d]$ which is defined by

$$NBV_+[c, d] = \{w \in NBV[c, d] : w(x) \geq C > 0 \, \text{for} \, c \leq x \leq d, \text{for a constant} \, C\}.$$

The theory used by Lorenz [54, 55] to solve the problem (3.13), requires the functions

$f = -F$ and D to be defined for all w. In [54], Lorenz studies a problem similar to the one in (3.13) but in the interval $[0, 1]$ (a change of variables can easily transform the original problem to one in the latter interval) and assuming that

$$D_w(x, w) \geq \delta > 0 \quad \forall (x, w) \in [0, 1] \times I\!R.$$

In order to apply the theory of Lorenz to a viscous problem relevant to the steady Saint-Venant problem for prismatic channels, MacDonald [59, 60] modifies the theory used by Lorenz [54] to restrict it to positive solutions (depth positive) and to allow some less restrictive conditions on D_w. The starting point is the following theorem.

Theorem 1 ([59]) *Consider the problem P_ϵ given by*

$$\epsilon \frac{d^2 w_\epsilon}{dx^2} - \frac{d}{dx} f(w_\epsilon) = b(x, w_\epsilon), \quad w_\epsilon > 0, \quad 0 \leq x \leq 1,$$
$$w_\epsilon(0) = \gamma_0, \quad w_\epsilon(1) = \gamma_1, \tag{3.16}$$

where $\epsilon, \gamma_0, \gamma_1 > 0$, $f \in C^2(0, \infty)$, $b_x, b_w, b_{wx} \in C([0, 1] \times (0, \infty))$ and

$$b_w > 0 \tag{3.17}$$

for all $w > 0$ and $x \in [0, 1]$. Additionally, suppose that there are positive constants m, M such that

$$b(x, m) \leq 0 \quad \text{and} \quad b(x, M) \geq 0 \quad \forall x \in [0, 1]. \tag{3.18}$$

Then the following hold:

(i). Problem P_ϵ has a unique solution $w_\epsilon \in C^2[0, 1]$ for all $\epsilon > 0$ which satisfies the bounds

$$0 < \underline{w} \leq w_\epsilon \leq \bar{w} \quad (0 \leq x \leq 1), \tag{3.19}$$

where $\underline{w} = \min\{\gamma_0, \gamma_1, m\}$ and $\bar{w} = \max\{\gamma_0, \gamma_1, M\}$.

(ii). $\|w'_\epsilon\| \leq K_1$ for all $\epsilon > 0$ where K_1 is independent of ϵ.

(iii). There is a unique function $W \in NBV_+[0, 1]$ such that $w_\epsilon \to W$ in L_1 as $\epsilon \downarrow 0$. The function W satisfies the bounds

$$0 < \underline{w} \leq W \leq \bar{w} \quad (0 \leq x \leq 1). \tag{3.20}$$

(iv). $w = W$ is the only function in $NBV_+[0,1]$ which satisfies

(a) If I is an interval where w is continuous, then $f(w(x))$ is differentiable on I, one-sided at end points, and the differential equation

$$-\frac{d}{dx}f(w) = b(x,w) \tag{3.21}$$

holds on I.

(b) If w is continuous at $x \in (0,1)$, then

$$\begin{aligned} f(w_l) = f(w_r) \geq f(k) \qquad if \qquad w_l > w_r \\ f(w_l) = f(w_r) \leq f(k) \qquad if \qquad w_l < w_r, \end{aligned} \tag{3.22}$$

for all k between $w_l = w(x-)$ and $w_r = w(x+)$.

(c) For $j = 0,1$ and k between $w(j)$ and γ_j we have

$$(-1)^{j+1}\operatorname{sgn}(w(j) - \gamma_j)(f(w(j)) - f(k)) \geq 0, \tag{3.23}$$

where $\operatorname{sgn}(x) = -1, 0, 1$ *for* $x < 0, = 0, > 0$, *respectively.*

Then, MacDonald [59] proceeds to show that the theory can be applied to the steady flow of water for a certain type of prismatic channels under certain smooth assumptions on the conveyance (2.14) and on the bed slope S_0 and also assuming the restrictive assumption that the bed slope is positive. He also proves that this assumption on the bed slope is needed for the theory to hold. Under those assumptions on the prismatic channels, MacDonald is able to prove that there exists at most one limit solution W (depth) satisfying any set of boundary values and also a weak existence result of a solution satisfying a entropy condition similar to Oleinik's (3.7) by assuming positive γ_0 and γ_1. By assuming that the channel has a single critical depth, then one can prove that a more physical entropy condition on the energy holds.

This theory is only applicable to prismatic channels and further research is needed in order to be able to prove similar properties for a flux function of the type $f(x,w)$. In [59], MacDonald refers to a paper by Lorenz and Sanders [58] that has some theory for that type of flux function that could possibly be adapted to apply to the steady flow problem.

As we have seen, the vanishing viscosity theory for the scalar case can be applied to the reduced equations obtained from the steady Saint-Venant equations (for prismatic channels), yielding results concerning existence and uniqueness of entropy solutions [59]. This vanishing viscosity approach is based on the idea of studying a steady problem with possibly discontinuous solutions, through the study of a family of problems having smooth solutions with these solutions tending to the discontinuous solutions of the original problem in some limit.

Under certain conditions the physical solutions of the steady flow problem are exactly the steady state entropy satisfying solutions of the scalar conservation law (3.2) with $\alpha = -1$ and, as Osher recognised, one can use finite difference schemes in conservation form (homogeneous problem) to approximate this nonhomogeneous scalar conservation law with a pointwise discretisation of the source term. In particular, three-point monotone conservative schemes (homogeneous problem) such as Godunov and Engquist-Osher can be used. (The Roe scheme is not a monotone scheme.) Both Lorenz [54] and MacDonald [59] studied the problem of existence and uniqueness of the solution (when $\epsilon \geq 0$ and $\Delta x > 0$) of the system of difference equations arising from using conservative monotone approximations of the spatial derivative. Lorenz's proof is based on the fact that the system of equations forms a M-function and MacDonald used the contraction mapping theorem which yields a practical algorithm for computing solutions of the system of difference equations. Moreover, it can be proved that to approximate the solution of the reduced problem ($\epsilon = 0$) in practice, it is not necessary to carry out the limit process on ϵ but simply to solve the problem resulting from setting $\epsilon = 0$ and to be concerned only with the limit as Δx vanishes.

A result, by Lorenz [54], on existence and uniqueness of solutions of the system of difference equations, is also adapted by MacDonald to hold only for positive solutions. The assumptions are the consistency of the numerical flux function, the monotonocity of the time dependent scheme which includes an appropriate *CFL-condition* and also some conditions on the numerical flux, namely, being Lipschitz continuous and non-increasing in its first argument and non-decreasing in its second argument.

In Section 3.3 we discuss the possible modification of similar theory to study the steady scalar problems presented in Chapter 2, obtained by reducing systems of conservation laws with source terms in the steady case. The focus is problems arising from breadth or width variation which lead to flux functions of the form $f(x, w)$.

3.3 Possible Application of Vanishing Viscosity Theory to Steady Problems with Breadth Variation

As described in Section 3.2, it is possible to use the vanishing viscosity theory to prove results (in the vanishing viscosity limit) about existence and uniqueness of entropy satisfying solutions of the limit problem in certain cases. Furthermore, the theory of monotone numerical schemes can help also in proving that the system of finite difference equations obtained from a conservative discretisation of the spatial derivative, converges (it may also provide a convergence rate estimate).

MacDonald showed that it is possible to modify Lorenz' theory to hold for the case of the steady Saint-Venant problem with a very general source term (that includes breadth and bed slope variation and friction terms), for prismatic channels and with a restriction of w (depth) being positive.

The steady scalar equations we study in this thesis have a form $f(x, w)$ which is not contemplated in either the work of MacDonald [59] and Lorenz [54, 55, 56], although a paper [57] referred by MacDonald is relevant. In that paper the condition $b_w \geq \delta > 0$ is replaced by the condition $b_w - |f_{xw}| \geq \delta > 0$ and since the solutions we are looking at are positive, it might be expected to require, as MacDonald did, a slightly different condition of the form $D_w - |F(x, w)| > 0$. Despite this, the introduction of the x-dependence on F (or f) for the case of the general form of the steady Saint-Venant equations (with breadth variation, bed slope and friction terms) raises new questions since, for example, there now will be a critical function at each x-cross section, hence depending on x. Nevertheless, we solved numerically some test problems for the Saint-Venant equations using upwind schemes based in Engquist-Osher and Roe schemes with a time stepping iteration.

The gas problem, raises different questions. One is that the condition on b_w can be violated by a quasi-one dimensional duct flow, for example in the case of a converging-diverging duct. In [56], Lorenz discusses the use of the Engquist-Osher and Godunov schemes in the case of one-dimensional duct flow described by a steady differential equation (in the variable u) whose flux function does not depend explicitly on the area variation. That equation was studied previously by Stephens and Shubin [87]. Furthermore, it may be possible to use a different reduction of the steady Euler equations that does not depend explicitly on K but still maintains the features leading to different types of solution (supersonic flow, sonic flow, subsonic flow) (see [87, 56]). Also, it may be possible in this case to remove the x dependence from the flux function (see [56]). Nevertheless, one should keep in mind that not all transformations will render a equation with the necessary physical features. Although equivalent for smooth solutions, different conservative formulations may not make any physical sense and in the presence of shock waves may produce wrong shock speeds and the wrong solutions (see Chapter 4).

Our choice of reduction in the case of the steady Euler equations is dependent on K and that brings extra difficulties since K, although being constant for isentropic flow, has a jump if a normal shock occurs, which in a nozzle occurs in the diverging section. Nevertheless, the source terms considered in this case are just width variation (no friction) and do not seem as complicated as the ones studied by MacDonald, although they depend on K and this may be relevant.

In Chapter 4 we present some theory on conservative schemes (homogeneous problem) and describe the upwind schemes of Engquist-Osher and Roe. In Chapter 5 we study conservation laws with source terms and modify the schemes studied in the previous chapter to include source terms.

Chapter 4

Theory on the Upwind Schemes of Engquist-Osher and Roe

In this chapter we describe some of the background theory for conservative first-order numerical schemes, for the case (2.2) where the function depends only on the conserved variables and the case (2.1) where the flux function depends also on the space variable independently.

Many numerical conservative schemes can be obtained for different choices of the numerical flux function. We are particularly interested in applying the upwind schemes of Engquist-Osher and Roe, which are Godunov-type methods, to a pseudo-time scalar PDE obtained from reducing steady systems of hyperbolic PDEs (as shown in Chapter 2). The theory will be presented in some cases for the general case of a system of hyperbolic PDEs and also for the scalar case.

Since the upwind schemes of Engquist-Osher and Roe are Godunov-type methods we will present firstly the latter scheme in some detail and then proceed to explain its relation to the former schemes. A review on the class of numerical schemes known as Godunov methods is given in [91].

In Section 4.1 we present the general concepts needed for the application of first-order conservative schemes to the hyperbolic problems described in Chapter 2. The cases of the flux function depending only on the conserved variable and also of the flux function depending on both the conservative variable and x are discussed in Sections 4.2 and 4.3. The upwind schemes of Godunov, Engquist-Osher and Roe are described in both cases but the details of the algorithms used are delayed until Chapter 6.

The splitting into two terms of the x-derivative of the flux function is studied in Section 4.4. There we discuss the possibility of obtaining a conservative scheme starting from a quasi-linear (nonconservative) form of the equations.

4.1 Background on Conservative Methods

For homogeneous systems of conservation laws it is known that two systems of conservation laws that are equivalent for smooth solutions do not necessarily remain so for weak solutions. In physical applications it is clear which form to choose: the form of the equations that comes directly from the integral form of the physical conservation law. This leads to *conservative* numerical schemes which give an approximation to shocks in a “correct” location ([49]) and at the right speed .

One way to derive numerical methods in conservative form is to use finite difference discretisations starting from the conservative form of the conservation law. Another way is to start from a quasilinear form (e.g., see [49, 5]).

For smooth solutions it is known that consistency and stability imply convergence. But these conditions are not sufficient in the case of weak solutions. Convergence (if it exists) to weak solutions satisfying jump conditions is guaranteed if the (consistent) numerical scheme used is in conservation form (*Lax-Wendroff theorem* [46, 49]), although this is not a guarantee that only the schemes in conservation form can converge to the correct weak solution. Schemes that are conservative and stable in the presence of discontinuities are known as *shock capturing schemes*. Furthermore, if the numerical scheme satisfies an additional condition known as *entropy condition* convergence is guaranteed to the unique physically relevant solution (Harten [34]). Examples of conservative schemes generating numerical solutions violating the entropy condition are given e.g. in Leveque [49] or Wesseling [101].

Nevertheless, as Toro [95] has pointed out, since Hou and LeFloch [41] proved that, in the case of a shock, if a scheme not written in conservative form converges it will converge to the solution of a new conservation law with a source term, this is another argument in support of using a conservative scheme when approximating discontinuities (or near a discontinuity). In [41] Hou and LeFloch derive the equation which the nonconservative schemes approximate and show that a local correction of any high-order accurate scheme in nonconservative form can be used to ensure its convergence to the correct solution. Actually, they used an hybrid scheme that switches to a conservative scheme near a point of discontinuity of the solution. The implementation of nonconservative schemes which fit the shock waves by explicitly computed discontinuities, may incorporate upwinding (of source terms as well) very cheaply [77]. Some references on using adaptive primitive-conservative schemes are given in Toro [95].

Moreover, the differential equations under study might include source terms. Examples of conservation laws with source terms were given in Chapter 2. The inclusion of source terms, which might have a dominant effect in the nonhomogeneous problem, raises some questions about how to discretise these terms adequately. The idea is to use the underlying physical (integral) conservation law to extend the notion of a conservative scheme to include source terms by properly approximating those source terms (see, e.g. Burguete and Garcia Navarro [5]). The emphasis of the work of this thesis is also on how to discretise these source terms when a conservative finite volume scheme (based on Roe and Engquist-Osher schemes) is adopted to approximate the flux terms.

The numerical techniques used to approximate the source terms are a *pointwise approach* and an *upwind approach*, the latter involving the upwinding of an average value of the source term and can be seen as a more physical approach. The discretisation of the source terms and the extension of conservative schemes to the nonhomogeneous case will be explained in more detail in the next chapter (Chapter 5).

In addition, the flux function depends on the conserved variables and might also vary spatially. This spatial variation of the flux introduces new difficulties, namely, in how to express this variation at a discrete level.

In Chapter 5 we will derive numerical schemes in conservation form for the nonhomogenous sytems of conservation laws, starting from both the conservation and nonconservation form of the PDE. Now we proceed to study in more detail the homogeneous case, studying both the case of a flux function depending on the conserved variable and the case where the flux function depends also on the space variable.

4.2 Flux Function of the Form $\mathbf{F}(\mathbf{U})$

In this section the notion of a conservative numerical scheme is introduced and the upwind schemes of Godunov, Roe and Engquist-Osher are presented.

4.2.1 Theory on Conservative Methods

Consider a homogeneous system of conservation laws of the general form

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = \mathbf{0}. \tag{4.1}$$

We shall consider a uniform grid in (x, t) space with Δx and Δt denoting the grid spacing in space and time, respectively. We denote by $\mathbf{U}_k^n$ an approximation of $\mathbf{U}(x_k, t^n)$ at the point $(x_k = k\Delta x, t^n = n\Delta t)$.

Integrating the conservation law (4.1) over the rectangle $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [t^n, t^{n+1}]$ we obtain the *integral form of the conservation law*, i.e.

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} [\mathbf{U}]_{t^n}^{t^{n+1}} dx + \int_{t^n}^{t^{n+1}} [\mathbf{F}(\mathbf{U})]_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} dt = \mathbf{0}, \tag{4.2}$$

or

$$\begin{aligned}
&\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{U}(x, t^{n+1}) dx = \\
&\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{U}(x, t^n) dx - \left[\int_{t^n}^{t^{n+1}} \mathbf{F}(\mathbf{U}(x_{j+\frac{1}{2}}, t)) dt - \int_{t^n}^{t^{n+1}} \mathbf{F}(\mathbf{U}(x_{j-\frac{1}{2}}, t)) dt \right].
\end{aligned} \tag{4.3}$$

From the integral form (4.3), introducing some integral averages, it is possible to derive the formula which constitutes the basis of *conservative numerical methods* (see, e.g., [95, 44]). Indeed, if we consider the integral cell average in space of $\mathbf{U}(x, t^n)$ and $\mathbf{U}(x, t^{n+1})$ over the interval $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ we have

$$\mathbf{U}_j^n = \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{U}(x, t^n) dx \tag{4.4}$$

and similarly

$$\mathbf{U}_j^{n+1} = \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{U}(x, t^{n+1}) dx. \tag{4.5}$$

We also consider time integral averages of the flux function $\mathbf{F}(\mathbf{U}(x, t))$ at positions $x = x_{j-\frac{1}{2}}$ and $x = x_{j+\frac{1}{2}}$, namely

$$\mathbf{F}^*_{j-\frac{1}{2}} = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} \mathbf{F}(\mathbf{U}(x_{j-\frac{1}{2}}, t)) dt, \tag{4.6}$$

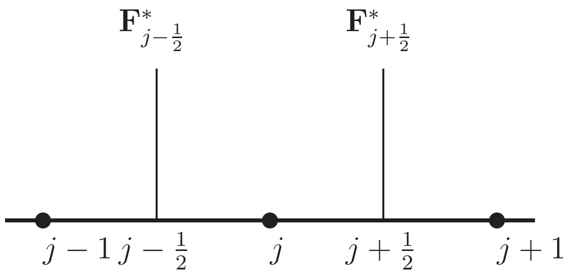
$\mathbf{F}^*_{j-\frac{1}{2}}$
$\mathbf{F}^*_{j+\frac{1}{2}}$
$j-1$
$j-\frac{1}{2}$
j
$j+\frac{1}{2}$
$j+1$

Figure 4.1: Numerical fluxes in j^{th}-cell

and similarly,

$$\mathbf{F}^*_{j+\frac{1}{2}} = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \mathbf{F}(\mathbf{U}(x_{j+\frac{1}{2}}, t))dt. \tag{4.7}$$

With these definitions, we can write equation (4.3) in the form

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\left(\mathbf{F}^*_{j+\frac{1}{2}} - \mathbf{F}^*_{j-\frac{1}{2}}\right). \tag{4.8}$$

The advantage of defining $\mathbf{U}_j^n$ as an integral average, rather than an approximation to the state $\mathbf{U}(i\Delta x, n\Delta t)$, is that equation (4.8) can be regarded as an integral law instead of a differential law. Methods based in this averaging technique are called *Finite Volume Schemes*.

When using a uniform grid, our spatial domain is split in cells of the form $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ with $j = 1, 2, \ldots, N-1$ (see Fig. 4.1) and $\mathbf{F}^*_{j+\frac{1}{2}}$ is called the *numerical flux function* (corresponding to the intercell boundary $x = x_{j+\frac{1}{2}}$) and is continuous and specified differently according to different numerical methods.

The numerical scheme is said to be in *conservation form* (or *conservative*) if it is written in the form (4.8). A discussion on the conservation form can be found in [47] and in later references like [49, 29].

These conservative methods satisfy a *telescopic property* expressing a more global form of conservation. Indeed, in the discrete case, if we sum up both sides of equation (4.8), the value of $\mathbf{F}^*_{j+\frac{1}{2}}$ used to update $\mathbf{U}_j^n$ when added with the value of $\mathbf{F}^*_{j-\frac{1}{2}}$ used to update $\mathbf{U}_{j+1}^n$ cancels out, leaving only the fluxes at the extreme cell boundaries, i.e.

$$\sum_{j=1}^{N-1} \mathbf{U}_j^{n+1} = \sum_{j=1}^{N-1} \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\left(\mathbf{F}^*_{N-\frac{1}{2}} - \mathbf{F}^*_{\frac{1}{2}}\right) \tag{4.9}$$

with $\mathbf{F}^*_{N-\frac{1}{2}}$ and $\mathbf{F}^*_{\frac{1}{2}}$ being, respectively, the rightmost and leftmost intercell boundaries fluxes. An advantage of equation (4.9) is the correct description of the integral laws (4.2) (see Roe [77]).

In general, the numerical flux function is a Lipschitz continuous function that can be written in the form (e.g. see [49, 95])

$$\mathbf{F}^*_{j+\frac{1}{2}} = \mathbf{F}^*_{j+\frac{1}{2}}(\mathbf{U}_{j-p_L}, \ldots, \mathbf{U}_{j-p_R}) \tag{4.10}$$

where p_L and p_R depend on the particular choice of the numerical flux.

For explicit methods, the approximations of the conserved variables are taken from the previous iteration at time level t^n.

The scheme is said to be *consistent* with the system of differential equations (4.1) if the numerical flux function computed at constant values coincides with the value of the flux function computed at those values, i.e. if $\mathbf{U} = \hat{\mathbf{U}}$, say, then we expect to have (see equation (4.10))

$$\mathbf{F}^*_{j+\frac{1}{2}}(\hat{\mathbf{U}}, \ldots, \hat{\mathbf{U}}) = \mathbf{F}(\hat{\mathbf{U}}). \tag{4.11}$$

Hence, the discretisation error (assuming that F is smooth) goes to zero when Δx goes to zero.

It is worth remarking that the scheme given by equation (4.8) can be thought of as an approximation of the original system by using an explicit Euler scheme for the time discretisation together with a finite volume method for the space discretisation.

Numerical conservative schemes have been obtained for many different choices of the numerical flux function. In this thesis we are concerned with the Godunov, Engquist-Osher and Roe schemes.

4.2.2 The Godunov Method

The Godunov method [30] is a first-order upwind method that comes from the observation that the numerical solution $\mathbf{U}^n_j$ satisfies the integral form of the conservation law (4.3) exactly if the averages (4.4) and (4.6) hold. The intercell numerical fluxes $\mathbf{F}^*_{j+\frac{1}{2}}$ are computed by using solutions of local Riemann problems. A Riemann problem consists of solving the conservation law for a single jump, i.e solving the system

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = \mathbf{0} \quad \text{on} \quad I\!R \times [t^n, t^{n+1}] \tag{4.12}$$

together with piecewise initial data of the form

$$\mathbf{U}(x,t^n) = \begin{cases} \mathbf{U}_j^n & x < x_{j+\frac{1}{2}} \\ \mathbf{U}_{j+1}^n & x > x_{j+\frac{1}{2}} \end{cases} . \tag{4.13}$$

The solution of the Riemann problem with left data state $\mathbf{U}_j^n$ and right data state $\mathbf{U}_{j+1}^n$ is denoted by $\mathbf{U}_{j+\frac{1}{2}}^n$.

Typically the solution of the Riemann problem in the case of systems is composed of m waves (m is the size of the system). Riemann solutions have been found for certain well known systems of conservation laws. For example, in the case of the Saint-Venant equations a reference is [100] and for the Euler equations some references are [49], [86] and [94].

The method proceeds in the following manner: at a time level n we use the numerical solution $\mathbf{U}^n$ to build a piecewise constant function $\tilde{\mathbf{U}}^n(x,t^n)$ which is equal to $\mathbf{U}_j^n$, given by equation (4.4), on the grid cell $(x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}})$ (see Fig. 4.2). This piecewise function is not constant over $t^n \leq t < t^{n+1}$ but it is used as the initial data of the conservation law, leading on to a sequence of Riemann problems to be solved, each one at a grid interface.

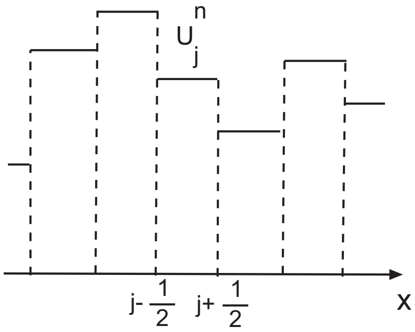

Figure 4.2: The piecewise constant distribution of data at time level n

The sequence of Riemann problems can be solved exactly at each cell interface yielding a solution $\tilde{\mathbf{U}}^n(x,t)$ for $t^n \leq t \leq t^{n+1}$ which is obtained by piecing together these Riemann solutions on the condition that the waves from neighbouring problems do not interact. (To prevent this from happening, for a given Δx one has to limit the size of the time step Δt.) Then we use the scheme in conservation form (4.8) to compute the cell averages at the next time step, i.e. $\mathbf{U}^{n+1}$ is computed by averaging the exact solution at time t^{n+1} (note that we need to compute the numerical flux function). The resulting value is in turn used to define a new piecewise constant function, and the process repeats.

Actually, the integration needed to compute the numerical flux function at a cell interface, for example $\mathbf{F}^*_{j+\frac{1}{2}}$ given by (4.7), is trivial. The integrand function is constant at the point $x_{j+\frac{1}{2}}$ over the interval (t^n, t^{n+1}) and hence does not depend on the whole Riemann solution but only on the flux of the state at $x_{j+\frac{1}{2}}$. This happens because the solution of the Riemann problem at $x_{j+\frac{1}{2}}$ is a similarity solution in x/t, i.e. is of the form $\mathbf{U}(x,t) = \mathbf{W}(x/t)$ (see [49]). So we can rewrite the numerical flux function as

$$\mathbf{F}^*_{j+\frac{1}{2}} = \mathbf{F}(\mathbf{W}^n_{j+\frac{1}{2}}(0)). \tag{4.14}$$

Likewise $\mathbf{F}^*_{j-\frac{1}{2}} = \mathbf{F}(\mathbf{W}^n_{j-\frac{1}{2}}(0))$.

Since the details of the Riemann problem solutions are not important within the cell for the calculation of the flux (what is needed is a cell average), they may be allowed to interact provided the interaction is contained within a grid cell, that is, the solution at $x_{j+\frac{1}{2}}$ does not influence the state at $x_{j-\frac{1}{2}}$ and vice-versa. Seeing that the wave speeds are bounded by the eigenvalues of the Jacobian of the system, $\frac{\partial \mathbf{F}}{\partial \mathbf{U}}$, and that the neighbouring Riemann problems are at a distance Δx away, the Riemann solution will be constant over $[t^n, t^{n+1}]$ by choosing Δt satisfying

$$\left| \frac{\Delta t}{\Delta x} \lambda_k(\mathbf{U}^n_j) \right| \leq 1 \tag{4.15}$$

for all eigenvalues λ_k at each $\mathbf{U}^n_j$. Alternatively, if in a particular problem we consider the maximum of the left-hand side of (4.15), that quantity is called the *Courant number* or the *CFL*[1] *coefficient* and its value lies between 0 and 1.

After obtaining this solution over the interval $[t^n, t^{n+1}]$ the new updated solution U^{n+1}_j is computed by averaging the exact solution at time t^{n+1} and the resulting value is in turn used to define a new piecewise constant function, and the process repeats.

In the next sections we study in more detail the scalar case.

4.2.3 The Scalar Problem

The scalar problem with smooth initial data

Consider a scalar Cauchy problem (or initial value problem (IVP)) with smooth initial data where the flux function depends only on the conserved variable w, i.e.

$$w_t + F(w)_x = 0 \tag{4.16}$$

[1] Courant, Friedrichs, Lewy [12] whom firstly recognized the importance of a condition like (4.15)

with

$$w(x,0) = w_0(x). \tag{4.17}$$

A solution of the IVP can be constructed (for small t) by following characteristic curves $x = x(t)$. The characteristics satisfy

$$\frac{dx}{dt} = F'(w) = \lambda(w), \quad x(0) = x_0. \tag{4.18}$$

and w is constant along those characteristic curves because if we consider both w and x as functions of t, the total derivative of w along the curve $x(t)$ we obtain

$$\frac{dw}{dt} = w_t + \frac{dw}{dx}\lambda(w) = 0. \tag{4.19}$$

Additionally, since w is constant on each characteristic, the slope $x'(t)$ as a function of w only is also constant. Hence the characteristics are straight lines with the slope determined by the initial data. If the initial data is smooth, we can solve the IVP (4.18) yielding

$$x = x_0 + w(x_0,t)t \tag{4.20}$$

and then

$$w(x,t) = w(x_0,0). \tag{4.21}$$

<u>The Riemann problem</u>

In the one-dimensional case the Riemann problem is an IVP where initial data is in the form of a jump, so no longer smooth. If the flux function depends only on w and is convex ($F'' > 0$), the solution of the Riemann problem (4.12)-(4.13) is either

(i). a *shock* (discontinuity) propagating with speed s, i.e.

$$w(x,t^{n+1}) = \begin{cases} w_j^n & x < x_{j+\frac{1}{2}} + st \\ w_{j+1}^n & x > x_{j+\frac{1}{2}} + st \end{cases} \tag{4.22}$$

where the *shock speed* s is given by the *jump condition*

$$s = \frac{F(w_{j+1}^n) - F(w_j^n)}{w_{j+1}^n - w_j^n}; \tag{4.23}$$

in addition we require that the entropy condition

$$F'(w_j^n) > s > F'(w_{j+1}^n) \tag{4.24}$$

holds (i.e., for convex F, the characteristics always enter in a shock and never emanate from it) or (ii). an *expansion wave* given by

$$w(x,t^{n+1}) = \begin{cases} w_j^n & x < F'(w_j^n)t \\ w_j^n + \frac{w_{j+1}^n - w_{j+1}^n}{F'(w_{j+1}^n) - F'(w_j^n)} \left(\frac{x}{t} - F'(w_j^n)\right) & F'(w_j^n)t < x < F'(w_{j+1}^n)t \\ w_{j+1}^n & x > F'(w_{j+1}^n)t \end{cases} \tag{4.25}$$

(see, e.g. [90] or [86]).

If F is concave ($F'' < 0$) similar conclusions can be drawn. In the cases of F being convex or concave the discontinuity in the Riemann solution is separated either by a shock or by a fan, but not by both. The case of F being neither convex or concave is more complicated but a solution can still be found and it may involve both a shock and a rarefaction wave (e.g. see [72, 101]).

An expression for the Godunov numerical flux which works even with nonconvex fluxes and that leads to entropy satisfying solutions of the Riemann problem is given by

$$F^*_{j+\frac{1}{2}} = F^*(w_{j+1}, w_j) = \begin{cases} \max_{w_{j+1} \leq w \leq w_j} F(w) & \text{for} \quad w_{j+1} < w_j \\ \min_{w_j \leq w \leq w_{j+1}} F(w) & \text{for} \quad w_{j+1} \geq w_j \end{cases} \tag{4.26}$$

(see LeVeque [49]).

As we have seen, Godunov's method solves the Riemann problem at each cell interface, exactly. Since the use of a known exact solution can be computationally expensive, it can be more efficient to use only an approximation to the Riemann solution. That is the idea behind the *approximate Riemann solvers* methods which solve the Riemann problem approximately and use the resulting value to compute the numerical flux. We will look at two approximate Riemann solvers for a scalar problem, namely the Roe scheme and the Engquist-Osher scheme.

Although it is possible and sometimes useful to think of how these methods work in the case of a system of conservation laws, the steady case for particular sorts of systems of conservation laws discussed in Chapter 2 allows us to reduce the system to a singular scalar ODE which can be thought of as the steady case of an unsteady scalar equation (see also Chapter 3).

In the following subsections the approximate Riemann solvers of Roe and Engquist-Osher are presented.

4.2.4 The Upwind Scheme of Roe

One of the simplest and most used Riemann solvers is that due to Roe [75] and involves a linearisation of the system of conservation laws (2.2) written in quasi-linear form (2.3). The scheme was introduced for systems of hyperbolic conservation laws but can be interpreted also for a scalar conservation law. We first introduce its general description and then proceed to apply it to a scalar conservation law and to the system of two ODEs of Section 2.3.5 obtained by reducing a different form of the Euler equations.

Consider the system of conservation laws in quasi-linear form (2.3) with $\mathbf{D}(x, \mathbf{U}) = 0$, i.e.

$$\mathbf{U}_t + J\mathbf{U}_x = \mathbf{0}. \tag{4.27}$$

Roe's approach linearises the system (4.27) by replacing the Jacobian matrix J by constant matrices in each interval. That is, by replacing J in each interval (x_j, x_{j+1}) by a matrix $\tilde{J} = \tilde{J}(\mathbf{U}_j, \mathbf{U}_{j+1})$. With this approach, the original Riemann problem is substituted by an approximate linear problem, with the same initial data, that is then solved exactly. The solution of the linear Riemann problems can be found in [49, 94], for example. Their solution contains only discontinuities and not expansion fans, leading to non entropy satisfying solutions. Different entropy fixes for Roe's scheme have been proposed, though. See for example, [79] and [37] (the latter is also outlined in [49]).

For any two adjacent states $\mathbf{U}_L$ and $\mathbf{U}_R$ the matrices $\tilde{J} = \tilde{J}(\mathbf{U}_L, \mathbf{U}_R)$ should satisfy:

(i). $\tilde{J}(\mathbf{U}_L, \mathbf{U}_R)$ is diagonalisable (Hyperbolicity)

(ii). $\tilde{J}(\mathbf{U}_L, \mathbf{U}_R) \to J(\mathbf{U})$ as $\mathbf{U}_L, \mathbf{U}_R \to \mathbf{U}$ (Consistency)

(iii). (Conservation)

$$\Delta\mathbf{F} = \mathbf{F}(\mathbf{U}_R) - \mathbf{F}(\mathbf{U}_L) = \tilde{J}(\mathbf{U}_L, \mathbf{U}_R)(\mathbf{U}_R - \mathbf{U}_L) \tag{4.28}$$

The first two conditions are satisfied if $\tilde{J}$ is taken to be the Jacobian evaluated at an averaged state $\tilde{\mathbf{U}}$, i.e. $\tilde{J}(\mathbf{U}_L, \mathbf{U}_R) = \tilde{J}(\tilde{\mathbf{U}})$. In general, an arithmetic average does not satisfy the last condition (see [49, 91] and also [39]) and a particular kind of geometric average is often used instead. This geometric average can be written in the form of an arithmetic mean of a parameter vector (see [75, 79, 27]). In [75] Roe showed how to build this matrix for the Euler equations. Later Roe and Pike [79] presented an approach where the explicit construction of the matrix $\tilde{J}$ can be avoided. The application of the Roe scheme to the Shallow Water Equations can be found in [25].

