

UNIVERSITY OF READING
DEPARTMENT OF MATHEMATICS

**ANALYSIS AND COMPUTATION OF STEADY OPEN CHANNEL
FLOW**

by

IAN MACDONALD

This thesis is submitted for the degree of
Doctor of Philosophy

SEPTEMBER 1996

Abstract

The Saint-Venant equations provide a one-dimensional model of free surface water flow in a channel. This thesis is concerned with both analytical and numerical aspects of steady state solutions to this model, with particular emphasis on the subject of transcritical flows.

Under certain conditions it is shown that there is at most one physically allowable steady solution for given boundary conditions, and when a solution exists, we demonstrate the convergence of a certain family of numerical methods to the solution as the grid size vanishes.

The numerical schemes are obtained from applying a family of monotone shock capturing schemes to a scalar conservation law which has identical steady solution

Acknowledgements

Firstly I would like to thank my acad

3.8.1	Boundary Conditions	51
3.8.2	Modifications to Roe's Scheme	52

4 Theory for the Steady Flo

7.3.2	High Order TVD Schemes	152
7.4	Conclusions	157
8	omputational Efficiency	15
8.1	The Time Stepping Iteration	159
8.2	Newton's Method	164
8.3	The Implicit Time Stepping Iteration	170
8.4	Conclusions	175
	Non-Prismatic Channels	176
9.1	Scalar Schemes	176
9.2	Roe's Approximate Riemann Solver	187
9.3	Conclusions	192
10	onclusions and Further Work	1 5
A	Theory for the Second Order Modification to Engquist-Osher	208
B	Test Problems for Non-Prismatic Channels	214

Notation for the Saint-Venant Model

x	Distance along channel (m)
t	Time (s)
L	Length of channel (m)
z_b	Bed level (m)
η	Height relative to the bed level (m)
σ	Width of channel as a function of x and η (m)
g	Acceleration due to gravity (ms^{-2})
ρ	Density (kgm^{-3})
h	Depth (m)
Q	Discharge (m^3s^{-1})
A	Wetted area (m^2)
T	Free surface width (m)
P	Wetted perimeter (m)
$F = \frac{Q^2}{A} + gI_1$	Momentum flux per unit density (m^4s^{-2})
$I_1 = \int_0^h (h - \eta)\sigma d\eta$	Hydrostatic pressure term (m^3)
$D = gA(S_0 - S_f) + gI_2$	Source term (m^3s^{-2})
$S_0 = -z'_b$	Bed slope
$S_f = \frac{ Q Q}{K^2}$	Friction slope
K	Conveyance (m^3s^{-1})
n	Friction coefficient
$I_2 = \int_0^h (h - \eta)\sigma_x d\eta$	Side reaction term for a non-prismatic channel (m^2)
$u = \frac{Q}{A}$	Component of fluid velocity in x direction (ms^{-1})
$c = \left(\frac{gA}{T}\right)^{\frac{1}{2}}$	Wave celerity (ms^{-1})
h_c, h_n	Critical and normal depths (m)
S_{0c}	Critical bed slope
$F_r = \left(\frac{Q^2 T}{g A^3}\right)^{\frac{1}{2}}$	Froude number
$E = \frac{Q^2}{2A^2} + gh$	(m^2s^{-2})
$B,$	Width (m) and side slope for a trapezoidal channel
$\mathbf{w} = (A, Q)^T$	State vector

Chapter 1

Introduction

The study of free-surface water flow in channels has many important applications, one of the most significant being in the area of river modelling. With major river engineering projects, such as flood prevention measures, becoming ever more common and ambitious, there is an increasing need to be able to model and predict the far ranging consequences on the environment as a whole of any potential project. A major part of this process is to predict the new hydraulic characteristics of the system. For example constricting the river at some point may result in an increased risk of flooding at a point upstream. The basic equations expressing hydraulic principles were formulated in the 19th century by de St Venant and Boussinesq. Properties of these relationships were studied in the first half of this century, but application to real river engineering projects awaited the advent of electronic computers. The hydraulic equations are also of great importance in the modelling and design of networks of artificial channels, as for example may occur in industrial plants or sewage systems.

The original hydraulic model of de St Venant[11] is written in the form of a system of two partial differential equations, known as the Saint-Venant equations. These are derived under the hypothesis that the flow is one-dimensional. One-dimensional flows do not actually exist in nature, but the equations remain valid provided the flow is approximately one-dimensional. Until recently, two or three-dimensional models have been too computationally expensive to be practical. Even now it is often prohibitively expensive to obtain the amount of survey data for a river network necessary to make

use of the added realism of a higher dimensional model. For this reason the bulk of river modelling still makes use of a one-dimensional model, with key parts of the network perhaps modelled with a higher-dimensional model. Empirical correction factors are often included in the one-dimensional model to

most one physically possible steady solution for any given boundary conditions. The proof relies on a novel formulation of the steady flow problem, with the solutions constructed as the vanishing viscosity limit of solutions to a singular perturbation problem. Properties of the smooth solutions of the singular perturbation problem give information about the not necessarily continuous solutions of the steady flow problem.

As well as theoretical results for the steady flow problem, this thesis is also concerned with numerical computation of solutions. The steady flow differential equation can be accurately and efficiently integrated in order to compute the free surface profile. This is in general only useful for computing smooth solutions, although Humpidge and Moss[26] present an algorithm for discontinuous solutions which works by com-
 $\int_{\Gamma} \mathbf{t} \cdot \mathbf{T} \cdot \mathbf{j} \mathbf{T} \mathbf{w} \cdot \mathbf{u} \cdot \mathbf{i} \cdot \mathbf{n} \cdot \mathbf{f} \mathbf{T} \mathbf{D} \cdot \mathbf{T} \mathbf{T} \mathbf{d} \mathbf{T} \mathbf{e} \cdot \mathbf{u} \mathbf{T} \mathbf{D} \mathbf{D} \cdot \mathbf{e} \cdot \mathbf{d} \mathbf{T} \mathbf{D} \cdot \mathbf{i} \mathbf{T} \mathbf{D} \cdot \mathbf{b} \mathbf{g} \mathbf{r} \mathbf{T} \mathbf{a}$

In this thesis we attempt a new approach for improving efficiency of the computation of steady solutions. Instead of applying shock capturing method to the Saint-Venant system as it stands, we apply shock capturing methods to a suitable scalar partial differential equation which is chosen so as to have identical steady solutions to the Saint-Venant model. The first benefit of this approach (which we refer to as the “scalar approach”) is that analysis for scalar methods is much simpler than for the case of systems, and in Chapter 5 we present theory for a particular family of schemes. Under identical conditions to the theory in Chapter 4 we show that at steady state the system of difference equations has a unique solution and we also demonstrate convergence to the unique physical solution of the steady flow problem (as the grid spacing vanishes).

In Chapter 6 we give a relatively simple technique for constructing test problems with known exact solutions. Although analytic solutions have previously been constructed for idealised problems, this appears to be the first time that solutions for problems with realistic features have been made available[38]. Such features include varying channel geometries and discontinuous solutions. Details are given for a wide selection of test cases so as to allow other researchers to compare their

schemes and Roe's scheme to the case of non-prismatic channels and compare the accuracy of the various schemes.

Chapter 2

The Saint-Venant Equations

In this chapter the Saint-Venant equations are introduced and some of their properties discussed. Attention is then fixed on the steady s.n

control volume technique can be found, for example, in [9].

Before we introduce the Saint-Venant equations, we introduce the notation used to describe the channel geometry. We let x, y, z denote a Cartesian coordinate system with z pointing vertically upwards, and we consider a channel of length L along the x direction. For simplicity th

as follows.

- (1) The fluid is incompressible, homogeneous and internal stresses are negligible.
- (2) The flow is one-dimensional with the fluid velocity depending solely on x and time (t).
- (3) At each cross-section the free surface is represented by a horizontal line.
- (4) The streamline curvature is small and the vertical accelerations are negligible so the pressure can be taken as hydrostatic.

The *depth* $h(x, t)$ is the level of the free surface above the bed level and is illustrated in Figure 2.1. The *discharge* $Q(x, t)$ is defined to be the total volume flux through a given cross-section. If $u(x, t)$ is the x component of the fluid velocity then

$$Q = \int_0^h \int_{-\frac{\sigma}{2}}^{\frac{\sigma}{2}} u dy d\eta = Au, \quad (2.2)$$

where the *wetted area* $A(x, t)$ (the instantaneous area of the flow through any cross-section) is given by

$$A = \int_0^h \sigma d\eta. \quad (2.3)$$

Using the above assumptions the Saint-Venant equations can be derived **from** the continuity and momentum equations.

Applying conservation of momentum (x component) to the same control region and same time interval yields the equation

$$\rho \int_{x_1}^{x_2} [Q(x, \cdot)]_{t_1}^{t_2} dx + \rho \int_{t_1}^{t_2} [F(x, \cdot)]_{x_1}^{x_2} dt = \rho \int_{t_1}^{t_2} \int_{x_1}^{x_2} D(x, \cdot) dx dt . \quad (2.5)$$

Here $F(x, \cdot)$ is given by

$$F = \frac{Q^2}{A} + gI_1, \quad (2.6)$$

where I_1 is given by

$$I_1 = \int_0^h (h - \eta) \sigma d\eta,$$

and g is the acceleration due to gravity. ρF represents the momentum flux through a cross-section and is composed of the advected momentum and a contribution from the hydrostatic pressure forces over the cross-section. $\rho D dx$ represents the instantaneous external forces acting on the fluid at a cross-section due to the channel boundary. It is composed of frictional forces and the reaction forces from hydrostatic pressure acting on the boundary.

is continuous, then it may be shown that the following differential equations hold on this region:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0, \quad (2.9)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial F}{\partial x} = D. \quad (2.10)$$

These are the *differential form* of the Saint-Venant equations.

2.1.1 Discontin

If functions h and Q satisfy the differential form of the Saint-Venant equations except at discontinuities, where the Rankine-Hugoniot conditions (2.11) and (2.12)

Now (2.15) can be written as

$$\frac{\partial \mathbf{w}}{\partial t} + J \frac{\partial \mathbf{w}}{\partial x} = \hat{\mathbf{D}}, \quad (2.16)$$

where J is the Jacobian given by

$$J = \frac{\partial \mathbf{F}}{\partial \mathbf{w}} = \begin{pmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{pmatrix},$$

c is the wave celerity given by

$$c = \frac{\overline{gA}}{T},$$

and $T = \sigma(x, h)$ is the free surface width. The modified source term is given by

$$\hat{\mathbf{D}} = \begin{pmatrix} 0 \\ gA(S_0 - S_f) + \frac{gA}{T} \int_0^h \sigma_x d\eta \end{pmatrix}.$$

The Jacobian J has real and distinct eigenvalues

$$\lambda_1 = u - c, \quad \lambda_2 = u + c,$$

which give the characteristic speeds. The theory of characteristics can be found in [5] and [64]. The system of equations can be decomposed into two ordinary differential equations which hold along characteristic curves given by $dx/dt = \lambda_1$ and $dx/dt = \lambda_2$, respectively. Examples of this type of decomposition are given in [54] and [31]. It is important to know the directions of λ_1 and λ_2 , since information is transmitted along these curves. The flow is classified according to the *Froude number*

$$F_r = \frac{|u|}{c} = \frac{\overline{Q^2 T}}{gA} \quad {}^3 \quad {}^2 \quad {}^2$$

of flow is known as *supercritical flow* and occurs when inertial forces dominate over gravitation

$$[F]_{x_1}^{x_2} = \int_{x_1}^{x_2} Ddx, \tag{2.19}$$

must be satisfied for all $0 \leq x_1 \leq x_2 \leq L$. These are the integral form of the steady flow equations. Equation (2.18) clearly implies that Q is constant throughout the entire reach. Without loss of generality the constant discharge Q is assumed positive, since if the discharge is negative then the direction can be reversed.

denominator of the right-hand side vanishes, i.e. when

$$F_r^2 = \frac{Q^2 T}{g A^3} = 1, \quad (2.26)$$

which corresponds to critical flow.

2.2.1 The Hydraulic Jump

In this section we consider the

$$(2) \quad F \rightarrow \infty \text{ as } h \rightarrow \infty.$$

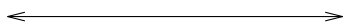
$$(3) \quad \partial F / \partial h = 0 \text{ at } h = h_c.$$

$$(4) \quad \partial F / \partial h < 0 \text{ for } h < h_c.$$

$$(5) \quad \partial F / \partial h > 0 \text{ for } h > h_c.$$

To interpret the implications of (2.23) and (2.24) we ask what depths $h_r \neq h_l$ satisfy both of these conditions. There are three cases to consider

- If $h_l < h_c$, then there is exactly one $h_r \neq h_l$ satisfying (2.23). This depth $h_r > h_c$ is called the *sequent depth* of h_l and is denoted by



so that

$$\frac{d}{dh} \left(\frac{T}{A^3} \right) = \frac{-3B^2 - 10Bh - 10h^2}{h^4(B+h)^4} < 0,$$

hence satisfying (2.30(1)). Condition (2.30(2)) is also clearly satisfied.

Suppose now that equation (2.17) is used for the conveyance with $k_1 > 0$ and $k_1 > k_2$, which includes both the widely used Manning and Chezy formulae. The conveyance is now given by

$$K = \frac{(Bh + h^2)^{k_1}}{n(B + 2h\sqrt{1 + ^2})^{k_2}},$$

giving

$$\frac{dK}{dh} = \frac{K}{AP} \left(Bk_1(B + 2h) + 2h\sqrt{1 + ^2} (B(k_1 - k_2) + h(2k_1 - k_2)) \right) > 0, \quad (2.33)$$

satisfying (2.31(1)). For the case of a rectangular channel ($ = 0$, $B > 0$), we have

$$K =$$

The set of conditions (2.31) ensure that such a depth exists. For the case $S_0(x) \leq 0$ equation (2.34) has no solution, but for convenience we define $h_n(x)$ to have a value of infinity.

Conditions (2.30) and (2.31) mean that given any position x and any depth h , the corresponding sign of dh/dx can be determined solely from the position of h relative to h_c and $h_n(x)$. The depth range is divided into three zones and we summarise the information in Table 2.1.