Roe's scheme for the scalar ODE

For the case of a scalar conservation law of the form (4.16) the third condition determines uniquely $\tilde{\lambda} = \tilde{J}(w_L, w_R)$ as

$$\tilde{\lambda} = \frac{F(w_R) - F(w_L)}{w_R - w_L}. \tag{4.29}$$

Hence, the linearised problem is the *scalar advection equation*

$$w_t + \tilde{\lambda} w_x = 0 \tag{4.30}$$

whose Riemann problem solution is a moving shock (a jump from w_L to w_R with speed $\tilde{\lambda}$). Since the average speed (4.29) is the (Rankine-Hugoniot) jump condition this "approximate" Riemann solution is a weak solution which may not satisfy the entropy condition.

The numerical flux can be written in the form

$$F^*_{j+\frac{1}{2}}(w_{j+1}, w_j) = \frac{1}{2}\left(F(w_j) + F(w_{j+1})\right) - \frac{1}{2}|\tilde{\lambda}_{j+\frac{1}{2}}|(w_{j+1} - w_j) \tag{4.31}$$

where

$$\tilde{\lambda}_{j+\frac{1}{2}}(w_j, w_{j+1}) = \begin{cases} \frac{F(w_{j+1}) - F(w_j)}{w_{j+1} - w_j} & w_{j+1} \neq w_j \\ F'(w_j) & w_{j+1} = w_j \end{cases}. \tag{4.32}$$

The numerical flux can also be written in the form

$$F^*_{j+\frac{1}{2}}(w_{j+1}, w_j) = \begin{cases} F(w_j) & \tilde{\lambda}_{j+\frac{1}{2}} \geq 0 \\ F(w_{j+1}) & \tilde{\lambda}_{j+\frac{1}{2}} < 0 \end{cases}. \tag{4.33}$$

The scheme is also known as the *first-order upwind scheme* (FOU). We call it the *Roe scheme* although, as referred to in [59], Murman and Cole [66, 67] came up with a similar scheme earlier.

The term *upwind* or *upstream* refers to the direction from which characteristic information propagates, with the grid points used in the spatial finite differences discretisation chosen to be the ones on the side from which the information ("wind") flows.

A disadvantage of the Roe scheme is that the Riemann solution consists only of discontinuities with no rarefaction waves, which can lead to entropy-violating solutions.

In both the scalar conservation law and the system of conservation laws, a Riemann solution that does not satisfy the entropy condition can occur in the case of a sonic rarefaction wave. For the scalar case this corresponds to $F'(w_L) < 0 < F'(w_R)$. There are different ways of modifying the Roe scheme to obtain entropy satisfying solutions. The *sonic entropy fix* discussed by Harten and Hyman [37] and outlined in LeVeque [49] is the one used in the different Roe algorithms we use. This entropy fix substitutes the single jump propagating with speed $\tilde{\lambda}$ with two jumps propagating with speeds $F'(w_L)$ and $F'(w_R)$ separated by the state

$$w_m = w_L + \frac{F'(w_R) - \tilde{\lambda}}{F'(w_R) - F'(w_L)}(w_R - w_L). \tag{4.34}$$

The approximate Riemann solution (see 4.22), which can be written as a similarity solution as

$$\hat{W}(x/t) = \begin{cases} w_L & x/t < \tilde{\lambda} \\ w_R & x/t > \tilde{\lambda} \end{cases} \tag{4.35}$$

is thus substituted by

$$\hat{W}(x/t) = \begin{cases} w_L & x/t < F'(w_L) \\ w_m & F'(w_L) < x/t < F'(w_R) \\ w_R & x/t > F'(w_R) \end{cases} \tag{4.36}$$

(see [49]).

Geometrically the state w_m is the abscissa of the point of intersection of the two tangent lines to the curve $F(w)$ at the points w_L and w_R. Its coordinate gives the numerical flux at this intermediate state (see Fig. 4.3). Hence, when $F'(w_L) < 0 < F'(w_R)$ the numerical flux is given by

$$F^*(w_R, w_L) = F(w_L) + F'(w_L)\frac{F'(w_R) - \tilde{\lambda}}{F'(w_R) - F'(w_L)}(w_R - w_L). \tag{4.37}$$

Roe's scheme for the reduced Euler system of two ODEs

For the system of two ODEs (2.110) derived in Section 2.3.5 it is possible to derive the Roe's matrix and also Roe's averages in a way similar to the one in [75] (see also [101]). Following Roe [75], we are going to express everything in terms of a parameter vector

$$\mathbf{z} = \sqrt{\rho}\begin{pmatrix} 1 \\ H \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}. \tag{4.38}$$

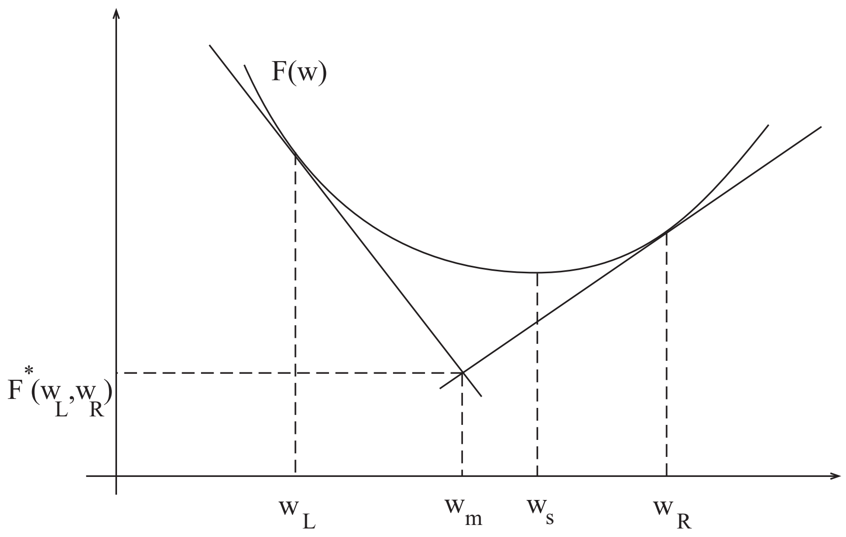

Figure 4.3: The geometric interpretation of the entropy fix for the Roe method in the sonic rarefaction case

By using equations (2.48), (2.50) and the equation of state (2.59) one can get the equations

$$E = \frac{1}{\gamma}\rho H + \frac{\gamma-1}{2\gamma}\rho u^2$$

and

$$p = \frac{\gamma-1}{\gamma}\left(\rho H - \frac{1}{2}\rho u^2\right)$$

which are useful to express **w** and **F** in terms of the vector **z**. Another equation which is useful is (2.86).

Thus, the vector of variables (2.113) can be written as

$$\mathbf{w} = \begin{pmatrix} z_1^2 \\ \frac{1}{\gamma}z_1z_2 + \frac{\gamma-1}{2\gamma}\frac{m^2}{z_1^2} \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{4.39}$$

and the flux function can also be written as

$$\mathbf{F} = \begin{pmatrix} \rho u^2 + p \\ u(E+p) \end{pmatrix} = \begin{pmatrix} \frac{\gamma-1}{\gamma}z_1z_2 + \frac{\gamma+1}{2\gamma}\frac{m^2}{z_1^2} \\ m\frac{z_2}{z_1} \end{pmatrix}. \tag{4.40}$$

Defining the difference

$$\Delta a = a_R - a_L$$

and the average

$$\bar{a} = \frac{a_L + a_R}{2}$$

we have

$$\Delta(a+b) = \Delta a + \Delta b$$

$$\Delta(ab) = \bar{a}\Delta b + \bar{b}\Delta a$$

$$\Delta\left(\frac{a}{b}\right) = \frac{\bar{b}\Delta a - \bar{a}\Delta b}{(b_L b_R)^2}.$$

Using these identities, it is easy easy to verify that

$$\Delta\mathbf{w} = \bar{B}\Delta\mathbf{z} \tag{4.41}$$

with

$$\bar{B} = \begin{pmatrix} 2\bar{z}_1 & 0 \\ \frac{1}{\gamma}\bar{z}_2 - \frac{\gamma-1}{\gamma}m^2\frac{\bar{z}_1}{(z_1^L z_1^R)^2} & \frac{1}{\gamma}\bar{z}_1 \end{pmatrix} \tag{4.42}$$

and

$$\Delta\mathbf{F} = \bar{C}\Delta\mathbf{z} \tag{4.43}$$

with

$$\bar{C} = \begin{pmatrix} \frac{\gamma-1}{\gamma}\bar{z}_2 - \frac{\gamma+1}{\gamma}m^2\frac{\bar{z}_1}{(z_1^L z_1^R)^2} & \frac{\gamma-1}{\gamma}\bar{z}_1 \\ -m\frac{\bar{z}_2}{z_1^L z_1^R} & m\frac{\bar{z}_1}{z_1^L z_1^R} \end{pmatrix}. \tag{4.44}$$

Moreover, using the inverse of the matrix $\bar{B}$

$$\bar{B}^{-1} = \begin{pmatrix} \frac{1}{2\bar{z}_1} & 0 \\ -\frac{1}{2}\frac{\bar{z}_2}{\bar{z}_1^2} + \frac{\gamma-1}{2}m^2\frac{1}{\bar{z}_1(z_1^L z_1^R)^2} & \frac{\gamma}{\bar{z}_1} \end{pmatrix} \tag{4.45}$$

it can be shown that

$$\Delta\mathbf{F} = \bar{C}\Delta\mathbf{z} = \bar{C}(\bar{B}^{-1}\Delta\mathbf{w}) \tag{4.46}$$

Hence, we have seen that condition (iii) in (4.28) is satisfied and that Roe's matrix is given by

$$\tilde{J}(\mathbf{w}_L, \mathbf{w}_R) = \bar{C}\bar{B}^{-1} = \begin{pmatrix} \frac{\gamma-3}{2}m^2\frac{1}{\tilde{\rho}^2} & \gamma-1 \\ -m\tilde{H}\frac{1}{\tilde{\rho}} + \frac{\gamma-1}{2}m^3\frac{1}{\tilde{\rho}^3} & \gamma m\frac{1}{\tilde{\rho}} \end{pmatrix} \tag{4.47}$$

where

$$\tilde{\rho} = z_1^L z_1^R = \sqrt{\rho_l \rho_R} \tag{4.48}$$

$$\tilde{H} = \frac{\bar{z}_2}{\bar{z}_1} = \frac{\sqrt{\rho_L}H_L + \sqrt{\rho_R}H_R}{\sqrt{\rho_L} + \sqrt{\rho_R}}. \tag{4.49}$$

The equations (4.48) and (4.49) are called the *Roe averages*. From the average (4.48) we can define the average

$$\tilde{u} = \frac{m}{\tilde{\rho}} \tag{4.50}$$

which can be thought of as a particular case of the well-know Roe average for the Euler system of three equations

$$\tilde{u} = \frac{\sqrt{\rho_L}u_L + \sqrt{\rho_R}u_R}{\sqrt{\rho_L} + \sqrt{\rho_R}} \tag{4.51}$$

obtained by considering $u_R = m/\rho_R$ and $u_L = m/\rho_L$.

A comparison between the linearised Jacobian matrix $\tilde{J}(\mathbf{U}_L, \mathbf{U}_R)$ given by equation (4.47) and the original Jacobian matrix $J = \frac{d\mathbf{F}}{d\mathbf{w}}$ shows that

$$\tilde{J}(\mathbf{w}_L, \mathbf{w}_R) = \mathbf{F}'(\tilde{\mathbf{w}}) = \frac{d\mathbf{F}}{d\mathbf{w}}(\tilde{\mathbf{w}}). \tag{4.52}$$

Hence, conditions (i)-(iii) in (4.28) are satisfied by this linearised Jacobian matrix and Roe's (numerical) flux can be written as

$$\begin{aligned} \mathbf{F}_{j+\frac{1}{2}} &= \tilde{J}_{j+\frac{1}{2}}\tilde{\mathbf{w}} = \\ &= \frac{1}{2}\left(\mathbf{F}(\mathbf{w}_j) + \mathbf{F}(\mathbf{w}_{j+1})\right) - \frac{1}{2}\sum_{k=1}^{2} |\tilde{\mu}_k|\tilde{\alpha}_k\tilde{\mathbf{r}}_k. \end{aligned} \tag{4.53}$$

where we used the notation

$$J_{j+\frac{1}{2}} = \tilde{J}(\mathbf{w}_j, \mathbf{w}_{j+1}).$$

Because of equation (4.52), the eigenvalues $\tilde{\mu}_k$ follow immediately from equations (2.117)-(2.118), so we can write

$$\tilde{\mu}_1 = \frac{\tilde{\phi} - \sqrt{\tilde{\phi}^2 + 4\tilde{\lambda}_1\tilde{\lambda}_2\tilde{\lambda}_3}}{2} \tag{4.54}$$

$$\tilde{\mu}_2 = \frac{\tilde{\phi} + \sqrt{\tilde{\phi}^2 + 4\tilde{\lambda}_1\tilde{\lambda}_2\tilde{\lambda}_3}}{2} \tag{4.55}$$

where

$$\tilde{\phi} = \frac{\gamma - 3}{2}\tilde{u}^2 + \gamma\tilde{u} \tag{4.56}$$

$$\tilde{\lambda}_1 = \tilde{u} - \tilde{c} \tag{4.57}$$

$$\tilde{\lambda}_2 = \tilde{u} \tag{4.58}$$

$$\tilde{\lambda}_3 = \tilde{u} + \tilde{c} \tag{4.59}$$

$$\tilde{c}^2 = (\gamma - 1)\left(\tilde{H} - \frac{1}{2}\tilde{u}^2\right) \tag{4.60}$$

and $\tilde{u}$ is given by equation (4.50). The eigenvectors follow from equations (2.122)-(2.123) giving

$$\tilde{\mathbf{r}}_1 = \begin{pmatrix} \gamma - 1 \\ \tilde{\mu}_1 - \frac{\gamma-3}{2}\tilde{u}^2 \end{pmatrix} \tag{4.61}$$

and

$$\tilde{\mathbf{r}}_2 = \begin{pmatrix} \gamma - 1 \\ \tilde{\mu}_2 - \frac{\gamma-3}{2}\tilde{u}^2 \end{pmatrix}. \tag{4.62}$$

The coefficients $\tilde{\alpha}_k$ in the Roe flux (4.53) are obtained from

$$\Delta\mathbf{w} = \sum_{k+1}^{2} \tilde{\alpha}_k \tilde{\mathbf{r}}_k \tag{4.63}$$

yielding

$$\tilde{\alpha}_1 = \frac{\Delta E - \frac{\Delta\rho}{\gamma-1}\left(\tilde{\mu}_2 - \frac{\gamma-3}{2}\tilde{u}^2\right)}{\tilde{\mu}_1 - \tilde{\mu}_2} \tag{4.64}$$

$$\tilde{\alpha}_2 = \frac{-\Delta E + \frac{\Delta\rho}{\gamma-1}\left(\tilde{\mu}_1 - \frac{\gamma-3}{2}\tilde{u}^2\right)}{\tilde{\mu}_1 - \tilde{\mu}_2} \tag{4.65}$$

$$\tag{4.66}$$

4.2.5 The upwind Scheme of Engquist-Osher (scalar case)

An important Riemann solver was introduced by Engquist and Osher [17, 18]. In the case where the flux function depends only on the conserved variable w, the numerical flux function is given by

$$F^*_{j+\frac{1}{2}}(w_{j+1}, w_j) = F_-(w_{j+1}) + F_+(w_j) + F(c) \tag{4.67}$$

where the functions $F_\pm$ are given by

$$F_-(w) = \int_c^w \min_\theta\{F'(\theta), 0\}d\theta \tag{4.68}$$

$$F_+(w) = \int_c^w \max_\theta\{F'(\theta), 0\}d\theta \tag{4.69}$$

and c is arbitrary.

If F is *strictly convex* ($F'' > 0$) then we can define the functions F_+ and F_- as

$$F_-(w) = F\left(\min\{w, w_c\}\right) \tag{4.70}$$

$$F_+(w) = F\left(\max\{w, w_c\}\right) \tag{4.71}$$

where w_c is the unique sonic point, i.e. $F'(w_c) = 0$. Then we have

$$F(w) = F_-(w) + F_+(w) + F(c) \tag{4.72}$$

and also

$$|F'(w)| = F'_-(w) + F'_+(w). \tag{4.73}$$

Similar conclusions can be drawn for the case of F being *strictly concave* ($F'' < 0$). Thus, for a convex or concave flux function, the numerical source function is equivalent to the Godunov numerical flux for a rarefaction wave and only differs in the case of a shock. Furthermore, in domains where the sign of F' is constant the numerical flux function coincides with the standard first-order upwind scheme (see [59]).

4.3 Flux Function of the Form $\mathbf{F}(x, \mathbf{U})$

In this section we follow a similar path to Section 4.2 for the case where the flux function depends on the conservative variables and on the space variable as well. The notion of a conservative numerical scheme is introduced for this case. Some extensions of the upwind schemes of Godunov, Roe and Engquist-Osher are also presented.

4.3.1 Theory on Conservative Methods

If the flux function depends on both x and $\mathbf{U}$, the homogeneous system of conservation laws can be written in the form

$$\mathbf{U}_t + \mathbf{F}(x, \mathbf{U})_x = \mathbf{0}. \tag{4.74}$$

In order to model the dependence of the flux function on x we allow the numerical flux function to depend also on x and define a conservative numerical scheme for this case.

Let us consider a uniform grid in the $x - t$ space as in Section 4.2 and proceed in a similar way to define a *conservative numerical scheme* in the case where the flux function $\mathbf{F}$ also depends on x.

Thus, by integrating the conservation law (4.74) over the rectangle $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [t^n, t^{n+1}]$ we obtain the *integral form of the conservation law*, i.e.

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} [\mathbf{U}]_{t^n}^{t^{n+1}} dx + \int_{t^n}^{t^{n+1}} [\mathbf{F}(x, \mathbf{U})]_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} dt = \mathbf{0}, \tag{4.75}$$

or

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{U}(x,t^{n+1})dx = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{U}(x,t^{n})dx - \\ - \left[\int_{t^n}^{t^{n+1}} \mathbf{F}(x_{j+\frac{1}{2}}, \mathbf{U}(x_{j+\frac{1}{2}},t))dt - \int_{t^n}^{t^{n+1}} \mathbf{F}(x_{j-\frac{1}{2}}, \mathbf{U}(x_{j-\frac{1}{2}},t))dt\right]. \qquad (4.76)$$

Equation (4.76) can be written in the form

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\left(\mathbf{F}^*_{j+\frac{1}{2}} - \mathbf{F}^*_{j-\frac{1}{2}}\right). \qquad (4.77)$$

by defining $\mathbf{U}(x,t^n)$ and $\mathbf{U}(x,t^{n+1})$ as the integral averages (over the interval $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$) given by equations (4.4) and (4.5) and by introducing time integral averages of the flux function $\mathbf{F}(x,\mathbf{U}(x,t))$ at positions $x = x_{j-\frac{1}{2}}$ and $x = x_{j+\frac{1}{2}}$ of the form

$$\mathbf{F}^*_{j-\frac{1}{2}} = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \mathbf{F}(x_{j-\frac{1}{2}}, \mathbf{U}(x_{j-\frac{1}{2}},t))dt, \qquad (4.78)$$

and

$$\mathbf{F}^*_{j+\frac{1}{2}} = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \mathbf{F}(x_{j+\frac{1}{2}}, \mathbf{U}(x_{j+\frac{1}{2}},t))dt. \qquad (4.79)$$

The function $\mathbf{F}^*_{j+\frac{1}{2}}$ is called the *numerical flux function* (corresponding to the intercell boundary $x = x_{j+\frac{1}{2}}$) and is, in general, Lipschtiz continuous.

The numerical scheme is said to be in *conservation form* (or *conservative*) if it is written in form (4.77).

As we have seen the conservative methods satisfy a discrete *telescopic property* yielding an equation of the form (4.9).

In general, the numerical flux function is of the form

$$\mathbf{F}^*_{j+\frac{1}{2}} = \mathbf{F}^*_{j+\frac{1}{2}}(x_{j-q_L}, \ldots, x_{j+q_R}, \mathbf{U}_{j-p_L}, \ldots, \mathbf{U}_{j-p_R}) \qquad (4.80)$$

where q_L, q_R, p_L and p_R depend on the particular choice of the numerical flux. The choice of q_L and q_R is related to the grid points, turning out to be grid points if their values are integers.

The scheme is said to be *consistent* with the system of differential equations (4.74) if the numerical flux function computed at constant values coincides with the value of the flux function computed at those values, i.e., if $x = \hat{x}$ and $\mathbf{U} = \hat{\mathbf{U}}$, say, then we expect to have (see equation (4.80))

$$\mathbf{F}^*_{j+\frac{1}{2}}(\hat{x}, \ldots, \hat{x}, \hat{\mathbf{U}}, \ldots, \hat{\mathbf{U}}) = \mathbf{F}(\hat{x}, \hat{\mathbf{U}}). \qquad (4.81)$$

Hence, the discretisation error (assuming that F is smooth) goes to zero when Δx goes to zero.

Methods that are of particular interest are *upwind methods*, the Godunov method being particularly relevant. These methods detect the correct direction from which each characteristic propagates and often require solving Riemann problems in order to accomplish the appropriate splitting between wave propagation to the left and right.

Modifications to the Godunov, Roe and Engquist-Osher schemes to include x-dependence are discussed in Section 4.3.2.

4.3.2 The Godunov, Roe and Engquist-Osher Schemes Extended to Include x-dependence

The discretisation of the derivative of the flux function can be thought of in two ways. One way (starting from the conservative form of the hyperbolic system) is to take in account the explicit dependence of the derivative of $\mathbf{F}$ in x directly in the discretisation. The other way is to split the derivative of F by the chain rule and to treat the terms coming from partial differentiation with respect to x (with $\mathbf{U}$ fixed) as source terms, keeping the other resulting terms on the left-hand side of the system equations. This latter (indirect) approach corresponds to starting from a quasi-linear form of the system (nonconservative) and brings some questions on conservativeness forward. A discussion of both approaches is the subject of Section 4.4. The direct approach is used in [59] and the indirect approach is explored, for example, in [20, 42, 5]. Both approaches are used in [48]. The idea of altering the form of the source term (corresponding to a indirect approach) is suggested in [74].

In a direct approach we look at ways of including the x-dependence of the flux function by simply adding the argument x to any evaluation of the function F and its derivatives. The numerical flux function will also depend on x and a natural choice is to consider

$$\mathbf{F}^*_{j+\frac{1}{2}}(x_{j+\frac{1}{2}}, \mathbf{U}_{j+1}, \mathbf{U}_j) \tag{4.82}$$

owing to the integral form (4.75) and the averages (4.78)-(4.79).

In this way the conservative scheme (4.8) takes the form

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\left(\mathbf{F}^*_{j+\frac{1}{2}}(x_{j+\frac{1}{2}}, \mathbf{U}_{j+1}, \mathbf{U}_j) - \mathbf{F}^*_{j-\frac{1}{2}}(x_{j-\frac{1}{2}}, \mathbf{U}_j, \mathbf{U}_{j-1})\right). \tag{4.83}$$

A more general possibility (see [59]) is to apply a conservative method of the form

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\left(\mathbf{F}^*_{j+\frac{1}{2}}(x_{j+q}, \mathbf{U}_{j+1}, \mathbf{U}_j) - \mathbf{F}^*_{j-\frac{1}{2}}(x_{j+q-1}, \mathbf{U}_j, \mathbf{U}_{j-1})\right). \tag{4.84}$$

for any real q, where $x_{j+q} = (j+q)\Delta x$. Note that when $q = 1/2$ we get (4.83) and that a choice of q integer may be more efficient computationally, since quantities will be mainly computed at grid points. These schemes are still consistent with the conservation law.

Conservative numerical schemes can be modified to include the x-dependence in the way described above. However, this approach corresponds to take approximations of the flux function, respectively, $\mathbf{F}_{j-1}$ associated with the $j-1$th cell and $\mathbf{F}_j$ associated with the jth cell. The Riemann problem at a cell interface $x_{j-\frac{1}{2}}$ is of the form

$$\begin{aligned} \mathbf{U}_t + \mathbf{F}_{j-1}(\mathbf{U})_x = 0 \qquad &\text{if} \quad x < x_{j-\frac{1}{2}} \\ \mathbf{U}_t + \mathbf{F}_j(\mathbf{U})_x = 0 \qquad &\text{if} \quad x > x_{j-\frac{1}{2}}. \end{aligned} \tag{4.85}$$

with initial data U_{j-1} and U_j.

For example, a direct approach combined with the Engquist-Osher scheme for scalar equations is taken in [59] and [48].

Applying the direct approach to Roe scheme is not so straightforward (especially for systems). Indeed, for a flux function depending on x as well as $\mathbf{U}$, Roe's linearisation of the system in each interval (x_L, x_R), with x_L and x_R representing adjacent states, leads to an approximation of the derivative of the flux function satisfying (4.28) that is not identical to the one obtained for a function depending only on $\mathbf{U}$ (see equation (2.5)). We have

$$\Delta\mathbf{F} = \mathbf{F}(\mathbf{U}_R) - \mathbf{F}(\mathbf{U}_L) = \tilde{J}(\mathbf{U}_L, \mathbf{U}_R)(\mathbf{U}_R - \mathbf{U}_L) + \tilde{\mathbf{V}} \tag{4.86}$$

where $\tilde{J} \approx \frac{\partial \mathbf{F}}{\partial \mathbf{U}}$ is the linearised Jacobian matrix and $\tilde{\mathbf{V}} \approx \frac{\partial \mathbf{F}}{\partial x}\Delta x$ (see, e.g. [19]). In [42], Hubbard and Garcia-Navarro choose to include this last term in the numerical flux via a characteristic decomposition, presenting both fluctuation-signal and flux-based forms using the Roe scheme.

The scalar case is simpler since the linearised Jacobian matrix is just a scalar. Nevertheless, the dependence of the flux function on x still yields an extra term due to x dependence. Indeed, we have $\Delta F = \tilde{\lambda}\Delta w + \tilde{V}$ where $\tilde{\lambda} \approx \frac{\partial F}{\partial w}$ and $\tilde{V} \approx \frac{\partial F}{\partial x}\Delta x$.

In a direct approach we seek to incorporate the dependence of the numerical flux on x within the evaluation of the numerical flux. Hence we include the extra term $\tilde{V}$ in the definition of the numerical flux. In an indirect approach we include the discretisation of $\tilde{V}$ in the source terms.

By including the x-dependence on the numerical flux function (scalar case) in the way given by (4.84), the Roe flux can be written in the form

$$\begin{aligned} F^*_{j+\frac{1}{2}} &= F^*(x_{j+q}, w_{j+1}, w_j) \\ &= \frac{1}{2}\left(F(x_{j+q}, w_j) + F(x_{j+q}, w_{j+1})\right) - \frac{1}{2}\operatorname{sgn}(\tilde{\lambda}_{j+\frac{1}{2}})\left(F(x_{j+q}, w_{j+1}) - F(x_{j+q}, w_j)\right) \\ &= \frac{1}{2}\left(F(x_{j+q}, w_j) + F(x_{j+q}, w_{j+1})\right) - \frac{1}{2}\left(|\tilde{\lambda}_{j+\frac{1}{2}}|\Delta w_{j+\frac{1}{2}} + \operatorname{sgn}(\tilde{\lambda}_{j+\frac{1}{2}})\tilde{V}_{j+\frac{1}{2}}\right) \qquad (4.87) \end{aligned}$$

where $\Delta w_{j+\frac{1}{2}} = w_{j+1} - w_j$ and

$$\tilde{\lambda}_{j+\frac{1}{2}} = \tilde{\lambda}(x_{j+q}, w_{j+1}, w_j) = \begin{cases} \frac{F(x_{j+q},w_{j+1})-F(x_{j+q},w_j)}{w_{j+1}-w_j} & w_{j+1} \neq w_j \\ \frac{\partial F}{\partial w}(x_{j+q}, w_j) & w_{j+1} = w_j \end{cases}. \qquad (4.88)$$

Likewise for $F^*_{j-\frac{1}{2}}$ and $\tilde{\lambda}_{j-\frac{1}{2}}$.

The choice of an approximation for $\tilde{V}_{j+\frac{1}{2}}$ will be discussed in more detail in Chapter 5 although, as an example, we can give

$$\tilde{V}_{j+\frac{1}{2}} = F(x_{j+q}, w_j) - F(x_{j+q-1}, w_j). \qquad (4.89)$$

In order to obtain entropy-satisfying solutions when using Roe scheme in the case of a sonic rarefaction wave ($\frac{\partial F}{\partial w}(x, w_L) < 0 < \frac{\partial F}{\partial w}(x, w_R)$) we tried, without success, an entropy fix similar to the one presented in Section 4.2.4. That is, if $\frac{\partial F}{\partial w}(x, w_L) < 0 < \frac{\partial F}{\partial w}(x, w_R)$, we write

$$w_m = w_L + \frac{\frac{\partial F}{\partial w}(x, w_R) - \tilde{\lambda}}{\frac{\partial F}{\partial w}(x, w_R) - \frac{\partial F}{\partial w}(x, w_L)}(w_R - w_L). \qquad (4.90)$$

and use an expression for the numerical flux given by

$$F^*(x, w_R, w_L) = F(x, w_L) + \frac{\partial F}{\partial w}(x, w_L)\frac{\frac{\partial F}{\partial w}(x, w_R) - \tilde{\lambda}}{\frac{\partial F}{\partial w}(x, w_R) - \frac{\partial F}{\partial w}(x, w_L)}(w_R - w_L). \qquad (4.91)$$

The need of an entropy fix when using the Roe scheme in a direct approach is confirmed by the numerical results we obtained (see Chapter 8).

Furthermore, by including x-dependence in the numerical flux in the way given by (4.84), the Godunov (numerical) flux for the scalar case can be written in the form

$$F^*_{j+\frac{1}{2}} = F^*(x_{j+q}, U_{j+1}, U_j) = \begin{cases} \max_{U_{j+1} \leq w \leq U_j} F(x_{j+q}, w) & \text{for} \quad U_{j+1} < U_j \\ \min_{U_j \leq w \leq U_{j+1}} F(x_{j+q}, w) & \text{for} \quad U_{j+1} \geq U_j \end{cases} \qquad (4.92)$$

(see (4.26)).

In addition, the Engquist-Osher flux can be written in the form

$$F^*_{j+\frac{1}{2}}(x_{j+q}, w_{j+1}, w_j) = F_-(x_{j+q}, w_{j+1}) + F_+(x_{j+q}, w_j) + F(x_{j+q}, c) \tag{4.93}$$

where the functions $F_\pm$ are given by

$$F_-(x, w) = \int_c^w \min_\theta\{F_\theta(x, \theta), 0\}d\theta \tag{4.94}$$

$$F_+(x, w) = \int_c^w \max_\theta\{F_\theta(x, \theta), 0\}d\theta \tag{4.95}$$

and c is arbitrary.

If F is *strictly convex* in the variable w ($\frac{\partial^2 F}{\partial w^2} > 0$) then we can define the functions F_+ and F_- as

$$F_-(x, w) = F\left(x, \min\{w, w_c\}\right) \tag{4.96}$$

$$F_+(x, w) = F\left(x, \max\{w, w_c\}\right). \tag{4.97}$$

We assume that in each cross-section there is a unique critical function $w_c(x)$, i.e. $\frac{\partial F}{\partial w}(x, w_c) = 0$. Then we have

$$F(x, w) = F_-(x, w) + F_+(x, w) + F(x, c) \tag{4.98}$$

and also

$$|F'(x, w)| = F'_-(x, w) + F'_+(x, w). \tag{4.99}$$

Similar conclusions can be drawn for the case of F being *strictly concave* ($\frac{\partial^2 F}{\partial w^2} < 0$).

The three upwind schemes described have the property that

$$\begin{aligned} w_j, w_{j+1} > w_c(x_{j+q}) &\Rightarrow F^*(x_{j+q}, w_{j+1}, w_j) = F(x_{j+q}, w_{j+1}) \\ w_j, w_{j+1} < w_c(x_{j+q}) &\Rightarrow F^*(x_{j+q}, w_{j+1}, w_j) = F(x_{j+q}, w_j). \end{aligned} \tag{4.100}$$

4.4 Nonconservative Form of the Equations

We remark that the class of conservative systems is a more restricted class than that of quasi-linear (nonconservative) systems of the form (4.101). Indeed, a homogeneous system of equations written in quasi-linear form

$$\mathbf{U}_t + J(\mathbf{U})\mathbf{U}_x = \mathbf{0} \tag{4.101}$$

is a conservative system if $J(\mathbf{U}) = \frac{d\mathbf{F}}{d\mathbf{U}}$.

For smooth solutions it is possible to construct different equivalent formulations of the conservation laws, based in nonconservative variables, that are conservative purely in a mathematical sense without regarding to their physical meaning (an example for the Shallow Water equations is presented in [94]). Nonconservative formulations are also not unique. On the contrary, for discontinuous solutions the physical background is important. Nonconservative schemes or conservative schemes built under mathematical considerations rather than physical considerations produce wrong solutions satisfying different jump conditions and thus having wrong speeds and the wrong shock position (see [94]). That does not happen with a conservative scheme based on conservative variables: a solution can be a shock smeared out but it will be at the correct location.

Convergence (if it exists) of numerical schemes to weak solutions satisfying the jump condition is guaranteed for conservative and consistent schemes (Lax-Wendroff Theorem [46]). The theorem does not guarantee that the weak solutions are unique. The physical relevant weak solutions can be obtained either via an entropy condition or as limits of an associated viscous problem as the viscosity vanishes (vanishing viscosity solutions).

If the system is conservative, a necessary condition for the existence of physically relevant solution is the jump condition. For a more general nonconservative system like (4.101), other approaches have been taken to define discontinuous solutions to (4.101) (see [29]). One approach is to use physical considerations to guide the choice of a viscous problem whose limit is the weak shock solution wanted. This approach comes naturally if the equations come from physics. Another approach to define discontinuous solutions is by looking at the product $J(\mathbf{U})(\mathbf{U})_x$ and to give it a meaning, in some way, when $\mathbf{U}$ is, say, a Heaviside function (a single jump). This can be done, as described in [29] following the theory of Dal Maso, Le Floch and Murat [14] who extended the work of Volpert (see [41, 92] for references). The definition of the jump $[J(\mathbf{U})(\mathbf{U})_x]_\phi$ depends on a path ϕ connecting the left state $\mathbf{U}_L$ and the right state $\mathbf{U}_R$. A different approach is to use generalised functions.

The work of Hou and LeFloch [41] for scalar unsteady equations is focused on the error introduced by using nonconservative finite difference schemes for the approximation of conservation laws. Their study has shown that nonconservative schemes do not in general converge to the correct solution if a shock occurs. Instead, the limit solution satisfies an inhomogeneous conservation law. Additionally, a nonconservative scheme does not necessarily converge to the correct solution of an homogeneous scalar conservation law even if it contains the same numerical viscosity as that of a conservative scheme ([41]). They point out that the error introduced by using nonconservative finite difference schemes for the approximation of conservation laws can be small for short times if the initial data is close to a constant but it will grow with time. Furthermore, Hou and LeFloch [41] show that a local correction of a nonconservative scheme can be made, ensuring its convergence to the entropy satisfying solution of the homogeneous scalar conservation law. This correction is implemented through an hybrid (nonconservative) scheme that switches to a conservative scheme in the neighbourhood of discontinuities of the solution.

Summarizing, in regions where the solution is smooth, the conservative forms (4.1) and (4.74) are equivalent to the nonconservative form (4.27) and to the one given by

$$\mathbf{U}_t + \bar{J}\mathbf{U}_x = -\frac{\partial \mathbf{F}}{\partial x}, \tag{4.102}$$

respectively.

At a discontinuity, the conservation forms (4.1) and (4.74) should be approximated by a conservative numerical scheme. If we start from a nonconservative form of the equations (which may be easier to deal with), such as the indirect approach in the x dependent flux function, some sort of nonconservative numerical scheme should be used (as suggested in [41] for scalar unsteady conservation laws). Furthermore, the use of an indirect approach may introduce source terms which have to be dealt with. If the solution is smooth it may be convenient to write the system of conservation laws in primitive variables. But, for example, at an interface when applying the Roe scheme we have a discontinuity, so the conservative form is preferred, especially when the jump is large.

4.5 Nonlinear Stability and Higher Order Extensions for the Scalar Equations

We recall that the Lax-Wendroff theorem does not say anything about whether the numerical method converges. It only asserts that if a conservative consistent scheme converges, then the limit is a weak solution of the homogeneous unsteady nonlinear system of conservation laws (see, e.g. [49]). To guarantee convergence an appropriate notion of (nonlinear) stability is needed.

When studying the stability of a numerical scheme in the scalar case one can use the concepts of *monotonicity preserving*, *monotone* and *total variation diminishing*.

A scheme is said to be *monotonicity preserving* if given any monotone initial data, the solution remains monotone for all time. This property prevents oscillations from occurring near discontinuities.

Another form of stability is that realted to *total variation* (TV) (see, e.g. [49]). It can be shown that the total variation of a solution of a scalar conservation law does not increase in time. Therefore numerical schemes should have this property. This gives rise to the class of *total variation diminishing* or TVD methods.

It can be shown that any conservative TVD scheme is convergent (see [49], Chapter 15) although it is not guaranteed its convergence to the (unique) entropy satisfying solution of the conservation law. Furthermore, TVD implies monotonicity preserving and hence prevents spurious numerical oscillations occurring near discontinuities.

A more restrict form of stability is associated with *monotone schemes* and can be seen as mimicking a monotone property of entropy satisfying solutions of a conservation law. A conservative scheme of the form (4.8) is said to be *monotone* if for any two numerical solutions u_j^n and v_j^n we have

$$v_j^n \geq u_j^n \Rightarrow v_j^{n+1} \geq u_j^{n+1} \quad \text{for all } j.$$

If we write a (scalar) conservative scheme (4.8) in the form

$$w_j^{n+1} = G(w_{j-k+1}^n, \ldots, w_{j+k}^n) \tag{4.103}$$

(the function G is a discrete solution operator), then the scheme is monotone if the function G is a nondecreasing function of all its arguments.

It can be shown that monotone conservative schemes are TVD and more importantly, that any conservative monotone scheme converges to the unique entropy satisfying solution of the conservation law (see [49, 29, 13, 38]). The drawback is that monotone schemes are first order accurate [38]. As described in [92], the limitation of first-order accuracy for monotone approximations can be avoided if L^1-contractive solutions are replaced with the (weaker) requirement of bounded variation solutions.

A classical second-order scheme in both space and time (which is not TVD) is the Lax-Wendroff scheme which despite yielding accurate approximations to smooth solutions develops oscillations in the vicinity of discontinuities.

One approach to obtain higher order accuracy scalar schemes (at least away from discontinuities) and to avoid smearing discontinuities is to use the TVD criteria. Some references on constructing high order TVD schemes using limiter functions can be found in [59]. We mention the work of Harten [35] and Sweby [89].

Second order schemes can be obtained by solving a Generalized Riemann problem defined by assuming piecewise linear data in each interval $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ with midpoint value w_j^n and slopes determined according to certain rules (c.f. van Leer [98]).

Other approaches can be taken, for example, the *essentially non-oscillatory* or ENO approach [36] (see also [85]) and *artificial viscosity schemes* (see [101] for references).

The Godunov, Engquist-Osher and Roe schemes under the hypothesis that an appropriate CFL condition holds, are examples of TVD schemes. Furthermore, it can be shown that the Godunov and Engquist-Osher schemes are monotone if a CFL condition holds. Hence they are first-order schemes.

As described in [59], for steady state computations implicit schemes can be useful since they allow a larger time step to be taken and therefore the steady state is reached faster, although each step is more expensive. Implicit schemes can relax or even avoid the time step restriction given by the CFL condition (see [59] for more details on the schemes and references).

4.6 Scalar Flux Function Depending on a Discontinuous Coefficient

The equation (2.91) obtained in Chapter 2 by reducing the steady (isentropic) Euler equations has a flux function depending on the entropy coefficient K which, although constant for smooth flow, has a jump if a shock occurs (being constant in both sides of the shock). In other words, the definition of the flux function depends on x also through a coefficient which is discontinuous when the solution has a jump.

In [97], Towers establishes convergence of scalar finite difference schemes based on the Godunov and Engquist-Osher schemes for a homogeneous unsteady scalar problem of the form

$$u_t + (k(x)f(u))_x = 0, \quad u(x,0) = u_0(x) \tag{4.104}$$

where the flux $k(x)f(u)$ has a possibly discontinuous spatial dependence through the coefficient k, which is allowed to have jump discontinuities. This type of equations arises in traffic flow problems such as the traffic flow in a highway (see [97]).

Actually, the steady form of equation (4.104) is not of the same form as the one we are interested in, (2.91). The latter, besides being nonhomogeneous (with a source depending on the coefficient K as well), does not have the coefficient K in all the flux terms. Hence, extra difficulties arise when studying equation (2.91).

When using Godunov and Engquist-Osher based schemes to solve the unsteady problem (4.104), Towers was able to construct a explicit time-marching consistent algorithm in conservation form where the monotonocity of the scheme is maintained (with a numerical flux defined as $k_{j+\frac{1}{2}}F^*_{j+\frac{1}{2}}(u_{j+1}, u_j)$ if $k_{j+\frac{1}{2}} \geq 0$ and with arguments reversed when k is negative). Furthermore, Towers uses a piecewise continuous discretisation of $k(x)$ that has jumps at the cell centres as opposed to cell boundaries, staggering in this way the discretisation of k and u, and thus reducing the complexity of the problem. Further study is needed to study whether some of the techniques used in [97] can be extended to equation (2.91).

In the next chapter, Chapter 5, the discretisation of the source terms is studied in more detail. In Chapter 6 the numerical schemes used are presented.

Chapter 5

Source Terms

Many physical applications are modelled by (unsteady) systems of conservation laws with source terms. Some of the theory that exists for one-dimensional homogeneous systems of conservation laws can be extended to the inhomogeneous case, such as the definition of weak and entropy solution (see [29]). In the scalar case, existence (via the vanishing viscosity method) and uniqueness of solution can still be proved for nonhomogenous systems [29]. Moreover, the jump condition (or Rankine-Hugoniot condition) remains unchanged.

One possible approach to deal with the source terms is to split the nonhomogeneous problem into an advection problem (homogeneous) and a source problem (ordinary differential equation) and then treat the resulting problems independently (see, e.g. [94]). This is not the approach adopted here. Instead we use numerical schemes based in the Roe [75] and Engquist-Osher [18] schemes to solve the entire equation (2.25) numerically.

The presence of source terms and that of a flux dependent on both x and the conserved variables are important aspects of the discretisation of a system of conservation laws with source terms such as those studied in this work. An approach to discretise a spatially dependent flux function can start from a quasi-linear form (nonconservative form of the equation) where a spatial partial derivative of the flux derivative is included in the source terms. Once again the importance of "properly" discretising the source terms arises. A proper balance between the discretisation of the flux and source terms, which physically exists in the steady state case, has been sought by several authors. Some of them are referred to in [42] and in [95].

Roe's scheme has been used by many authors to solve systems of the form (2.2) numerically (e.g. [3, 19, 59, 74, 75]) and also of the form (2.1) (e.g. [20, 42, 59, 99, 5]). These works point towards a discretisation of the source terms corresponding to the way the derivative of the flux function is discretised (upwind).

We also mention the work of Roe [78], Glaister [22, 24, 23, 25, 28, 26], Sweby [90], LeVeque [50, 53], Emmerson [15], Gosse [31, 32], Greeenberg and LeRoux [33] and that of Jenny and Muller [43].

The use of the Engquist-Osher scheme [18] for problems of the form (2.1), even in the steady-state case, has not been so thoroughly studied. Special mention is due to the work of MacDonald [59], which was fundamental to our study. Some of the ideas presented in [59] are developed further in the present thesis. Although MacDonald used both the Engquist-Osher and the Roe schemes to solve problems of the form (2.2) and (2.1) in the steady-sate case, some questions remain unanswered. For example, which is the best discretisation of the source terms, particularly if the Roe method is used and if the flux function F depends on x.

Other interesting work is by LeVeque and co-authors (e.g. [51, 52]), Chinnayya and Le Roux [6] and also by Toro and coauthors (e.g. [96]).

In Chapter 2 we showed that it is possible to reduce, in the steady state case, some systems of conservation laws to a single, singular ODE. That is the case for the Saint-Venant equations and the Euler equations of gas dynamics under certain assumptions. A discretisation of this equation using finite differences combined with a time-stepping iteration can be thought of as the discretisation of an unsteady scalar equation of the general form (5.1) (see Chapter 3).

In Section 5.1 a general unsteady scalar conservation law with source terms is studied. The inclusion of source terms in a otherwise conservative scheme is discussed in Section 5.2 and the discretisation of source terms is the subject of Section 5.3. Modifications of the schemes of Roe and Engquist-Osher to include source terms is discussed in, respectively, Section 5.4 and Section 5.5.

5.1 The Scalar Unsteady Nonhomogeneous Equation

Consider the unsteady nonhomogeneous (nonlinear) scalar equation

$$w_t + F(w)_x = D(x, w) \tag{5.1}$$

with some smooth initial data given by (4.17). The source function D may depend on both x and w. Comparing with the homogeneous case presented in Section 4.2.3, two main difficulties arise. The solution w need no longer be constant along the characteristic of the equation and the slope of the characteristics changes as well. In fact, we have two ODEs:

$$\frac{dw}{dt} = D(x, w) \tag{5.2}$$

along paths

$$\frac{dx}{dt} = \frac{dF}{dw}(w), \quad x(0) = x_0. \tag{5.3}$$

The slope of the paths depends on w (see (5.3)) and need not be constant anymore since w is not constant along the characteristics (5.2).

If the flux function depends on both x and w, then instead of the equation (5.1) we have

$$w_t + F(x, w)_x = D(x, w) \tag{5.4}$$

with smooth initial data given by equation (4.17). A solution of the IVP can be constructed (for small t) by following characteristic curves $x = x(t)$. The characteristics satisfy

$$\frac{dx}{dt} = \frac{\partial F}{\partial w} = \lambda(x, w), \quad x(0) = x_0 \tag{5.5}$$

and are not straight lines anymore. Furthermore, w is no longer constant along these characteristic curves since we have

$$\begin{aligned} \frac{d}{dt} w(x(t), t) &= \frac{\partial w}{\partial t} + \frac{\partial w}{\partial x}\frac{dx}{dt} \\ &= D(x, w) - \frac{\partial}{\partial x}F(x, w) + \frac{\partial w}{\partial x}\frac{dx}{dt} = \left(\frac{dx}{dt} - \frac{\partial F}{\partial w}\right)\frac{\partial w}{\partial x} + D(x, w) - \frac{\partial F}{\partial x}. \end{aligned} \tag{5.6}$$

Therefore, from (5.6) we see that we have

$$\frac{dw}{dt} = D(x, w) - \frac{\partial F}{\partial x} \tag{5.7}$$

along the characteristics given by (5.5). As a consequence, it would be possible, for instance, for a solution starting as an expansion wave to shock during the course of the solution.

We can draw similar conclusions if we have $D(x, w) = 0$ in equation (5.4) which is equivalent (for smooth solutions) to the homogeneous equation written in quasi-linear form

$$\frac{\partial w}{\partial t} + \frac{\partial F}{\partial w}\frac{\partial w}{\partial x} = -\frac{\partial F}{\partial x}. \tag{5.8}$$

Note that, in the homogeneous case, if the flux function depends on both x and the solution w, a source term comes into place which corresponds to the partial derivative $-\frac{\partial F}{\partial x}$.

Since Godunov schemes approximate the solution at the boundary of cells using a piecewise constant function, and in the nonhomogeneous case (or homogeneous case with x-dependent flux function) the speed on the characteristics is no longer constant (with curved characteristics), small time steps may be needed. Furthermore, the solution in the cell may change between shock and expansion within the time step, restricting again the time step.

Nevertheless, the jump condition is verified by a discontinuity solution in both the homogeneous and the inhomogeneous case, although it gives only an instantaneous shock speed [90].

Furthermore, a complication which may rise due to source terms is stiffness [53, 90].

In the next Section, Section 5.2, we look at discretisations of the source terms when applying a conservative scheme.

5.2 Conservative Schemes and Numerical Source Terms

Consider a one-dimensional system of conservation laws with source terms of the general form (2.1), i.e.

$$\mathbf{U}_t + \mathbf{F}(x, \mathbf{U})_x = \mathbf{D}(x, \mathbf{U}). \tag{5.9}$$

Integrating the conservation law (5.9) over the rectangle $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}] \times [t^n, t^{n+1}]$ we

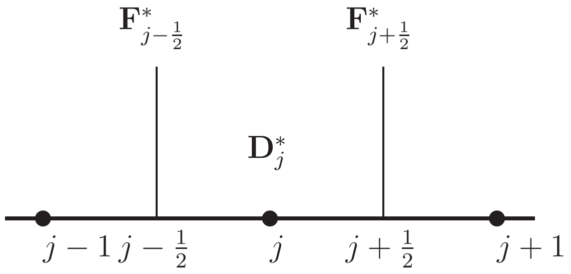
$\mathbf{F}^*_{j-\frac{1}{2}}$
$\mathbf{F}^*_{j+\frac{1}{2}}$
$\mathbf{D}^*_j$
$j-1$
$j-\frac{1}{2}$
j
$j+\frac{1}{2}$
$j+1$

Figure 5.1: Numerical fluxes and sources in j^{th}-cell

obtain the *integral form of the conservation law*, i.e.

$$\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} [\mathbf{U}]_{t^n}^{t^{n+1}} dx + \int_{t^n}^{t^{n+1}} [\mathbf{F}(x,\mathbf{U})]_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} dt = \int_{t^n}^{t^{n+1}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{D}(x,\mathbf{U}) dx dt, \tag{5.10}$$

or, equivalently,

$$\begin{aligned} & \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{U}(x,t^{n+1}) dx = \\ = \; & \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{U}(x,t^n) dx - \left[\int_{t^n}^{t^{n+1}} \mathbf{F}(x_{j+\frac{1}{2}}, \mathbf{U}(x_{j+\frac{1}{2}}, t)) dt - \int_{t^n}^{t^{n+1}} \mathbf{F}(x_{j-\frac{1}{2}}, \mathbf{U}(x_{j-\frac{1}{2}}, t)) dt \right] \\ + \; & \int_{t^n}^{t^{n+1}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{D}(x,\mathbf{U}) dx dt. \end{aligned} \tag{5.11}$$

If we introduce now the integral averages of **U** and **F** considered in Chapter 4 (see (4.4) and (4.78), respectively), equation (5.11) can be rewritten in the form

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x} \left(\mathbf{F}^*_{j+\frac{1}{2}} - \mathbf{F}^*_{j-\frac{1}{2}} \right) + \frac{1}{\Delta x} \int_{t^n}^{t^{n+1}} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \mathbf{D}(x,\mathbf{U}) dx dt. \tag{5.12}$$

For a uniform grid in space and time with grid spacing in space and in time being, respectively, Δx and Δt, the equation (5.12) can be thought of from a numerical point of view (see Fig. 5.1). Then we can write

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x} \left(\mathbf{F}^*_{j+\frac{1}{2}} - \mathbf{F}^*_{j-\frac{1}{2}} \right) + \frac{\Delta t}{\Delta x} \mathbf{D}^*_j \tag{5.13}$$

where $\mathbf{F}^*_{j+\frac{1}{2}}$ is the *numerical flux function* evaluated at the cell interface $j + \frac{1}{2}$ between control volumes (the grid cells) and $\mathbf{D}^*_j = \int \mathbf{D} dx$ is a *numerical source* integral over the cell j a term whose approximation will be discussed. The numerical flux function has the general form (4.80) presented in Chapter 4.

If we sum up both members of equation (5.13) the numerical flux terms over interior grid interfaces cancel out giving

$$\sum_{j=1}^{N-1} \Delta \mathbf{U}_j = \sum_{j=1}^{N-1} \left(\mathbf{U}_j^{n+1} - \mathbf{U}_j^n \right) = -\frac{\Delta t}{\Delta x} \left(\mathbf{F}^*_{N-\frac{1}{2}} - \mathbf{F}^*_{\frac{1}{2}} \right) + \frac{\Delta t}{\Delta x} \sum_{j=1}^{N-1} \mathbf{D}^*_j, \tag{5.14}$$

with $\mathbf{F}_{N-\frac{1}{2}}$ and $\mathbf{F}_{\frac{1}{2}}$ being respectively, the leftmost and rightmost intercell boundaries fluxes.

This equation (5.14) is a discrete version of the equation that can be obtained from the conservation law integrating spatially over the whole domain, i.e.,

$$\int_{x_{\frac{1}{2}}}^{x_{N-\frac{1}{2}}} (\mathbf{U}_t + \mathbf{F}_x)\, dx = \int_{x_{\frac{1}{2}}}^{x_{N-\frac{1}{2}}} \mathbf{D}(x, \mathbf{U}) dx, \tag{5.15}$$

or, equivalently, integrating the derivative of the flux terms,

$$\int_{x_{\frac{1}{2}}}^{x_{N-\frac{1}{2}}} \mathbf{U}_t\, dx = \mathbf{F}_{\frac{1}{2}} - \mathbf{F}_{N-\frac{1}{2}} + \int_{x_{\frac{1}{2}}}^{x_{N-\frac{1}{2}}} \mathbf{D}(x, \mathbf{U}) dx. \tag{5.16}$$

Hence the global variation of the conserved variable is due only to the contributions of the flux terms at the outer boundary grid interfaces and to the source terms.

We would like the source terms to have, at a discrete level, a similar cancellation property. Yet, this is not always possible to achieve. Source terms with derivatives can be approximated by finite differences yielding such a feature but source terms are not of this form in general.

The notion of a *consistent* numerical scheme approximating a system of conservation laws with source terms has to include the homogenous case (no source terms). Hence we can talk about *consistency* in relation to the way the flux function is approximated by the numerical flux function, and we accept the definition given by equation (4.11) of Chapter 4, i.e., the numerical flux function is consistent with the continuous flux when, if computed at constant values, it coincides with the value of the flux function computed at those values. We can also define *consistency* of the discretisation of the source terms in a similar way. Therefore, we say that the discretisation of the source terms is *consistent* if it satisfies

$$\lim_{\substack{\Delta x \to 0 \\ \mathbf{U}_1, \ldots, \mathbf{U}_k \to \mathbf{U}}} \mathbf{D}^*_j(x, \ldots, x, \mathbf{U}_1, \ldots, \mathbf{U}_k) - \mathbf{D}(x, \mathbf{U}) = 0. \tag{5.17}$$

In [99], Vázquez-Cendón calls *conservative* a numerical scheme that, when applied to the shallow water equations (with source terms), it approximates exactly or with order greater than one, a stationary solution. This definition makes use of the "C-property" and of the "approximate C-property" introduced by Bermúdez and Vázquez [3]. Their work on the Q-schemes of van Leer and Roe for inhomogeneous problems is generalised by Vázquez-Cendón [99] to nonuniform meshes and applied to the shallow water equations for channels with rectangular cross-section taking account bed slope, breadth variation and bottom friction.

Studying nonhomogeneous systems of conservations laws of the form (2.2), Gascón and Corberán [21] (see also [10]) take a different approach of that explained above (see equation (5.13)) by including the primitive of the source (the numerical source) in the definition of a new numerical flux function, i.e the new numerical flux function is constructed through the difference between the numerical flux function associated with the physical flow (conservative form) and the primitive of the source term. In this way the original nonhomogeneous system of conservation laws is rewritten in a homogeneous form. They proceed to construct a scheme in conservation form which is an adapted second-order one-step Lax-Wendroff scheme reducing to the original scheme if there are no source terms. The choice of approximation of the primitive of the source maintains the balance between discretisation of the flux function and the discretisation of the source terms in the steady state case. This evaluation of the source terms can be interpreted with the concept of the *exact C-property* or *approximate C-property* introduced by Bermúdez and Vázquez [3] when studying upwind methods based on a flux-difference or flux-splitting discretisation of the flux combined with an upwind discretisation of the source terms. Numerical results are presented for quasi-one dimensional flow test problems.

5.2.1 The Scalar Case

In this section we consider conservative schemes modified with the inclusion of a numerical source, similar to those described for systems (see equation (5.13)). Hence, in the scalar case we have

$$w_j^{n+1} = w_j^n - \frac{\Delta t}{\Delta x}\left(F^*_{j+\frac{1}{2}} - F^*_{j-\frac{1}{2}}\right) + \frac{\Delta t}{\Delta x} D^*_j \tag{5.18}$$

where $F^*_{j+\frac{1}{2}}$ is a numerical flux function at interface $j+\frac{1}{2}$, and D^*_j is an approximation of the integral of the source term. More appropriately, the scheme (5.18) takes the form

$$w_j^{n+1} = w_j^n - \Delta t \frac{F^*_{j+\frac{1}{2}}(x_{j+q}, w_{j+1}, w_j) - F^*_{j-\frac{1}{2}}(x_{j+q-1}, w_j, w_{j-1})}{\Delta x} + \frac{\Delta t}{\Delta x} D^*_j \tag{5.19}$$

(q is a parameter) if the flux function depends on both x and w (see the corresponding homogeneous scheme (4.84)) but simplifies to

$$w_j^{n+1} = w_j^n - \Delta t \frac{F^*_{j+\frac{1}{2}}(w_{j+1}, w_j) - F^*_{j-\frac{1}{2}}(w_j, w_{j-1})}{\Delta x} + \frac{\Delta t}{\Delta x} D^*_j \tag{5.20}$$

if the flux function depends solely on w.

The type of discretisation of the integral of the source term is the topic of discussion of the next section.

5.3 Types of Discretisation of the Source Terms

A pointwise and an upwind approach for the discretisation of the source terms is discussed. A *pointwise* discretisation of the source terms consists of a simple evaluation at a grid point or some kind of average between neighbouring grid points (centred discretisation). An upwind discretisation of the source terms uses an average of the source term on the left and right of the cell interface agreeing with the type of flux discretisation adopted. If any of the source terms is in the form of a derivative, then the pointwise approach corresponds to taking a centred finite differences approximation whereas a upwind discretisation of the derivative corresponds to a one-sided finite differences approximation of the derivative.

An upwind discretisation of the source terms takes into account the way the flux terms are discretised. The idea is to discretise the source terms in such a way that the numerical model preserves the balance present in the mathematical model. Since the source terms do not have an inherent "wind" direction, that direction is picked up from the conservative discretisation of the flux terms. We defer the description of both pointwise and the upwind discretisation of the source terms used in the thesis to discuss it in connection with Roe and Engquist-Osher discretisations of the flux.

Many authors pointed out advantages of doing an upwind discretisation of the source terms over a pointwise discretisation.

In [78], Roe shows the necessity to modify the upwind schemes for nonhomogeneous (one-dimensional) hyperbolic conservation laws. He presents some schemes where the source terms are taken into account through an integration along the characteristics argument. He points out that for linear systems of conservation laws the source terms should be upwinded in the same way as the flux.

Roe's scheme has been extended by several authors to include source terms. For example, Glaister applied Roe's ideas to the shallow water equations [23, 25] and to the Euler equations [22] by projecting the source terms onto the local eigenvectors of the Jacobian matrix (Roe matrix) and upwinding them.

For shallow water equations, several authors have extended Roe Riemann solver to nonhomogeneous problems (e.g. [3, 99, 19, 42, 5]) where the discrete form of the source terms is constructed in a way similar to the numerical fluxes, seeking an equilibria that exists in steady conservation laws with source terms. Seeking a "balance of source terms and flux gradients" (which often occur for steady nonhomogeneous conservation laws), LeVeque [50] proposed the quasi-steady wave-propagation algorithm. The idea consists in introducing a Riemann problem in the centre of each cell in such a way that the cell average is unchanged and the effect of the source terms in the cell is cancelled by the waves resulting from solving the Riemann problem.

Jenny and Müller [43] introduced a characteristic based Riemann solver which takes into account source terms and viscous fluxes. The Rankine-Hugoniot-Riemann (RHR) solver's basic idea is to transform the volume integrals of the source terms (which include viscous terms) into surface integrals. They show that, for one-dimensional nonhomogeneous linear hyperbolic schemes, the RHR solver coincides (for a particular choice of a parameter measuring the fraction of the spatial increment taken in the the cell where the conserved variable is approximated) with the scheme introduced by LeVeque [50] and the scheme introduced by Roe [78]. The schemes differ in the nonlinear case, though.

Difficulties arise when the source terms include spatial derivatives and their discrete representation has to be chosen. Such is the case for the source terms arising from bed slope and variable width.

5.4 Roe Scheme Modified to Include Source Terms

In [3], Bermúdez and Vázquez showed the importance of upwinding the bed slope source term of the Saint-Venant equations modelling the flow of water in a constant breadth channel. They compared several schemes (the Roe scheme being one of them) by means of a *C-Property* related to a stationary solution ('water at rest') and showed that spurious numerical waves can appear when this property does not hold. Such is the case for centred discretisation of the source terms. The motivation was the work of Roe [78] for linear and nonlinear systems suggesting that the source terms should be upwinded in the same way as the flux function and the work of Glaister [23] applying these ideas to the Saint-Venant equations.

Vázquez-Cendón [99] extend this idea of discretising the source terms and the flux function in the 'same way' for the one-dimensional shallow water equations in channels with variable bed slope to consider also the source terms arising from variable breadth function (rectangular cross-section) and bottom friction. The discretisation via the Q-schemes of Roe [75] and van Leer [39] is studied. An upwind discretisation of the source term due to bed variation is adopted since this was shown (in [3]) to provide a better numerical approximation by avoiding spurious oscillations. The discretisations of the source terms due to breadth variation and bottom friction are studied by introducing a new stationary solution, since that corresponding to 'water at rest' could not be used in this case (it eliminated exactly the source terms that had to be studied). For this new stationary solution an upwind and a centred discretisation of the source terms are again compared in terms of satisfying exactly or approximately a *C-property*. It is shown that an upwind discretisation of some source terms (such as the bed slope) should be used despite its analytical expression being known. The use of an analytical expression instead of an upwind discretisation can destroy the C-property exactness. If there is no source term the schemes proposed satisfy exactly the C-property. However if the source term due to friction is considered the schemes satisfy an approximate C-property. The consistency of the schemes is also analyzed.

Note that because the channel is assumed to have rectangular cross-section Vázquez-Cendón is able to use a conservative form of the Saint-Venant equations that does not depend explicitly on x.

The explicit dependence of the flux function on x in the context of Roe's scheme and the C-property introduced by Bermúdez and Vázquez [3] was studied by Garcia-Navarro and Vázquez-Cendón (cf. [20, 19]). They showed that for this type of flux function the local linearisation of Roe scheme does not hold as $\Delta\mathbf{F} = J\Delta\mathbf{U}$ ($J$is the Jacobian matrix at an averaged state) but as $\Delta\mathbf{F} = J\Delta\mathbf{U} + \mathbf{V}$ among spatial increments of the variables $\mathbf{F}$ and $\mathbf{U}$. The term $\mathbf{V}$ corresponds to a discretisation of the partial derivative $\frac{\partial\mathbf{F}}{\partial x}$ which comes forward when applying the chain rule (under smoothness assumptions) to the term $\mathbf{F}(x, \mathbf{U})_x$.

When using Roe's scheme to solve the problem of a rectangular channel of variable width, Priestley [74] chooses to include this term into the source terms. However, in doing so the conservation form of the system of partial differential equations is destroyed. MacDonald [59] chooses to add a discretisation of $\frac{\partial\mathbf{F}}{\partial x}$ to each cell besides the upwind distribution of the wave strengths.

Garcia-Navarro and Vázquez-Cendón [20, 19] discuss two ways of presenting Roe's scheme for nonhomogeneous systems of conservation laws in the one-dimensional case with x-dependent flux function and possible extensions of two-dimensional upwinding via finite volumes methods. In a *fluctuation-signal formulation* [76], the extra term V is thought of as a source term that is subtracted from the existing ones whereas in a *numerical flux formulation* this term V is included in the definition of the numerical flux function with necessary adaptations unless these corrections are also passed to the right-hand side.

The *upwind* discretisation of the source terms approach taken in [20, 19] corresponds to the projection of the source terms onto the basis of eigenvectors. The idea is to enforce the balance between the discrete fluxes and the source terms discretisation and is related with the work of Bermúdez and Vázquez [3] and the *well-balanced scheme* of Greenberg and LeRoux [33]. The fluctuation-signal formulation of the upwind approach adopted is presented and compared with the corresponding matrix notation introduced by Bermúdez and Vázquez [3]. It becomes clearer that the upwind discretisation of the source term determines the amount that is added to each grid point coming from the left-half cell and from the right-half cell.

Both pointwise and upwind discretisation choices adopted in [20, 19] are analyzed in the context of satisfying a C-property (see [3, 99]) for the stationary solution of water at rest. Results are presented for the Saint-Venant equations in a nonprismatic channel with rectangular cross-section and variable width.

Hubbard and Garcia-Navarro [42] use modifications of Roe scheme to approximate nonhomogeneous conservation laws. Special care is taken to approximate terms arising from a spatially dependent flux function (with the ideas of [19]) and to keep the balance between flux and source terms existent in the steady state case. They study besides first-order discretisations, flux and slope-limited high-resolution corrections. The method is extended to two-dimensional flow. Numerical results are presented, in one and two dimensions, for the shallow water equations.

Burguete and Garcia-Navarro [5] extend high-resolution TVD schemes to nonhomogeneous conservation laws. A new technique is proposed that includes the source terms (upwind discretised) in the flux limiter functions in order to mantain the balance with the fluxes. They discuss ways of preserving the conservative character of the schemes, that is, seeking an exact balance between the flux gradient and the source terms discretisation starting from conservative forms with source terms and nonconservative forms of the conservation laws. The case where the flux function is spatially dependent is studied. Numerical results are presented for a nonconservative formulation of the shallow water equations.

5.5 Engquist-Osher Scheme Modified to Include Source Terms

A natural modification of Engquist-Osher scheme to include source terms is to discretise the source terms pointwise. Nevertheless, a upwind discretisation of the source terms is also possible yielding higher order accuracy in certain types of flow (see MacDonald's [59] and Lorenz's [59, 54] work).

In [59], MacDonald used a generalisation of the Engquist-Osher scheme with source terms upwinded to solve the steady problem with a time stepping iteration. The theory extends the work of Lorenz [54] on second-order accurate Engquist-Osher based schemes to solve a singular perturbation problem. The upwinding is performed in a smoothing manner, that is, with the help of a smooth function providing the switch between subcritical and supercritical flow. The rate of switching is controlled by a parameter [59, 54]. With the help of this auxiliary function, the Engquist-Osher flux can be written in a new form that includes source terms.

MacDonald discusses and presents results (for the steady Saint-Venant equations). For the case where the flux function depends solely on the conserved variable, higher order accuracy can be achieved if the source terms are upwinded. The higher order accuracy for the Engquist-Osher extended schemes studied does not happen in all regions of the solution if the flux function depends also spatially (see [59]). We also applied this modification of Engquist-Osher scheme which allowed the comparison between the Roe scheme and the Engquist-Osher scheme with upwind discretisation of the source terms. The details on the particular schemes used are explained in Chapter 6.

Chapter 6

Details of the Numerical Schemes Used

In this Chapter we focus once again on the scalar case and discuss the application of the modified Roe and Engquist-Osher upwind schemes studied in Chapters 4 and 5 to the scalar nonlinear ODEs obtained, in Chapter 2, from reducing the steady Saint-Venant equations and the steady Euler equations, which are of the form

$$\frac{d}{dx}F(x,w) = D(x,w). \tag{6.1}$$

The numerical schemes are based in the use of a conservative finite volume scheme to approximate the flux terms combined with a pointwise and upwind discretisation of the source terms. This discretisation leads to a nonlinear system of difference equations which is solved by a *time stepping iteration* with a chosen initial approximation. As we have seen in Chapter 3, this time stepping iteration can be seen as a pseudo time discretisation of the associated unsteady scalar problem.

We focus also on the search for the discretisation of the source terms that may balance the flux term discretisation for both water and gas flow applications (steady Saint-Venant equations and steady Euler equations) described in Chapter 2.

In Section 6.1 we discuss the use of the pseudo time stepping iteration to solve the finite difference nonlinear systems. Then, in Section 6.2 we present the schemes studied in this work, based on a direct approach (Section 6.2.1) and an indirect approach (Section 6.2.2). These two approaches correspond to a conservation law written in conservation from with source terms included and a quasi-linear form. The ideas of achieving conservation at a discrete level and of a *well-balanced scheme* are discussed in Section 6.3 and applied to the scalar equations obtained from reducing the Steady Saint-Venant equations.