Zone	
------	--

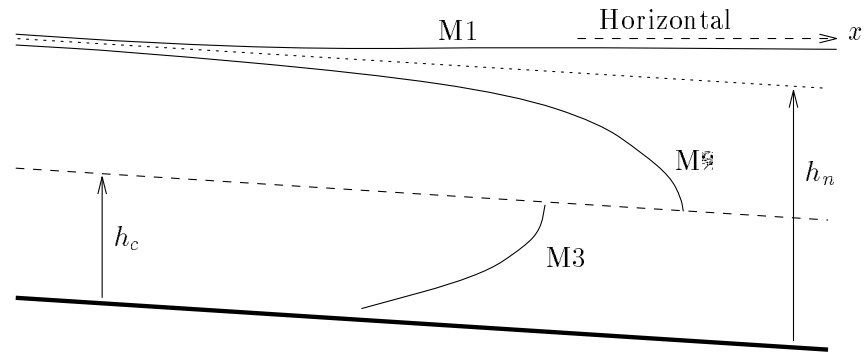
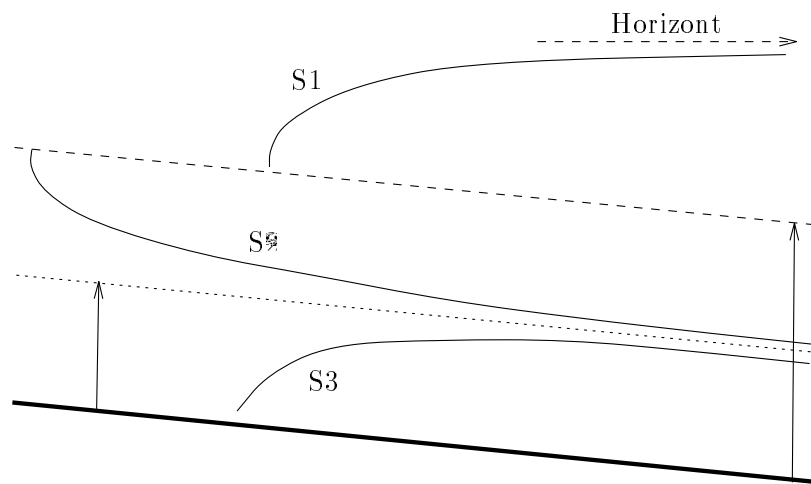


Figure 2.3: Behaviour of free surface for a channel with constant mild bed slope



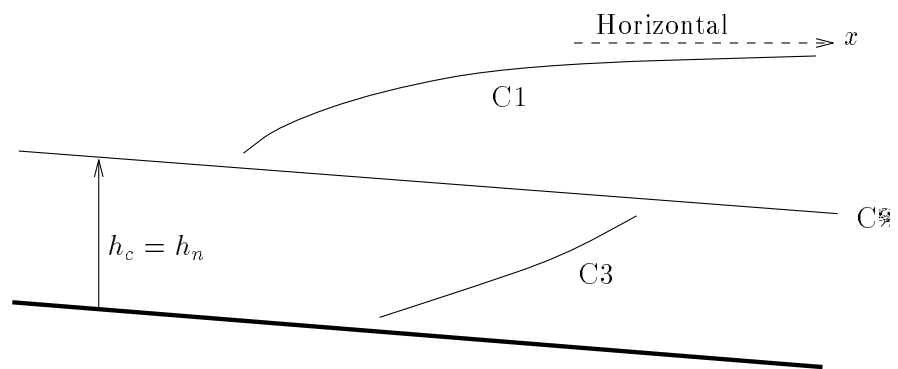


Figure 2.5: Beha

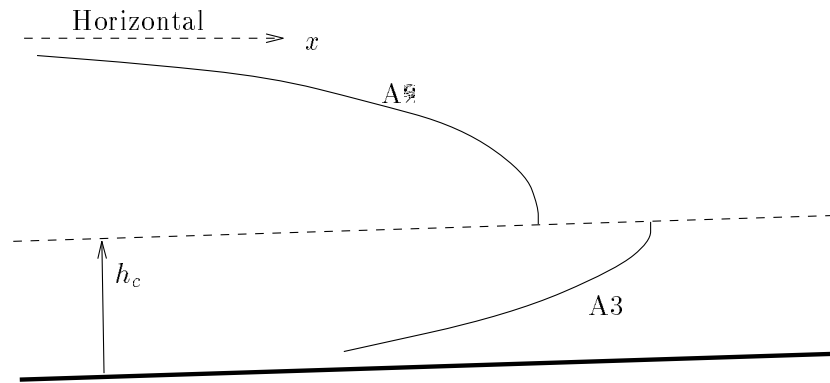


Figure 2.7: Behaviour of free surface for a channel with constant adverse bed slope

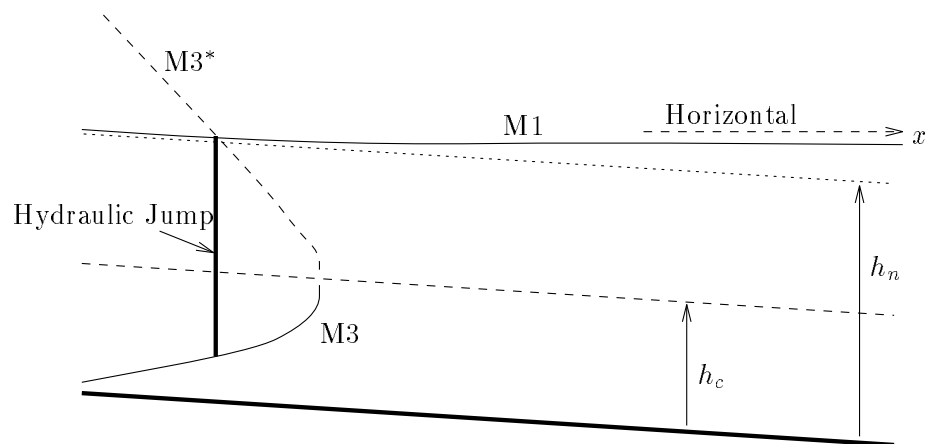


Figure 2.8: Example of problem with hydraulic jump

Varying Bed Slope

We now consider the situat

Weo7TalTj7esiann7c7ledTcephgff7aTD7curr7TTf7e.7TTD7Tc7eHoh

At such a point it is possible for equation (2.29) to have a solution which passes smoothly through the critical depth $h_c = h_n(x_c)$. Equation (2.29) cannot immediately be used to calculate the gradient when passing through the critical depth, but we can apply L'Hôpital's rule [63], [65] to the right hand side of (2.29) to obtain

$$\frac{dh}{dxn}$$

so in this case the flow cannot change from subcritical to supercritical at such a point. If the value under the square root in (2.44) is zero there is only one value, and this is again posit

due to contraction or expansion of the channel cross-section must compete with each other. A new more complex definition of normal depth requires taking into account the variation of the channel cross-section with x . The theory of singular points also becomes more complicated. The theory for non-prismatic channels is beyond the scope of this thesis, but it is discussed in for example [4].

2.2.4

steady state is said to be *stable* if for any such small perturbation, the flow eventually tends back to the original steady state. If for some small perturbation the flow does not tend back to the original state, then it is said to be *unstable*. In any practical situation small perturbations are always present, so any unstable steady state will not be maintainable indefinitely. Such steady state

o t a p a n t t o s s u r g e , t e

Chapter 3

Shock Capturing Methods

The Saint-Venant equations form a system of conservation laws of hyperbolic type. Such systems of equations occur frequently in applied mathematics and

giving details of how this method can be implemented for the Saint-Venant equations.

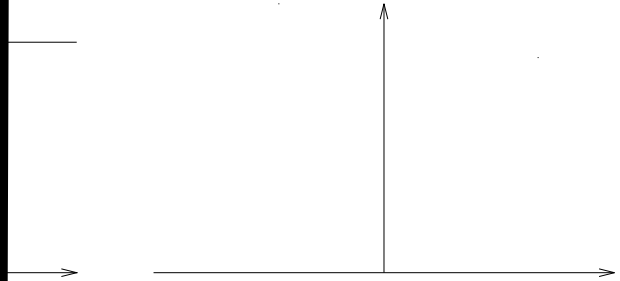
Consider the system of equations

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial}{\partial x}$$

Fortunately a simple requirement exists that if satisfied ensures that a scheme does approximate the correct conservation law. This requirement is that the scheme can be written in conservation form. A scheme is in *conservation form* if it is written in the form

$$\mathbf{u}$$

$$f'(\hat{w}(\xi)) = \xi.$$



for $w(0)$ can still be found (see [48]), giving the general form for the numerical flux

$$g_{j+\frac{1}{2}}^{\text{God}} = \begin{cases} \max\{f(s) : u_{j+1} \leq s \leq u_j\} & \text{for } u_{j+1} \leq u_j \\ \min\{f(s) : u_j \leq s \leq u_{j+1}\} & \text{for } u_{j+1} > u_j. \end{cases} \quad (3.12)$$

The notation we use is to omit the n superscripts in the definition of quantities when it is clear that the definition at a particular time level n is simply obtained by introducing n superscripts to all the time dependen _

the Saint-Venant equations and uses this as part of a Godunov approach. Even when the structure of the Riemann solution is known, it tends to be computationally expensive to compute the actual solution, since intersections of the Hugoniot and integral curve must be found (usually numerically). For this reason it is often not practical to base a numerical scheme on the exact solution of the Riemann problem. Instead an attempt is made to compute an approximation

where the functions $f_{\pm}$ are given by

$$\begin{aligned} f_{-}(u) &= \int_c^u \min\{f'(s), 0\} ds, \\ f_{+}(u) &= \int_c^u \max\{f'(s), 0\} ds \end{aligned}$$

and c is arbitrary. For a convex or concave flux function this form is equivalent to the Godunov flux for a rarefaction wave and only differs in the case of a shock.

In the case of systems of conservations laws the need for an approximate Riemann solver is more pressing. By far the most used approximate Riemann solver is that due to Roe[55]. This works by linearising the system of equations at each cell interface and then calculating the flux at the interface by exactly solving the resulting linear Riemann problem. Solving a linear Riemann problem is straightforward and is described in [30]. At the interface at $x_{j+\frac{1}{2}}$ and at time n the linearised system is given by

$$\frac{\partial \mathbf{u}}{\partial t} + \tilde{J}_{j+\frac{1}{2}}^n \frac{\partial \mathbf{u}}{\partial x} = 0, \quad (3.17)$$

where $\tilde{J}_{j+\frac{1}{2}} = \tilde{J}(\mathbf{u}_{j+1}, \mathbf{u}_j)$ is a constant matrix which approximates the Jacobian $J(\mathbf{u}) = \partial \mathbf{f} / \partial \mathbf{u}$ at the interface. Roe gives the properties that the matrix $\tilde{J}$ should satisfy. These are:

- (1) $\mathbf{f}(\mathbf{u}_r) - \mathbf{f}(\mathbf{u}_l) = \tilde{J}(\mathbf{u}_r, \mathbf{u}_l)(\mathbf{u}_r - \mathbf{u}_l)$ for all $\mathbf{u}_l, \mathbf{u}_r$.
- (2) $\tilde{J}(\mathbf{u}_r, \mathbf{u}_l)$ is diagonalisable for each $\mathbf{u}_l, \mathbf{u}_r$
- (3) For each $\mathbf{u}$, $\tilde{J}(\mathbf{u}_r, \mathbf{u}_l) \rightarrow J(\mathbf{u})$ as $\mathbf{u}_l, \mathbf{u}_r \rightarrow \mathbf{u}$.

Condition (1) is necessary to ensure the resulting scheme is conservative. Condition (2) ensures that the linearised system is hyperbolic and hence solvable. Condition (3) ensures that the the system (3.17) is a true linearisation of the nonlinear system so that the scheme is valid for smooth solutions. A matrix satisfying the above three conditions is often called a *Roe matrix*. There is not in general a unique choice for the Roe matrix for a particular problem. In his paper Roe demonstrates how to calculate a Roe matrix via an intermediate variable called a parameter vector. We specify a Roe matrix for the Saint-Venant system at the end of this chapter. Before

we state the numerical flux function we must define some notation. For the scalar quantity s , we define

$$s^+ = \frac{s + |s|}{2} = \begin{cases} 0 & s \leq 0 \\ s & s > 0, \end{cases}$$

$$s^- = \frac{s - |s|}{2} = \begin{cases} s & s \leq 0 \\ 0 & s > 0, \end{cases}$$

$$\Gamma(s) = \begin{cases} 0 & s \leq 0 \\ 1 & s > 0. \end{cases}$$

To generalise these quantities to the matrix $\tilde{J}$ we diagonalise this matrix. Suppose $\tilde{\lambda}_1, \dots, \tilde{\lambda}_m$ denote the eigenvalues of $\tilde{J}$ and $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_m$

The value of u_j^{n+1} depends only on the values of u_j^n and u_{j+1}^n and so again the discrete solution only depends on the solution at upwind points. We observe that the schemes switch their behaviour depending on the on the local wave direction. Schemes which exhibit this behaviour are known as *upwind schemes*. In the case of a system of equations, upwind schemes, such as Roe's approximate Riemann solver, essentially decompose the solution into its component waves and apply a scalar upwind scheme to each individual wave. As opposed to upwind schemes, schemes such as Lax-Wendroff and Lax-Friedrichs are known as *symmetric schemes*. These schemes have a constant stencil regardless of the wave direction of the solution. Upwind schemes are in general found to be far superior for computing discontinuous solutions.

3.4 Nonlinear Stability

The Lax-Wendroff Theorem[29] mentioned in section 3.1 shows that any convergent sequence of solutions to a conservative difference method must converge to a weak solution of the conservation law. However it does not guarantee that a sequence of solutions with $\Delta x, \Delta t \rightarrow 0$ will converge to a weak solution of the conservation law. However it does not guarantee that a sequence of solutions with $\Delta x, \Delta t \rightarrow 0$ will converge to a weak solution of the conservation law.

If we write the scheme in the form

$$u_j^{n+1} = G(u_{j+k}^n, \dots, u_{j+1}^n, u_j^n, \dots, u_{j-k+1}^n),$$

then the scheme is monotone if the function G is an increasing function of all of its arguments. It is shown in [8] and [23] that any conservative, monotone scheme converges to the unique entropy satisfying solution of the conservation law. Examples of monotone schemes include the Engquist-Osher, Godunov and Lax-Friedrichs schemes. Note again that this form of stability also requires the appropriate CFL condition to hold. We conclude that monotone schemes have some very nice properties. Not only are they convergent to the unique entropy solution of the conservation law, but since they must also be TVD they are non-oscillatory. The one big drawback is that they can at most be first order accurate

flux

$$\tilde{g}_{j+\frac{1}{2}} = g$$

3.6 Implicit Schemes

The methods discussed so far in this chapter are all subject to the CFL condition of the type (3.13) in order that the scheme be stable. This condition restricts the size of the time step that may be used. This may not be too much of a restriction for transient computations where the time step must also be kept small to achieve the required accuracy in time. However for steady state computations, where the accuracy of the transient solution is of no importance, we wish to take as large a time step as possible. The larger the time step we can take, the fewer time steps it takes to reach the steady state and the more economical the method is on computer CPU time.

To relax or even remove the time step restriction one can consider implicit methods. Consider the family of schemes

$$\frac{\mathbf{u}_j^{n+1} - \mathbf{u}_j^n}{\Delta} + \theta \left(\frac{\mathbf{g}_{j+\frac{1}{2}}^{n+1} - \mathbf{g}_{j-\frac{1}{2}}^{n+1}}{\Delta x} \right) + (1 - \theta) \left(\frac{\mathbf{g}_{j+\frac{1}{2}}^n - \mathbf{g}_{j-\frac{1}{2}}^n}{\Delta x} \right) = 0, \quad (3.19)$$

where $0 \leq \theta \leq 1$. This can be written in the conservative form (3.3) with the numerical flux

$$\tilde{\mathbf{g}}_{j+\frac{1}{2}}^n = \theta \mathbf{g}_{j+\frac{1}{2}}^{n+1} + (1 - \theta) \mathbf{g}_{j+\frac{1}{2}}^n.$$

We can re-write the scheme (3.19) as

$$L_j \mathbf{u}^{n+1} = R_j \mathbf{u}^n,$$

where

$$\begin{aligned} L_j \mathbf{u} &= \mathbf{u}_j + \theta \frac{\Delta}{\Delta x} (\mathbf{g}_{j+\frac{1}{2}} - \mathbf{g}_{j-\frac{1}{2}}) \\ R_j \mathbf{u} &= \mathbf{u}_j - (1 - \theta) \frac{\Delta}{\Delta x} (\mathbf{g}_{j+\frac{1}{2}} - \mathbf{g}_{j-\frac{1}{2}}). \end{aligned}$$

For the case $\theta = 0$, the scheme reduces to (3.3). In this case L_j is a linear operator, in fact $L_j \mathbf{u} = \mathbf{u}_j$ so that

$$\mathbf{u}^{n+1} = R_j \mathbf{u}^n,$$

and the numerical solution at the next time level is given explicitly as a function of the solution at the current time level. The scheme is hence called *explicit*. In the case $\theta \neq 0$, L_j is now a nonlinear operator (except for a linear problem where

3.7 Inhomogeneous Conservation Laws

Man

This section has described all the source term discretisations used throughout the rest of this thesis. We will see that the given upwind discretisation with the choice

$$\tilde{b}$$

where

$$\tilde{c}^2 = \left\{ \begin{array}{l} g \left(I_1. \right. \end{array} \right.$$

- (1) For each j set $\hat{\boldsymbol{w}}_j = \boldsymbol{w}_j^n$.
- (2) Compute and store $\left(\boldsymbol{\Phi}_{j+\frac{1}{2}}^\pm\right)^n$ for each cell interface.
- (3) At each cell interface $x_{j+\frac{1}{2}}$ carry out

$$\begin{aligned}\hat{\boldsymbol{w}}_j &= \hat{\boldsymbol{w}}_j + \left(\boldsymbol{\Phi}_{j+\frac{1}{2}}^-\right)^n \\ \hat{\boldsymbol{w}}_{j+1} &= \hat{\boldsymbol{w}}_{j+1} + \left(\boldsymbol{\Phi}_{j+\frac{1}{2}}^+\right)^n.\end{aligned}$$

- (4) For each j set $\boldsymbol{w}_j^{n+1} = \hat{\boldsymbol{w}}_j$

Solving this equation yields

$$\tilde{\alpha}$$

$\tilde{\lambda}_{i,-\frac{1}{2}}^n$ ($i = 1, 2$). The simplest way to obtain values for these wave speeds is to extrapolate from inside the domain, and in particular to take

$$\tilde{\lambda}_{i,-\frac{1}{2}}^n = \tilde{\lambda}_{i,\frac{1}{2}}^n, \quad i = 1, 2.$$

There are three possible situations depending on the signs of these wave speeds. For the case $\tilde{\lambda}_{i,-\frac{1}{2}}^n \leq 0$ ($i = 1, 2$), neither characteristic enters the domain and so no boundary conditions may be specified. The situation is very straightforward since now from (3.22) we have

$$\left(\Phi_{-\frac{1}{2}}^+\right)^n = 0.$$

If $\tilde{\lambda}_{i,-\frac{1}{2}}^n > 0$ ($i = 1, 2$), then both characteristics enter the domain and so both flow variables must be specified at the boundary. In this case we simply overwrite both flow variables at the boundary with the appropriate values. In the case where only one wave speed, say $\tilde{\lambda}_{i,-\frac{1}{2}}^n$ is positive, only one characteristic enters the domain so that only one flow variable must be specified on the boundary. Equation (3.22) gives

$$\left(\Phi_{-\frac{1}{2}}^+\right)^n = -\frac{\Delta}{\Delta x} \tilde{\lambda}_{i,-\frac{1}{2}}^n \tilde{\alpha}_{i,-\frac{1}{2}}^n \tilde{\mathbf{r}}_{i,-\frac{1}{2}}^n.$$

If the boundary condition is $A = A^0(\)$, then this is satisfied at time level $n + 1$ if we choose $\tilde{\alpha}_{i,-\frac{1}{2}}^n$ to satisfy

$$A^0(_{n+1}) = (1, 0) \left(\hat{\mathbf{w}}_0 - \frac{\Delta}{\Delta x} \tilde{\lambda}_{i,-\frac{1}{2}}^n \tilde{\alpha}_{i,-\frac{1}{2}}^n \tilde{\mathbf{r}}_{i,-\frac{1}{2}}^n \right).$$

If the boundary condition is $Q = Q^0(\)$, then this is satisfied at time level $n + 1$ if we choose $\tilde{\alpha}_{i,-\frac{1}{2}}^n$ to satisfy

$$Q^0(_{n+1}) = (0, 1) \left(\hat{\mathbf{w}}_0 - \frac{\Delta}{\Delta x} \tilde{\lambda}_{i,-\frac{1}{2}}^n \tilde{\alpha}_{i,-\frac{1}{2}}^n \tilde{\mathbf{r}}_{i,-\frac{1}{2}}^n \right).$$

Roe's scheme can also be modified to give second order accuracy. The increment at each cell interface is then of the form

$$\hat{\boldsymbol{w}}_j = \hat{\boldsymbol{w}}_j + \left(\boldsymbol{\Phi}_{j+\frac{1}{2}}^- \right)^n - \boldsymbol{B}_{j+\frac{1}{2}}^n$$

Chapter

Theory for the Steady Flow Problem using Vanishing Viscosity

In this chapter we present some theory for the steady state Saint-Venant problem. The theory arises from a novel formulation of the problem and is applicable to a large number of cases.

4.1 Vanishing Viscosity

The Saint-Venant equations are a hyperbolic system of conservation laws. These suffer from two main difficulties, namely solutions may be discontinuous and secondly not all of these so-called weak solutions are physically possible.

Hyperbolic systems of conservation laws often arise from models of physical processes which ignore effects due to viscous or dispersive mechanisms. The next level of accuracy for any such model is to include these effects. The differential equations are modified by the addition of higher order derivatives which are multiplied by small coefficients called *viscosity coefficients*. For the original model to be consistent with the more complete model which includes the viscous or diffusive effects, it is required that the solutions of the two models are “close” in some sense. In particular any solution of the first order system must be the limit of the corresponding solution of the higher order system as the visce

Unfortunately the two models are not generally consistent in the above sense, in that not all weak solutions of the first order system will be vanishing viscosity solutions. It is clearly only the vanishing viscosity solutions which have physical relevance.

In general the higher order system is parabolic and so always has smooth solutions. The apparent discontinuities (which form actual discontinuities in the vanishing viscosity limit) are actually narrow regions where the solution changes extremely rapidly. These regions are called *shock layers*.

The above concept is illustrated by the Euler model of gas dynamics. The Euler equations arise from neglecting terms which model the effects of fluid viscosity from the Navier-Stokes equations, the general model of fluid flow. This is done when the effects of viscosity are thought to be of only secondary importance relative to the effects of inertia. Solutions of the Euler equations, which include discontinuous solutions, are hoped to model the vanishing viscosity limit of solutions to the Navier-Stokes equations. However neglecting the viscous terms introduces solutions which are not vanishing viscosity solutions. Even though the effects of viscosity are small throughout almost all of the flow, they are sometimes still important. In particular their effects are always strong in shock layers. Viscosity prevents the solutions from becoming discontinuous and is also the mechanism for discriminating against unphysical discontinuities. There is a parallel here between the Euler equations and the Saint-Venant equations, since both systems can be derived from the Navier-Stokes equations and both models ignore viscous and diffusive effects. Extensions of the Saint-Venant system which include some of the effects of fluid viscosity are discussed in [58].

By considering the limit of solutions of “some” system of parabolic equations as the viscosity coefficients vanish, we may obtain results concerning the existence and uniqueness of physical solutions to a hyperbolic system. This approach is called the *vanishing viscosity method*. The parabolic problem will have only smooth solutions so that these may be easier to construct. The more difficult step is to then obtain estimates which are independent of the viscosity coefficients and allow passage to the limit.

Consider the scalar Cauchy problem

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0, \tag{4.1}$$

$$u > 0, \quad -\infty < x < \infty, \quad u(x, 0) = U_0(x).$$

Equation (4.1) arises from the conservation of a quantity u transported with flux $f(u)$ and can be written in the integral form

$$\int$$

for all h between h_l and h_r . If we take $m = -1$ then it is not difficult to see that this condition implies

$$E(x, h_r) \leq E(x, h_l), \quad (4.5)$$

because of the relationship (2.27). Thus we conclude that at steady state, any entropy satisfying solution of (4.3) (with $m = -1$) must also be a physical solution of the Saint-Venant equations. The converse is not necessarily true, h

Shocks may occur along a particular reach of channel, $0 \leq x \leq L$, and hence this equation will not in general hold everywhere. Motivated by the previous section we choose to study t

Consider the problem

$$\epsilon u''_\epsilon + a u'_\epsilon = 0, \quad 0 \leq x \leq 1, \quad u_\epsilon(0) = 1, \quad u_\epsilon(1) = 0, \quad (4.9)$$

where $\epsilon > 0$ and $a \neq 0$ is a constant. The solution to this problem is given by

$$u_\epsilon(x) = 1 - \frac{e^{-\frac{ax}{\epsilon}} - 1}{e^{-\frac{a}{\epsilon}} - 1}.$$

First consider the case $a > 0$, where for small ϵ the solution decreases rapidly from one to zero near $x = 0$. In fact as ϵ tends to zero we have

$$u_\epsilon(x) \rightarrow \begin{cases} 1 & x = 0 \\ 0 & x \neq 0. \end{cases}$$

The nonuniform behaviour at $x = 0$ is known as a boundary layer and is characterised by the property:

$$1 = \lim_{\epsilon \downarrow 0} \lim_{x \downarrow 0} u_\epsilon(x) \neq \lim_{x \downarrow 0} \lim_{\epsilon \downarrow 0} u_\epsilon(x) = 0.$$

For the case $a < 0$ we have

$$u_\epsilon(x) \rightarrow \begin{cases} 1 & x \neq 1 \\ 0 & x = 1, \end{cases}$$

as ϵ tends to zero. This corresponds to a boundary layer at $x = 1$.

Next consider the problem:

$$\epsilon u''_\epsilon + x u'_\epsilon = 0, \quad -1 \leq x \leq 1, \quad u_\epsilon(-1) = 1, \quad u_\epsilon(1) = 2, \quad (4.10)$$

where $\epsilon > 0$. The solution to this problem is given by

$$u_\epsilon(x) = 1 + \frac{\operatorname{erf}(\frac{x}{\sqrt{2\epsilon}}) + \operatorname{erf}(\frac{1}{\sqrt{2\epsilon}})}{2\operatorname{erf}(\frac{1}{\sqrt{2\epsilon}})},$$

where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds.$$

As ϵ tends to zero we have

$$u_\epsilon(x) \rightarrow \begin{cases} 1 & x < 0 \\ 3/2 & x = 0 \\ 2 & x > 0. \end{cases}$$

In this case the nonuniformity is in the interior of the domain and as ϵ vanishes the solution tends to a discontinuity at $x = 0$. For this reason, this type of nonuniformity is known as a shock layer. There are other types of nonuniformity that are possible, for example corner layers where the limit is continuous but has a discontinuous first derivative.

The examples given above are all linear problems and the theory for such problems is well understood (for example see [50]). It is usually possible to predict in advance from the differential equation, the type and the position of the nonuniformities. For nonlinear problems this is not the case and the situation is considerably more complicated. Analysis of simple nonlinear problems can be found in [45] and [25]. These make use of asymptotic techniques and usually rely on being able to integrate the reduced differential equation.

Integration of the reduced differential equation is not possible for problem (4.8) and so another approach is required. It happens that theory exists for a class of problems which are very closely related to problem (4.8). This theory comes from a functional analysis approach (as opposed to an asymptotic approach) and applies to a general class of problems. The theory requires some adaptation before it can be applied to (4.8).

4. Functions of Bounded Variation

The theory in this chapter will make use of the class of functions which have bounded total variation. The term bounded total variation was defined in section 3.4, and we define $BV[c, d]$ to be the set of real functions on $[c, d]$ which have bounded total variation. A function $u \in BV[c, d]$ has the following properties:

- (1) The function is bounded.
- (2) All points of discontinuity are simple ($u(x-)$ and $u(x+)$ exist) and the set of discontinuities is countable. Also $u(c+)$ and $u(d-)$ exist.

We consider functions in $BV[0, L]$ which satisfy the integral relationship (2.19). A more common method of defining weak solutions is through the use of test functions

(see [30], [62]). A function h is then a weak solution of the steady flow problem if

$$\int_0^L \phi'(x)F(x, h(x)) + \phi(x)D(x, h(x))dx = 0, \quad (4.11)$$

for all $\phi \in$

4.6 Theory of Lorenz

In this section we adapt theory from the literature so that it can be applied to problem (4.8). The argument is based on work by Lorenz in [32] and can be summarised in the following theorem.