6.1 Time-stepping Iteration

We consider a uniform grid on a channel of length $L > 0$ ($x \in [0, L]$). We have $x_i = i\Delta x$, $i = 0, 1, \ldots, N$ with $\Delta x = L/N$ ($N \in I\!N$).

The numerical solution of (6.1) is sought by using a pseudo time stepping iteration to solve the nonlinear system of difference equations arising from using a conservative finite volume scheme to approximate the flux terms and a pointwise or upwind discretisation of the source terms.

Hence, if $\mathcal{T}_j$ represents the finite difference operator approximating the differential operator $\mathcal{T}$ given by

$$\mathcal{T}w = \frac{df(x,w)}{dx} + D(x,w), \tag{6.2}$$

where $f = -F$, the pseudo time stepping iteration has the form

$$\frac{w_j^{n+1} - w_j^n}{\Delta t} + \mathcal{T}_j w^n = 0 \qquad\qquad n = 0, 1, ... \quad . \tag{6.3}$$

The superscript notation indicates the iteration and the subscript notation indicates the grid point or the cell index, whichever is more convenient to use according to the numerical scheme formulation.

The initial approximation w^0 is taken to be the linear depth profile joining the values of the boundary conditions (when provided) or/and the value of the critical function (i.e., depth h in the water test problems) if a numerical boundary condition is needed (as suggested in [59] for the Saint-Venant equations). A similar initial approximation could be tried in the gas problem (with dependent variable the density ρ) even if it remains to be proven that this is adequate in this case.

6.1.1 Convergence of the Time-stepping Iteration

As we have seen in Chapter 3 for prismatic channels we may consider the limit of the steady viscosity problem , when $\epsilon = 0$, and be only concerned with the limit as Δx

vanishes. That is, we can concentrate on the discrete problem given by

$$\mathcal{T}_j w \equiv \tfrac{f^*(w_{j+1},w_j)-f^*(w_j,w_{j-1})}{\Delta x} + D(x_j, w_j) = 0$$

$$w_0 = \gamma_0, \quad w_N = \gamma_1, \tag{6.4}$$

with $j = 1, \ldots, N-1$.

In [59], MacDonald modifies the theory of Lorenz to consider only positive solutions over a finite range (which one would like to be as small as possible since a CFL condition holds for $\alpha \leq w \leq \beta$, where α and β depedn on the boundary condions and on the critical depth). He proves that under the conditions stated in Theorem 1 and under some assumptions on the numerical flux (consistency, nonincreasing in the first variable and nondecreasing on the second variable, Lipschitz continuous), the difference equations have a unique solution in a finite range, which is bounded. Moreover, a piecewise constant extension of the discrete solution given by $W^{\Delta x} = w_j^n$ converges to W in L_1 as $\Delta x \to 0$, where $W \in NBV_+[0,1]$ is the limit solution of problem (3.16) as $\epsilon \downarrow 0$. Additionally, under the same assumptions and assuming Δt satisfies a certain form of CFL condition, the mapping

$$\mathbf{G} : [\alpha, \beta] \to I\!R^{N+1} \tag{6.5}$$

given by

$$\mathbf{G}(\mathbf{w}) = \begin{pmatrix} \gamma_0 \\ w_1 - \Delta t \mathcal{T}_1 w \\ \vdots \\ w_j - \Delta t \mathcal{T}_j w \\ \vdots \\ w_{N_1} - \Delta t \mathcal{T}_{N-1} w \\ \gamma_1 \end{pmatrix} \tag{6.6}$$

where $\Delta t > 0$ and α and β are constant vectors, has only one fixed point $\mathbf{w}$, which is a solution of the difference equations (6.4).

The theory used by MacDonald enabled him to establish the allowed time steps (CFL conditions) that are sufficient to guarantee convergence of the time stepping iteration when using Engquist-Osher, Godunov and Lax-Friedrichs numerical fluxes. Hence, for Engquist-Osher flux (which is differentiable) the CFL condition reduces to

$$\Delta t \left(\frac{|f'(w)|}{\Delta x} + D_w(x_j, w) \right) \leq 1, \quad 0 \leq j \leq N \tag{6.7}$$

for all $w \in [\alpha, \beta]$, where w is the depth h. For small Δx and if the source term is not dominant, we require

$$\frac{\Delta t}{\Delta x}|f'(w)| \leq 1 \tag{6.8}$$

at all times.

Therefore, using this CFL condition and the theory described, MacDonald obtains efficient and robust algorithms for computing solutions of the steady flow problem. Nevertheless, the theory does not hold for the Roe/FOU scheme or for nonprismatic channels.

Although it has not been proved that a similar CFL condition would hold for the case where the flux function is of the form $f(x, w)$, which arises in connection with variable witdth/breadth channels, we also used it in the Engquist-Osher algorithms but with necessary adaptations corresponding to each x-cross section (the soucer term was not included).

Even if it may be difficult to build a theory such as the one developed by MacDonald [59, 60] for prismatic channels, it may be possible to prove convergence of the time stepping iteration through the contraction mapping theorem under slightly different assumptions. Further study is needed here.

In the next Section we describe the numerical upwind methods used in the thesis and discuss the use of other first-order variants.

6.2 Numerical Schemes and Source Terms Discretisation

In a direct approach we consider conservative schemes with a numerical source (see Chapter 5) that can be written in a *flux-based* form as

$$w_j^{n+1} = w_j^n - \frac{\Delta t}{\Delta x}\left(f^*_{j+\frac{1}{2}} - f^*_{j-\frac{1}{2}}\right) - \frac{\Delta t}{\Delta x} D_j^* \tag{6.9}$$

where $f^*_{j+\frac{1}{2}}$ is a numerical flux function at interface $j + \frac{1}{2}$ and D_j^* is an approximation of the integral of the source term.

When using the Roe scheme it is sometimes convenient to write the numerical schemes (5.20) and (5.19) in a *fluctuation-signal* formulation [76]. Both schemes, (5.20) and

(5.19), can be written as

$$w_j^{n+1} = w_j^n - \frac{\Delta t}{\Delta x}\left(\Delta f_{j+\frac{1}{2}}^- + \Delta f_{j-\frac{1}{2}}^+\right) - \frac{\Delta t}{\Delta x} D_j^* \tag{6.10}$$

where j is a cell index.

When the flux function depends on both the conservative variable w and x, an indirect approach is often used and can be looked at as coming from a nonconservative form of the differential equations (in the absence of source terms). In the indirect approach we seek to approximate the differential operator

$$\mathcal{T} w = \frac{\partial f}{\partial w}\frac{dw}{dx} + \frac{\partial f}{\partial x} + D(x, w), \tag{6.11}$$

obtained from (6.2) by using the chain rule under smooth assumptions.

Taking these two approaches we study different schemes based on the Roe and Engquist-Osher schemes with different types of discretisation of the source terms. We describe some of those schemes in Sections 6.2.1 and 6.2.2. Furthermore, results concerning the application of these schemes to the scalar equation obtained in Chapter 2 by reducing the steady Saint-Venant equations are presented in Chapter 8.

6.2.1 Direct Approach

We are concerned with first order accurate approximations of (6.1) given by

$$\mathcal{T}_j w = \frac{f_{j+\frac{1}{2}}^*(x_{j+q}, w_{j+1}, w_j) - f_{j-\frac{1}{2}}^*(x_{j+q-1}, w_j, w_{j-1})}{\Delta x} + \hat{D}_j = 0, \tag{6.12}$$

for any real q, where $x_{j+q} = (j+q)\Delta x$ and

$$f^*(x, w, w) = f(x, w) = -F(x, w),$$

for all x and w. Possible choices of q are 0, 1 and 1/2. Also, we used the notation $\hat{D}_j = \frac{D_j^*}{\Delta x}$.

Hence, we call Scheme 1 the scheme obtained from using algorithm (6.9) with the Engquist-Osher numerical flux given by (4.93)-(4.95) and a pointwise discretisation of the source terms

$$\hat{D}_j = D(x_j, w_j). \tag{6.13}$$

Therefore, Scheme 1 with a pointwise discretisation of the source term function D can be written as

$$w_j^{n+1} = w_j^n - \frac{\Delta t}{\Delta x}\left(f^*_{j+\frac{1}{2}} - f^*_{j-\frac{1}{2}}\right) - \Delta t D(x_j, w_j) \tag{6.14}$$

with numerical flux function

$$f^*_{j+\frac{1}{2}}(x_{j+q}, w_{j+1}, w_j) = f_-(x_{j+q}, w_{j+1}) + f_+(x_{j+q}, w_j) + f(x_{j+q}, c) \tag{6.15}$$

where the functions $f_\pm$ are given by

$$f_-(x, w) = \int_c^w \min_\theta\{f_\theta(x, \theta), 0\} d\theta \tag{6.16}$$

$$f_+(x, w) = \int_c^w \max_\theta\{f_\theta(x, \theta), 0\} d\theta \tag{6.17}$$

and c is arbitrary (f is concave).

The algorithm of Scheme 1 was implemented with the help of the numerical flux function. In order to update the boundary conditions artificially imposed, we extended the Engquist-Osher scheme to those boundaries by taking the wind direction from the leftmost or the rightmost interior cell in a way similar to the way Roe's scheme operates. Thus, the left (artificial) boundary value can be updated by considering $w_{-1} = w_0$ in the numerical flux function $f^*_{-\frac{1}{2}}$ so that we have

$$f^*_{-\frac{1}{2}} = f^*(x_0, w_0, w_0). \tag{6.18}$$

Similarly, an artificial right-boundary value, can be updated by computing the Engquist-Osher numerical flux $f^*_{N+\frac{1}{2}}$ with $h_{N+1} = h_N$ so that we have

$$f^*_{N+\frac{1}{2}} = f^*(x_N, w_N, w_N). \tag{6.19}$$

Another scheme, Scheme 2 is obtained by using equation (6.9) with a Roe numerical flux given by equations (4.87)-(4.88) where the spatially dependence of the flux function is taken into consideration by adding a discretisation of the partial derivative in order to x to the numerical flux function. It is also possible to write the scheme in a fluctuation-signal formulation of the form (6.10). If a pointwise discretisation of the source term function D is considered, we have

$$w_j^{n+1} = w_j^n - \frac{\Delta t}{\Delta x}\left(\Delta f^-_{j+\frac{1}{2}} + \Delta f^+_{j-\frac{1}{2}}\right) - \Delta t D(x_j, w_j) \tag{6.20}$$

where

$$\Delta f^{-}_{j-\frac{1}{2}} = \tilde{\lambda}^{-}_{j-\frac{1}{2}} \Delta w_{j-\frac{1}{2}} + \frac{1}{2}\left(1 - \operatorname{sgn}(\tilde{\lambda}_{j-\frac{1}{2}}\right) \tilde{V}_{j-\frac{1}{2}} \tag{6.21}$$

$$\Delta f^{+}_{j-\frac{1}{2}} = \tilde{\lambda}^{+}_{j-\frac{1}{2}} \Delta w_{j-\frac{1}{2}} + \frac{1}{2}\left(1 + \operatorname{sgn}(\tilde{\lambda}_{j-\frac{1}{2}}\right) \tilde{V}_{j-\frac{1}{2}} \tag{6.22}$$

where $\Delta w_{j-\frac{1}{2}} = w_j - w_{j-1}$ and $\lambda^{\pm} = \frac{1}{2}(\lambda \pm |\lambda|)$.

An alternative pointwise discretisation of D is to take an average of D on the jth cell of the form

$$D_{j-\frac{1}{2}} = \frac{D(x_{j-1}, w_{j-1}) + D(x_j, w_j)}{2} \tag{6.23}$$

(or even the value of the function D computed at averaged points). Thus, scheme (6.20) can be rewritten as

$$w_j^{n+1} = w_j^n - \frac{\Delta t}{\Delta x}\left(\Delta f^{-}_{j+\frac{1}{2}} + \Delta f^{+}_{j-\frac{1}{2}}\right) - \frac{1}{2}\Delta t\left(D_{j+\frac{1}{2}} + D_{j-\frac{1}{2}}\right). \tag{6.24}$$

If a upwind discretisation of the source function D is considered, scheme 2 can be written as equation (6.20) except that the source term function discretisation $D(x_j, w_j)$ is of the form

$$\hat{D}_j = \frac{1}{2}(1 - \operatorname{sgn}(\lambda_{j+\frac{1}{2}}))\tilde{D}_{j+\frac{1}{2}} + \frac{1}{2}(1 + \operatorname{sgn}(\lambda_{j-\frac{1}{2}}))\tilde{D}_{j-\frac{1}{2}}. \tag{6.25}$$

However, the choice of $\tilde{D}_{j\pm\frac{1}{2}}$ remains open. Often a cell average of the form (6.23) is taken but other choices are possible.

The discretisation of the terms $\tilde{V}_{j\pm\frac{1}{2}}$ corresponding to the partial derivative of the flux function in order to x with w constant, was considered in the form

$$\hat{V}_{j+\frac{1}{2}} = f(x_{j+1}, w_{j+k}) - f(x_j, w_{j+k}) \tag{6.26}$$

where w_{j+k} represents a value of w in the jth cell. If k is not an integer $(k \neq 0, 1)$ an average can be taken.

If we do not think about splitting ΔF with the help of a Roe-type linearisation, we can still think about creating other FOU schemes that approximate the total derivative as a difference with both a variation on x and a variation on w. We considered a scheme, Scheme 2* with numerical flux given by

$$f^{*}_{j+\frac{1}{2}}(x_{j+q}, w_{j+1}, w_j) = \frac{1}{2}\left[f(x_{j+q}, w_j) + f(x_{j+q}, w_{j+1}) - |\lambda_{j+\frac{1}{2}}|\,(w_{j+1} - w_j)\right] \tag{6.27}$$

where $\lambda_{j+\frac{1}{2}}$ is given by equation (4.88) (the same λ as in scheme 2). Similarly to scheme 2, a pointwise and a upwind discretisation of the source function D can be considered (see (6.13) and (6.25)).

If there is a unique critical function (depth $h_c(x)$ or density $\rho_c(x)$) at each channel cross section and in the Saint-Venant problem, if the width does not approach zero as the depth becomes large, the schemes 1 and 2* verify the properties (4.100). It can be shown that in a region where the flow is subcritical, i.e. $w_{j-1}, w_j > w_c(x_{j+q-1})$ and $w_j, w_{j+1} > w_c(x_{j+q})$ (with w_c is the critical function), scheme 1 (Engquist-Osher) and scheme 2*(FOU) reduce to

$$\frac{f(x_{j+q}, w_{j+1}) - f(x_{j+q-1}, w_j)}{\Delta x} + D(x_j, w_j) = 0 \tag{6.28}$$

with truncation error given by

$$\begin{aligned} T.E. &= \frac{\Delta x}{2}\Big(w'' f_w + (w')^2 f_{ww} + 2qw' f_{wx} + (2q-1) f_{xx}\Big) + \mathcal{O}(\Delta x^2) \qquad (6.29) \\ &= \frac{\Delta x}{2}\frac{d}{dx}\left(-D + 2(q-1) f_x\right) + \mathcal{O}(\Delta x^2). \end{aligned}$$

In a region of the solution where the flow is supercritical, i.e. $w_{j-1}, w_j < w_c(x_{j+q-1})$ and $w_j, w_{j+1} < w_c(x_{j+q})$, the scheme 1(Engquist-Osher) and scheme 2*(FOU) reduce to

$$\frac{f(x_{j+q}, w_j) - f(x_{j+q-1}, w_{j-1})}{\Delta x} + D(x_j, w_j) = 0 \tag{6.30}$$

and the truncation error is given by

$$\begin{aligned} T.E. &= \frac{\Delta x}{2}\Big(-w'' f_w - (w')^2 f_{ww} + 2(q-1)w' f_{wx} + (2q-1) f_{xx}\Big) + \mathcal{O}(\Delta x^2)(6.31) \\ &= \frac{\Delta x}{2}\frac{d}{dx}\left(D + 2q f_x\right) + \mathcal{O}(\Delta x^2). \end{aligned}$$

From the truncation error formulas we see that the error is $\mathcal{O}(\Delta x)$ for all values of q.

An upwind dsicretisation of the source function D was also tried for Engquist-Osher scheme 1. A possible approach of upwinding the source function D similarly to Roe's scheme is not what is sought in the case of scheme 1. We would like to build a switch function that can be included in the Engquist-Osher flux function and that provides the switch to upwind the source term. We looked at the work of MacDonald [59] and Lorenz [54] concerning upwind discretisation of the source terms when using the Engquist-Osher scheme.

In [54], Lorenz shows how to construct an upwind discretisation of the source terms using a smooth function. MacDonald [59] extends Lorenz's ideas to a x-dependent flux function by using similar expressions but with the derivative of the flux function substituted by a partial derivative in order to w. The partial derivatives have to be computed at an x value that may or may not correspond to a grid point. In [59], MacDonald uses the grid point corresponding to the value of w computed, but the results obtained were not as accurate as one should expect at first. The second-order accuracy attained by Lorenz is not obtained in all types of flow. Further study is needed in ways of upwinding the source terms in this case. Nevertheless, we describe the modification discussed by MacDonald [59].

The upwind discretisation of the source terms when using the Engquist-Osher scheme in the case of a flux function of the form $f(x, w)$ (see equation 6.9) is given by

$$D_j^* = \Delta x \left(\chi_j^- D_{j-1} + \chi_j^0 D_j + \chi_j^+ D_{j+1} \right) \tag{6.32}$$

where $D_j = D(x_j, w_j)$ and

$$\begin{aligned} \chi_j^- &= \chi\left(\frac{\alpha f_w(x, w_{j-1})}{\sqrt{\Delta x}}\right) && (6.33)\\ \chi_j^+ &= \chi\left(\frac{-\alpha f_w(x, w_{j+1})}{\sqrt{\Delta x}}\right) && (6.34)\\ \chi_j^0 &= 1 - \chi_{j+1}^- - \chi_{j-1}^+, && (6.35) \end{aligned}$$

$\alpha \geq 0$ is a parameter and χ is a smooth arbitrary increasing function connecting the values of 0 and $\frac{1}{2}$ given by

$$\chi(y) = \begin{cases} 0 & y < 0 \\ y^2 & 0 \leq y \leq \frac{1}{2} \\ \frac{1}{2} - (1-y)^2 & \frac{1}{2} \leq y \leq 1 \\ \frac{1}{2} & y > 1 \end{cases} . \tag{6.36}$$

The selection of $\alpha = 0$ corresponds to a pointwise discretisation of the source term.

Since we chose to discretise x in the same form as in the numerical flux, we are able to rewrite the Engquist-Osher flux in a new form,

$$f_{j-\frac{1}{2}}^{*new}(x_{j+q-1}, w_j, w_{j-1}) = f_{j-\frac{1}{2}}^* + \Delta x \left[\chi_{j-1}^+ D(x_j, w_j) - \chi_j^- D(x_{j-1}, w_{j-1}) \right]. \tag{6.37}$$

Therefore this new scheme can be written with the help of the original Engquist-Osher flux as

$$w_j^{n+1} = w_j^n - \frac{\Delta t}{\Delta x}\left(f^*_{j+\frac{1}{2}} - f^*_{j-\frac{1}{2}}\right) - \Delta t\left(\chi_j^- D_{j-1} + \chi_j^0 D_j + \chi_j^+ D_{j+1}\right) \tag{6.38}$$

or, with the help of the new form (6.37), as

$$\begin{aligned} w_j^{n+1} = w_j^n \quad &- \quad \frac{\Delta t}{\Delta x}\left[f^*_{j+\frac{1}{2}}(x_{j+q}, w_{j+1}, w_j) - f^*_{j-\frac{1}{2}}(x_{j+q-1}, w_j, w_{j-1})\right] \\ &- \quad \Delta t D(x_j, w_j) \end{aligned} \tag{6.39}$$

The parameter $\alpha > 0$ controls the rate at which the source terms switch between subcritical and supercritical flow across a transition. The higher the value of α, the greater the speed the switching occurs. (If $\alpha = 0$ we have a pointwise discretisation of the source terms.)

In a region of the solution where the flow is subcritical, the scheme reduces to

$$\frac{f(x_{j+q}, w_{j+1}) - f(x_{j+q-1}, w_j)}{\Delta x} + \frac{D(x_j, w_j) + D(x_{j+1}, w_{j+1})}{2} = 0 \tag{6.40}$$

and in a region where the flow is supercritical, it reduces to

$$\frac{f(x_{j+q}, w_j) - f(x_{j+q-1}, w_{j-1})}{\Delta x} + \frac{D(x_j, w_j) + D(x_{j-1}, w_{j-1})}{2} = 0. \tag{6.41}$$

The truncation error of both schemes, (6.40) and (6.41) can be obtained by adding, respectively,

$$\frac{D(x_j, w_j) + D(x_{j+1}, w_{j+1})}{2} - D(x_j, w_j) = \frac{\Delta x}{2}(D_x + w' D_w) + \mathcal{O}(\Delta x^2)$$

and

$$\frac{D(x_j, w_j) + D(x_{j-1}, w_{j-1})}{2} - D(x_j, w_j) = -\frac{\Delta x}{2}(D_x + w' D_w) + \mathcal{O}(\Delta x^2)$$

to the truncation error formulas (6.29) and (6.31). The resulting truncation errors for regions of subcritical and supercritical flow, which are given, respectively, by

$$T.E. = 2(q-1)\Delta x \frac{d}{dx} f_x + \mathcal{O}(\Delta x^2) \tag{6.42}$$

and

$$T.E. = 2q\Delta x \frac{d}{dx} f_x + \mathcal{O}(\Delta x^2), \tag{6.43}$$

show that a choice of $q = 1$ for subcritical flow and that of $q = 0$ for supercritical flow lead to second order accurate schemes. This suggests the introduction of a switch in the values of q depending on the type of flow. To guarantee conservation the switch of the parameter q has to be achieved so that a discrete telescopic property is verified (see Section 4.3).

6.2.2 Indirect Approach

In an indirect approach we seek approximations of the differential operator (6.11). (Note that in this case, the finite volume discretisation reflects a variation only on the the dependent variable w since all flux terms are computed at a fixed x.)

We consider a numerical scheme (Scheme 3) based in the Engquist-Osher numerical flux (4.67)-(4.69) and given by

$$\frac{f^*(x_{j+q}, w_{j+1}, w_j) - f^*(x_{j+q}, w_j, w_{j-1})}{\Delta x} + \hat{D}_j + \hat{V}_j = 0. \tag{6.44}$$

In a pointwise approach, we chose to discretise $\hat{D}_j$ in the form (6.13) and $\hat{V}_j$ as a centred (half-point) approximation of the partial derivative $\frac{\partial f}{\partial x}$, i.e.

$$\hat{V}_j = \frac{f(x_{j+\frac{1}{2}}, w_{j+k}) - f(x_{j-\frac{1}{2}}, w_{j+k})}{\Delta x}. \tag{6.45}$$

The choice of k is related to the cell and if k is not an integer, and an average of neighbouring values can be taken for w_{j+k}. In most of the results shown related to this scheme, $q = 0$ and $k = 0$. Another centred first-order approximation of the partial derivative $\frac{\partial f}{\partial x}$ based in grid points x_{i+1} and x_{i-1} can also be applied.

In a upwind approach we have already a way of upwinding the source D which was introduced by Lorenz [55, 56] and is described in Section 6.2.1. The upwinding is done with the help of a smooth function which allows the switch. In order to have a discretisation of the $\hat{V}$ term similar to the flux term discretisation, a one-sided discretisation of the derivative $\hat{V}$ (only variation on x, w fixed) is considered. The switch yielding the one-sided approximation to be caken is associated with the sign of the partial derivative $\frac{\partial f}{\partial w}$. For example, for subcritical flow we would like to have

$$\frac{f(x_{j+q}, w_{j+1}) - f(x_{j+q}, w_j)}{\Delta x} + \frac{f(x_{j+1}, w_k) - f(x_j, w_k)}{\Delta x} + \frac{D_j + D_{j+1}}{2} = 0, \tag{6.46}$$

whereas for supercritical flow we would like to have

$$\frac{f(x_{j+q}, w_j) - f(x_{j+q}, w_{j-1})}{\Delta x} + \frac{f(x_j, w_k) - f(x_{j-1}, w_k)}{\Delta x} + \frac{D_j + D_{j-1}}{2} = 0, \tag{6.47}$$

It is possible to discretise the derivative source term $\hat{V}$ in a upwind manner of the form

$$\hat{V}_j = \frac{1}{\Delta x}\left[\left(1-\psi_j^+\right)V_{j-\frac{1}{2}} + \left(1-\psi_j^-\right)V_{j+\frac{1}{2}}\right] \tag{6.48}$$

where

$$V_{j-\frac{1}{2}} = \frac{f(x_j,w) - f(x_{j-1},w)}{\Delta x} \tag{6.49}$$

and

$$V_{j+\frac{1}{2}} = \frac{f(x_{j+1},w) - f(x_j,w)}{\Delta x}, \tag{6.50}$$

with

$$\psi_j^- \;=\; \psi\left(\frac{\beta f_w(x_{j-1},w)}{\sqrt{\Delta x}}\right) \tag{6.51}$$

$$\psi_j^+ \;=\; \psi\left(\frac{-\beta f_w(x_{j+1},w)}{\sqrt{\Delta x}}\right). \tag{6.52}$$

The w is fixed, $\beta \geq 0$ is a parameter and ψ is a smooth arbitrary increasing function connecting the values of 0 and 1 given by

$$\psi(t) = \begin{cases} 0 & t \geq 0 \\ t^2 + \frac{1}{2}t & 0 \leq t \leq \frac{1}{2} \\ -t^2 + \frac{5}{2}t - \frac{1}{2} & \frac{1}{2} \leq t \leq 1 \\ 1 & t \geq 1 \end{cases}. \tag{6.53}$$

Other choices of the function ψ are possible. The selection of $\beta = 0$ corresponds to a centred discretisation of the source term with the exception of a coefficient 2 that should be introduced in the denominator. Thus, possibly, this choice of the ψ function can be improved in order to have a centred discretisation on the derivative term when $\beta = 0$.

Therefore, a modification of the Engquist-Osher is built which can be written in the form

$$\frac{f^*(x_{j+q},w_{j+1},w_j) - f^*(x_{j+q},w_j,w_{j-1})}{\Delta x} + \frac{1}{\Delta x}\left[\left(1-\psi_j^+\right)V_{j-\frac{1}{2}} + \left(1-\psi_j^-\right)V_{j+\frac{1}{2}}\right]$$
$$+ \quad \chi_j^- D_{j-1} + \chi_j^0 D$$

and pseudo-time iterated.

The algorithm corresponding to scheme 3 was implemented through the numerical flux function and an update of artificial boundary values was done similarly to scheme 1 (direct approach), that is, by using special flux function extensions (6.18) and (6.19).

The truncation error of the pointwise source discretisation scheme in regions of subcritical flow and supercritical flow is given by, respectively,

$$\begin{aligned} T.E. &= \frac{\Delta x}{2}\left(w'' f_w + (w')^2 f_{ww} + 2(q+k)w' f_{wx}\right) + \mathcal{O}(\Delta x^2) \\ &= \frac{\Delta x}{2}\left[\frac{d}{dx}\left(-D + (2(q+k)-3)f_x\right) + 2(q+k-1)f_{xx}\right] + \mathcal{O}(\Delta x^2) \end{aligned} \tag{6.55}$$

and

$$\begin{aligned} T.E. &= \frac{\Delta x}{2}\left(-w'' f_w - (w')^2 f_{ww} - (q+k)w' f_{wx}\right) + \mathcal{O}(\Delta x^2) \\ &= \frac{\Delta x}{2}\left[\frac{d}{dx}\left(D + (2-q-k)f_x\right) + (q+k-1)f_{xx}\right] + \mathcal{O}(\Delta x^2). \end{aligned} \tag{6.56}$$

A choice of k and p such that $k+p=1$ eliminates the term with f_{xx} in both truncation error equations, (6.55) and (6.56). Nevertheless, we considered $q=0$ and $k=0$ which corresponds to the chosen grid point x_j and the corresponding value of the approximate solution, w_j.

We also considered an upwind scheme based on the Roe scheme (indirect approach), <u>Scheme 4</u>. The scheme can be written in the form

$$\frac{1}{\Delta x}\left(\Delta f^-_{j+\frac{1}{2}} + \Delta f^+_{j-\frac{1}{2}}\right) + \hat{V}_j + \hat{D}_j = 0 \tag{6.57}$$

with

$$\Delta f^\pm_{j+\frac{1}{2}} = \lambda^\pm_{j+\frac{1}{2}} \Delta w_{j+\frac{1}{2}}, \tag{6.58}$$

or in the form

$$\lambda^-_{j+\frac{1}{2}} \frac{w_j - w_{j-1}}{\Delta x} + \lambda^+_{j-\frac{1}{2}} \frac{w_{j+1} - w_j}{\Delta x} + \hat{V}_j + \hat{D}_j = 0 \tag{6.59}$$

where

$$\lambda_{j+\frac{1}{2}} = \begin{cases} \frac{f(x_{j+q}, w_{j+1}) - f(x_{j+q}, w_j)}{w_{j+1} - w_j} & w_{j+1} \neq w_j \\ f_w(x_{j+q}, w_j) & w_{j+1} = w_j \end{cases} \tag{6.60}$$

and $\lambda^\pm = \frac{1}{2}(\lambda \pm |\lambda|)$.

The pointwise discretisation of the flux terms considered corresponds to take an approximation for the V term in form (6.45) and an approximation of the source term D like the one in (6.23). A upwind discretisation of both source terms, $\hat{V}_j$ and $\hat{D}_j$, was also tried with the upwind direction of the source terms taken from the one of the corresponding flux. Hence an upwind discretisation of the source terms yields

$$\hat{V}_j = \frac{1}{2}(1 - \mathrm{sgn}(\lambda_{j+\frac{1}{2}}))\hat{V}_{j+\frac{1}{2}} + \frac{1}{2}(1 + \mathrm{sgn}(\lambda_{j-\frac{1}{2}}))\hat{V}_{j-\frac{1}{2}} \tag{6.61}$$

$$\hat{D}_j = \frac{1}{2}(1 - \mathrm{sgn}(\lambda_{j+\frac{1}{2}}))\hat{D}_{j+\frac{1}{2}} + \frac{1}{2}(1 + \mathrm{sgn}(\lambda_{j-\frac{1}{2}}))\hat{D}_{j-\frac{1}{2}} \tag{6.62}$$

with

$$\hat{V}_{j+\frac{1}{2}} = \frac{f(x_{j+1}, w_{j+k}) - f(x_j, w_{j+k})}{\Delta x} \tag{6.63}$$

$$\hat{D}_{j+\frac{1}{2}} = \frac{D(x_j, w_j) + D(x_{j+1}, w_{j+1})}{2}. \tag{6.64}$$

and $\lambda_{j+\frac{1}{2}}$ given by equation (6.60).

The truncation error in this case is

$$T.E. = \Delta x(p + k - 1)w' f_{wx} + \mathcal{O}(\Delta x^2) \tag{6.65}$$

in a region where the flow is subcritical and

$$T.E. = \Delta x(p + k + 1)w' f_{wx} + \mathcal{O}(\Delta x^2) \tag{6.66}$$

in a region where the flow is supercritical. A choice of $q = 0$ and $k = 1$ or vice-versa, when the flow is subcritical and a choice of $q = -1$ and $k = 0$ (or vice-versa), renders the scheme second order accurate (in regions where the solution is smooth).

6.3 Well-balanced schemes

The work of several authors [3, 99, 20, 19, 5, 42] suggests that a more careful discretisation of the source terms may improve the accuracy of the schemes. If a balance between the source terms discretisation and the flux terms discretisation is achieved, i.e. if a reproduction at a discrete level of the continuous problem equation is achieved this may improve the accuracy of the schemes ([3, 99, 42]).

In [42], Hubbard and Garcia-Navarro describe how to balance the source terms and the flux terms discretisation when using the Roe scheme. They proceed to apply these ideas to the Saint-Venant equations. We discuss how to apply these ideas to the reduced steady singular differential equations.

In fact, when using the Roe scheme to approximate a conservation law with a flux function f depending on both w and x, we have

$$\Delta f_{j+\frac{1}{2}} = \widetilde{\left(\frac{\partial f}{\partial w}\right)}_{j+\frac{1}{2}} \Delta w_{j+\frac{1}{2}} + \widetilde{\left(\frac{\partial f}{\partial x}\right)}_{j+\frac{1}{2}} \Delta x_{j+\frac{1}{2}} \tag{6.67}$$

where “∼” represents an averaged quantity.