Theorem 1 ([32]) *Consider the two point boundary value problem*

$$\begin{aligned} \epsilon u_\epsilon'' - f(u_\epsilon)' &= b(x, u_\epsilon), \quad 0 \leq x \leq 1, \\ u_\epsilon(0) &= \gamma_0, \quad u_\epsilon(1) = \gamma_1, \end{aligned} \quad (4.12)$$

where $\epsilon > 0$. Suppose that $f \in C^2(-\infty, \infty)$, $b \in C^1([0, 1] \times (-\infty, \infty))$ and that for some constant δ

$$b_u \geq \delta > 0, \quad (4.13)$$

for all u and all $x \in [0, 1]$, then under these conditions the following hold:

- (1) The problem has a unique solution $u_\epsilon \in C^2[0, 1]$ for all $\epsilon > 0$.
- (2) The solution is uniformly bounded in ϵ , i.e. $\|u_\epsilon\|_\infty \leq K_0$ for all $\epsilon > 0$, where K_0 is independent of ϵ .
- (3) The solution has total variation bounded in ϵ , i.e. $\|u_\epsilon'\|_1 \leq K_1$ for all $\epsilon > 0$, where K_1 is independent of ϵ .
- (4) There is a unique function $U \in NBV[0, 1]$ such that $u_\epsilon \rightarrow U$ in L_1 as $\epsilon \downarrow 0$.
- (5) $u = U$ is the only function in $NBV[0, 1]$ which satisfies the following:

$$\left. \begin{aligned} (i) \quad & \text{If } I \text{ is an interval where } u \text{ is continuous, then } f(u(x)) \text{ is dif-} \\ & \text{ferentiable on } I, \text{ one-sided at end points, and the differential} \\ & \text{equation} \\ & -f(u)' = b(x, u), \\ & \text{holds on } I. \end{aligned} \right\} \quad (4.14)$$

$$\left. \begin{aligned} (ii) \quad & \text{If } u \text{ is discontinuous at } x \in (0, 1), \text{ then} \\ & f(u_l) = f(u_r) \geq f(k) \quad \text{if } u_l > u_r, \\ & f(u_l) = f(u_r) \leq f(k) \quad \text{if } u_l < u_r, \\ & \text{for all } k \text{ between } u_l = u(x-) \text{ and } u_r = u(x+). \end{aligned} \right\} \quad (4.15)$$

$$\left. \begin{aligned}
(iii) \quad & \text{For } j = 0, 1 \text{ and } k \text{ between } u(j) \text{ and } \gamma_j \\
& (-1)^{j+1} \operatorname{sgn}(u(j) - \gamma_j)(f(u(j)) - f(k)) \geq 0, \\
& \text{where } \operatorname{sgn}(x) = -1, 0, 1 \text{ for } x < 0, = 0, > 0, \text{ respectively.}
\end{aligned} \right\} \quad (4.16)$$

The above theory relates to a problem closely resembling problem (4.8). This is made clearer by a transformation onto the unit interval given by

$$\begin{aligned}
u_\epsilon(x) &\equiv h_\epsilon(xL), \\
f(x, u) &\equiv -LF(xL, u), \\
b(x, u) &\equiv L^2 D(xL, u).
\end{aligned} \quad (4.17)$$

Theorem 1 will be adapted to apply to this problem under certain conditions. To do this requires some understanding of Theorem 1 and how it is constructed.

Part 1 of the theorem gives the existence and uniqueness of the solution to the singular perturbation problem for each positive ϵ . The existence proof relies on Nagumo's Lemma ([43],[27]), which uses the fact that the problem has both upper and lower solutions. The functions $\bar{u}(x)$, $\underline{u}(x)$ are upper and lower solutions, respectively, if the following hold for all x in the interval $[0, 1]$:

- (1) $\underline{u} \leq \bar{u}$
- (2) $\epsilon \underline{u}'' - f(\underline{u})' - b(x, \underline{u}) \leq 0$
- (3) $\underline{u}(0) \leq \gamma_0, \quad \underline{u}(1) \leq \gamma_1$
- (4) $\epsilon \bar{u}'' - f(\bar{u})' - b(x, \bar{u}) \geq 0$
- (5) $\bar{u}(0) \geq \gamma_0, \quad \bar{u}(1) \geq \gamma_1$

The condition $b_u \geq \delta > 0$ holds if

$$u' > \langle \rangle < \bar{u}'$$



The uniform bound of part 2 of the theorem comes directly from the existence proof, since the upper and lower solutions are independent of ϵ . This bound and the uniform bound on the total variation, from part 3 of the theorem, gives that the set $\{u_\epsilon\}_{\epsilon>0}$ is precompact in $L_1[0, 1]$. Thus for any positive null sequence $S = \{\epsilon_n\}$, there is a subsequence $S' = \{\epsilon_{n_k}\}$ and a function $U \in NBV[0, 1]$ such that

$$u_\epsilon \rightarrow U \text{ in } L_1 \text{ as } \epsilon \downarrow 0, \epsilon \in S'.$$

Part 5 of the theorem gives the properties of the limit function and states that there is exactly one function in $NBV[0, 1]$ with these properties.

4.7 The Modified Theory

The main difference between D_5 -Tc5NB-T-Tc5ceT-VT5er,TivnfuT5-TD5ill

() Problem P_ϵ has a unique solution $u_\epsilon \in C^2[0, 1]$ for all $\epsilon > 0$ and this satisfies the bounds

$$0 < \underline{u}$$

- (1) Problem $P_{\epsilon}^{\alpha,\beta}$ satisfies the conditions of Theorem .

The function $f^{\alpha,\beta}$ is constructed to be continuous and have continuous first and second derivatives at both $u = \alpha$ and $u = \beta$, hence $f^{\alpha,\beta} \in C^2(-\infty, \infty)$.

Also

$$b_u^{\alpha,\beta}(x, u) = \begin{cases} b_u(x, \beta) & u > \beta \\ b_u(x, u) & \alpha \leq u \leq \beta \\ b_u(x, \alpha) & u < \alpha, \end{cases}$$

and

$$b_x^{\alpha,\beta}(x, u) = \begin{cases} b_x(x, \beta) + (u - \beta)b_{ux}(x, \beta) & u > \beta \\ b_x(x, u) & \alpha \leq u \leq \beta \\ b_x(x, \alpha) + (u - \alpha)b_{ux}(x, \alpha) & u < \alpha, \end{cases}$$

so since b_x, b_u, b_{ux} are continuous on $[0, 1] \times (0, \infty)$, it follows that $b^{\alpha,\beta} \in C^1([0, 1] \times (-\infty, \infty))$. Finally

$$b_u^{\alpha,\beta} \geq \delta > 0,$$

where

$$\delta = \min_{\substack{0 \leq x \leq 1 \\ \alpha \leq u \leq \beta}} \{b_u(x, u)\}.$$

- (2) $\partial_{\epsilon} \{ \epsilon \in \mathbb{R} \mid \epsilon \in \mathbb{R} \setminus \mathbb{Q} \mid \epsilon \times \downarrow \epsilon \mid \epsilon \in \mathbb{R} \mid \epsilon \downarrow \}$

Next suppose that $u_\epsilon < u$

- (6) *There is a unique function $U \in NBV_+[0, 1]$ such that $u_\epsilon \rightarrow U$ in L_1 as $\epsilon \downarrow 0$. The function U satisfies the bounds (4.22).*

Applying part 4 of Theorem 1 to $P_\epsilon^{\alpha, \beta}$ gives that there is a unique function $U \in NBV[0, 1]$ such that $u_\epsilon \rightarrow U$ in L_1 as $\epsilon \downarrow 0$. We show that U satisfies the bounds (4.22) and hence is in $NBV_+[0, 1]$.

We can choose a positive sequence $S = \{\epsilon_n\}$ such that $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$ and

$$u_\epsilon \rightarrow U \text{ a.e. as } \epsilon \downarrow 0, \epsilon \in S.$$

Define the set

$$X = \{x \in [0, 1] : u_\epsilon(x) \rightarrow U(x) \text{ as } \epsilon \downarrow 0, \epsilon \in S\}. \quad (4.23)$$

The set $[0, 1] \setminus X$ has zero measure and the bounds (4.22) clearly hold for all $x \in X$. Now for arbitrary x in $[0, 1]$ by definition of the set NBV we have

$$\begin{aligned} U(x) &= \lim_{\substack{s \downarrow x \\ s \in X}} U(s), \\ U(1) &= \lim_{\substack{s \uparrow 1 \\ s \in X}} U(s), \end{aligned}$$

and thus the bounds (4.22) hold at all points, giving part 3 of Theorem 2.

- (7) *Prove that T^3 is a T -invariant measure.*

Firstly we have that

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

so that

$$F'(h) = -\frac{Q^2 T(h)}{A(h)^2} + gI_1'(h)$$

Hence h_n is bounded and in particular bounded below above zero. Now taking m and M as in (4.30) and observing that $D(x, h_n(x)) \equiv 0$, we have that

$$D(x, m) \leq D(x, h_n(x)) = 0$$

and

$$D(x, M) \geq D(x, h_n(x)) = 0,$$

for $0 \leq x \leq L$, showing that (4.20) holds. We have now shown that the functions f and b given by (4.30) satisfy all the conditions of Theorem 2. Theorem 3 follows by simply writing Theorem 2 in terms of the functions F , D and using the transformations

$$h_\epsilon(x) \equiv u_\epsilon(xL), \quad H(x) \equiv U(xL).$$

Interpretation of Theorem 3 The above theorem is only an intermediate step, however it is extremely important to this thesis, because it defines the conditions under which we can make progress. The conditions of the theorem will be assumed to hold in what follows.

Consider a function $h \in NBV_+[0, L]$ which satisfies

$$[F(h(x))]_{x_1}^{x_2} = \int_{x_1}^{x_2} D(x, h(x))dx, \quad \text{for all } x_1, x_2 \in [0, L]. \quad (4.31)$$

First we observe that the integral is mathematically sensible since $D(x, h(x))$ is in $L_1[0, L]$ and is bounded. For fixed x_1 the right hand side is continuous in x_2 (see [65] p.319), so it follows that $F(h(x))$ must be continuous on $[0, L]$. Thus if h is discontinuous at $x \in (0, L)$ then the jump condition

$$F(h(x-)) = F(h(x+)) \quad (4.32)$$

must hold. Following [65](p.319) it can be shown that for $x \in [0, L)$

$$\lim_{s \downarrow 0} \left(\frac{F(h(x+s)) - F(h(x))}{s} \right) = D(x, h(x+)),$$

and that for $x \in (0, L]$

$$\lim_{s \uparrow 0} \left(\frac{F(h(x+s)) - F(h(x))}{s} \right) = D(x, h(x-)).$$

Thus for an interval where h is continuous, at interior points the reduced differential equation must hold since both of the one-sided derivatives equal $D(x, h(x))$. At any end points the corresponding one-sided differential equation clearly holds. We can now give the precise mathematical definitions of what we mean by a solution of the steady flow problem for a prismatic channel.

Definition 4.1 (Type-I Solution) *A function $h \in NBV_+[0, L]$ is a type-I solution of the steady flow problem if (4.31) holds and at any discontinuity $x \in (0, L)$*

$$E(h(x+)) \leq E(h(x-)), \quad (4.33)$$

where E is given by (2.14).

Definition 4.2 (Type-II Solution) *A function $h \in NBV_+[0, L]$ is a type-II solution of the steady flow problem if (4.31) holds and at any discontinuity $x \in (0, L)$*

$$\frac{F(k) - F(h(x-))}{k - h(x-)} \leq 0, \quad \text{for all } k \text{ between } h(x-) \text{ and } h(x+). \quad (4.34)$$

The definition of type-I solutions corresponds to the definition of physical solutions of the steady flow problem as introduced in section 2.2. The definition of the type-II solutions is stronger and arises naturally from our theory. Condition (4.34) is simply Oleinik's condition for a steady shock for the problem (4.6). We observed in section 4.2 that any type-II solution is also type-I solution, since $F(h(x-)) = F(h(x+))$ along with (4.34) implies (4.33) (by (2.27)). The converse of this is not necessarily true, i.e. a type-I solution is not necessarily a type-II solution. In the next section we introduce further assumptions in order that these two definitions are equivalent.

We can now give the following theorem.

Theorem 4 *For $\gamma_0, \gamma_1 > 0$ and under the conditions of Theorem 3, the function $h = H$ is the only type-II solution which satisfies*

$$\left. \begin{aligned} (1) \quad & \text{For all } k \text{ between } \gamma_0 \text{ and } h(0) \\ & \text{sgn}(h(0) - \gamma_0)(F(h(0)) - F(k)) \geq 0. \end{aligned} \right\} \quad (4.35)$$

$$\left. \begin{aligned} (2) \quad & \text{For all } k \text{ between } \gamma_1 \text{ and } h(L) \\ & \text{sgn}(h(L) - \gamma_1)(F(h(L)) - F(k)) \leq 0. \end{aligned} \right\} \quad (4.36)$$

Proof of Theorem 4 This theorem is proved in two parts. We start by demonstrating that the function $h = H$ is a type-II solution.

Firstly $H \in NBV_+[0, L]$. Using property 2 of Theorem 3 we have that $\epsilon h'_\epsilon \rightarrow 0$ in L_1 as $\epsilon \downarrow 0$. Hence thw

and therefore we must ha

Region	Subregion	$H(0)$	$H(L)$
$\gamma_0 \geq h_c, \quad \gamma_1 \leq h_c$		α_0	α_1
$\alpha_0^* \leq \gamma_0 < h_c, \quad \gamma_1 \leq h_c$		α_0	α_1
$\gamma_0 < \alpha_0^*, \quad \gamma_1 \leq h_c$		γ_0	$\beta_1(\gamma_0)$
$\gamma_0 \geq h_c, \quad h_c < \gamma_1 \leq \alpha_1^*$		α_0	α_1
$\gamma_0 \geq h_c, \quad \gamma_1 > \alpha_1^*$		$\beta_0(\gamma_1)$	γ_1
$\alpha_0^* \leq \gamma_0 < h_c, \quad h_c < \gamma_1 \leq \alpha_1^*$		α_0	α_1
$\gamma_0 < \alpha_0^*, \quad h_c < \gamma_1 \leq \alpha_1^*$	$\gamma_1 \leq \beta_1(\gamma_0)^*$	γ_0	$\beta_1(\gamma_0)$
	$\gamma_1 > \beta_1(\gamma_0)^h$		

Flow supercritical at inflow and subcritical at outflow If γ_0 and γ_1 satisfy one of the conditions below, then there is a solution which satisfies $h(0) = \gamma_0$ and $h(L) = \gamma_1$.

- (1) $0 < \gamma_0 \leq \alpha_0^* \leq h_c$ and $h_c \leq \beta_1(\gamma_0)^* < \gamma_1 \leq \alpha_1^*$
- (2) $h_c \leq \alpha_0^* \leq \gamma_0 < \beta_0(\gamma_1)^*$ and $\gamma_1 > \alpha_1^* \geq h_c$
- (3) $0 < \gamma_0 < \min\{\alpha_0^*, \beta_0(\gamma_1)^*\} \leq h_c$ and $\gamma_1 > \max\{\alpha_1^*, \beta_1(\gamma_0)^*\} \geq h_c$.

From the above we observe that in order to specify the depth at inflow with any degree of freedom the depth specified must at the very minimum correspond to supercritical flow (and even then there will only be solutions for certain ranges of depth). Similarly to specify the depth at outflow with any degree of freedom requires this depth to correspond to subcritical flow. This observation agrees with the theory of characteristics discussed in section 2.2.4.

We end this section by demonstrating that practical problems exist which do satisfy the conditions required by the theory. The major restrictions placed by theory are as follows:

- (1) The channel must be prismatic
- (2) The bed slope must be positive.
- (3) The conveyance must satisfy (4.24).
- (4) There must be only one critical depth.