Using the notations

$$\widetilde{\lambda}_{j+\frac{1}{2}} = \widetilde{\left(\frac{\partial f}{\partial w}\right)}_{j+\frac{1}{2}} \tag{6.68}$$

for the advection velocity and

$$\widetilde{V}_{j+\frac{1}{2}} = \widetilde{\left(\frac{\partial f}{\partial x}\right)}_{j+\frac{1}{2}} \Delta x_{j+\frac{1}{2}}, \tag{6.69}$$

for the extra term due to spatial variation of the flux function, equation (6.67) can be written as

$$\Delta f_{j+\frac{1}{2}} = \widetilde{\lambda}_{j+\frac{1}{2}} \Delta w_{j+\frac{1}{2}} + \widetilde{V}_{j+\frac{1}{2}}. \tag{6.70}$$

An upwind scheme is constructed by setting

$$\Delta f^{\pm}_{j+\frac{1}{2}} = \widetilde{\lambda}^{\pm}_{j+\frac{1}{2}} \Delta w_{j+\frac{1}{2}} + \frac{1}{2}\left(1 \pm \operatorname{sgn}(\widetilde{\lambda}_{j+\frac{1}{2}})\right) \widetilde{V}_{j+\frac{1}{2}} \tag{6.71}$$

where $\widetilde{\lambda}^{\pm} = \frac{1}{2}(\widetilde{\lambda} \pm |\widetilde{\lambda}|)$. The discretisation of D^*_j is sought by requiring that the balance existent at steady state, i.e. $\frac{d}{dx} f(x,w) = -D$, is maintained at a discrete level. Hence, we would like to have

$$-D^*_j = \Delta f^+_{j-\frac{1}{2}} + \Delta f^-_{j+\frac{1}{2}} \tag{6.72}$$

and this can be achieved if we consider

$$D^*_j = \Delta x(\tilde{D}^+_{j-\frac{1}{2}} + \tilde{D}^-_{j+\frac{1}{2}}), \tag{6.73}$$

with

$$\tilde{D}^{\pm}_{j\pm\frac{1}{2}} = \frac{1}{2}\left(1 \mp \operatorname{sgn}(\tilde{\lambda}_{j\pm\frac{1}{2}})\right) \tilde{D}_{j\pm\frac{1}{2}}. \tag{6.74}$$

The resulting scheme can be written in a flux-based finite volume form with a numerical flux that includes the term $\tilde{V}$ and also in the fluctuation signal formulation as

$$w^{n+1}_j = w^n_j - \frac{\Delta t}{\Delta x}(\Delta f^-_{j+\frac{1}{2}} + \Delta f^+_{j-\frac{1}{2}}) - \frac{\Delta t}{\Delta x} D^*_j. \tag{6.75}$$

There are two ideas to bear in mind. Firstly, the idea of having some sort of conservation, at least at a discrete level. Secondly, the idea of achieving a well-balanced scheme. The first idea leads, in the case of the Roe scheme, for example, to

$$(\Delta f)_{j+\frac{1}{2}} = \left(\tilde{\lambda}\Delta w + \tilde{V}\right)_{j+\frac{1}{2}}. \tag{6.76}$$

The second idea is that of having $f(x,w)_x = -D$ at a discrete level, which corresponds in the case of Roe's scheme, to equation (6.72). This approach is the one taken in

[42] when using the Roe scheme to solve the (unsteady) Saint-Venant equations. A different approach, yielding the same concepts is given in [5]. There, Burguete and Garcia-Navarro discuss the idea of discrete conservation coming from both what we call a direct (conservation law with source terms) and an indirect approach (quasi-linear form). When starting from a system of conservation laws in a quasi-linear form, we should have

$$\left(D - \frac{\Delta F}{\Delta x}\right)_{j+\frac{1}{2}} = \left(\bar{D} - \tilde{\lambda}\Delta w\right)_{j+\frac{1}{2}} . \tag{6.77}$$

where $\bar{D} = D - \frac{\tilde{V}}{\Delta x}$. Both the left-hand side and right-hand side of equation (6.76) lead to the concept of a balanced scheme, i.e. discretising flux and source terms in a similar way. Hence, when using upwind schemes, the source terms should be upwind discretised and the wind direction may be taken from the corresponding flux.

We would like to apply the ideas of preserving discrete conservation and that of a balanced discretisation of the source and flux terms to the scalar reduced steady Saint-Venant equation (2.25) introduced in Chapter 2 when using scheme (6.75).

Inspired by Burguete and Garcia-Navarro [5], the discrete conservation of the quasi-linear form of the equations is achieved if equation (6.76) is verified.

For the reduced Saint-Venant equation (2.25), $f = -F$ is obtained through equation (2.26) and D is given by equation (2.27). Furthermore, the partial derivatives of F in order to x and h can be written as

$$\lambda = \frac{\partial F}{\partial h} = \left(g\frac{A}{b} - \frac{Q^2}{A^2}\right) b = (c^2 - u^2)b \tag{6.78}$$

$$\frac{\partial F}{\partial x} = -\frac{Q^2}{A^2}\frac{\partial A}{\partial x} + g\frac{\partial I_1}{\partial x}. \tag{6.79}$$

Since $\frac{\partial I_1}{\partial x} = I_2$ and

$$\frac{dA}{dx} = \frac{\partial A}{\partial x} + \frac{\partial A}{\partial h}\frac{dh}{dx},$$

the partial derivatives in equation (6.79) can be avoided yielding

$$\frac{\partial F}{\partial x} = -\frac{Q^2}{A^2}\left(\frac{\partial A}{\partial x} - b\frac{dh}{dx}\right) + gI_2.$$

Therefore using equation (6.77) we have

$$\left[gI_2 + gA(S_0 - S_f) - \frac{\Delta}{\Delta x}\left(\frac{Q^2}{A^2} + gI_1\right)\right]_{j+\frac{1}{2}} =$$

$$= \left[gA(S_0 - S_f) - \frac{Q^2}{A^2}b\frac{\Delta h}{\Delta x} + \frac{Q^2}{A^2}\frac{\Delta A}{\Delta x} - \left(g\frac{A}{b} - \frac{Q^2}{A^2}\right)b\frac{\Delta h}{\Delta x}\right]_{j+\frac{1}{2}} \tag{6.80}$$

where we have considered that

$$\begin{aligned}\left(\frac{dh}{dx}\right)_{j+\frac{1}{2}} &\approx \left(\frac{\Delta h}{\Delta x}\right)_{j+\frac{1}{2}} \\ \left(\frac{dA}{dx}\right)_{j+\frac{1}{2}} &\approx \left(\frac{\Delta A}{\Delta x}\right)_{j+\frac{1}{2}}.\end{aligned}$$

Additionally, since $\frac{dI_1}{dx} = I_2 + A\frac{dh}{dx}$, we consider

$$(I_2)_{j+\frac{1}{2}} \approx \left(\frac{\Delta I_1}{\Delta x} - A\frac{\Delta h}{\Delta x}\right)_{j+\frac{1}{2}},$$

yielding

$$\left[-\frac{\Delta}{\Delta x}\left(\frac{Q^2}{A}\right) - gA\frac{\Delta h}{\Delta x}\right]_{j+\frac{1}{2}} = \left[-\frac{Q^2}{A^2}b\frac{\Delta h}{\Delta x} + \frac{Q^2}{A^2}\frac{\Delta A}{\Delta x} - (c^2 - u^2)b\frac{\Delta h}{\Delta x}\right]_{j+\frac{1}{2}}. \tag{6.81}$$

Equation (6.81) is satisfied by the averages given by

$$c_{j+\frac{1}{2}} = \sqrt{g\frac{A_{j+\frac{1}{2}}}{b_{j+\frac{1}{2}}}}, \quad u_{j+\frac{1}{2}} = \frac{Q}{\sqrt{A_i A_{i+1}}}. \tag{6.82}$$

The former average is similar to the one obtained in [5] for the full equations and the latter is a particular case obtained when Q is constant.

The choice of other discrete averages such as $A_{j+\frac{1}{2}}$, $(S_0)_{j+\frac{1}{2}}$ and $(S_f)_{j+\frac{1}{2}}$ is open. Nevertheless, keeping in mind that the flux terms discretisation and the source terms discretisation should balance, possible averaging values arise.

Hence, aiming to achieve this balance, we look at

$$\left(\tilde{\lambda}\Delta h + \tilde{V}\right) = \Delta x \tilde{D}_{j+\frac{1}{2}}. \tag{6.83}$$

Considering the average values (6.82) and the definitions of $\tilde{\lambda}$ (6.68) and $\tilde{V}$ (6.69), the computation of

$$\tilde{\lambda}_{j+\frac{1}{2}} = \left(-\frac{Q^2}{\tilde{A}^2} + g\bar{h}\right)_{j+\frac{1}{2}} \bar{b}_{j+\frac{1}{2}} \tag{6.84}$$

and

$$\tilde{V}_{j+\frac{1}{2}} = \left(-\frac{Q^2}{\tilde{A}^2}\bar{h} + \frac{1}{2}g\tilde{h^2}\right)_{j+\frac{1}{2}} \Delta b \tag{6.85}$$

can be carried out by choosing the following averages

$$\begin{aligned}\tilde{A}_{j+\frac{1}{2}} &= \sqrt{A_j A_{j+1}}, & \bar{b}_{j+\frac{1}{2}} &= \frac{b_j + b_{j+1}}{2}, \\ \tilde{h^2}_{j+\frac{1}{2}} &= \frac{h_j^2 + h_{j+1}^2}{2}, & \bar{h}_{j+\frac{1}{2}} &= \frac{h_j + h_{j+1}}{2}.\end{aligned} \tag{6.86}$$

The source function $D(x, h)$ for a nonprismatic channel with rectangular cross-section, is given by

$$D(x, h) = I_2 + gbh(S_0 - S_f) = \frac{1}{2}gh^2b'(x) + gbhS_0 - g\frac{Q^2n}{b(x)h}\left(\frac{b + 2h}{bh}\right)^{\frac{4}{3}}. \tag{6.87}$$

In order to study a possible balanced discretisation of (6.87) similar to the flux terms discretisation, a zero velocity steady state is assumed ("water at rest") with bed level and width variations (i.e $u = Q = 0$, h, z_b and b not constant and $h + z_b =$ constant).

Considering a discrete approximation of (6.87) of the form

$$\Delta x\tilde{D}_{j+\frac{1}{2}} = \left[\left(\frac{1}{2}g\tilde{h^2}\Delta b + g\bar{b}\bar{h}(\tilde{S}_0 - \tilde{S}_f)\right)\Delta x\right]_{j+\frac{1}{2}}, \tag{6.88}$$

equilibrium can be reached if there is no friction term ($S_f = 0$) and we take the averages (6.86) and $\tilde{S}_0\Delta x = \Delta h = h_{j+1} - h_j$.

When including the friction term, the (discrete) balance can not be achieved and the scheme will satisfy a "approximate C-property" instead of an " exact C-property" (see [3]). Possible choices of $\tilde{S}_f$ are simply an arithmetic average or one (bearing in mind other chosen averages) of the form

$$(\tilde{S}_f)_{j+\frac{1}{2}} = \left[-g\bar{b}\bar{h}\frac{Q^2n^2}{\tilde{A}^2}\left(\frac{\bar{b} + 2\bar{h}}{\bar{b}\bar{h}}\right)^{\frac{4}{3}}\right]_{j+\frac{1}{2}}. \tag{6.89}$$

This scheme was not fully implemented.

Chapter 7

Description of the Test Problems

The test problems studied were taken from MacDonald's [59] (Saint-Venant equations) and Wixcey's [103] (Euler equations) Ph.D. theses.

In [61], MacDonald *et al.* introduced a technique for constructing test problems for the steady Saint-Venant (with friction term included) with known analytical solutions including solutions with hydraulic jumps. The test problems were created using an "inverse" approach where the bed slope analytical expression is determined from a desired water depth and specified flow rate. This technique was also employed in later work (see [60, 59, 63, 64]). Although the method described does not provide an analytical expression for the bed level (the integral of the bed slope) it is possible to use numerical methods to obtain this value (see [62]). In this thesis the test problems regarding the Saint-Venant equations were taken from the nonprismatic channel test problems in [59].

The test problems for the Euler equations, a diverging section duct and a nozzle, were taken from [103]. Not all types of flow are analyzed in this thesis. Those we choose to study are presented in Section 7.2.

In the next sections, Sections 7.1 and 7.2, we describe in more detail the test problems studied.

7.1 Test Problems for the Steady Saint-Venant Equations

The test problems were chosen from those presented in [59] such that different types of flow (hydraulic jumps inclusive) are illustrated. These test problems are constructed in such a way that an analytical solution for the full steady Saint-Venant equations is known. The principle is that if a particular depth profile h is known, it is possible to compute (analytically) the bed slope S_0 that makes this profile an actual solution of the steady equation (see [59] for more details).

The four test problems chosen illustrate different flow features and correspond to the flux of a flow in an open channel of rectangular cross-section and with variable breadth function given by

$$b(x) = 10 - 5\exp\left\{-10\left(\frac{x}{200} - \frac{1}{2}\right)^2\right\}. \tag{7.1}$$

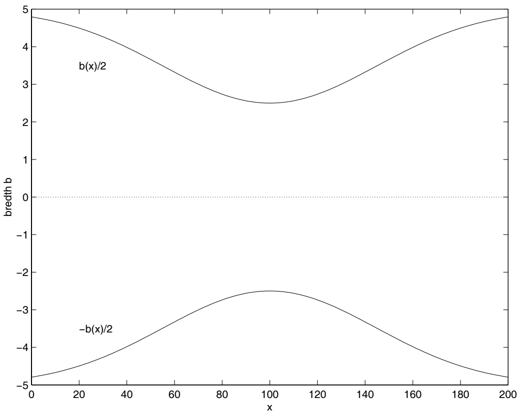

Figure 7.1: Graph of the horizontal cross-section of the channel

Manning's friction law was used with coefficient $n = 0.03$. The details of the test problems are given in Tables 7.1 and 7.2.

The test problems were chosen from those given in [59] by MacDonald (test problems 9-12 in Appendix B) and correspond to different types of flow. The boundary conditions needed (see section 2.2.5) are given in table 7.2. Other boundary conditions (if needed) are taken as the values of the critical depth function h_c at the corresponding endpoints, as suggested by MacDonald [59].

The depth profile in test problem 1 is subcritical whereas in test problem 2 it is entirely supercritical. In test problem 3 the flow is subcritical until approximately one

Prob.	Type of flow	Analytical depth profile $\hat{h}$
1	Subcritical	$\hat{h}(x) = 0.9 + 0.3\exp\left(-20\left(\frac{x}{200} - \frac{1}{2}\right)^2\right)$
2	Supercritical	$\hat{h}(x) = 0.5 + 0.5\exp\left(-20\left(\frac{x}{200} - \frac{1}{2}\right)^2\right)$
3	Smooth Trans.	$\hat{h}(x) = 1.0 - 0.3\tanh\left(4\left(\frac{x}{200} - \frac{1}{3}\right)\right)$
4	Hydraulic Jump	$\hat{h}(x) = \begin{cases} 0.7 + 0.3\exp\left(\frac{x}{200} - 1\right) & x \leq 120 \\ \exp(-0.1(x-120))\sum_{i=0}^{2} k_i \left(\frac{x-120}{200-120}\right)^i + \phi(x) & x > 120 \end{cases}$
		where $k_0 = -.274406$, $k_1 = -.948343$, $k_2 = 4.89461$
		and $\phi(x) = 1.5\exp\left(0.1\left(\frac{x}{200} - 1\right)\right)$

Table 7.1: Test problem details: type of flow and analytical solution

Prob.	L/m	$Q/(m^3 s^{-1})$	h_{in}/m	h_{out}/m
1	200	20		0.902021
2	200	20	0.503369	
3	200	20		
4	200	20	0.7	1.49924

Table 7.2: Test problem details: length of channel, discharge and boundary conditions

third of the length of the channel and then changes smoothly to supercritical. In test problem 4, a hydraulic jump occurs at $x = 120$m and the depth profile jumps there from supercritical to subcritical.

7.2 Test problems for the Steady Euler Equations

In this section we describe some test problems for compressible flow in ducts with axi-symmetric geometries governed by the steady Euler equations (2.98-(2.100). The test problems studied are a diverging cone and a *de Laval nozzle*, i.e. a combination of a converging cone followed by a diverging cone connected through a location of minimum area called the *nozzle throat*. Nozzles are important for example, in the design of turbines and wind tunnels, since the gases passing through them increase velocity.

The flow is assumed to be isentropic on each streamline except when a normal shock happens (in a diverging section) but assumed isentropic before and after the shock. On each streamline where the flow is assumed to be isentropic we give analytical expressions depending on the speed u and on the total specific enthalpy H and entropy function K (H and K constants). These analytical expressions allows the drawing of graphs through the parametrization of the speed u whose limits are known (see (2.104)).

In Section 7.2.1 it is shown how to obtain expressions for some of the flow variables, depending on H, K and u and the particular values used for H and K are introduced. It is also shown how to obtain graphs of the flow variables as functions of u and the graphs of other relations between the flow variables by using u as an intermediary parameter. Then, in Section 7.2.2 we describe the test problems for isentropic flow in a nozzle. The particular boundary values yielding the different types of flow studied are given in each respective section.

7.2.1 The Analytical Expressions of the Flow Variables and Some Graphs

In Section 2.3.4 we used the fact that the quasi one-dimensional flow of a gas in a pipe when assumed to be isentropic in the whole domain is represented by a single streamline. Furthermore, we have seen that some quantities are constant for steady smooth flow, namely the entropy function K and that others like the enthalpy H and mass flow m are constant for either smooth or discontinuous flow. Other quantities remain constant only across shocks, like Q and the *flow stress* P which is defined by

$$P = p + \rho u^2. \tag{7.2}$$

Note that although the entropy function K is not constant across a shock (the flow is not isentropic there), it is constant before and after the shock.

The values of H and K are prescribed in the test problems. In particular, if there is a jump both the values of K before and after the jump will be given.

It is possible to obtain ρ as a function of H, K and u by using equation (2.89) yielding

$$\rho(H, K, u) = K^{\frac{1}{1-\gamma}} \left(\frac{\gamma - 1}{\gamma} \left(H - \frac{u^2}{2} \right) \right)^{\frac{1}{\gamma-1}}. \tag{7.3}$$

Combining equations (2.66) and (7.3) we obtain an expression for p, i.e.

$$p(H, K, u) = K^{\frac{1}{1-\gamma}} \left(\frac{\gamma - 1}{\gamma} \left(H - \frac{u^2}{2} \right) \right)^{\frac{\gamma}{\gamma-1}}. \tag{7.4}$$

Also, by using (7.4) and (7.3) in (2.54) and solving the resulting equation in order to get T we obtain

$$T(H, u) = \frac{\gamma - 1}{\gamma \mathcal{R}} \left(H - \frac{u^2}{2} \right). \tag{7.5}$$

From (2.87) and (7.2) by using equations (7.4) and (7.3) we obtain, respectively,

$$Q(H, K, u) = K^{\frac{1}{1-\gamma}} u \left(\frac{\gamma - 1}{\gamma} \left(H - \frac{u^2}{2} \right) \right)^{\frac{1}{\gamma-1}}. \tag{7.6}$$

and

$$P(H, K, u) = K^{\frac{1}{1-\gamma}} \left[\frac{\gamma - 1}{\gamma} \left(H - \frac{u^2}{2} \right) + u^2 \right] \left[\frac{\gamma - 1}{\gamma} \left(H - \frac{u^2}{2} \right) \right]^{\frac{1}{\gamma-1}}. \tag{7.7}$$

In order for the expressions (7.3)-(7.7) to define real quantities the velocity on the streamline must satisfy $u \leq (2H)^{1/2} = u_{\max}$.

Using the value of the critical speed (defined in Section 2.3.4 and automatically defined if we prescribe γ and H) it is possible to obtain the critical values of other flow variables simply by taking $u = C_*$ in equations (7.3)-(7.7). These critical values are denoted p_*, T_*, ρ_*, Q_* and P_*, respectively.

We choose for the case of isentropic flow the same (approximate) values for H and K as given in [82] (p. 364) and [103]. Hence a representative streamline for which the magnitude of the entropy function (constant) is specified as the one at standard temperature (273K) and standard pressure (1.01×10^5Nm^{-}2), namely

$$K = 7.08 \times 10^4 \tag{7.8}$$

and the total enthalpy value as that at standard temperature and zero fluid speed, i.e.

$$H = 2.74 \times 10^5. \tag{7.9}$$

These values of H and K yield the maximum speed

$$u_{\max} = 740.3 \tag{7.10}$$

and the critical speed

$$c_* = 302.5 \tag{7.11}$$

(these were the approximate values taken in [103] and are the ones taken here unless otherwise stated).

As explained before it is now possible to obtain the critical values for the flow variables by substituting $u = C_*$ into equations (7.3)-(7.7). Hence we obtain

$$p_* = 53100.1370 \tag{7.12}$$

$$T_* = 227.2659 \tag{7.13}$$

$$\rho_* = 0.8143 \tag{7.14}$$

$$Q_* = 246.3112 \tag{7.15}$$

$$P_* = 127609.2860 \tag{7.16}$$

(the last value is slightly different from the one presented in [103]).

For smooth flow it is possible to draw graphs of the flow variables p, T, ρ, Q and P versus u by taking a parametrization of u knowing that $0 \leq u \leq u_{\max}$. For example

$$u_j = j * u_{\max}/N \quad j = 1, 2, \ldots, N. \tag{7.17}$$

This form of building the graphs can also be found in [83] and will be used in Section 7.2.2.

Other graphs can be drawn between the flow variables, e.g. a graph of p against Q, but in this case a parametrization for the speed u like (7.17) has to be used as an auxiliary step. For this set of values of u we obtain the corresponding set of values of p and Q given by (7.4) and (7.6) and plot them, one versus the other.

If there is a (stationary) normal shock Q, P and H are conserved across the shock but K will increase in magnitude. In [103], Wixcey shows how, knowing the values of K before and after the shock, it is possible to obtain graphically values of the flow variables Q and P at the shock and the jumps in the remaining variables ρ, u and T. This is based in the work of Sewell [81].

7.2.2 The Gas Flow Test Problems

The first test problem we present is of gas flow in a *de Laval* nozzle and was considered in [103]. A diverging duct problem can be considered by studying the diverging section of the nozzle. This is a particular test problem where the duct is not smooth at the throat. Other test problems may be chosen which do not have this trait.

The de Laval Nozzle considered, is defined by the cross-section area function

$$A(x) = \begin{cases} 1.1 - 0.125x & 0.0 \leq x \leq 0.8 \\ \frac{2.6}{3.0} + \frac{x}{6} & 0.8 \leq x \leq 2.0 \end{cases} \tag{7.18}$$

where $0.0 \leq x \leq 2.0$, the inlet occurring at $x = 0$, the outlet at $x = 2.0$ and the throat happening at $x = 0.8$ (see Fig. 7.2).

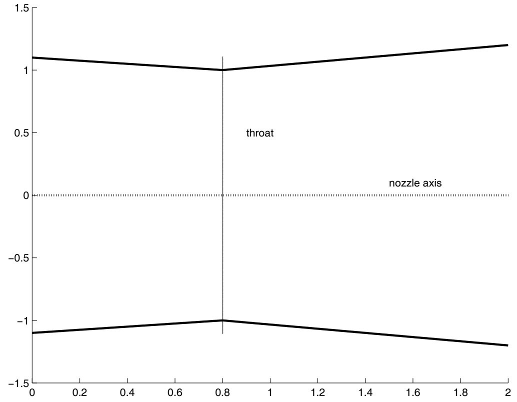

Figure 7.2: Graph of the de Laval nozzle

From the definition of $A(x)$ it is to see that

$$A_{in} = 1.1$$

$$A_t = 1.0$$

$$A_{out} = 1.2 \tag{7.19}$$

If the mass flow at inlet, Q_{in}, is prescribed (under certain restrictions to prevent *choked* flow) then the magnitude of the mass flow at outlet, Q_{out}, can also be computed since $m = AQ$ is constant and we have

$$Q_{out} = \frac{Q_{in} * A_{in}}{A_{out}}. \tag{7.20}$$

Another way of thinking is to give the value of m. In fact, for steady flow we have

$$A_{out}Q_{out} = Q_{in} * A_{in} = A_t * Q_t = m. \tag{7.21}$$

Hence, knowing the shape of the nozzle (i.e. $A(x)$) it is easy to compute Q_{in}, Q_{out} and Q_t.

Furthermore since $m = AQ$ is constant (see equation(2.86)), we see that if Q has a maximum, Q_*, there exists a corresponding cross-section area A_* through which the flow may take place. This can only happen (if it happens) at the outlet of a convergent nozzle (place of minimum area) and that place corresponds to the throat of the de Laval nozzle ($A_* = A_t$).(A de Laval nozzle can be thought of as a convergent cone connected to a diverging cone through a throat.) Otherwise, a flow cannot theoretically exist and the nozzle is said to be *choked*.

This implies a restriction on the value of the mass flow Q at inlet, i.e.

$$Q_{in} = \frac{Q_t * A_t}{A_{in}} \leq \frac{A_t * Q_*}{A_{in}}. \tag{7.22}$$

Using the values of the total specific enthalpy H and of the constant entropy function K given by (7.9) and (7.8), respectively, it is possible to draw the graphs of the flow variables, versus u (see Fig. 7.3) or between other flow variables, for example, ρ and p versus Q (see Fig. 7.4). In order to draw these graphs we use a parametrization of u (as an auxiliary step in the latter case).

Note that in order to draw the graphs of the flow variables the computed value of $u_{\max}$ we use is slightly smaller and more accurate than the one in [103].

From the graph of Q versus u (only needed in the convergent part of the nozzle) it is possible to see that there are two possible values for u_{in} corresponding to a given value Q_{in}, one corresponding to subcritical flow at inlet and the other corresponding to supercritical flow at inlet. In order to visualise this, just draw a line in the fourth

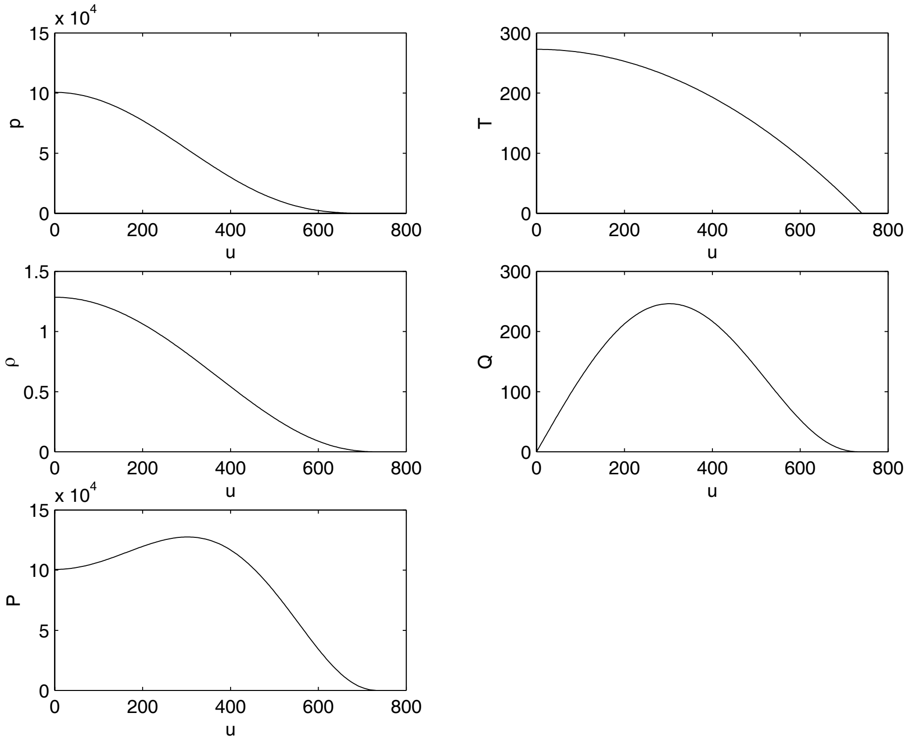

Figure 7.3: Graphs of flow variables vs. u $(N = 201)$

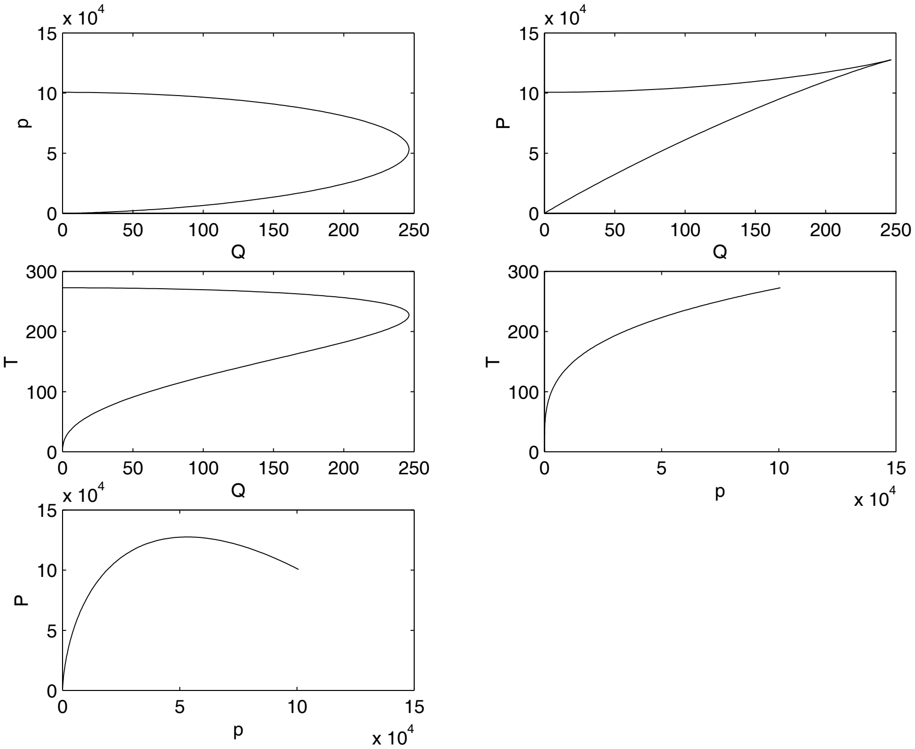

Figure 7.4: Graphs of some flow variables vs. Q or P $(N = 201)$

graph of Fig. 7.3 parallel to the x-axis at the height equal to the given inlet mass flow magnitude. The former value of u_{in} is the one we consider since in our test cases we only consider flow that is subcritical at inlet. Instead of using a graph, the two referred values of u can be obtained by numerically solving the equation

$$Q_{in} = Q(H, K, u) \tag{7.23}$$

by using, for example, the Newton method with adequate initial iterations (see [103] for more details). Note that the expression for $Q(H, K, u)$ is given by equation (7.6).

The test problems were chosen from those given in [103] and reflect different types of smooth flow. These types of flow are completely described by drawing the graphs between the flow variables.

As we have seen, for a given nozzle, there is a maximum value of the mass flow at inlet that can be specified (see (7.22)). Choosing the inlet mass flow Q_{in} such that the mass flow at the throat Q_t is less than the critical value corresponding to specify the mass flow at inlet such that the inequality in (7.22) is satisfied strictly. In this case, the flow is subsonic at the throat and also throughout the divergent cone with its behaviour being determined by the pressure/density at inlet. Therefore the flow is wholly subsonic (could be wholly supersonic if the pressure/density at inlet were supersonic, but this case is not considered in the thesis). Usually the variable prescribed discriminating the type of flow is the pressure but theoretically the density is a possible choice as well. Given one it is possible to get the other since pressure and density are related through the isentropic equation of state. In our case, since we chose to reduce, in Section 2.3.4, the steady Euler equations to an ODE in the dependent variable ρ, boundary values for the density are needed and its choice will determine the type of flow. It is relatively easy, though, to obtain the corresponding values for the pressure.

In Fig. 7.5 we show how it is possible to find the values for the density ρ at inlet, at the throat and at outlet corresponding to the mass flow value prescribed at inlet, Q_{in} and computed at the throat Q_t and at outlet Q_{out}.

Although it is common to choose the pressure at outlet to differentiate certain types of flow, here we choose to use the density ρ instead since we also chose to reduce, in Section 2.3.4, the steady Euler equations to an ODE in the dependent variable ρ. The correspondent values for the pressure can be obtained either through the isentropic equation of state (2.66) or by a graph (see Fig. 7.6).

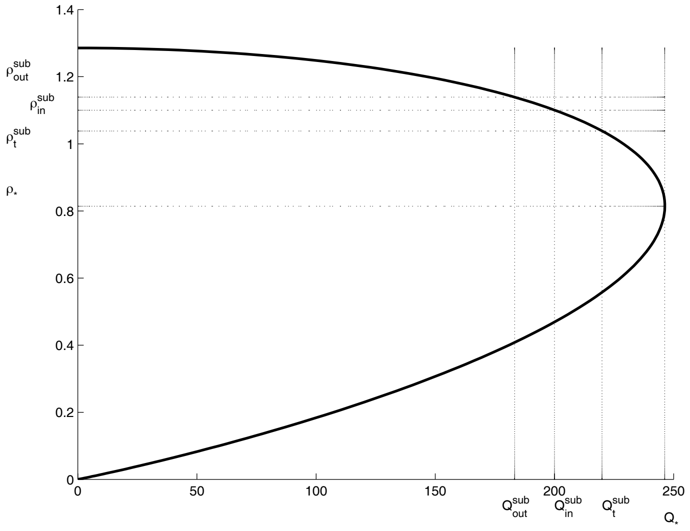

Figure 7.5: The relationship between the density and the mass flow in subsonic flow

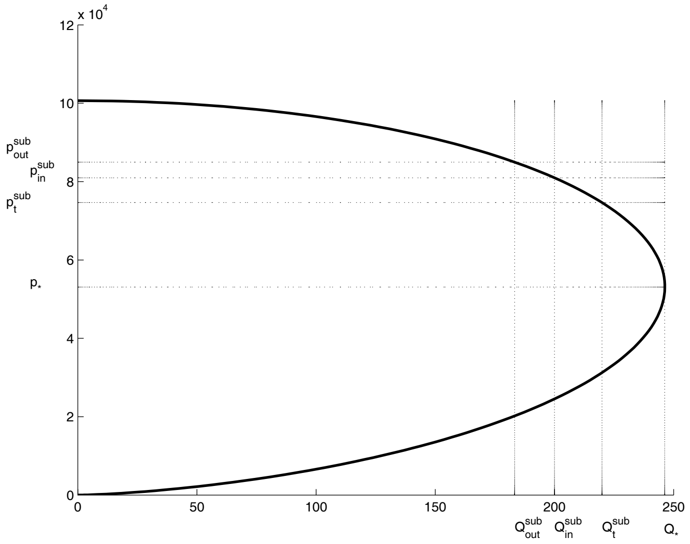

Figure 7.6: The relationship between the pressure and the mass flow in subsonic flow

In Table 7.3 we present the values of some of the flow variables.

It is possible to draw these graphs for other types of flow, like for example critical flow (see Fig. 7.7 and Fig. 7.8).

The implementation of an algorithm to these test problems was not concluded. The discretisation of the scalar equation obtained from the Euler equations in Chapter 2 is more complicated that the scalar equation we dealt with in the Saint-Venant problem since the flux function depends on the constant K which takes different values across a

Type of flow	Inlet	Throat	Outlet	Notes
Subsonic flow	$Q_{in} = 200$	$Q_t = 220$	$Q_{out} = 183.333$	$Q_t < Q_* = 246.311$
	$\rho_{in} = 1.10065$	$\rho_t = 1.03828$	$\rho_{out} = 1.13905$	$\rho_* = 0.814252$
	$u_{in} = 181.71$	$u_t = 211.889$	$u_{out} = 160.952$	$C_* = 302.5$
	$p_{in} = 80973.7$	$p_t = 74623$	$p_{out} = 84956.2$	$p_* = 53100.1$

Table 7.3: Theoretical values of some flow variables for isentropic wholly subsonic flow

shock. Two possibilities of overcoming this problem were thought. The first one was to make use of the jump conditions in a new algorithm but the idea was not fully developed. The second idea is simply to try to get a simpler scalar equation in the reduction process, without the K appearing explicitly in the definition of the flux function and this latter approach is under study.