The condition that the bed slope is positive appears to be the most restrictive. However, as we demonstrate later, when this condition is violated the uniqueness conclusions of the theory may not hold. This condition on the conveyance is only a slightly stronger version of the condition (2.31) which is used in section 2.2.2. If we again take the form (2.17) for the conveyance and now require that

$$k_1 \geq 1/2 \text{ and } 0 \leq k_2 \leq k_1 - 1/2, \quad (4.38)$$

which includes both the Manning and Chezy forms, then conditions (4.24) are satisfied for rectangular, trapezoidal and triangular channels. This can be seen by

using (2.33) with k_1 replaced with $k_1 - 1/2$. Such cross-sections also have a unique critical depth (see section 2.2.2). There is no obvious way of showing that these conditions hold for a wider class of cross-sections and friction laws, other than testing each individual case.

4.10 Extension of the Theory

The theory derived in this chapter has certain limitations on the situations it can be applied to. In this section we discuss whether these limitations may be overcome.

Theorem 3 requires that the bed slope is positive and that (4.24) holds, in order that $D_h > 0$ for all $h > 0$ and all $0 \leq x \leq L$. If this condition is violated, then are the conclusions of the theorem still true? We demonstrate that in general they are not.

Consider a “well-behaved” channel in the sense of section 2.2.1. The channel, which need not be prismatic, has a single critical depth $h_c(x)$ at each cross-section, and a jump is allowable at x if and only if

$$h(x-) < h_c(x) < h(x+) = h(x-)^*.$$

Suppose that $\gamma_0 < h_c(0)$, $\gamma_1 > h_c(L)$ and that the following two problems have solutions:

$$\left. \begin{aligned} F(x, h^1(x))' &= D(x, h^1(x)), \quad h^1(x) < h_c(x), \quad 0 \leq x \leq L, \\ h^1(0) &= \gamma_0, \end{aligned} \right\} \quad (4.39)$$

$$F(x, h^2(x))' = D(x, h^2(x)), \quad h^2(x) > h_c(x), \quad 0 \leq x \leq L,$$

which satisfies both $h(0) = \gamma_0$ and $h(L) = \gamma_1$. Thus if J has more than a single root, then there is more than one physical solution satisfying the same boun

We conclude that when $D_h > 0$ is violated, the conclusions of the theory may no longer be true, i.e. there may be more than one physical weak solutions satisfying identical boundary values. In a case where this happens, which of the solutions is the solution we require? It may only be possible to answer this question by examining the transieny p tet T 5 ? ? T l e o n T 5 ? ? T D 5 ? T D s

Chapter 5

A Class of Numerical Methods

In the previous chapter we demonstrated that under certain conditions there is at most one physical solution to the steady Saint-Venant problem, for any given boundary values, and that this solution is the vanishing viscosity solution of a second order two-point boundary value problem. In this chapter we follow on from these ideas to consider a family of finite difference approximations to the steady flow problem. As before we consider only prismatic channels, although the schemes will be extended to non-prismatic channels in Chapter 9. The basis of the theory, as in the previous chapter, is the work by Lorenz[32] although other authors, notably Abrahamsson and Osher[1] and Osher[47], have made significant contributions. Other closely related work by Lorenz can be found in [35], [34] and [33].

The steady flow equation (2.21) for a prismatic channel can be written as

$$\frac{d}{dx}f(h) = -D(x, h), \quad (5.1)$$

where $f(h) \equiv -F(h)$ and the functions F and D are given by (2.6) and (2.7) respectively. We consider approximations to this equation of the form

$$\frac{g(h_{j+1}, h_j) - g(h_j, h_{j-1})}{\Delta x} = -D(x_j, h_j), \quad (5.2)$$

where $x_j = j\Delta x$, $h_j \approx h(x_j)$ and Δx is the uniform grid spacing. We require that

$$g(h, h) = f(h) \quad (5.3)$$

for all positive h , in order that the scheme be consistent with the differential equation. A motivation for considering such a scheme comes from the previous chapter

where we observed that under certain conditions the physical solutions of the steady flow problem are exactly the steady state entropy satisfying solutions of the scalar conservation law (4.6). Applying a three-point conservative finite difference scheme to this scalar conservation law and using a pointwise discretisation of the source term yields the scheme

$$\frac{h_j^{n+1} - h_j^n}{\Delta} + \frac{g(h_{j+1}^n, h_j^n) - g(h_j^n, h_{j-1}^n)}{\Delta x} = -D(x_j, h_j^n), \quad (5.4)$$

where $h_j^n \approx h(j\Delta x, n\Delta t)$ and again (5.3) is required for consistency. At steady state this reduces to (5.2). In theory, almost any of the vast amount of numerical methods for scalar conservation laws (some of which are described in Chapter 3) could be applied to (4.6) and so be used to compute solutions of the steady flow problem, and in Chapter 7 we apply some specimen schemes from the literature.

From the viewpoint of theory we consider only simple schemes of the above type. The following theorem is due to

reason we may perform the limit as ϵ vanishes immediately to obtain the scheme

$$\frac{g(u_{j+1}, u_j) - g(u_j, u_{j-1})}{\Delta x} + b(x_j, u_j) = 0, \quad j = 1, \dots, N-1 \quad (5.6)$$

$$u_0 = \gamma_0, \quad u_N = \gamma_1,$$

and we are now only concerned with the limit as Δx vanishes. This scheme is clearly of the same form as (5.2) for the appropriate functions f and b and a transformation onto the interval $[0, L]$. The theory of Lorenz described above is summarised by the following theorem. This is essentially the discrete analogue of Theorem 1 with the role of ϵ replaced by Δx . The proof can be found in [32].

Theorem 6 ([32]) *Suppose the situation is as in Theorem 1 and consider the difference equations*

$$\begin{aligned} \mathcal{T}_j u &= 0, & j &= 1, 2, \dots, N-1 \\ u_0 &= \gamma_0, & u_N &= \gamma_1, \end{aligned} \quad (5.7)$$

where

$$\mathcal{T}_j u = \frac{g(u_{j+1}, u_j) - g(u_j, u_{j-1})}{\Delta x} + b(x_j, u_j),$$

$\Delta x = 1/N$, $x_j = j\Delta x$ and the function g has the following properties:

$$(1) \quad g(u, v) = g(v, u).$$

The first condition placed on the numerical flux function g is simply the consistency condition. The second and third conditions on g ensure that that the scheme is monotone. To be precise, when we say that the scheme is monotone, w

$$\mu_1(u) = \begin{cases} \beta & u > \beta \\ u & \alpha \leq u \leq \beta \\ \alpha & u < \alpha, \end{cases}$$

$$\mu_2(u) = \begin{cases} \alpha & u > \alpha \\ u & u \leq \alpha, \end{cases}$$

and

$$\mu_3(u) = \begin{cases} \beta & u < \beta \\ u & u \geq \beta. \end{cases}$$

We start by showing that conditions (1)-(4) of Theorem 6 hold for $g^{\alpha,\beta}$. Firstly we have that

$$g^{\alpha,\beta}(. \\$$

is Lipschitz continuous. The first term is Lipschitz continuous since for $j = 1, 2, 3$ and for each $u_1, u_2 \in \mathcal{R}$ we have

$$|\mu_j(u_2) - \mu_j(u_1)|$$

equation for $j = j_1$ yields

$$-b^{\alpha,\beta}(x$$

.2 The Tim Stopping Iteration

In this section we consider a method for solving the system of difference equations.

The theory relies on the following lemma which is simply an application of the con-

traction mapping theoremEimpTJ3D3tTj~TD3EL~~D3Tj3~mTJ3~32]~TD3TTj3~mCida~~he~

This demonstrates that $\mathbf{G}$ maps onto its own domain. The contractivity of $\mathbf{G}$ then gives the uniqueness and existence of the fixed point by the contraction mapping theorem (for example see ref. [46] section 5.1.3). The convergence of the sequence given by (5.11) arises from observing observing that

$$\|\mathbf{u}^n - \mathbf{u}\|_1 = \|\mathbf{G}(\mathbf{u}^{n-1}) - \mathbf{G}(\mathbf{u})\|_1 \leq k \|\mathbf{u}^{n-1} - \mathbf{u}\|_1,$$

and hence by induction

$$\|\mathbf{u}^n - \mathbf{u}\|_1 \leq k^n \|\mathbf{u}^0 - \mathbf{u}\|_1 \leq \|\mathbf{u}^0 - \mathbf{u}\|_1 e^{-n(1-k)}.$$

Here we have used the fact that $k \leq e^{-(1-k)}$ for $k \in [0, 1]$. This completes the proof.

Under the conditions of Theorem 7 we apply the above lemma to the m h e m

since $\alpha \leq \underline{u} = \min\{\gamma_0, \gamma_1, m\}$ and hence $b(x_j, \alpha) \leq 0$. Also we have

$$\mathbf{G}(\underline{u}) = \begin{bmatrix} \gamma_0 \\ \beta - \Delta b(x_1, \beta) \\ \vdots \\ \beta - \Delta b(x_j, \beta) \\ \vdots \\ \beta - \Delta b(x_{N-1}, \beta) \\ \gamma_1 \end{bmatrix} \leq \underline{u},$$

since $\beta \geq \bar{u} = \max\{\gamma_0, \gamma_1, M\}$ and hence $b(x_j, \beta) \geq 0$.

We next investigate the circumstances under which $\mathbf{G}$ satisfies condition (2) of Lemma 5.1. For $u_1, u_2, v_1, v_2 \in [\alpha, \beta]$ we define the functions

$$l^u(u_2, u_1; v_1) = \begin{cases} \frac{g(u_2, v_1) - g(u_1, v_1)}{u_2 - u_1} & \text{if } u_2 \neq u_1 \\ 0 & \text{if } u_2 = u_1, \end{cases} \quad (5.13)$$

and

$$l^v(u_2, u_1; v_1) = \begin{cases} \frac{g(v_1, u_2) - g(v_1, u_1)}{u_2 - u_1} & \text{if } u_2 \neq u_1 \\ 0 & \text{if } u_2 = u_1, \end{cases} \quad (5.14)$$

which from the properties of g are bounded and satisfy

$$l^u(u_2, u_1; v_1) \leq 0, \quad l^v(u_2, u_1; v_1) \geq 0.$$

We can now write

$$\begin{aligned} g(u_2, v_2) - g(u_1, v_1) &= g(u_2, v_2) - g(u_1, v_2) + g(u_1, v_2) - g(u_1, v_1) \\ &= l^u(u_2, u_1; v_2)(u_2 - u_1) + l^v(v_2, v_1; u_1)(v_2 - v_1). \end{aligned}$$

Using this relationship and applying the mean value theorem to the difference in the term involving b , we can for $\mathbf{u}, \mathbf{v} \in [\boldsymbol{\alpha}, \boldsymbol{\beta}]$ write

$$\mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v}) = M(\mathbf{u} - \mathbf{v}),$$

wh

since $r_0 \geq 0$. For the j^{th} column ($3 \leq j \leq N-2$) the sum is given by

$$r_{j-2} + q_{j-1} + p_j = 1 - \Delta b_u(x_{j-1}, \hat{u}_{j-1}) \leq 1 - \Delta \delta.$$

The same argument shows that the remaining two column sums satisfy the same bound, hence we conclude that

$$\|M\|_1 \leq 1 - \Delta \delta < 1.$$

We can obtain a slightly less restrictive requirement on the parameter Δ than given by (5.16) if the function $g(u, v)$ is assumed to be continuously differentiable for all $u, v \in [\alpha, \beta]$. In this case the function $\mathbf{G}$ is Frechet-differentiable and for $\mathbf{u}, \mathbf{v} \in [\boldsymbol{\alpha}, \boldsymbol{\beta}]$ we can write

$$\mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v}) = M(\mathbf{u} - \mathbf{v}),$$

where

$$M = \int_0^1 \mathbf{G}'(\mathbf{u} + s(\mathbf{v} - \mathbf{u})) ds,$$

(see [46], sections 3.2.6 and 3.2.8). The Jacobian $\mathbf{G}'(\mathbf{u})$ is again of the form (5.15)

where now

$$\begin{aligned} p_j &= \frac{\Delta}{\Delta x} g_v(u_j, u_{j-1}), \\ q_j &= 1 - \Delta \left(\frac{g_v(u_{j+1}, u_j) - g_u(u_j, u_{j-1})}{\Delta x} + b_u(x_j, u_j) \right) \\ &= 1 - \Delta b_u(x_j, u_j) - p_{j+1} - r_{j-1}, \\ r_j &= -\frac{\Delta}{\Delta x} g_u(u_{j+1}, u_j). \end{aligned}$$

As before $p_j, r_j \geq 0$, but in this case the condition

$$\Delta \left(\frac{g_v(u_1, u_2) - g_u(u_2, u_3)}{\Delta x} + b_u(x_j, u_2) \right) \leq 1, \quad (5.18)$$

for all $u_1, u_2, u_3 \in [\alpha, \beta]$ and $0 \leq j \leq N$,

is sufficient to ensure $q_j \geq 0$.

We estimate the L_1 norm of the matrix $\mathbf{G}'(\mathbf{u})$ by computing the sum of each column. The sum of the first column is p_1 , and using condition (5.18) with the correct values we obtain

$$p_1 \leq 1 - \Delta b_u(x_0, u_0) + \frac{g_u(u_0, u_0)}{\Delta x} \leq 1 - \Delta \delta,$$

since $g_u \leq 0$. The sum of the second column is given by

$$q_1 + p_2 = 1 - \Delta b_u(x_1, u_1) - r_0 \leq 1 - \Delta \delta,$$

since $r_0 \geq 0$. For the j^{th} column ($3 \leq j \leq N - 2$) the sum is given b.

$j = 1, 2, \dots, N - 1$, where $u_0^n = \gamma_0$ and $u_N^n = \gamma_1$. This is a first order time accurate approximation to the partial differential equation

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = -b(x, u), \quad 0 \leq x \leq 1,$$

and the condition on the time step Δt plays the role of the CFL condition.

3 Application to the Steady Flow Problem

In this section we apply the theory from the previous section to the steady flow problem. Application of Theorem 7 under the conditions of Theorem 3 yields the following theorem. Any discussion throughout the remainder of this chapter assumes that the problem satisfies the conditions of Theorem 3 and also the additional assumptions of Theorem 5.

Theorem *Suppose the situation is as in Theorem 3 and consider the difference equations*

$$\begin{aligned} \mathcal{T}_j h &= 0, & j &= 1, 2, \dots, N - 1 \\ h_0 &= \gamma_0, & h_N &= \gamma_1, \end{aligned} \tag{5.20}$$

where

$$\mathcal{T}_j h = \frac{g(h_{j+1}, h_j) - g(h_j, h_{j-1})}{\Delta x} + D(x_j, h_j),$$

$\Delta x = L/N$ and $x_j = j\Delta x$. Suppose that for some $0 < \alpha \leq \underline{h}$, $\beta \geq \overline{h}$ the function g has the properties:

(1) $g(u, u) = -F(u)$ for all $u \in [\alpha, \beta]$.