Type of flow	Inlet	Throat	Outlet
Critical flow	$Q_{in} = 223.919$	$Q_t = Q_* = 246.311$	$Q_{out} = 205.259364$
critical subsonic	$\rho_{in}^{sub} = 1.02262$	$\rho_* = 0.814252$	$\rho_{out}^{sub} = 1.0864$
		$C_* = 302.5$	
	$p_{in}^{sub} = 73051.8$	$p_* = 53100.1$	$p_{out}^{sub} = 79509.5$
smooth transition	$\rho_{in}^{sub} = 1.02262$	$\rho_* = 0.814252$	$\rho_{out}^{sup} = .490112$
		$C_* = 302.5$	
	$p_{in}^{sub} = 73051.8$	$p_* = 53100.1$	$p_{out}^{sup} = 26088.3$

Table 7.4: Theoretical values of some flow variables for isentropic wholly subsonic flow

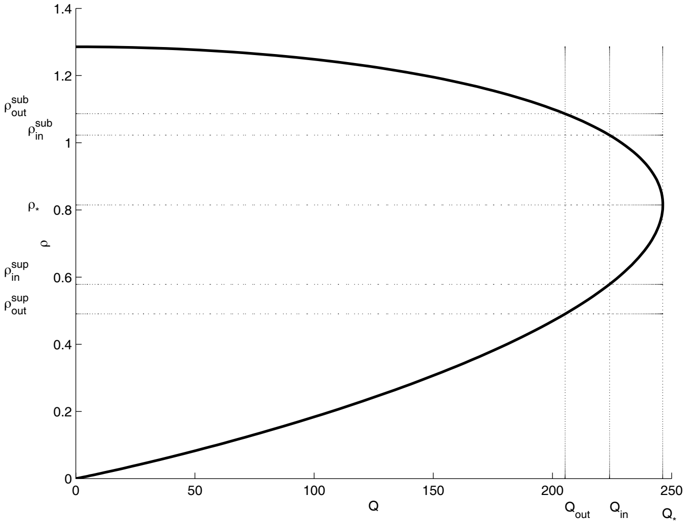

Figure 7.7: The relationship between the density and the mass flow in critical flow

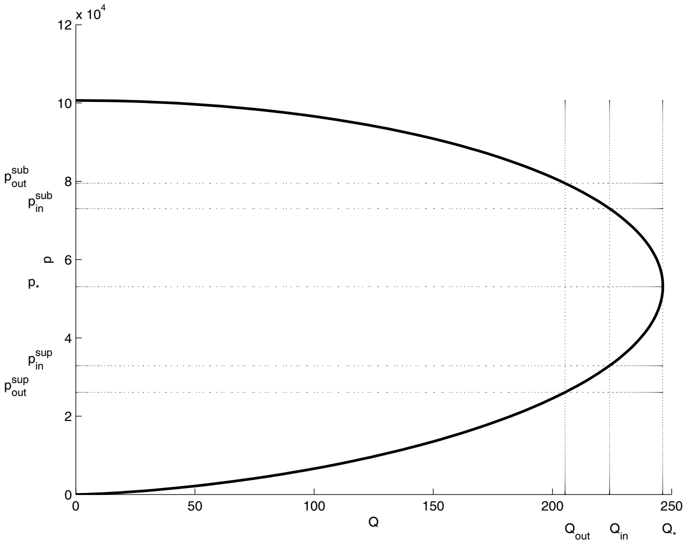

Figure 7.8: The relationship between the pressure and the mass flow in critical flow

Chapter 8

Results and Discussion

In this Chapter we present and discuss some results obtained by applying schemes discussed in Chapter 6 to scalar problems arising from steady conservation laws. The results were obtained under a CFL condition on the derivative of the flux, restricting the allowed (pseudo) time step but where the source term is not considered. If the source terms are dominant, though, one may have to consider a semi-implicit treatment of the source terms. The steady conservation laws studied in the thesis are nonhomogeneous and, thus, the presence of the source terms is an item to take in account. Furthermore, since the conservation laws with source terms lead to curved characteristics where the solution changes along them, a numerical domain of solution must be considered containing the analytical domain of solution. Otherwise the numerical solution is not useful to approximate a true solution that may fall out of the numerical domain of dependence considered. Hence, naturally, a CFL condition arises which includes a condition on the slope of the characteristic and yields a condition on the time step which in the theory presented in the thesis, corresponds to a restriction on the pseudo time iteration (a contraction mapping theorem may be used to prove convergence of the pseudo-time iteration process [59]). These considerations led us to build numerical algorithms with a CFL condition implemented. The Courant number was taken to be 0.9.

Moreover, a measure of accuracy (in the L_2 norm) for both Engquist-Osher and Roe schemes was used. The criterion of convergence for the iterative method we considered is

$$\sqrt{\frac{1}{N+1}\sum_{j=0}^{N}\left(\frac{h_j^{n+1}-h_j^{n}}{\Delta t}\right)^2} < TOL \tag{8.1}$$

with a tolerance $TOL = 10^{-8}$. The initial guess for the time-stepping was taken to be the one considered in [59] for the Saint-Venant equations, i.e. a linear profile of the dependent variable joining the end values γ_0 and γ_1.

The approximate solutions obtained from applying the numerical schemes described in Chapter 6 to the Saint-Venant test problems of Tables 7.1 and Table 7.2 and some error graphs are shown in Figures 8.1-8.29. These test problems feature different times of flow and a graph in each set of four graphs grouped together represents one of the test problems, being ordered from left to right, with test problem 1 placed on the top left-hand corner.

We recall that the solution to problem 1 represents entirely subcritical flow and has a depth profile with a hump which is symmetric about the center of the reach (see, e.g. Fig. 8.1). The solution to problem 2 has a similar depth profile but the flow is supercritical throughout the entire reach of the channel. A smooth transition solution occurs in test problem 3, with the flow being subcritical until approximately one third of the distance along the channel and then smoothly becoming supercritical. The solution of test problem 4 has a hydraulic jump inside the channel.

Furthermore, the schemes described in Chapter 6, whose results are presented and discussed in this section, correspond to modifications of the Engquist-Osher scheme and the Roe scheme. Two approaches were taken to discretise the total derivative of a explicitly x-dependent flux function. In a direct approach, a finite volume discretisation of the total derivative is done whereas, in the indirect approach, the total derivative is split by using the chain rule (assuming smoothness) and the partial derivative with respect to the independent variable is treated like a source term. The remaining derivative terms are discretised as in the direct approach but computed at a fixed value of the independent variable x (see Chapter 6 for more details).

The application of the Engquist-Osher method we called Scheme 1 with a pointwise discretisation of the source term (see equations (6.14)-(6.17)) yield the results shown in Fig. 8.1 (note that N is the number of subintervals). In the figure are shown different results corresponding to the choices of $q = 0, 1, \frac{1}{2}$ (related to x discretisation, as explained in Section 6.2.1). The results show that a choice of $q = 1/2$ is overall yielding the best results in test problems 1 (subcritical) and 3 (smooth transition) and the second best in test problems 2 (supercritical) and 4 (hydraulic jump) after the choice $q = 0$.

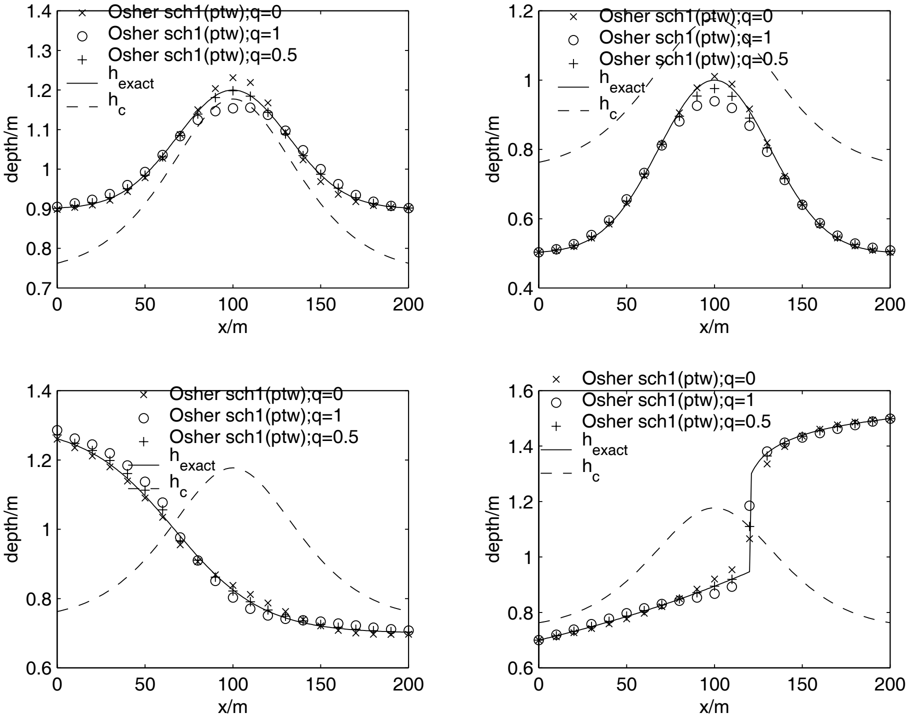

Figure 8.1: Engquist-Osher (direct approach) obtained by using scheme 1 with a pointwise discretisation of the source terms, $N = 20$ and $q = 0$, $q = 1$, $q = 1/2$ and $k = 0$).

We also tried a upwind discretisation of the source term D as explained in Section 6.2.1 and the results are shown in Fig. 8.2. A choice of the parameter $q = 0$ provides the better results in test problems 2, 3 and 4. In spite of the best results in test problem 1 being obtained whith $q = 0.5$, the convergence behaviour, as seen in the error graph of Fig. 8.3, indicates a rate of convergence slower than in the case $q = 1$ (second best results).

The graphs of the errors shown in Fig. 8.3- Fig. 8.6 were obtained with increasing number of subintervals and have a logarithmic scale. These graphs allow the comparison between the pointwise and the upwind discretisation of the source terms when using Engquist-Osher scheme 1. The upwind discretisation of the source terms does not provide the most accurate results in the different test problems studied. In the smooth transition case (test problem 3), it is the pointwise approach that provides the most accurate results (see Fig. 8.5) and the order of convergence of the upwind source discretisation is similar to the pointwise source discretisation. The upwind approach provides the most accurate results and shows a higher order of convergence than the pointwise approach in test

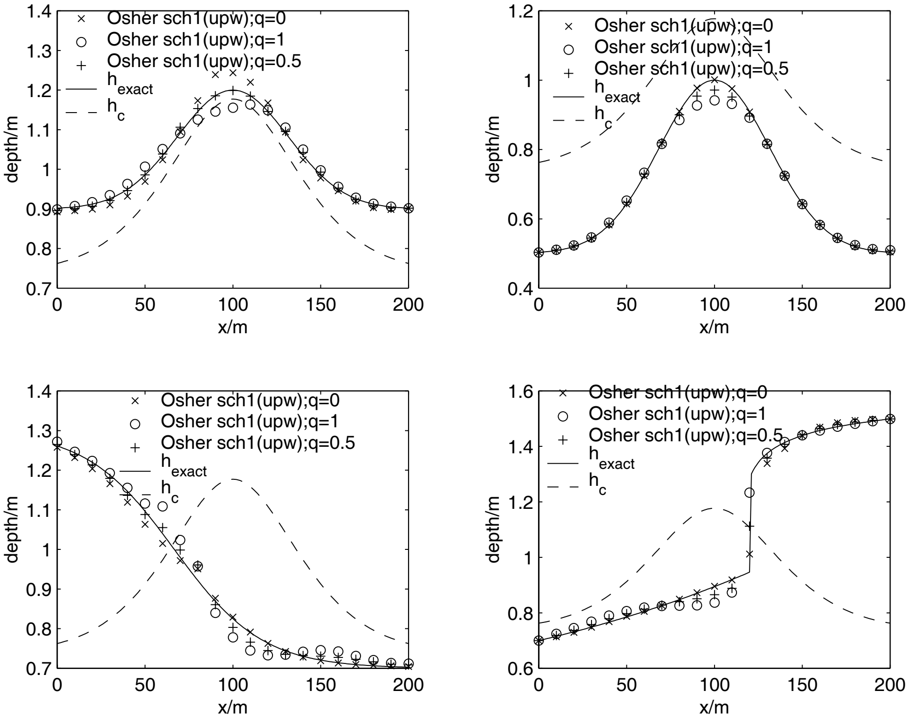

Figure 8.2: Engquist-Osher (direct approach) obtained by using scheme 1 with a upwind discretisation of the source terms, $N = 20$ and $q = 0$, $q = 1$, $q = 1/2$).

problem 2 (see Fig. 8.4). In relation to the choice of the parameter q in both source discretisation approaches, pointwise and upwind, the choice of $q = 1/2$ provides the most accurate results in test problem 1 (subcritical flow) (see Fig. 8.3); the choice $q = 1/2$ provides the most accurate results in test problem 2(see Fig. 8.4) and in test problem 4 (see Fig. 8.6). In the smooth transition test problem 3, the choice of either $q = 1/2$ or $q = 0$ with a pointwise discretisation of the source terms provides better results than all three choices of q with a upwind source discretisation (see Fig. 8.5).

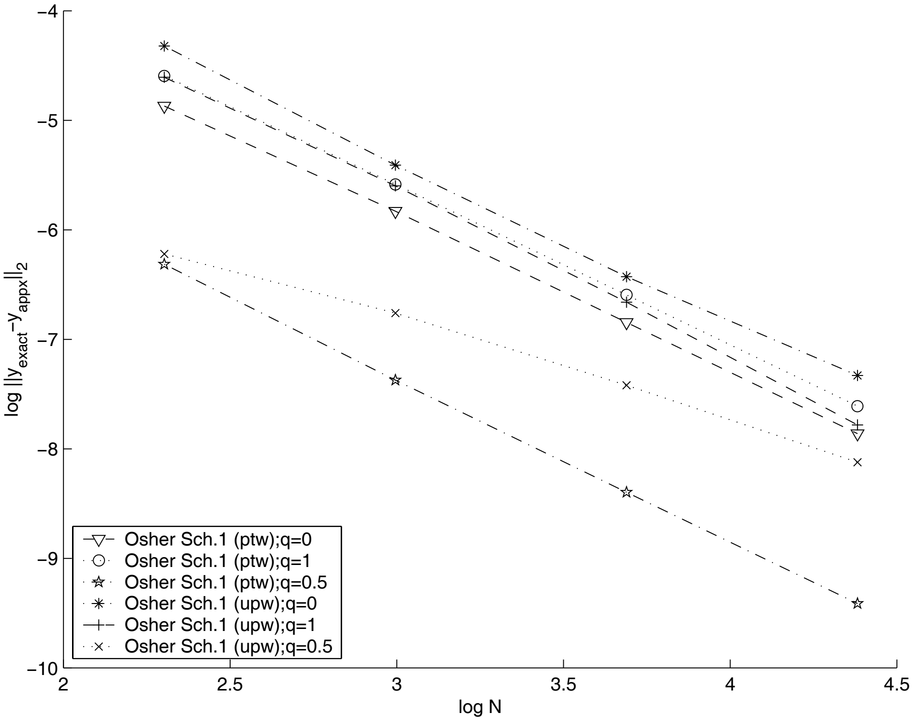

Figure 8.3: L_2 errors of the results obtained using scheme 1 in test problem 1 (subcritical flow) with $N = 10, 20, 40, 80$ subintervals.

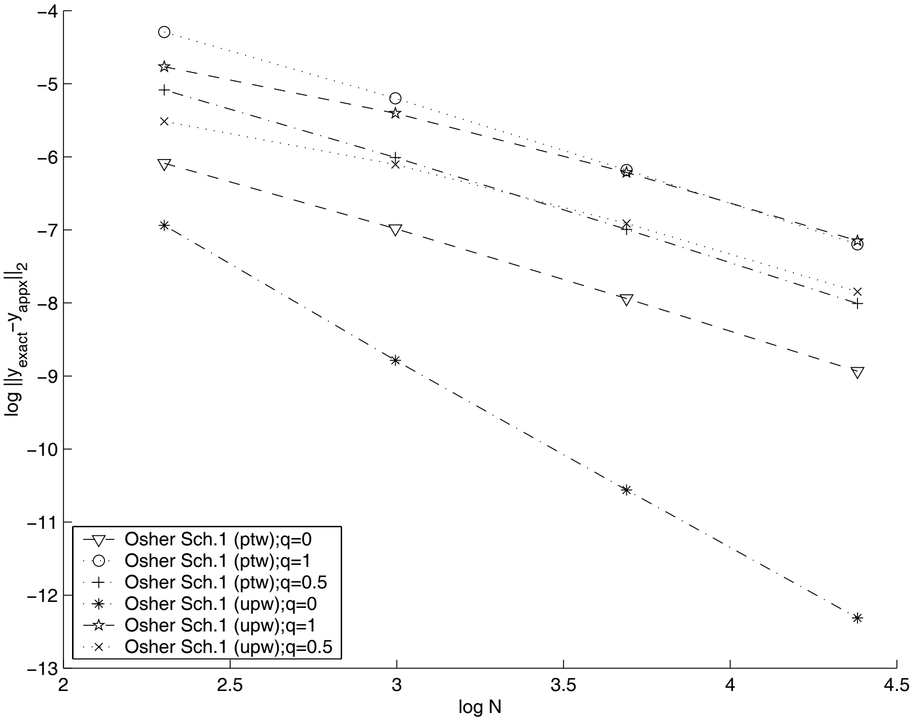

Figure 8.4: L_2 errors of the results obtained using scheme 1 in test problem 2 (supercritical flow) with $N = 10, 20, 40, 80$ subintervals.

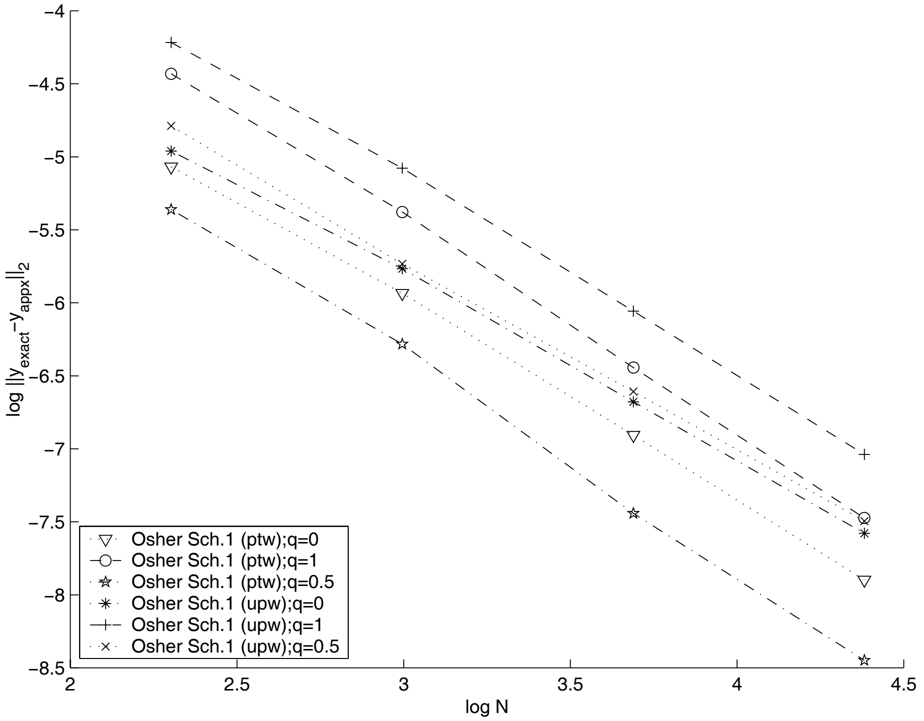

Figure 8.5: L_2 errors of the results obtained using scheme 1 in test problem 3 (smooth transition) with $N = 10, 20, 40, 80$ subintervals.

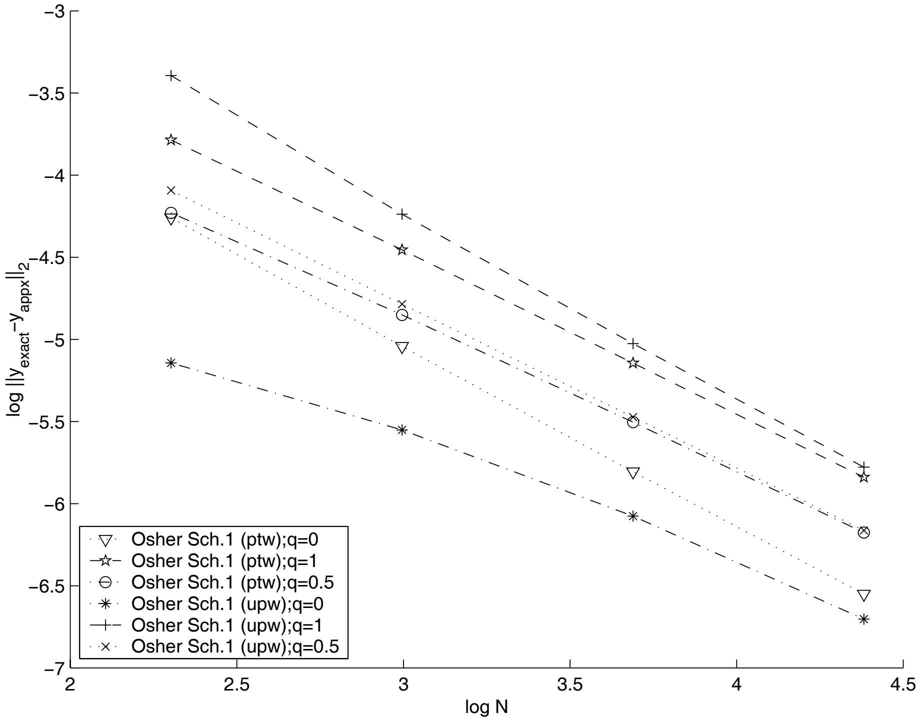

Figure 8.6: L_2 errors of the results obtained using scheme 1 in test problem 4 (hydraulic jump) with $N = 10, 20, 40, 80$ subintervals.

The solutions obtained from Roe scheme (direct approach) are less accurate than the solutions obtained form Engquist-Osher scheme 1, either when the source function is pointwise discretised or upwind discretised (see Fig. 8.7 and Fig. 8.8). The results shown in Fig. 8.7 and Fig. 8.8 were obtained without the use of an entropy fix. A modification of the well known entropy fix for flux functions of the form $F(w)$, which is described in the thesis, was tried for flux functions of the form $F(x, w)$ but did work. The resulting scheme does not deal well with the sonic transition crossing. A entropy fix is, in particular, very much in need in test problem 3 (a sonic transition) and also in test problem 1. In the latter problem, in spite of the flow being subcritical throughout the all domain, the sonic line is very near the exact solution we want to approximate. Even when a initial time stepping approximation which does not cross the sonic line was chosen, the iteration process yield approximations that crossed the sonic line. As an example, the numerical solution shown in Fig. 8.9 was obtained by using scheme 2 with a initial approximation that does not cross the sonic line (linear profile connecting the fixed right boundary condition and $h(0) = 1.7$, though subsequent approximations cross it.

We recall that the graphs corresponding to problem 1 in Fig. 8.7 and Fig. 8.8 were obtained with the initial approximation previously described in Section 7.1, that is, a linear profile connecting the right (fixed) boundary condition and the critical depth value at the left endpoint, thus, crossing the sonic line.

All the results shown for test problem 1 with a pointwise discretisation of the source function were achieved with a bigger tolerance, $TOL = 10^{-5}$ or $TOL = 10^{-4}$ (the latter when $q = 0$).

Although the experiments showed the need to use an entropy fix, a sonic transition did no occur in the test problems 2 and 4. In test problem 2 the initial approximation and the iterated approximations do not cross the sonic line and in test problem 4 the crossing verifies an entropy condition. Only the results obtained by using scheme 2 to solve test problem 2 will be used for comparison with other schemes.

For both upwind or pointwise discretisation of the source term, Scheme 2 renders a solution that is shifted to the right of the original exact solution and such that in the smooth transition test problem the results form a bump on the region of supercritical flow.

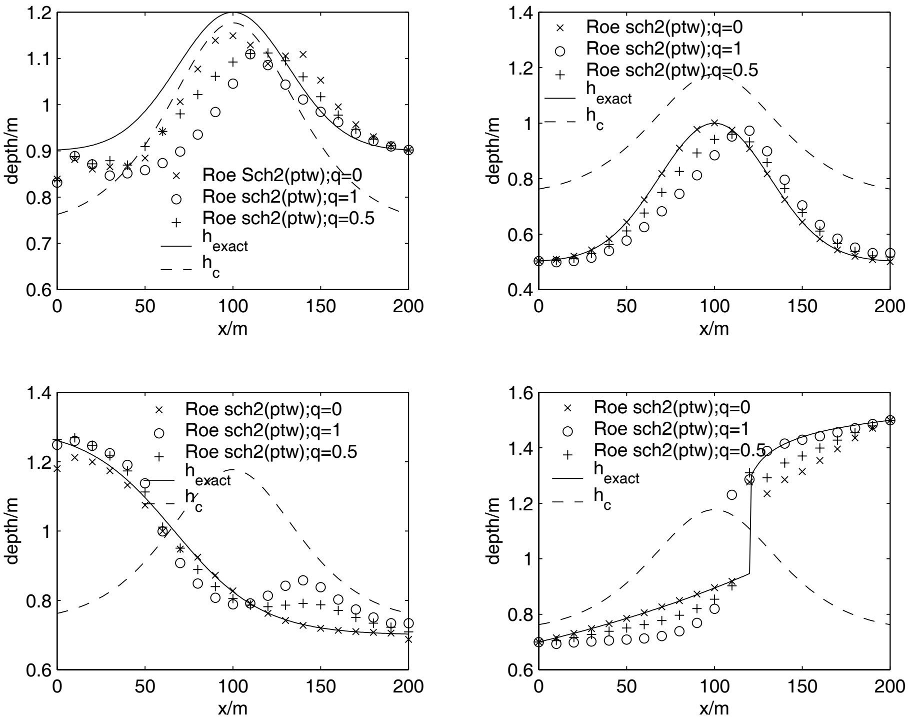

Figure 8.7: Roe solution (direct approach) obtained by using scheme 2 with a pointwise discretisation of the source term and $N = 20$ and $q = 0$, $q = 1$ and $q = 0.5$

In Fig. 8.10 we can see that decreasing the step size we still get poor accuracy and smeared results.

The importance of choosing the 'right' x evaluation (i.e. the best q) can be inferred from the analysis of the solution graphs (Fig. 8.7 - Fig. 8.8) and also from the error graph (Fig. 8.11).

For supercritical flow (test problem 2), a choice of $q = 0$ is the most adequate. This is confirmed by the results in test problem 4, before the jump occurs (flow is supercritical there). The choice of $q = 1$ being the most adequate for subcritical flow is not so easily seen since the scheme rendered a solution that is not well-behaved in the wholly subcritical test problem 1. Nevertheless, the results of test problem 4 (shock) show that the choice of $q = 1$ gives accurate results on the right-hand side of the jump, where the flow gets subcritical.

A comparison between a pointwise and a upwind discretisation of the source terms can be done by looking at the numerical solution graphs in Fig. 8.12 and the graphs of the errors provided by Fig. 8.11. Both approaches, pointwise and upwind discretisation

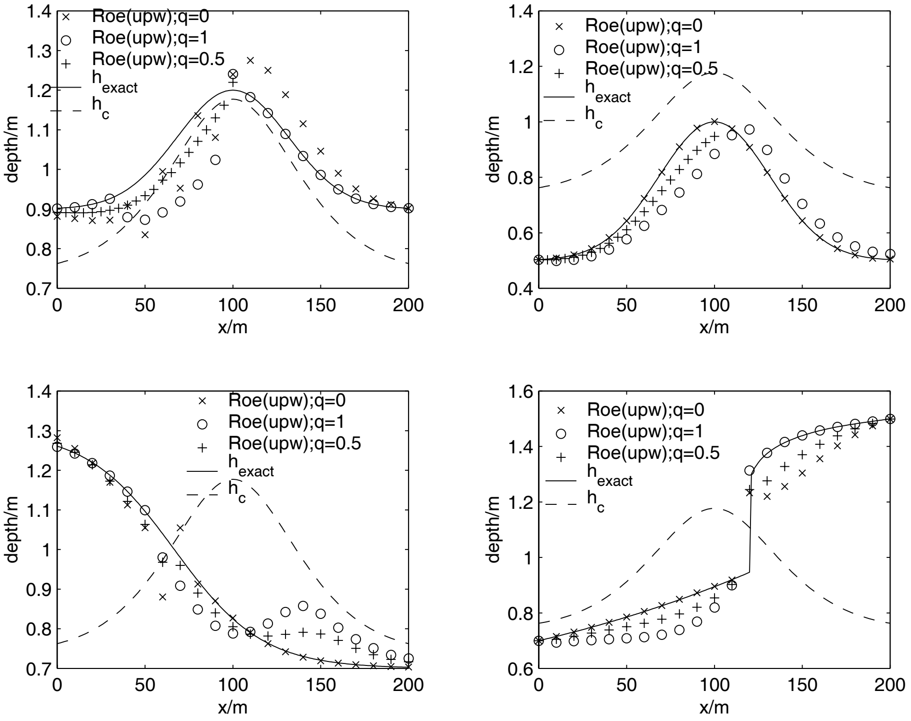

Figure 8.8: Roe solution (direct approach) obtained by using scheme 2 with an upwind discretisation of the source term and $N = 20$ and $q = 0$, $q = 1$ or $q = 0.5$.

of the source terms, provided very similar results and order of convergence with a choice of parameter $q = 0$.

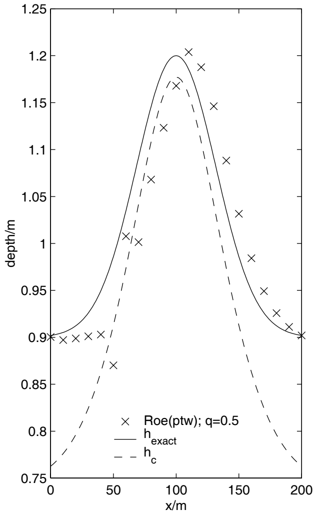

Figure 8.9: Roe solution (direct approach) of test problem 1 obtained by using scheme 2 with a pointwise and a upwind discretisation of the source term and $N = 20$ ($q = 0.5$) and a different (numerical) left boundary condition.

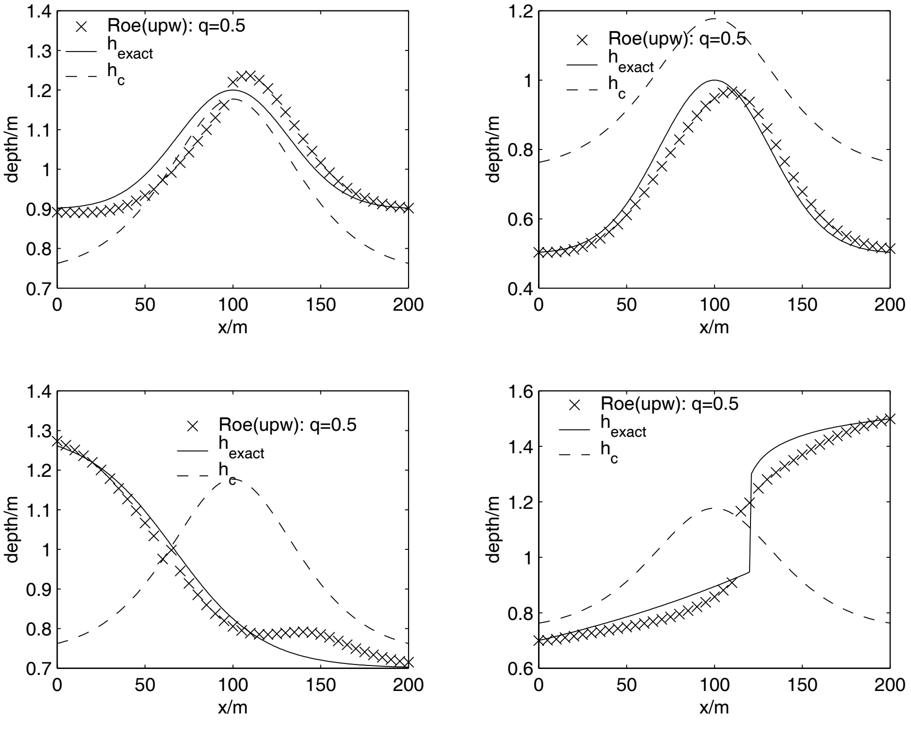

Figure 8.10: Roe solution (direct approach) obtained by using scheme 2 with a upwind discretisation of the source term and $N = 40$ and $q = \frac{1}{2}$.

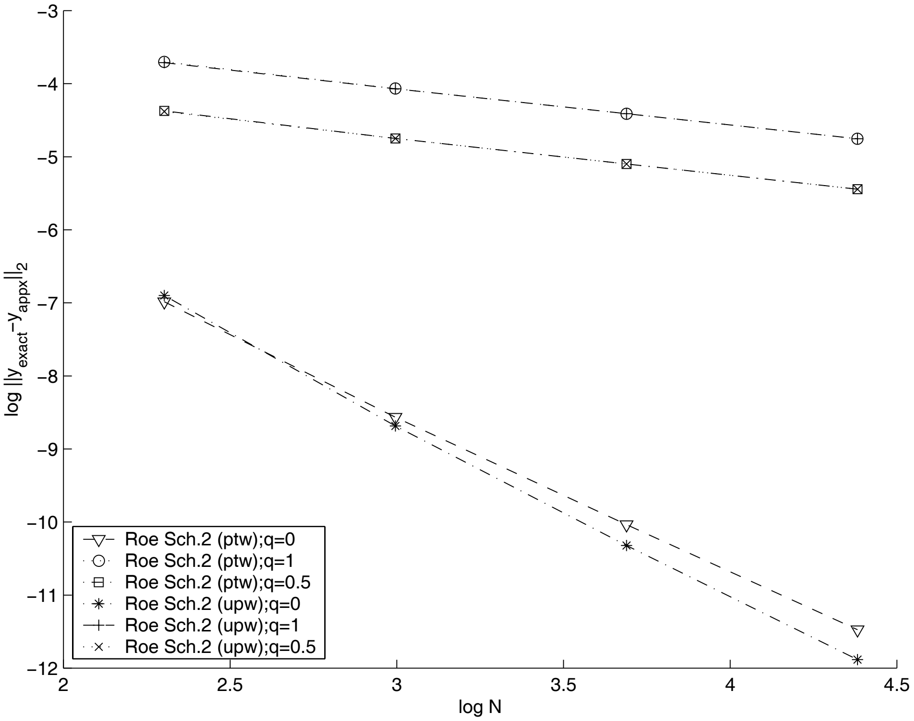

Figure 8.11: L_2 errors of the results obtained using scheme 2 in test problem 2.