(2) $u_1, u_2, v \in [\alpha, \beta]$ with $u_1 < u_2 < v$ and $u_1 < u_2 < v$ then $g(u_1, u_2) < g(u_2, u_1) < g(u_1, v) < g(v, u_1) < g(u_2, v) < g(v, u_2)$.

in $[\underline{\alpha}, \bar{\alpha}]$ for each $N \in \mathbb{N}$. This solution satisfies the bounds:

$$0 < \underline{h} \leq h_j^{\Delta x} \leq \bar{h}, \quad j = 0, 1, \dots, N. \quad (5.21)$$

If $H^{\Delta x} \in L_1[0, L]$ denotes the piecewise constant extension of this discrete solution given by

$$H^{\Delta x} = h_j^{\Delta x} \quad \text{for } j\Delta x \leq x < (j+1)\Delta x, \quad j = 0, 1, \dots, N \quad (5.22)$$

then $H^{\Delta x} \rightarrow H$ in L_1 as $\Delta x \rightarrow 0$, where $H \in NBV_+[0, L]$ is the limiting solution of problem (4.8) as $\epsilon \downarrow 0$.

Proof of Theorem As in the proof of Theorem 3 we write

$$\begin{aligned} f(u) &= -LF(u), \\ b(x, u) &= L^2 D(xL, u), \end{aligned}$$

to transform the continuous problem (4.8) onto the unit interval. This problem then satisfies the conditions of Theorem 2. Likewise transform the discret

Under these conditions the mapping (5.12) has exactly one red point $\mathbf{h}$ which is the only solution in [

and $c > 0$ is arbitrary. This choice of g is continuously differentiable and we have

$$g_u(u, v) = f'_-(u) = \min\{f'(u), 0\} \leq 0,$$

$$g_v(u, v) = f'_+(v) = \max\{f'(v), 0\} \geq 0.$$

Also we have $g(u, u) \equiv f(u)$ for positive u , so that conditions 1-4 of Theorem 9 are satisfied for any choice of α and β . Hence under the conditions of Theorem 3 the conclusions of Theorem 9 are true. Theorem 9 does not show the global uniqueness of the discrete solution, however since the conditions on g hold for arbitrary α and β , the corollary to the theorem is valid, giving the global uniqueness.

Using the fact that g is continuously differentiable, the stronger form of the CFL condition (condition 5.24) is sufficient to guarantee convergence of the time stepping iteration. This condition reduces to the one required in [14] if $\mathbf{A}$ is

and

$$0 \leq l^v(h_1, h_2; h_3) \leq \overline{|f'|}, \quad (5.30)$$

where

$$\overline{|f'|} = \max_{\alpha \leq h \leq \beta} \{|f'(h)|\}.$$

Since the consistency condition also holds then g satisfies conditions 1-4 of Theorem 9, hence under the conditions of Theorem 3 the conclusions of Theorem 9 are true. Again the conditions on g hold for arbitrary α and β , thus the corollary to the Theorem gives the global uniqueness of the discrete solution.

Since in this case the numerical flux function is not everywhere differentiable, we must use the weaker form of the CFL condition (condition 5.23) to ensure the convergence of the time stepping iteration. Using the bounds (5.29) and (5.30), the requirement that

$$\Delta \left(\frac{2|f'(h_1)|}{\Delta x} + D_h(x_j, h_2) \right) \leq 1, \quad (5.31)$$

for all $h_1, h_2 \in [\alpha, \beta]$ and $0 \leq j \leq N$ can be seen to be sufficient. The difference between this condition and that for the Engquist-Osher form is the addition of the factor two in the first term. Hence as Δx becomes small the condition will only allow a time step of half that allowed by the Engquist-Osher scheme. It is likely that more thorough analysis, using the fact that g is only non-differentiable on isolated curves in the u - v space, can eliminate this extra factor from the CFL condition.

The Lax-Friedrichs form of the numerical flux function is given by

$$g(u, v) = \frac{1}{2}$$

values of u and v . The best we can achieve is to enforce these conditions to hold over the finite range $\alpha \leq u, v \leq \beta$, by taking λ such that

$$\lambda \leq |f'(h)|, \quad \text{for all } \alpha \leq h \leq \beta. \quad (5.33)$$

Conditions 1-4 of Theorem 9 are then satisfied, and under the conditions of Theorem 3 the conclusions of Theorem 9 are true. In this case, however, the corollary to Theorem 9 is not valid and thus the discrete solution may not be globally unique. The theory does not preclude the existence of other solutions which are not contained in the set $[\alpha, c]$.

Th ory into Pr ctic

In this section we describe how to carry out the necessary steps to utilise the theory and obtain an efficient, robust and practical algorithm for computing solutions to the steady flow problem. We consider the following five steps.

- (1) Choose the values for γ_0 and γ_1 .
- (2) Determine bounds on the normal depth for the problem and hence find bounds on the exact solution.
- (3) Choose the starting vector $\mathbf{h}^0$ for the time stepping iteration and then appropriate values for α and β .
- (4) Ensure the numerical flux function satisfies conditions 1-4 of Theorem 9.
- (5) Find a value of Δ which satisfies the CFL condition and hence guarantees the convergence of the time stepping iteration.

For a given problem, the first step is to choose values for γ_0 and γ_1 in order to give the required solution. In section 4.9 we observed that

the corresponding γ_j as the critical depth since this minimises the range over which the CFL condition must hold.

The next step is to obtain bounds for the normal depth h_n . The normal depth is defined by

$$K(h_n(x)) = Q$$

The convergence rate estimate (5.25) indicates that a larger value of Δ can yield a faster rate of convergence, so it is of interest to find the greatest time step allowable by the CFL condition. For all the three forms of g discussed in the previous section, the CFL condition can be written as

$$\Delta \Gamma(x_j, h_1, h_2) \leq 1,$$

for $j = 0, 1, \dots, N$ and all $h_1, h_2 \in [\alpha, \beta]$. The greatest allowable time step is then given by

$$\Delta_{\text{opt}} = \left(\max_{\substack{0 \leq j \leq N \\ h_1, h_2 \in [\alpha, \beta]}} \{\Gamma(x_j, h_1, h_2)\} \right)^{-1}.$$

We are therefore required to maximise the

Chapter 6

Test Problems

makes this profile an actual solution of the steady equation is then found. The method can be used to construct test problems with almost any desired features, including hydraulic jumps. Hence these test problems can be used to compare the numerical results, for any algorithm, with an exact solution. The method is also useful for evaluating unsteady solvers, since, if an unsteady model is given steady boundary conditions, the limiting steady solution can be compared with the analytic steady solution. The method presented in this chapter fits in well with the validation documentation initiative of the European hydraulics laboratories (see [12]), since it enables the creation of benchmark test problems which can be used as a standard measure for the performance of commercial software packages.

6.1 Test Problems with Smooth Solutions

It is convenient to write equation (2.25) as

$$S_0(x) = f_1(x, h(x))h'(x) + f_2(x, h(x)), \quad (6.1)$$

where

$$f_1 = 1 - \frac{Q^2 T}{gA^3} = 1 - F_r^2 \quad (6.2)$$

and

$$f_2 = \frac{Q^2}{K^2} - \frac{Q^2}{gA^3} \int_0^h \sigma_x d\eta. \quad (6.3)$$

The crux of the work in this chapter depends on the following argument. Suppose that for some reach the function T representing channel width is arbitrarily defined. For example for a rectangular channel we would define $T = B$, where $B(x) > 0$ gives the width. If the conveyance function K is completely specified and a value for the discharge Q is given, then the functions f_1 and f_2 given by (6.2) and (6.3) are completely defined. The main part of the method is to choose a hypothetical depth profile $\hat{h}(x)$ for the reach, which at this stage we assume to be smooth. We then use the following formula to determine the bed slope for the reach:

$$S_0(x) = f_1(x, \hat{h}(x))\hat{h}'(x) + f_2(x, \hat{h}(x)). \quad (6.4)$$

It is not difficult to conclude that, for the above situation, the function $h = \hat{h}$ satisfies the differential equation (6.1) for the entire reach.

We can now use the above argument to specify a benchmark test problem for which the exact solution is known. The following information is required:

✓

—●≤

get round this difficulty. Suppose that the hypothetical depth profile is now only piecewise smooth, where all the discontinuities represent physical hydraulic jumps, i.e. satisfy (2.23) and (2.24). Consider a discontinuity at $x = x^*$. Using (6.4) the bed slope is not defined at x^* and this corresponds to a discontinuity in the bed slope, i.e.

$$S_0(x^*-) \neq S_0(x^*+).$$

This is not a great difficulty since this yields a perfectly realistic bed profile and one may go ahead and use this as test problem. However we feel that it is worthwhile taking further steps to improve the quality of the bed slope.

In general for a problem where a hydraulic jump is triggered by a bed slope discontinuity, the position of the jump is determined by the bed slope discontinuity. In this case the position of the jump is determined by the bed slope discontinuity.

of the jump can be restricted to the locality of the jump. The exact functional form still made it difficult to control the solution away from the jump. The examples in this thesis still use exponential functions, however the exact form allows more systematic control over the solution. The form can be written as

$$\hat{h}(x) = \exp(-p(x - x^*)) \sum_{i=0}^M k_i \left(\frac{x - x^*}{x^{**} - x^*} \right)^i + \phi(x). \quad (6.6)$$

The parameters $k_0, k_1, \dots, k_M$ are used satisfy the constraints at the jump. Calculating these values only involves solving a small linear system. The positive parameter p influences the rate at which the high derivatives and curvatures drop to zero. The parameter x^{**} is used to control the width of the jump.

Problem	B/m		L/m	n	$Q/(\text{m}^3\text{s}^{-1})$	h_{in}/m	h_{out}/m
1	10	0	150	0.03	20		0.800054
2	10	2	300	0.03	20		0.710000
3	10	2	200	0.03	20	0.400013	
4	10	2	200	0.03	20		
5	0	10	100	0.03	20	0.700000	1.900000
6	10	0	150	0.03	20		1.700225
7	5	5	200	0.03	20	0.750000	
8	5	5	650	0.03	20	0.850000	

Table 6.1: Information for test problems 1-8

where it only remains to specify the function $\hat{h}$, which is also the solution of the problem.

Problem 1 (subcritical flow) In this case we have

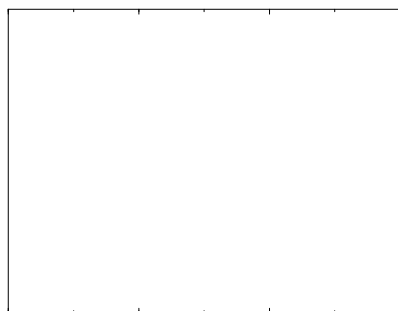
$$\hat{h}(x) = 0.8 + 0.25 \exp \left(-33.75 \left(\frac{x}{150} - \frac{1}{2} \right)^2 \right).$$

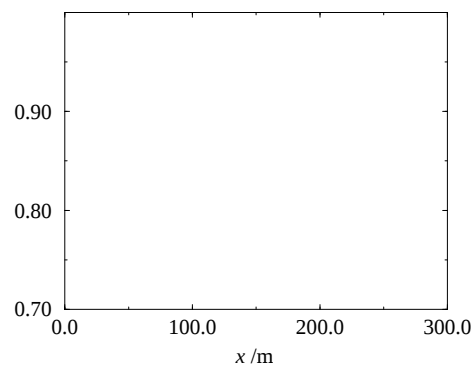
Figure 6.1(a) shows $\hat{h}$, Figure 6.1(b) shows the corresponding bed slope and Figure 6.1(c) shows the bed level and the free surface elevation. The channel flattens as we approach the mid-point of the reach, having the least gradient at this point. The channel then steepens again, returning to the initial gradient. The solution of this problem corresponds to entirely subcritical flow. The depth rises to a maximum at the center of the reach and approaches the critical depth at both ends.

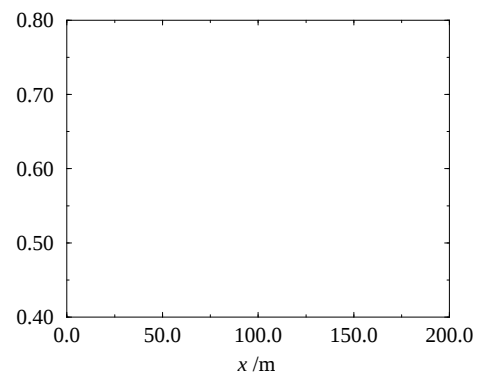
Problem 2 (subcritical flow) In this case we have

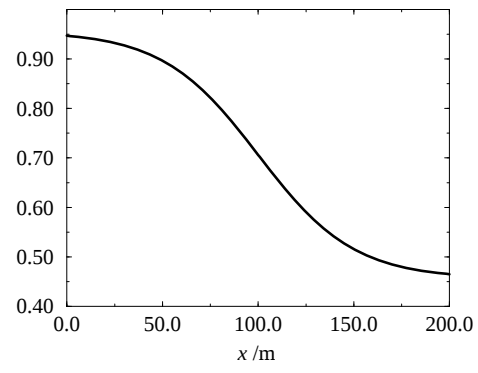
$$\hat{h}(x) = 0.71 + 0.25 \sin^2 \left(\frac{3\pi x}{300} \right)$$

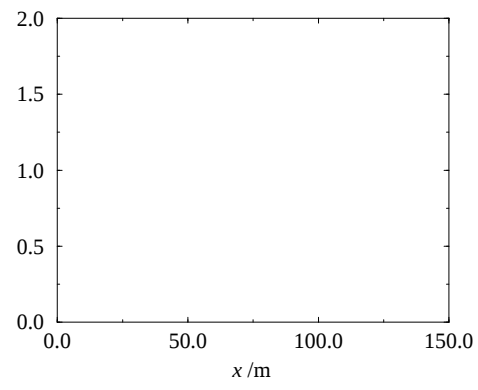
and the problem is illustrated by Figure 6.2. The gradient of the channel flattens and then steepens again three times. As in the previous case the solution corresponds to entirely subcritical flow. The depth has local maximim im

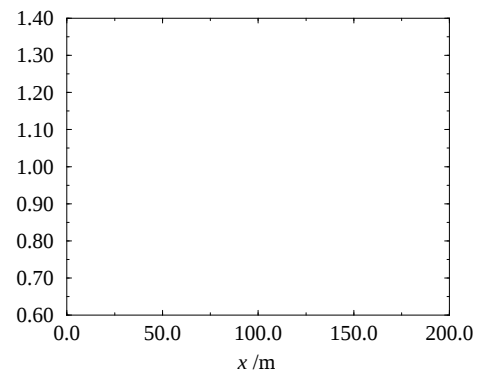












Chapter 7

Numerical Experiments

In Chapter 5 we presented theory for a family of numerical methods for computing steady solutions

wind scheme which is not a monotone scheme, but nevertheless is still well behaved. We use the strategy described in section 5.5 to compute a-priori bounds for the solution, and allowable time steps for the time stepping iteration.