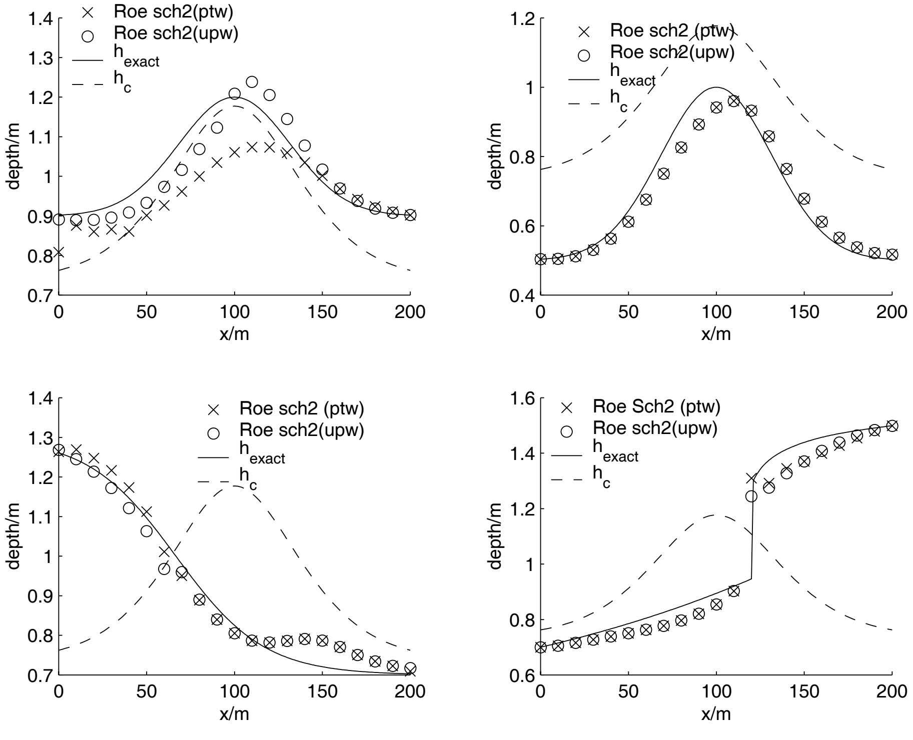

Figure 8.12: Roe scheme 1 (direct approach) with pointwise and upwind discretisation of source terms .

In Fig. 8.13 we show the results obtained using the FOU scheme 2*. We recall that this scheme is simply a version of scheme 2 where the total derivative of the flux function is discretised directly in a compact form without trying to express Δf in terms of the dependent variable, as a Roe-like approach does. In the particular case of a spatially dependent flux function like the one we are studying it seemed important to make the distinction. Nevertheless, the results are not so dissimilar from the scheme 2, previously discussed. We did not analyze this scheme any further and just show the results obtained in Fig. 8.13 which were obtained without the use of any entropy fix.

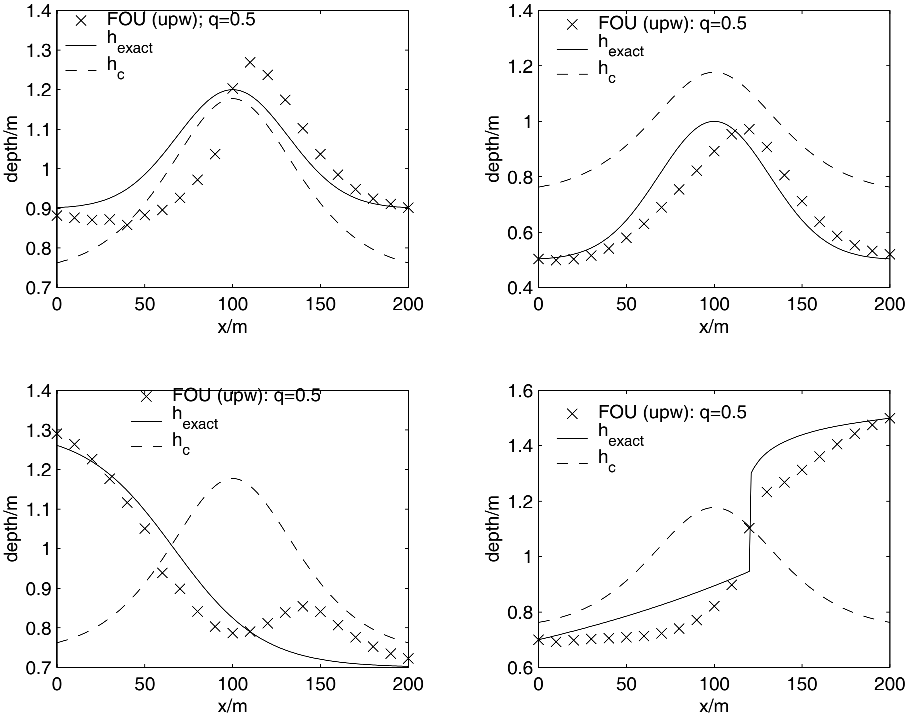

Figure 8.13: Solution (direct approach) obtained by using scheme 2* (FOU) with a upwind discretisation of the source term and $N = 20$ and $q = \frac{1}{2}$.

In Fig. 8.14 are plotted the results obtained by using $q = 1/2$ and the schemes 1 (Engquist-Osher) and 2 (Roe) with a pointwise discretisation of the source term. As expected from previous discussion, the Engquist-Osher scheme provides the most accurate results since Roe's scheme lacks an entropy fix.

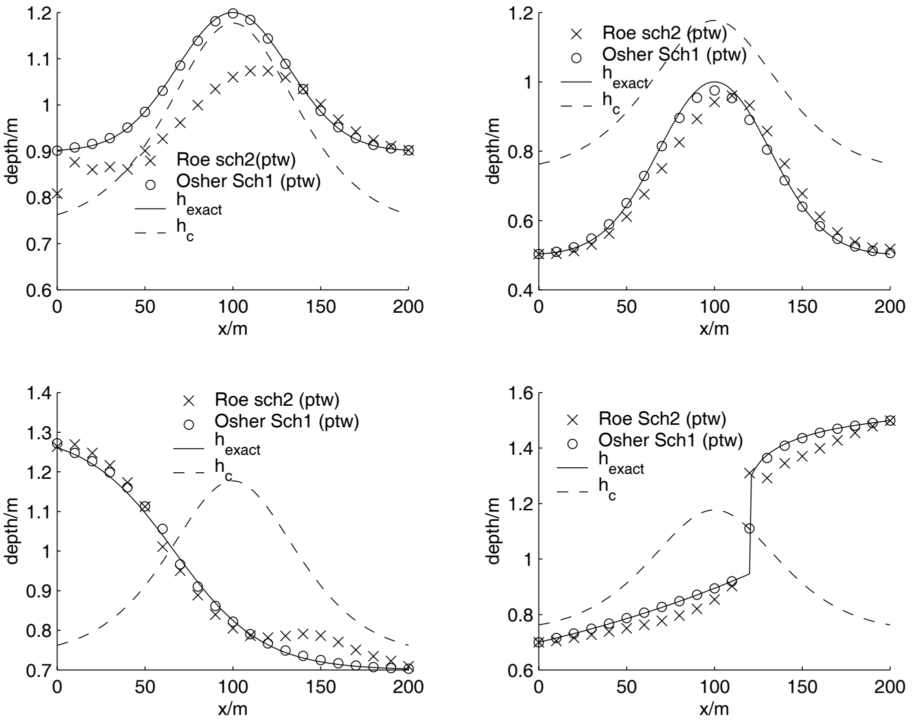

Figure 8.14: Engquist-Osher scheme versus Roe scheme results (direct approach) with source term pointwise discretised and $q = 1/2$.

The results of numerical schemes based on a indirect approach provide accurate solutions as well. We show results with different choices of discretisation of the source function and of the derivative term which, in this indirect approach, is treated like a source term.

In Fig. 8.15, the results obtained by using Engquist-Osher scheme 3 (indirect approach) are plotted together with the results obtained from the corresponding scheme in a direct approach (scheme 1) (with $q = 0$ and $k = 0$). The results shown correspond to a pointwise discretisation of the source term D and a centred (at half-point) discretisation of the derivative term V (only needed in scheme 3). The numerical schemes yield good accuracy results for all the four types of flow studied. Both schemes are accurate and a conclusion on the most accurate scheme in this case is answered by the graphs of the error (see Fig. 8.26 - Fig. 8.29). Those error graphs show that the scheme giving the best accuracy for supercritical and shock test problems is the one coming from a direct approach whereas for the smooth transition and subcritical test problems the scheme yielding the best accurate results is the one coming from an indirect approach.

The upwind Engquist-Osher scheme 3 was not fully implemented.

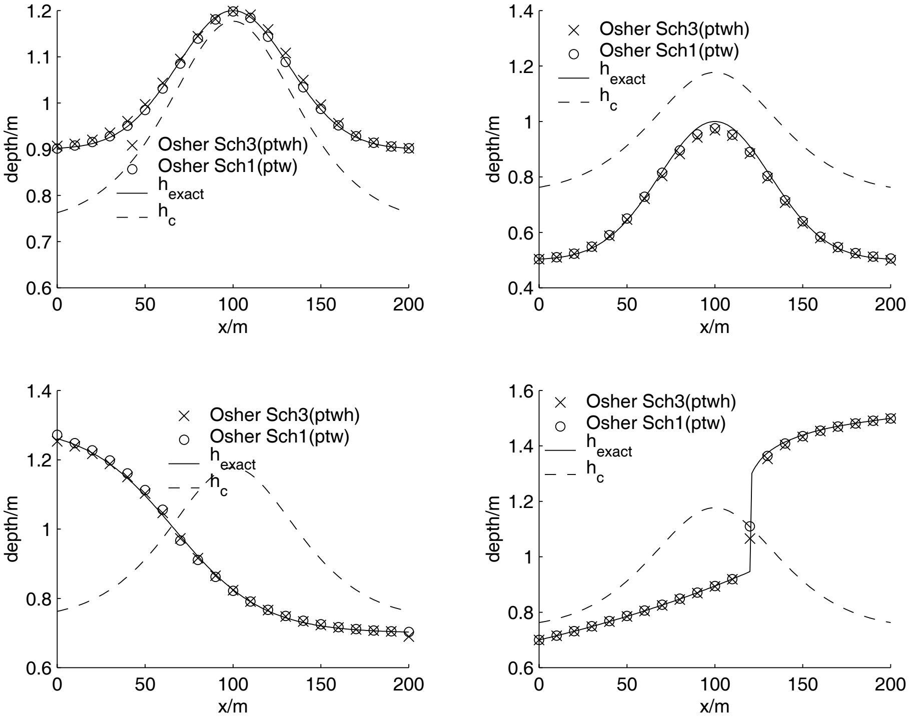

Figure 8.15: Results obtained by using the Engquist-Osher schemes 1 and 3 with a pointwise discretisation of the source term and a centred (half-point) discretisation of the derivative term V ($N = 20$)

The results obtained by using Roe's scheme 4 (indirect approach) with a pointwise approximation of the source term D and both a centred and half-centred discretisation of the derivative term V, are shown in Fig. 8.16 whereas the results coming from an upwind approximation of both source terms are shown in Fig. 8.17. Both schemes yield accurate results but the upwind scheme gives the better results and this can be seen particularly well in the subcritical and supercritical test problems (see Fig. 8.18).

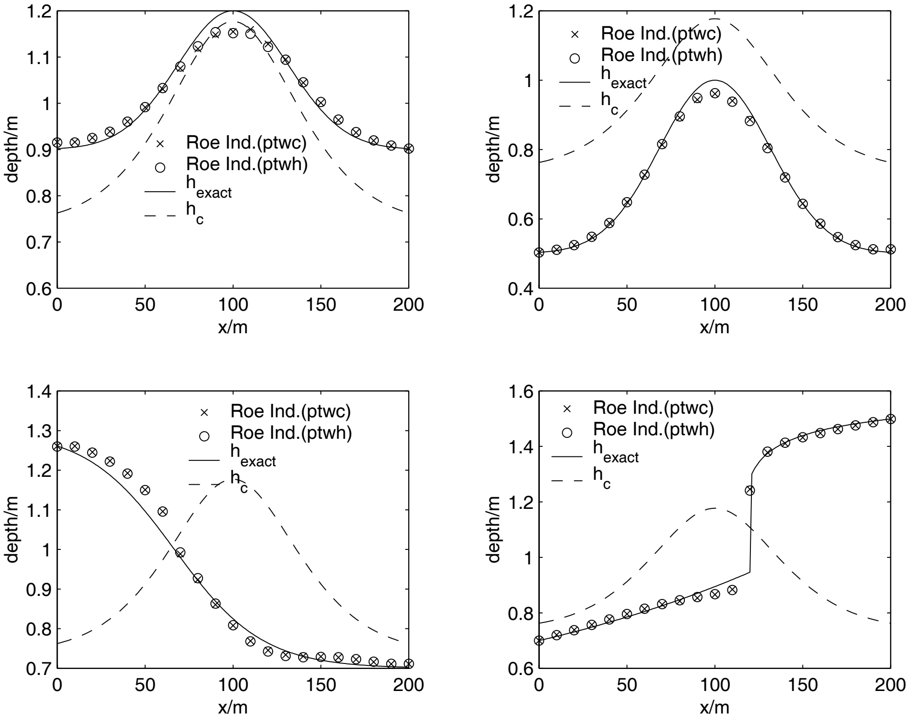

Figure 8.16: Results obtained using the Roe's scheme 4 (indirect approach) with a pointwise discretisation of the source term D and a centred and half-point discretisation of the derivative term V (respectively, ptwc and ptwh in legend) and $N = 20$ and $q = 0.5$.

The error graphs of the Roe scheme 4 solution in the different test problems studied are shown in Fig. 8.19 - Fig. 8.22. These graphs confirm that the upwind discretisation of the source terms yields the most accurate results in all the test problems. Furthermore, the order of convergence is improved in test problems 1 and 2 which correspond to, respectively, subcritical flow and supercritical flow throughout the whole domain (see Fig. 8.19 and Fig. 8.20). The smooth transition flow error graph shown in Fig. 8.21 reveals the fact that it is possible to get a higher order of convergence if the nonentropy satisfying sonic crossing is dealt with. The scheme's order of convergence decreased

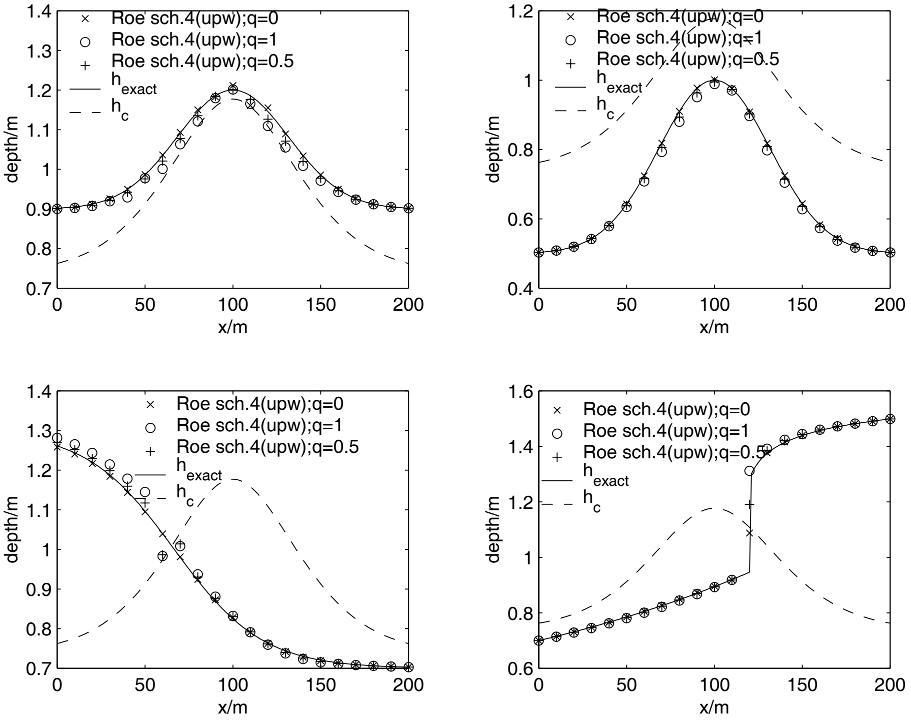

Figure 8.17: Results obtained using the Roe scheme 4 (indirect approach) with a upwind discretisation of both the source terms, D and V and $q = 0$, $q = 1$ and $q = 0.5$ ($N = 20$).

when a sonic transition occurred persistently in the iteration process when using $N = 40$ subintervals. For the hydraulic jump (test problem 4) all the schemes have a similar order the convergence, but the upwind scheme yields the most accurate results.

Comparing the possible choices of the parameter q in all the error graphs given in Fig. 8.19 - Fig. 8.22, the choice $q = 0$ yields the most accurate results in all test problems except in test problem 3 (smooth transition) where it is a choice of $q = 0.5$ that wins, providing the most accurate results in this particular case. It is worth mentioning that these results were obtained with the discretisation of the derivative term V also computed with $k = 0$, that is, the numerical flux function is computed at a fixed grid point and the derivative term V is computed with the corresponding (fixed) dependent variable (see Section 6.2.2).

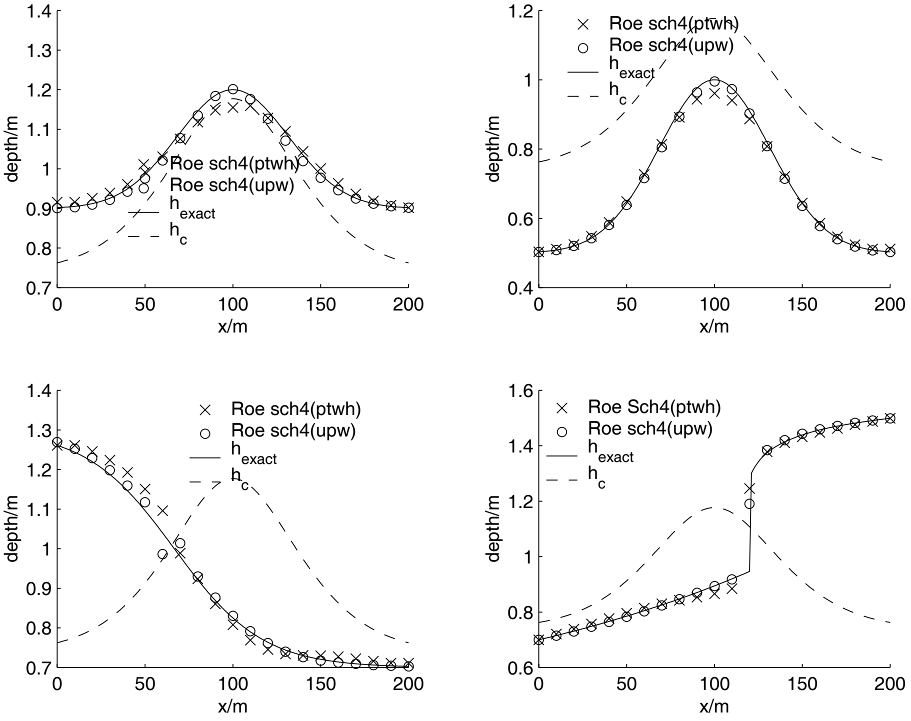

Figure 8.18: Roe's scheme 4 (indirect approach) with source terms pointwise and upwind discretised.

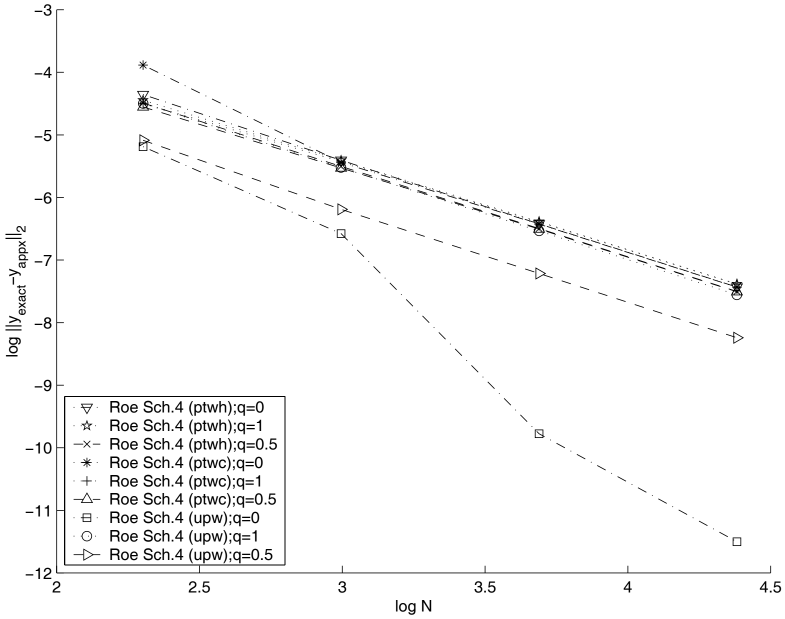

Figure 8.19: L_2 errors of the results obtained using scheme 4 in test problem 1 and $N = 10, 20, 40, 80$.

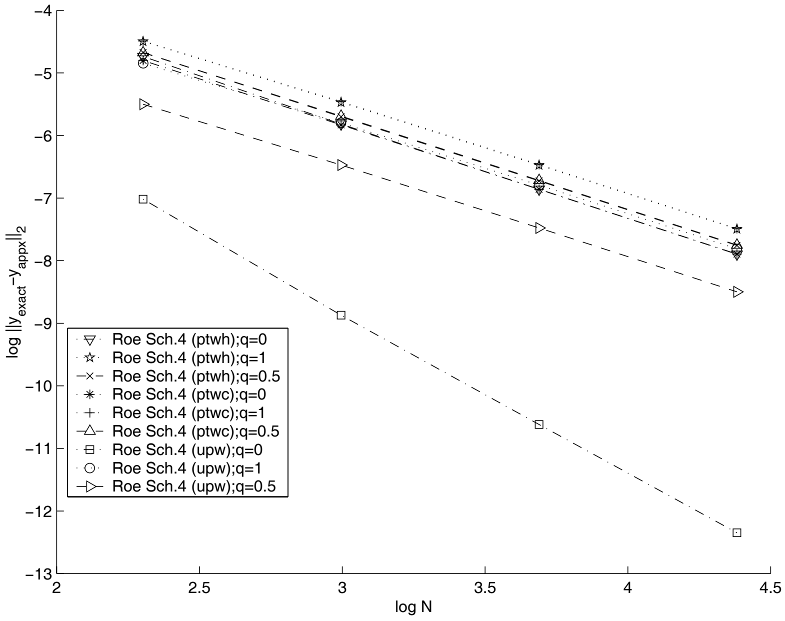

Figure 8.20: L_2 errors of the results obtained using scheme 4 in test problem 2 and $N = 10, 20, 40, 80$.

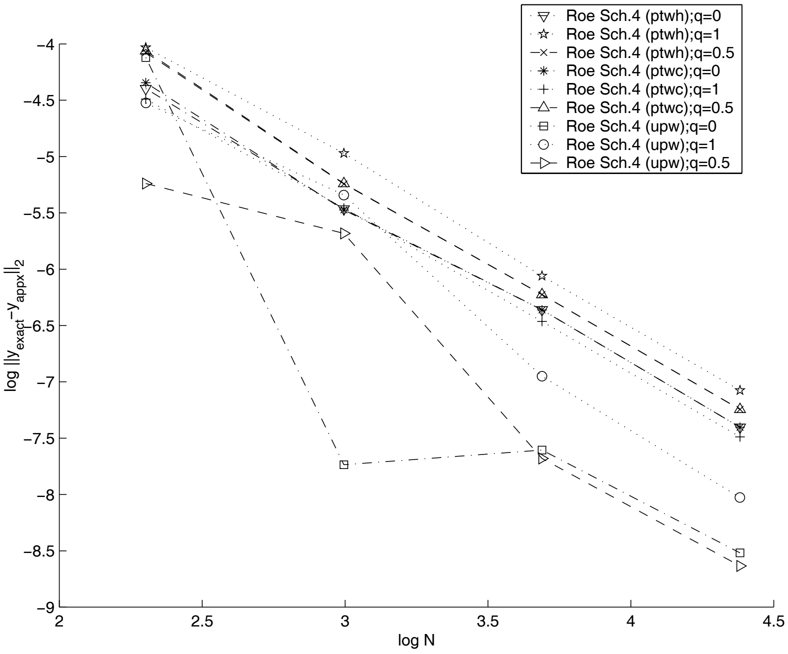

Figure 8.21: L_2 errors of the results obtained using scheme 4 in test problem 3 and $N = 10, 20, 40, 80$.

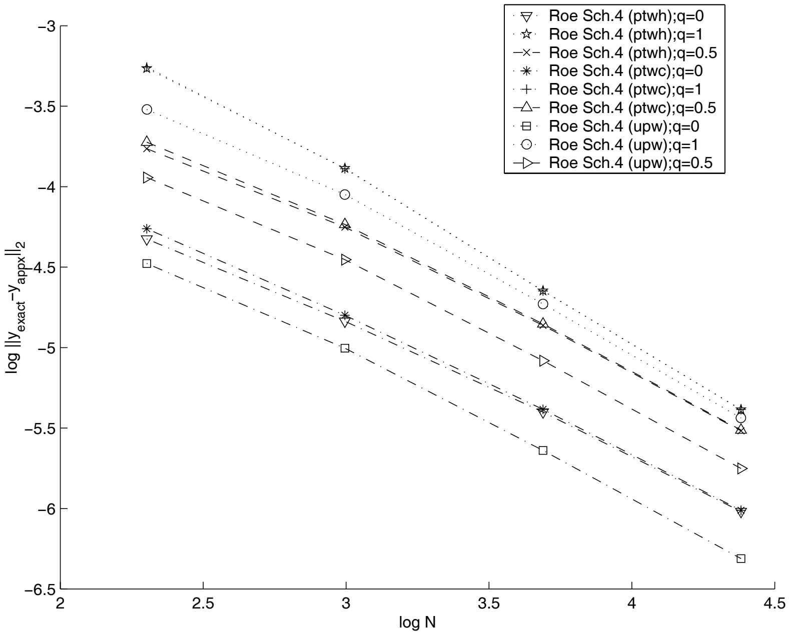

Figure 8.22: L_2 errors of the results obtained using scheme 4 in test problem 4 and $N = 10, 20, 40, 80$.

In order to be able to do a comparison between some of the schemes implemented in this work, the solutions of different numerical schemes are plotted against each other and some error graphs are drawn as well.

In Fig. 8.23 are shown the results obtained by using the Roe scheme in the direct and indirect approaches, respectively with a pointwise discretisation of the source terms and a centred (half-point) discretisation of the derivative V (the latter only needed in the indirect approach). From the graphs it can be seen that the most accurate results are obtained with the Roe scheme 4 (indirect approach).

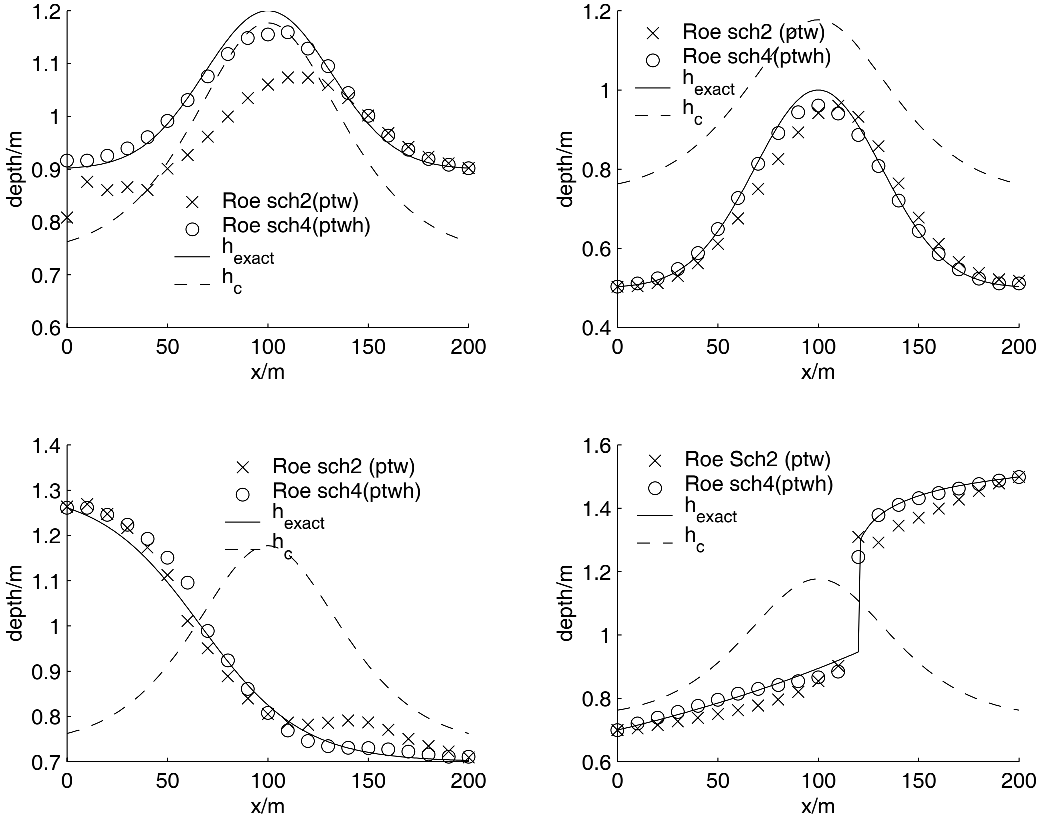

Figure 8.23: Roe scheme in direct approach versus Roe scheme in indirect approach with source terms pointwise discretised.

The comparison between results obtained from the Roe scheme in direct and indirect approaches with source terms upwind discretised is given in Fig. 8.24. Again, the indirect approach gives more accurate results.

In Fig. 8.18 the numerical results obtained from Roe's scheme 4 with a pointwise discretisation of the source terms (ptw) and a centred (half-point) discretisation of the

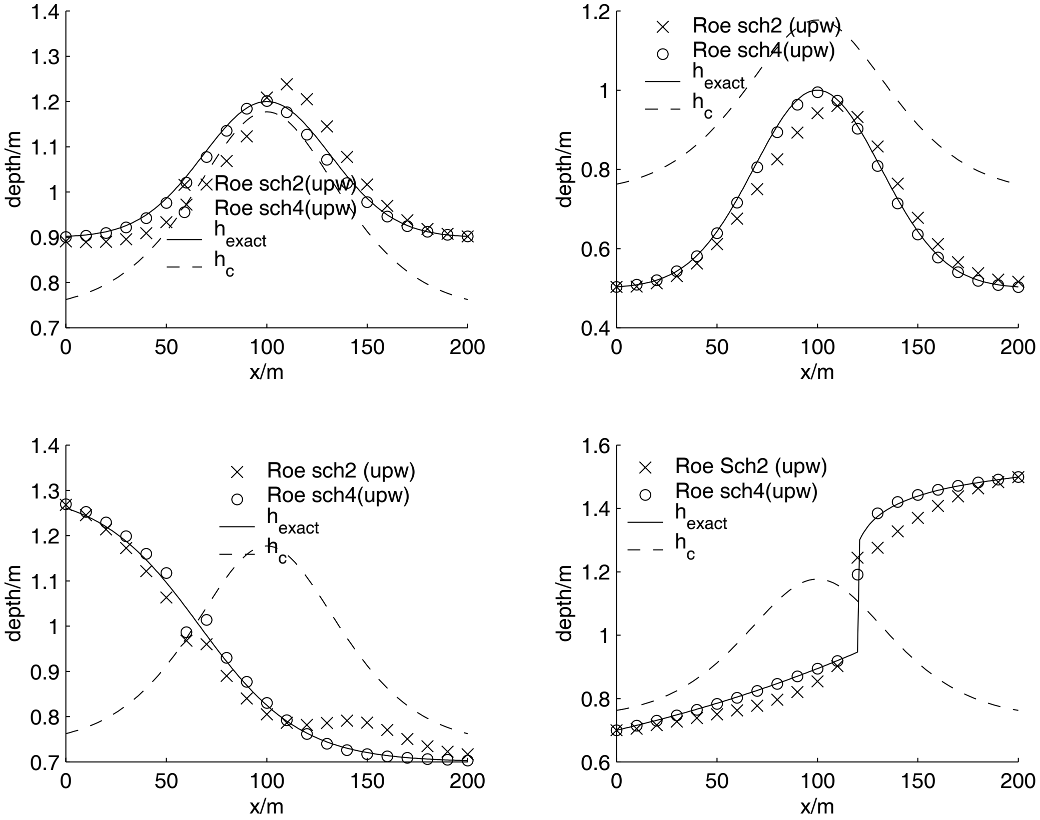

Figure 8.24: Roe scheme in direct approach versus Roe scheme in indirect approach with source terms upwind discretised.

derivative (ptwh) source term are compared with the results obtained with a upwind discretisation. The graphs show that the upwind discretisation provides the most accurate results. This is confirmed by the error graphs in Fig. 8.26 - Fig. 8.29.

We also plotted the results of Engquist-Osher scheme and Roe scheme (indirect approach) when using a pointwise discretisation of the source term and a centred (at half-point) discretisation of the derivative V. The results are shown in Fig. 8.25 and seem to indicate that the most accurate scheme is Engquist-Osher in this indirect approach, which is confirmed by the error graphs in Fig. 8.26 - Fig. 8.29.