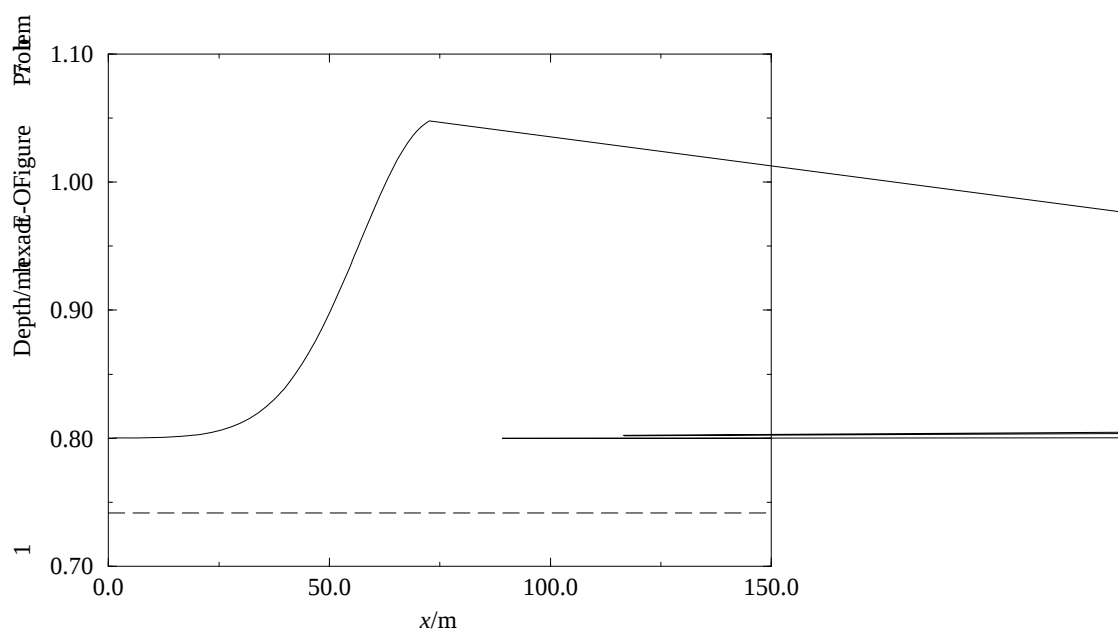
Problem 1 Consider the test problem 1. It is given that the flow at inflow is subcritical and that the depth at outflow is $\hat{h}(L) \approx 0.80\text{m}$. Following the strategy described in section 5.5, we take $\gamma_0 = h_c \approx 0.74\text{m}$ and γ_1 to be the depth specified at outflow. This yields the bounds on the solution

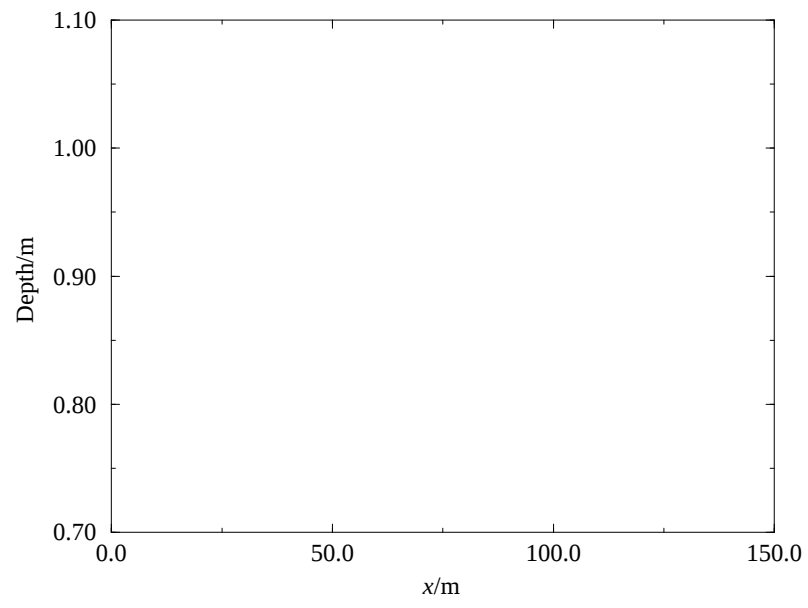
$$\underline{h} \leq h(x) \leq \overline{h}, \quad 0 \leq x \leq L,$$

where $\underline{h} = 0.74\text{m}$ and $\overline{h} = 2.64\text{m}$. Comparing the bounds and the actual solution (shown in Figure 7.1) we find that the upper bound is not at all tight. The actual solution does not rise above 1.05m . In general the bounds given by the theory cannot be expected to be tight, for they depend solely on the extreme values of the bed-slope. In the current example the upper bound must therefore take into account the worst case scenario for which the bed slope is at its minimum value for a great enough distance for the solution to asymptote to the corresponding normal depth. In reality though, the bed slope (see Figure 6.1(b)) is only close to its minimum value for a small fraction of the reach.

Figure 7.1 shows results for the Engquist-Osher scheme for problem 1 with $\Delta x = 10\text{m}$. The Godunov and the first-order upwind schemes give identical results because the difference equations reduce to an identical form for purely subcritical or purely supercritical solutions. The numerical solution gives a reasonable representation of the solution. The numerical solution is slightly skew, whereas the exact solution is symmetric about the middle of the reach. The numerical solution also fails to reach the correct maximum depth by a few centimeters.

The initial guess for the time-stepping iteration is $h = 0.74\text{m}$.





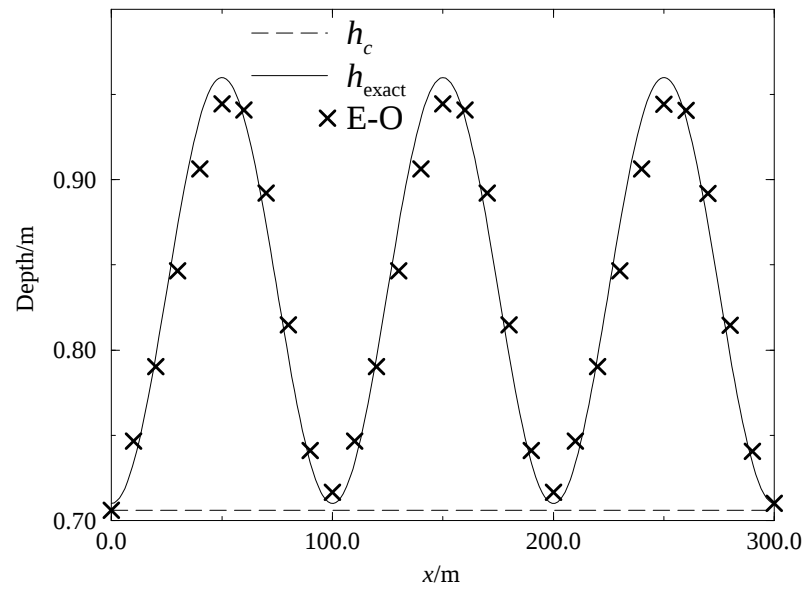


Figure 7.3: Prob

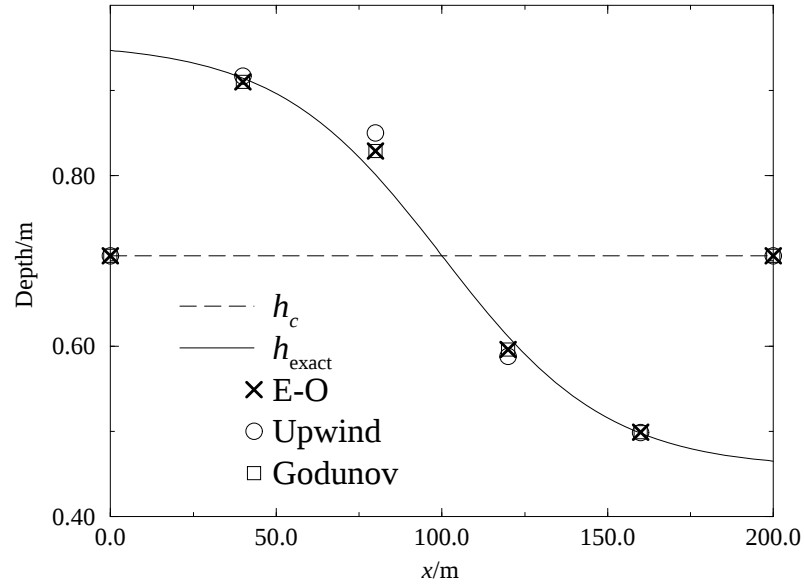


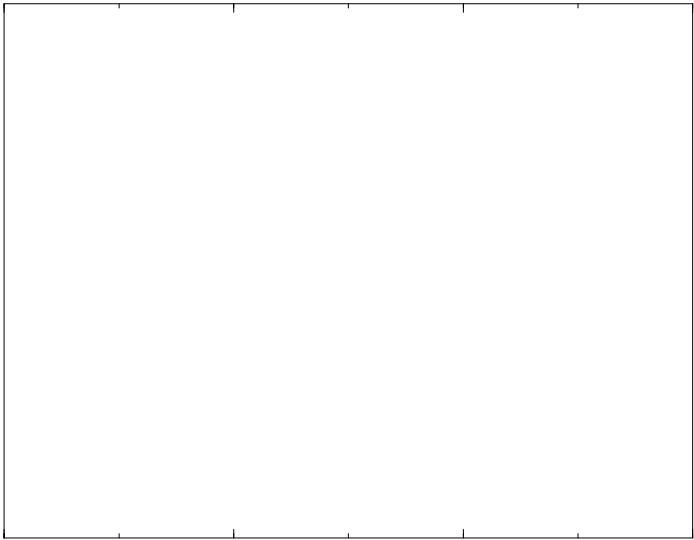
Figure 7.5: Comparison of E-O, Godunov and upwind schemes for problem 4 ($\Delta x = 40\text{m}$).

changes smoothly from subcritical to supercritical at the midpoint of the channel. The methods give a good representation of the solution even with so few grid points. The upwind scheme is found to be less accurate than for the Engquist-Osher or Godunov schemes at the grid points on either side of the transition. The smooth transition is a steady expansion wave for th

7.2 Comparison with Roe's Approximate Riemann Solver

We now compare the accuracy of the schemes in the previous section against that of Roe's approximate Riemann solver[55] which is described in sections 3.3 and 3.8. The latter scheme is a time accurate solver of the time dependent Saint-Venant system and we model the transient flow until steady state is attained. The scheme is a natural generalisation of the first-order upwind scheme to systems of equations and is designed specifically for the computation of discontinuous flows. In section 3.7 we discussed two different methods of discretising a source term. Here we apply both the pointwise discretisation and the upwind

where the function $\hat{h}$ is the exact solution. In order to allow a fair comparison of the two distinct approaches, the end points of the reach are not included in the error measures. This is because the solution is not in general approximated at these points for the scalar approach, since we fix $h_0 = \gamma_0$ and $h_N = \gamma_1$



(a)

be written as

$$\frac{\boldsymbol{w}_j^{n+1} - \boldsymbol{w}_j^n}{\Delta} + \left(\tilde{J}_{j+\frac{1}{2}}^- \right)^n \frac{(\boldsymbol{w}_{j+1}^n - \boldsymbol{w}_j^n)}{\Delta x} + \left(\tilde{J}_{j-\frac{1}{2}}^+ \right)^n \frac{(\boldsymbol{w}_j^n - \boldsymbol{w}_{j-1}^n)}{\Delta x} = \boldsymbol{D}_j^n,$$

where $\boldsymbol{D}_j^n = \boldsymbol{D}(x_j, \boldsymbol{w}_j)$. At steady state this reduces to

$$\tilde{J}_{j+\frac{1}{2}}^- \frac{(\boldsymbol{w}_{j+1} - \boldsymbol{w}_j)}{\Delta x} + \tilde{J}_{j-\frac{1}{2}}^+ (\boldsymbol{w}_j - \boldsymbol{w}_j)$$

The coefficients of $(A_{j+1} - A_j)$ and $(A_j - A_{j-1})$ must be positive, so this precludes the wetted area from having extrema in a subcritical region of flow. This is clearly nonsense, so we conclude that the difference equations are not in general consistent at steady state with a constant discharge solution.

and

$$I-\Gamma\left(\tilde{J}_{j+\frac{1}{2}}\right)=\frac{1}{\tilde{\lambda}_{2,j+\frac{1}{2}}-\tilde{\lambda}_{1,j+\frac{1}{2}}}\begin{pmatrix} & \tilde{\lambda}_{2,j+\frac{1}{2}} & \\ \tilde{\lambda}_{2,j+\frac{1}{2}} & & -1 \end{pmatrix}$$

accuracy observed for problem 1, since the trapezium rule is a second order accurate discretisation.