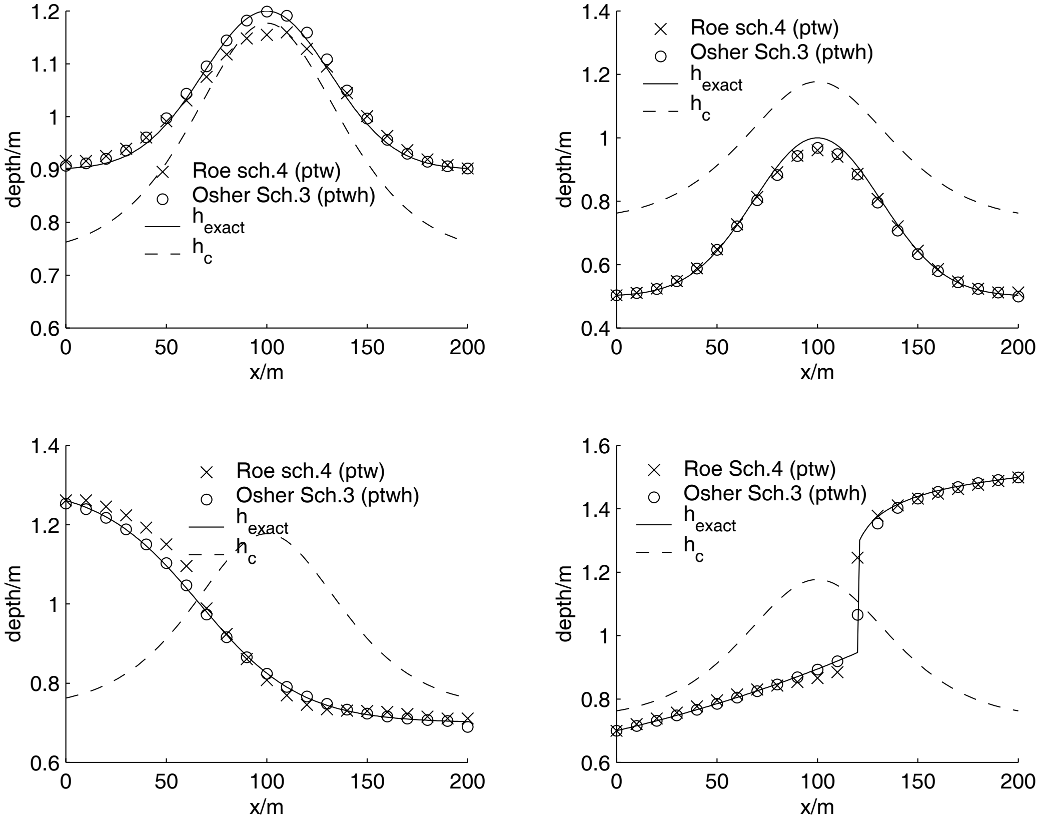

Figure 8.25: Engquist-Osher scheme versus Roe scheme (indirect approach) with source term pointwise discretised.

It can be seen from the graphs of the error, Fig. 8.26 - Fig. 8.29, that overall the approximate solution obtained by the Roe upwind scheme with a indirect approach is more accurate in the subcritical and supercritical test problems. The Engquist-Osher pointwise scheme 1 (direct) is the most accurate in the hydraulic jump and the indirect Engquist-Osher scheme 3 is the most accurate in the smooth transition but just because Roe's upwind scheme 4 loosed accuracy when N increases. This odd behaviour can be explained to this test problem being a smooth transition and the Roe scheme is not so well behaved in this case. The supercritical test problem does not cause any particular

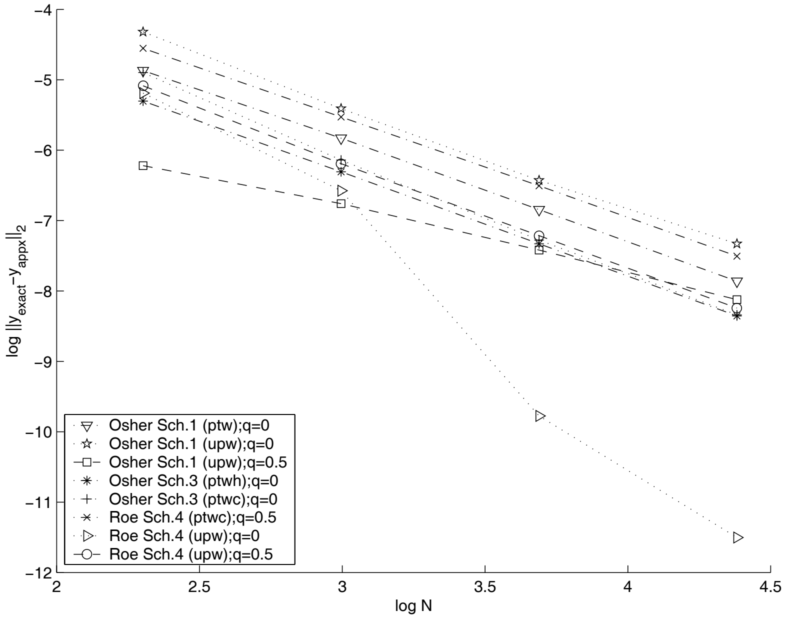

Figure 8.26: L_2 errors of the more accurate results obtained for test problem 1 (subcritical flow).

difficulty to these schemes since the sonic line is not very near the exact solution. Several schemes provide good accurate results with a similar (slight) higher order of convergence. In those schemes is included the Roe direct scheme 2 whose results did not suffer from the lack of an entropy fix in this case. The most difficult test problems, the smooth transition flow and the shock, are the ones where the Engquist-Osher scheme yields the best results even with just a pointwise discretisation of the source terms.

Once more, one can draw the conclusion that a more accurate solution is obtained, when using Roe scheme, from upwind discretisation of the source terms. This has been also pointed out in work done by other researchers when applying Roe scheme to approximate the unsteady Saint-Venant equations [3, 99, 19].

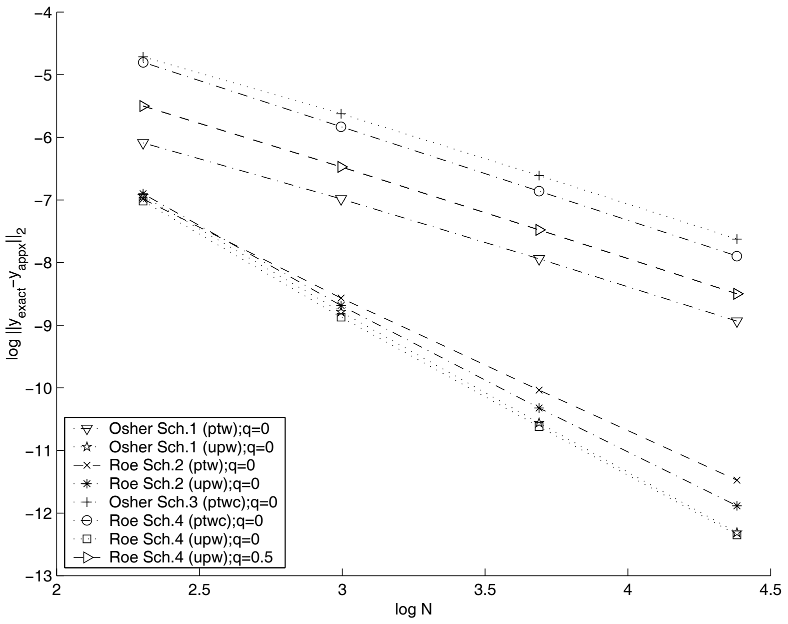

Figure 8.27: L_2 errors of the more accurate results obtained for test problem 2 (supercritical flow).

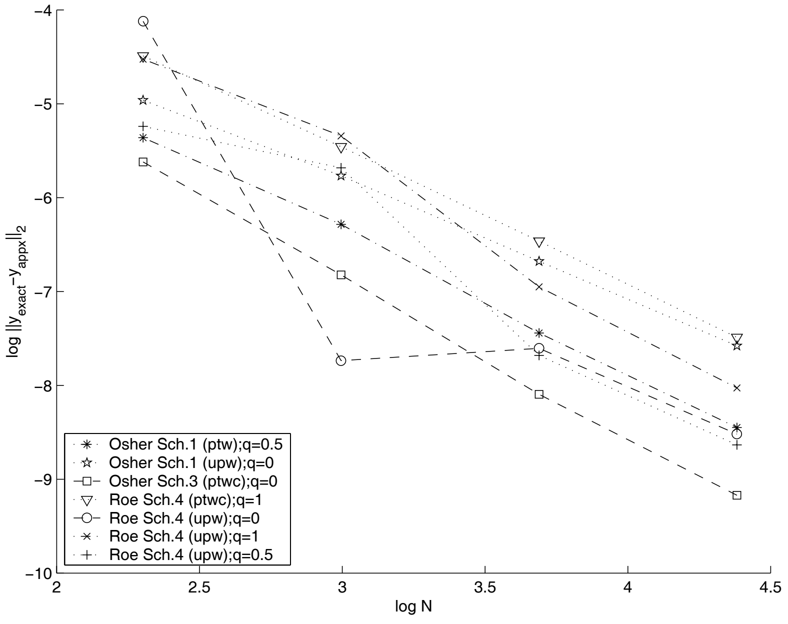

Figure 8.28: L_2 errors of the more accurate results obtained for test problem 3 (smooth transition).

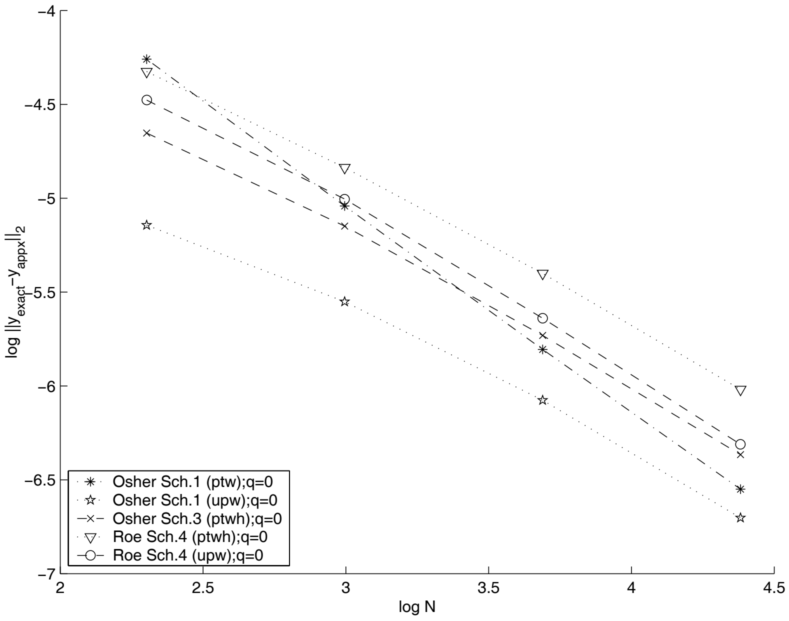

Figure 8.29: L_2 errors of the more accurate results obtained for test problem 4 (hydraulic jump).

Chapter 9

Conclusions and Further Work

In this thesis we have studied both analytical and numerical aspects of singular differential equations derived from the steady Saint-Venant equations and from the steady Euler equations. Their singular behaviour occurs in points where the flow features change. We have been particularly concerned with discontinuous solutions and in ways of discretising these singular equations with a switch providing the right "wind" direction. We would like to have convergence of those schemes to the physically relevant solution. The use of the Roe and Engquist-Osher shock capturing schemes to compute discontinuous solutions of steady flow problems is justified since one can look at the solutions of the steady flow problems we are interested in as steady entropy satisfying solutions of a particular (unsteady) scalar conservation law.

The work of MacDonald [59] on the steady Saint-Venant problem provided much of the mathematical background on the subject. It is also the source of the water test problems used in the thesis.

When addressing the problem of solving numerically the singular equation obtained from the steady Euler equations, we were confronted with the search of possible applications of these equations, and those were found in the theory of quasi-one dimensional gas flow. Wixcey's [103] work was fundamental to the understanding of the Euler equations problem. It gave a rather more mathematical approach than the one found, for example, in [84].

One question we would have liked to have answered was if, in the case of the Euler equations, the 'scalar approach' taken by MacDonald [59] when studying the Saint-Venant equations with prismatic channels could be applicable with some modifications.

A related question is if the theory can be extended to the Saint-Venant equations in nonprismatic channels. Unfortunately, the scalar equation we have derived in the gas problem has some particularities that do not allow such a theory to be applied in a straightforward manner. The theory for the nonprismatic case is also more complicated but is hoped that it would be possible to extend it to certain classes of nonprismatic channels.

As the work on these type of singular differential equations was developing a lot of questions and concerns arose, some were answered, some were abandoned, and others opened future research paths. Some progress was made and we describe it.

We have verified that, if there is no source term in the conservation of mass equation, the steady Saint-Venant equations and the Euler equations can be reduced to a scalar differential equation which preserve particular features of the original equations, namely, becoming singular when the Froude number or the Mach number become 1. This reduction is possible when the steady conservation of mass ODE is easily solved and, in the Euler equations case, if no source term in the energy equation is considered. If there is a source term in the energy equation, we showed that the Euler equations can be reduced to a system of two ODEs. This system also inherits features of the original equations which are of interest, namely, the Jacobian matrix has two eigenvalues which are related to the eigenvalues of the original matrix. For this particular system of two equations, all the background work (averages, eigenvectors, etc) was derived in order to apply the Roe scheme. At the time we finish this piece of work what was lacking was a test problem and the algorithm was not fully implemented. The research moved on to study the scalar ODE equation obtained from the steady Euler equations.

A physically relevant application of the scalar equation, in the gas case, is that of isentropic flow in a duct assumed to be slowly varying, so that we can consider flow in the x direction only. Our choice of a reduced equation maintaining the relevant physical features, led us to a reduced singular nonlinear equation whose flux definition and source function definition depend on the entropy K, which has a jump if a normal shock occurs. When a normal shock occurs, the flow is not isentropic across the shock but is still isentropic before and after the shock. Since we also want to study those discontinuous, physically relevant solutions, at first we felt very tempted to use well-known properties, which the variables of the problem satisfy, to help dealing with the jump in K. But those techniques would be more from the field of adaptive shock fitting methods than that of shock capturing methods. Recently, a technique was devised that possibly will allow us to overcome the discontinuity of K and simultaneously, allowing an independent switch in the "wind" direction (though using known properties of the flow). Another possible way to go is to use an idea similar to the one in [97] where the discontinuity of the coefficient K is thought as in occurring in the middle of the cell whereas if we are solving Riemann problems at the boundary. We are looking to the possibility of deriving a different form of the scalar equation which avoids this complication. Unfortunately, there was not enough time to complete those investigations.

Another concern arose from the particular form of the Saint-Venant equations studied, modelling water flow in nonprismatic channels. These equations have a flux function which depends explicitly on the spatial variable. We looked at ways of discretising these equations by using modifications of the Engquist-Osher and the Roe scheme with different types of discretisations of the source terms and two approaches were taken: direct and indirect. We have shown that modifications of the Roe scheme and of the Engquist-Osher scheme yield accurate results in both type of approaches. A higher order of convergence was achieved, in certain test problems, when using a upwind discretisation of the source terms combined with the Roe scheme. The upwind Engquist-Osher scheme used in the direct approach did not achieve the higher accuracy that is known to reach in the case of a flux function of the form $F(w)$. A upwind Engquist-Osher scheme for the indirect approach built in a similar way was shaped but it was not fully implemented. A switch for the derivative source term was built but still needs some refining. The choice of discretisation of this derivative term was similar to the adopted in Roe scheme, that is, a one-sided finite differences approximation for both supercritical and subcritical flow, except that the switch is provided by a smooth function. Nevertheless,without studying further this scheme we do not know how well the Engquist-Osher scheme will behave since, although similar to Roe's scheme in the subcritical and supercritical flow, it differs in the case of a shock or a smooth transition. Furthermore, when using Engquist-Osher scheme, a Newton iteration must be used to solve the nonlinear system of difference equations. But the same cannot be said of the Roe scheme due to lack of continuity.

The Roe scheme, in the direct approach, suffered from the lack of an entropy fix. A well known entropy fix was tried but did not work. The results could improve to be similar or even better than the ones obtained in the indirect approach if a entropy fix is devised.

When comparing similar order of accuracy methods (first-order) the Engquist-Osher scheme seems the better option providing good accurate results in all types of flow.

Further study is needed on the particular value of x used in the discretisations. A more exhaustive study has to be done to see if the adoption of different x discretisations still allows the scheme to verify a discrete telescopic property. Possibly, a x discretisation related with the type of flow may contribute for the scheme to pick up the correct 'wind' direction.

Besides the already discussed current lines of research, we discuss some other ideas that came up as possible future work.

An argument for using upwind methods for other singular nonlinear differential equations of the form

$$f_1(x, y_1(x))\frac{dy}{dx} = f_2(x, y),$$

without the physical background, is that this is a nonlinear equation which can be solved by iteration, but the iteration must be convergent. If we use Picard iteration

$$y_{n+1} - y_n = \tau\alpha(f_1(x, y(x))\frac{dy}{dx} - f_2(x, y(x)))$$

the iteration converges if α has the right sign and τ is sufficiently small to make the iteration a contraction. Conservation probably plays a role in the proof (Osher).

For general nonlinear singular differential equations of the form

$$f_1(x, y(x))y'(x) = f_2(x, y),$$

if we can write it as

$$\frac{dF(x, y(x))}{dx} = f_2(x, y) + \frac{\partial F}{\partial x}$$

then we can use the technique in this thesis (upwinding with source terms) to solve the equation.

Bibliography

[1] M. B. Abbott. *Computational Hydraulics - Elements of the theory of free surface flows.* Pitman Publishing Limited, London, 1979.

[2] J. D. Anderson, Jr. *Computational fluid Dynamics - the basics with applications.* Mechanical Engineering series. McGraw-Hill International Editions, New York, 1995.

[3] A. Bermúdez and M. E. Vázquez. Upwind methods for hyperbolic conservation laws with source terms. *Computers Fluids*, 8:1049–1071, 1994.

[4] J. Billingham and A.C. King. *Wave Motion.* Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2000.

[5] J. Burguete and P. Garcia-Navarro. Efficient construction of high-resolution TVD conservative schemes for equations with source terms: applications to shallow water flows. *Int. J. for Numer. Meth. Fluids*, 37:209–248, 2001.

[6] A. N. Chinnayya and A.-Y. Le Roux. A new general Riemann solver for the shallow water equations, with friction and topography. Conservation Laws Preprint Server: www.math.ntnu.no/conservation, 1999.

[7] A. J. Chorin and J. E. Marsden. *A mathematical introduction to Fluid Mechanics.* Number 4 in Texts in Applied Mathematics. Springer-Verlag, Berlin, 1993.

[8] V. T. Chow. *Open-Channel Hydraulics.* McGraw-Hill International Editions, London, 1959.

[9] B. Cockburn, C. Johnson, C.-W. Shu, and E. Tadmor. *Advanced numerical approximation of nonlinear hyperbolic equations.* Number 1697 in Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1998.

[10] J. M. Corberán and L. Gascón. Alternative treatment of strong source terms in nonlinear hyperbolic conservation laws. Application to unsteady 1D compressible flow in pipes with variable cross-section. In J. Henriette, P. Lybaert, and M. Hayek, editors, *Advanced Concepts and Techniques in Thermal Modelling*, page 275, New York, 1998. Elsevier.

[11] R. Courant and K. O. Friedrichs. *Supersonic Flow and Shock Waves*. Springer-Verlag, New York, 1948.

[12] R. Courant, K. O. Friedrichs, and H. Lewy. Uber die partiellen differenzengleichungen der mathematisches physik. *Math. Ann.*, 100:32–74, 1928.

[13] M. Crandall and A. Majda. The method of fractional steps for conservation laws. *Numerische Mathematik*, 34:285–314, 1980.

[14] G. Dal Maso, P. G. Le Floch, and F. Murat. Definition and weak stability of nonconservative products. *J. Math. Pures et Appliquées*, 74:483–548, 1995.

[15] S. Emmerson. *Modelling of Transient Dynamics of Gas Flow in Pipes*. PhD thesis, Department of Mathematics, University of Reading, 1991.

[16] B. Engquist and S. Osher. One-sided difference schemes and transonic flow. *Proc. Natl. Acad. Sci. U.S.A.*, 77:3071–3074, 1980.

[17] B. Engquist and S. Osher. Stable and entropy satisfying approximations for transonic flow calculations. *Mathematics of Computation*, 34:45–75, 1980.

[18] B. Engquist and S. Osher. One-sided difference approximations for nonlinear conservation laws. *Mathematics of Computation*, 36:321–351, 1981.

[19] P. Garcia-Navarro and M. E. Vázquez-Cendón. Some considerations and improvements on the performance of Roe's scheme for 1D irregular geometries. Pre-Publicacións do Departamento de Matemática Aplicada Internal Report 23, Universidade de Santiago de Compostela, 1997.

[20] P. Garcia-Navarro and M. E. Vázquez-Cendón. On numerical treatment of the source terms in the shallow water equations. *Computers and Fluids*, 29:951–979, 2000.

[21] L. Gascón and J. M. Corberán. Construction of second-order TVD schemes for nonhomogeneous hyperbolic conservation laws. *Journal of Computational Physics*, 172:261–297, 2001.

[22] P. Glaister. Flux difference splitting techniques for the Euler equations in non-cartesian geometry. Numerical Analysis Report 8/85, Department of Mathematics, University of Reading, 1985.

[23] P. Glaister. Difference schemes for the shallow water equations. Numerical Analysis Report 9/87, Department of Mathematics, University of Reading, 1987.

[24] P. Glaister. Second order difference schemes for hyperbolic conservation laws with source terms. Numerical Analysis Report 6/87, Department of Mathematics, University of Reading, 1987.

[25] P. Glaister. Approximate Riemann solutions of the shallow water equations. *Journal of Hydraulic Research*, 26:293–306, 1988.

[26] P. Glaister. Flux difference splitting for the Euler equations in one spatial coordinate with area variation. *J. Numer. Meth. Fluids*, 8:97, 1988.

[27] P. Glaister. Flux difference splitting for the Euler equations with axial symmetry. *Journal of Engineering Mathematics*, 22:107–121, 1988.

[28] P. Glaister. Prediction of supercritical flow in open channels. *Computers Math. Applic.*, 24:69–75, 1992.

[29] E. Godlewski and P.-A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Number 118 in Applied Mathematical Sciences. Springer-Verlag, New York, 1996.

[30] S. K. Godunov. A finite difference method for the computation of discontinuous solutions of the eqautions of fluid dynamics. *Math. Sbornik*, 47:271–306, 1959.

[31] L. Gosse. A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms. *Computers and Mathematics with Applications*, 39:135–159, 2000.

[32] L. Gosse. A nonconservative numerical approach for hyperbolic systems with source terms: the well-balanced schemes. In *Hyperbolic Problems: theory, numerics,applications*, volume I and II, pages 453–461. 8th International Conference on Hyperbolic Problems, Feb 27-Mar 03 (2000), 2001.

[33] J. M. Greenberg and A. Y. Le Roux. A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. *SIAM Journal of Numerical Analysis*, 33:1–16, 1996.

[34] A. Harten. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 49:357–393, 1983.

[35] A. Harten. On a class of high resolution total-variation-stable finite difference schemes. *SIAM J. Numer. Anal.*, 21:1–23, 1984.

[36] A. Harten, B. Engquist, S. Osher, and S. R. Chakravarthy. Uniformly high order essentially non-oscillatory schemes III. *Journal of Computational Physics*, 71:231–303, 1987.

[37] A. Harten and J. M. Hyman. Self-adjusting grid methods for one-dimensional hyperbolic conservation laws. *Journal of Computational Physics*, 50:235–269, 1983.

[38] A. Harten, J. M. Hyman, and P. D. Lax. On finite-difference approximations and entropy conditions for shocks. *Communications on Pure and Applied Mathematics*, 29:297–322, 1976.

[39] A. Harten, P. D. Lax, and B. van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. ICASE Report 82-5, Institute for Computer Applications in Science and Engineering, 1982.

[40] C. Hirsch. *Numerical computation of internal and external flows*, volume 1. John Wiley & Sons, New York, 1988.

[41] T. Y. Hou and P. G. Le Floch. Why nonconservative schemes converge to wrong solutions: Error analysis. *Mathematics of Computation*, 62(206):497–530, 1994.

[42] M. E. Hubbard and P. Garcia-Navarro. Flux difference splitting and the balancing of source terms and flux gradients. *Journal of Computational Physics*, 165:89–125, 2000.

[43] P. Jenny and B. Müller. Rankine-Hugoniot-Riemann solver considering source terms and multidimensional effects. *Journal of Computational Physics*, 145:575–610, 1998.

[44] D. Kröner. *Numerical Schemes for Conservation Laws.* Wiley & Teubner, Chichester, 1997.

[45] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics*, volume 6 of *Course of Theoretical Physics.* Pergamon Press, Oxford, 1959.

[46] P. Lax and B. Wendroff. Systems of conservation laws. *Communications on Pure and Applied Mathematics*, 13:217–237, 1960.

[47] P. D. Lax. *Hyperbolic Systems of Conservation laws and the Mathematical Theory of Shock Waves.* CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, 1973.

[48] A. C. Lemos, M. J. Baines, and N. K. Nichols. Numerical solution of hyperbolic systems of conservation laws with source terms - Part I. Numerical Analysis Report 1/2001, Department of Mathematics, University of Reading, 2001.

[49] R. J. LeVeque. *Numerical Methods for Conservation Laws.* Lectures in Mathematics. Birkhäuser, Basel, 1992.

[50] R. J. LeVeque. Balancing source terms and flux gradients in high-resolution Godunov methods: the quasi-steady wave-propagation algorithm. *Journal of Computational Physics*, 146:346–365, 1998.

[51] R. J. LeVeque and D. S. Bale. Wave propagation methods for conservation laws with source terms. Technical Report 98-13, University of Washington, 1998.

[52] R. J. LeVeque and M. Pelanti. A class of approximate Riemann solvers and their relation to relaxation schemes. *Journal of Computational Physics*, 172:572–591, 2001.

[53] R. J. LeVeque and H. C. Yee. A study of numerical methods for hyperbolic conservation laws with stiff source terms. *Journal of Computational Physics*, 86:187–210, 1990.

[54] J. Lorenz. Numerical solution of a singular perturbation problem with turning points. In H. W. Knoblock and K. Schmitt, editors, *Equadiff 82*, number 1017 in Lecture Notes in Applied Mathematics, pages 432–439, New York, 1983. Springer-Verlag.

[55] J. Lorenz. Analysis of difference schemes for a stationary shock problem. *SIAM J. Numer. Anal.*, 21:1038–1053, 1984.

[56] J. Lorenz. Convergence of upwind schemes for a stationary shock. *Mathematics of Computation*, 46:45–57, 1986.

[57] J. Lorenz and R. Sanders. On the rate of convergence of viscosity solutions and boundary value problems. *SIAM J. Math. Anal.*, 18:306–320, 1987.

[58] J. Lorenz and H. J. Schroll. Hyperbolic systems with relaxation: Characterization of stiff well-posedness and asymptotic expansions. *Journal of Mathematical Analysis and Applications*, 235:497–532, 1999.

[59] I. MacDonald. *Analysis and Computation of Steady Open Channel Flow*. PhD thesis, Department of Mathematics, University of Reading, 1996.

[60] I. MacDonald, M. J. Baines, and N. K. Nichols. Analysis and computation of steady open channel flow using a singular perturbation. Numerical Analysis Report 7/94, Department of Mathematics, University of Reading, 1994.

[61] I. MacDonald, M. J. Baines, and N. K. Nichols. Test problems with analytic solutions for steady open channel flow. Numerical Analysis Report 6/94, Department of Mathematics, University of Reading, 1994.

[62] I. MacDonald, M. J. Baines, N. K. Nichols, and P. G. Samuels. Comparison of some steady state Saint-Venant solvers for some test problems with analytic solutions. Numerical Analysis Report 2/95, Department of Mathematics, University of Reading, 1995.

[63] I. MacDonald, M. J. Baines, N. K. Nichols, and P. G. Samuels. Steady open channel test problems with analytic solutions. Numerical Analysis Report 3/95, Department of Mathematics, University of Reading, 1995.

[64] I. MacDonald, M. J. Baines, N. K. Nichols, and P. G. Samuels. Analytic benchmark solutions for open-channel flows. *Journal of Hydraulic Engineering*, 123:1041–1045, 1997.

[65] B. Massey and J. Ward-Smith. *Mechanics of Fluids*. Stanley Thornes Ltd., Cheltenham, seventh edition, 1998.

[66] E. M. Murman. Analysis of embedded shoclwaves calculated by relaxation methods. *AIAA Journal*, 12:636, 1974.

[67] E. M. Murman and J. D. Cole. Calculation of plane steady transonic flows. *AIAA Journal*, 9:114, 1971.

[68] O. Oleinik. Discontinuous solutions of non-linear differential equations. *Usp. Mat. Nauk.*, 12:3–73, 1957. Translation in Amer. Math. Soc. Transl. Ser. 2, 96, 95–172, 1963.

[69] R. E. O'Malley. *Introduction to Singular Perturbations*. Academic Press, New York, 1974.

[70] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[71] S. Osher. Nonlinear singular perturbation problems and one sided difference schemes. *SIAM J. Numer. Anal.*, 18:129–144, 1981.

[72] S. Osher. Riemann solvers, the entropy condition, and difference approximations. *SIAM J. Numer. Anal.*, 21:217–235, 1984.

[73] M. V. Papalexandris, A. Leonard, and P. E. Dimotakis. Unsplit schemes for hyperbolic conservation laws with source terms in one space dimension. *Journal of Computational Physics*, 134:31–61, 1997.

[74] A. Priestley. Roe-type schemes for super-critical flows in rivers. Numerical Analysis Report 13/89, Department of Mathematics, University of Reading, 1989.

[75] P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, 43:357–372, 1981.

[76] P. L. Roe. Fluctuations and signals - a framework for numerical evolution problems. In K. W. Morton and M. J. Baines, editors, *Numerical Methods for Fluid Dynamics*, pages 219–257, London, 1982. Academic Press.

[77] P. L. Roe. Characteristic-based schemes for the Euler equations. *Ann. Rev. Fluid Mech.*, 18:337–365, 1986.

[78] P. L. Roe. Upwind differencing schemes for hyperbolic conservation laws with source terms. In C. Carasso, P.-A. Raviart, and D. Serre, editors, *Nonlinear hyperbolic problems*, number 1270 in Lecture Notes in Mathematics. Springer-Verlag, Berlin/New York, 1986.

[79] P. L. Roe and J. Pike. Efficient construction and utilisation of approximate Riemann solutions. In R. Glowinski and J.L. Lions, editors, *Computing Methods in Applied Sciences and Engineering VI*. North Holland, 1984.

[80] D. Serre. *Systems of conservation laws 1 - hyperbolicity, entropies, shock waves.* Cambridge University Press, Cambridge, 1999.

[81] M. J. Sewell. Properties of a streamline in gas flow. *Phys. Technol.*, 16:127–133, 1985.

[82] M. J. Sewell. *Maximum and minimum principles: A unified approach, with applications.* Cambridge University Press, Cambridge, 1987.

[83] M. J. Sewell and D. Porter. Constitutive surfaces in fluid mechanics. *Math. Proc. Camb. Phil. Soc.*, 88:517–546, 1980.

[84] A. H. Shapiro. *The dynamics and thermodynamics of compressible fluid flow*, volume I and II. The Ronald Press Company, New York, 1953.

[85] C. W. Shu. Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In A. Quarteroni, editor, *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, number 1697 in Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1998.

[86] J. Smoller. *Shock waves and reaction-diffusion equations.* A series of comprehensive studies in Mathematics. Springer-Verlag, New York, second edition, 1994.

[87] A. B. Stephens and G. R. Shubin. Existence and uniqueness for an exponentially derived switching scheme. *SIAM J. Numer. Anal.*, 20:885–889, 1983.

[88] J. J. Stoker. *Water Waves: the mathematical theory with applications.* Wiley Classics Library. Wiley, New York, 1992 reprint edition, 1958.

[89] P. K. Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM J. Numer. Anal.*, 21:995–1011, 1984.

[90] P. K. Sweby. Source terms and conservation laws: a preliminary discussion. Numerical Analysis Report 6/89, Department of Mathematics, University of Reading, 1989.

[91] P. K. Sweby. Godunov methods. Numerical Analysis Report 7/99, Department of Mathematics, University of Reading, 1999.

[92] E. Tadmor. Approximate solutions of nonlinear conservation laws. In A. Quarteroni, editor, *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, number 1697 in Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1998.

[93] J. W. Thomas. *Numerical Partial Differential Equations: Conservation Laws and Elliptic Equations.* Number 33 in Texts in Applied Mathematics. Springer-Verlag, New York, 1999.

[94] E. F. Toro. *Riemann solvers and numerical methods for Fluid Dynamics - a practical introduction.* Springer, Berlin, second edition, 1999.

[95] E. F. Toro. *Shock-Capturing Methods for Free-Surface Shallow Flows.* Wiley, New York, 2001.

[96] E. F. Toro and V. A. Titarev. Solution of the generalised Riemann problem for advection-reaction equations. *Proc. R. Soc. Lond. A*, 458:271–281, 2002.

[97] J. D. Towers. Convergence of a difference scheme for conservation laws with a discontinuous flux. *SIAM J. Numer. Anal.*, 38:681–698, 2000.

[98] B. van Leer. Towards the ultimate conservative difference scheme v. a second order sequel to Godunov's method. *Journal of Computational Physics*, 32:101–136, 1979.

[99] M. E. Vázquez-Cendón. 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.

[100] J. P. Vila. Simplified Godunov schemes for 2x2 systems of conservation laws. *SIAM Journal Numerical Analysis*, 23:1173–1192, 1986.

[101] P. Wesseling. *Principles of computational Fluid Dynamics.* Springer-Verlag, Berlin, 2001.

[102] G. B. Whitman. *Linear and Nonlinear Waves.* Pure and Applied mathematics. Wiley-Interscience Publication, New York, 1999.

[103] J. R. Wixcey. *Stationary principles and adaptive finite elements for compressible flow in ducts.* PhD thesis, Department of Mathematics, University of Reading, 1990.