Figure 7.14 compares the L_1 accuracy of the three schemes. em5

at x

7.3 Higher Order Accuracy

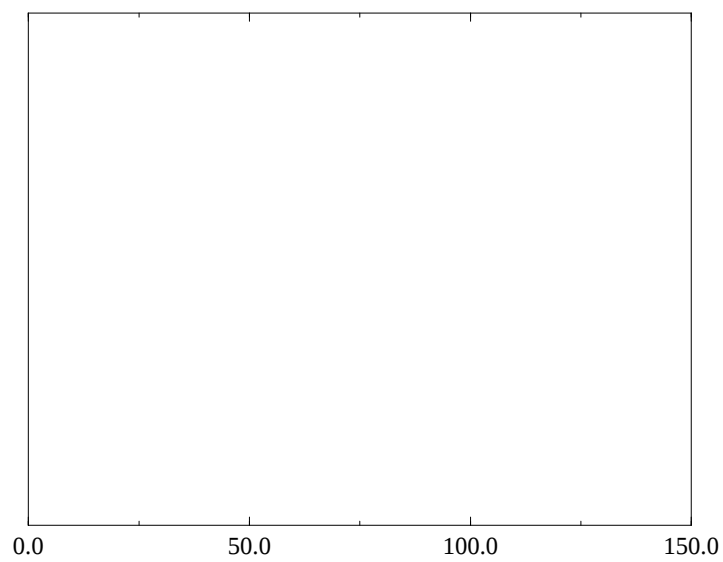
This section generalises the “scalar approach” in order to obtain higher order approximations to the steady state solutions. In the previous section we obtained second order accuracy using Roe’s approximate Riemann solver combined with a particular method of upwinding the source term. We apply the direct analogue of this scheme to the scalar equation

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x} f(h) = -D, \quad (7.4)$$

where $f(h) = -F(h)$. This gives the first-order upwind scheme with upwinded source term. Next a generalisation of the Engquist-Osher scheme is considered which again comes down to upwinding the source term. These methods achieve high order accuracy.

and $s_{j+\frac{1}{2}}$ is given by (3.14). We can also write

$$\begin{aligned} \mathcal{T}_j^{\text{UPW-2}} h &= \frac{g_{j+\frac{1}{2}}^{\text{FOU}} - g_{j-\frac{1}{2}}^{\text{FOU}}}{\Delta x} \\ &+ \Gamma \left(s_{j-\frac{1}{2}} \right) \sim \end{aligned}$$



channel as a function of p , for three different grid spacings ($N = 20$, $N = 80$ and $N = 320$). The behaviour is very similar at each of the four points. For a particular grid spacing there are two regions of constant error, separated b

7.3.2 High Order TVD Sch

of schemes with numerical flux function written as

$$g_{j+\frac{1}{2}} = \frac{1}{2} \left(f(h_j) + f(h_{j+1}) + \phi_{j+\frac{1}{2}} \right) ,$$

for different functions $\phi_{j+\frac{1}{2}}$. In general the resulting scheme is a five point scheme and this adds difficulties at the boundaries. In the case of the three point schemes, the values of the boundary nodes h_0^n and h_N^n are fixed regardless of whether these represent physical boundary conditions. We treat the boundaries for five-point schemes in the same unsophisticated manner, by now fixing the values of h_0^n , h_{-1}^n and h_N^n , h_{N+1}^n

where the minmod function is given by

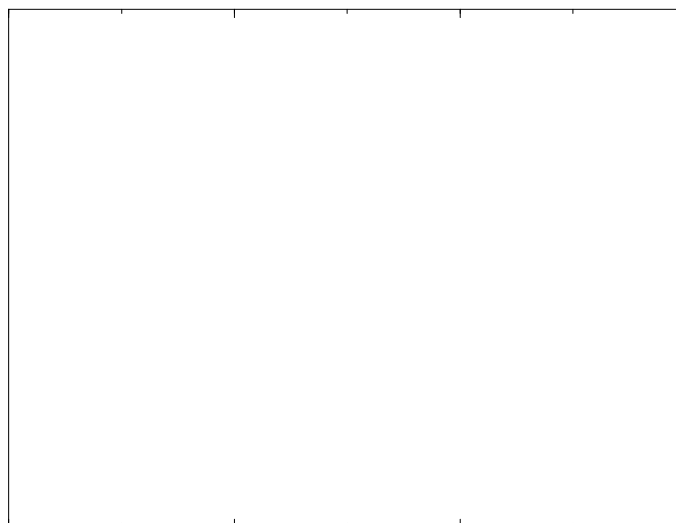
$$\text{minmod}(x,y) = \text{sgn}(x) \max\left\{0, \min\left\{|x|, y \cdot \text{sgn}(x)\right\}\right\}.$$

Other forms of the function η are given in [72].

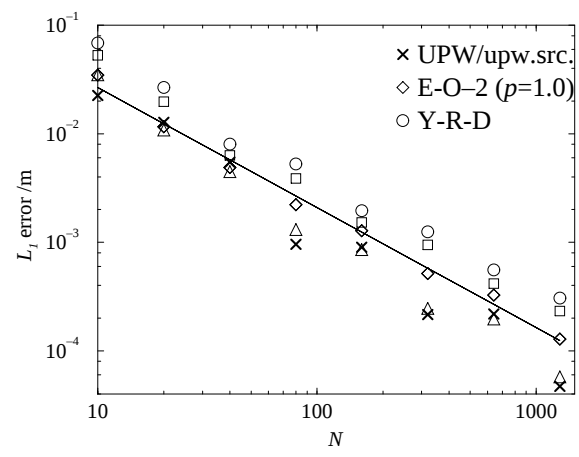
Yee-Roe-Davis Symmetric Scheme This scheme is a generalisation by Yee ([71], [70]) of the schemes of Roe[56] and Davis[10], and is given by

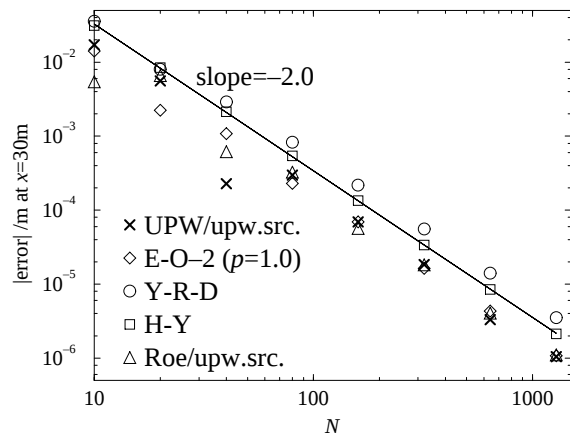
$$\phi_{j+\frac{1}{2}} = -\left|s_{j+\frac{1}{2}}\right| \left((h_{j+1} - h_j) - \psi_{j+\frac{1}{2}} \right),$$

where $s_{j+\frac{1}{2}}$ is given by $(\text{DT}\Gamma\Gamma\text{Tf}[\Gamma\Gamma\text{TD}[\Gamma\text{T}3.14\text{t}\Gamma\Gamma\text{TD}[\text{c}[\text{disa}\Gamma\Gamma\text{fM}\mathfrak{m}\Gamma\Gamma\Gamma\text{TD}\Gamma\Gamma\text{TD}[\text{yTj}[\Gamma\text{tff}(\text{lrj}_h$

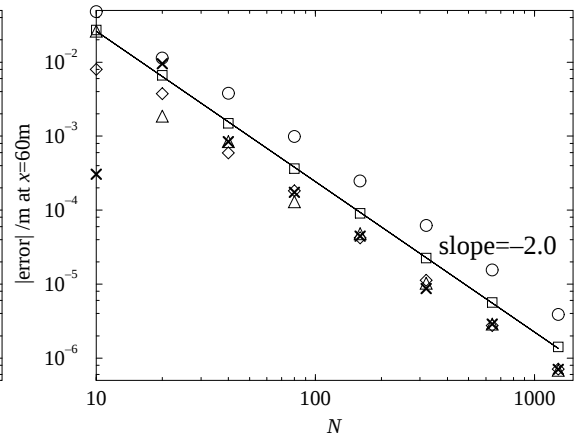


convergence at the info

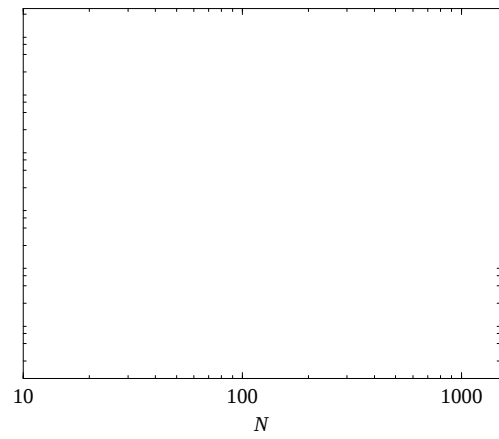




(a)



(b)



Chapter 8

Computation Efficiency

This chapter explores different meth

N	O timum			Monotonicity		
	Δt	No. Ste s	CPU time/s	Δt	No. Ste s	CPU time/s
10	0.084	48	0.108×10^{-1}	0.046	95	0.154×10^{-1}
20	0.058	56	0.151×10^{-1}	0.033	107	0.195×10^{-1}
40	0.035	96	0.511×10^{-1}	0.022	158	0.868×10^{-1}
80	0.015	104	0.116	0.011	181	0.311
160	0.0073	441	0.931	0.0059	548	0.115×10^1
320	0.0035	864	0.367×10^1	0.003	1010	0.408×10^1
640	0.0016	1911	0.161×10^2	0.0015	1049	0.190×10^2
1280	0.0008	3840	0.639×10^2	0.00076	4043	0.719×10^2

Table 8.1: Time st

$$\sqrt{\frac{1}{N+1}\sum_{j=0}^N\left(Q_j^{n+1}-Q_j^n\right)}$$

N	U wind-			E-O- ($p = 1$)		
	Δt	No. Ste s	CPU time/s	Δt	No. Ste s	CPU time/s
10	0.083	47	0.995×10^{-2}	0.078	51	0.159×10^{-1}
50	0.067	59	0.170×10^{-1}	0.074	50	0.156×10^{-1}
40	0.057	145	0.795×10^{-1}	0.036	100	0.710×10^{-1}
80	0.015	549	0.576	0.016	545	0.340
160	0.0074	470	0.116×10^1	0.0074	485	0.130×10^1
350	0.0035	976	0.450×10^1	0.0035	1035	0.535×10^1
640	0.0015	5495	0.535×10^2	0.0017	1904	0.189×10^2
1580	0.0008	3759	0.699×10^2	0.0008	3786	0.775×10^2

Table 8.3:8.3Gm

N	Yee-Roe-Davis			Otimum/Harten-Yee		
	Δt	No. Steps	CPU time/s	Δt	No. Steps	CPU time/s
10	0.086	119	0.35×10^{-1}	0.13	76	0.179×10^{-1}
20	0.046	175	0.748×10^{-1}	0.045	387	0.175
40	0.013	390	0.333	0.0103	1459	0.138×10^1
80	0.0063	664	0.18×10^1	Fails to converge for all Δt		
160	0.0031	1357	0.479×10^1	Fails to converge for all Δt		
320	0.0015	1967	0.145×10^2	Fails to converge for all Δt		
640	0.00075	5415	0.857×10^2	Fails to converge for all Δt		
1280	0.00036	847	0.53×10^3	Fails to converge for all Δt		

Table 8.4: Time stepping for the high order TVD schemes for problem 6 using optimal time steps

8.2 Newton's Method

The nonlinear system of difference equations can be written in vector notation as

$$\mathcal{T}(\mathbf{h}) = 0, \quad (8.3)$$

where $\mathcal{T}(\mathbf{h}) = (\mathcal{T}_1 h, \mathcal{T}_2 h, \dots, \mathcal{T}_{N-1} h)^T$, $\mathbf{h} = (h_1, h_2, \dots, h_{N-1})^T$ and $h_0 = h_{-1} = \gamma_0$, $h_N = h_{N+1} = \gamma_1$. The time stepping iteration is essentially a Picard iteration applied to this system. Such methods only give a linear convergence rate, i.e. the residual is inversely proportional to number of iterations. Newton's method however is well known to give a quadratic convergence rate, i.e. the residual is inversely proportional to the number of iterations squared. The drawback of Newton's method is that in general, global convergence is not obtained, i.e. convergence will not occur for all initial guesses of the solution. The theory of Newton's method and other related methods can be found in [46].

Applying Newton's method to the system of difference equations yields the following algorithm:

$$\mathbf{h}^{n+1} = \mathbf{h}^n + s^n \mathbf{d}^n, \quad (8.4)$$

where $\boldsymbol{d}^n = (d_1^n, d_2^n, \dots, d_{N-1}^n)^T$ solves the linear system

where

$$\begin{aligned} p_j &= -\frac{g_v(h_j, h_{j-1})}{\Delta x}, \\ q_j &= \frac{g_v(h_{j+1}, h_j) - g_u(h_j, h_{j-1})}{\Delta x} + D_h(x_j, h_j), \\ r_j &= \frac{g_u(h_{j+1}, h_j)}{\Delta x}. \end{aligned}$$

The Jacobian does not strictly exist in the case of Godunov and the first-order upwind schemes because of the switching, however on the curves where the function g is not differentiable, either the partial derivatives from the left or right can be used. The Jacobian exists at all points for the Engquist-Osher and Lax-Friedrichs schemes since the numerical flux functions are differentiable in these cases. For monotone schemes (i.e. under the conditions of Theorem 9 with $\alpha \leq \mathbf{h} \leq$) we have that g

N	E-O		E-O-
	No. iter	CPU t	

Hence we have

$$|q_j| = |p_{j+1}| + |r_{j-1}| + D_h(x_j, h_j) > |p_{j+1}| + |r_{j-1}|.$$

Hence the transpose of the Jacobian is strictly diagonally dominant and the Jacobian itself is non-singular, and as for the first order monotone schemes the Jacobian is an M-matrix.

Newton's method applied to the second order Engquist-Osher scheme is in practice found to be well-behaved and Table 8.6 shows the performance for problem 1 with $p = 1$. At best for this example the method converges in 10-15 iterations.

having a band-width of five, as would be the case if one attempted to use the full Jacobian. We expect the fact that only an approximate Jacobian is used to reduce the performance of the Newton's method. Applying the method to the Yee-Roe-Davis scheme and the Harten-Yee scheme we find that convergence is obtained if the number of grid-points is very small, although the method is more expensive than for the previous schemes such as the Engquist-Osher scheme. For even a moderate number of grid points and even for solutions without transitions, the method fails to converge for the Yee-Roe-Davis scheme. The Harten-Yee scheme converges for a higher number of grid-points, however for problems with hydraulic jumps fails for significantly smaller N than for example the Engquist-Osher scheme. The robustness of the method can in some cases be improved by use of a grid refining approach, but the method is not really suited to solving these schemes.

8.3 The Implicit Time Stepping Iteration

We now consider a generalisation of the time stepping algorithm. The particular implementation of the algorithm relates it closely to the Newton algorithm described in the previous section. The efficiency of the time stepping method is related to the

where $\boldsymbol{d}^{n+1,k} = (d_1^{n+1,k}, d_2^{n+1,k}, \dots, d_{N-1}^{n+1,k})^T$ solves the linear system

$$\mathcal{R}'(\boldsymbol{h}^{n+1,k})\boldsymbol{d}^{n+1,k} = -\mathcal{R}(\boldsymbol{h}^{n+1,k}), \quad (8.10)$$

$$\mathcal{R}' = I$$

expect the implicit algorithm to have no restriction on the time step. This is found to be the case in practice. A measure of the performance for this method is given by the number of Jacobian inversions required for convergence to occur. The number of inversions required decreases as the time step increases and approaches the number of inversions required for the Newton algorithm as Δt becomes very large. Even for large time steps where the num

where $\mathbf{d}^n = (d_1^n, d_2^n, \dots, d_{N-1}^n)^T$ solves the linear system

$$\mathcal{R}'(\mathbf{h}^n)\mathbf{d}^n = -\mathcal{R}(\mathbf{h}^n), \quad (8.12)$$

and s^n is taken as in Newton's method. Again for $\theta = 1$ this method approaches the Newton algorithm as Δt grows large. The method is known as a linearised implicit scheme (see section 3.6) since it can also be derived using Taylor's expansions to linearise the implicit part of the operator. The linearised method is more efficient to implement than the method allowing a variable number of inner iterations, and each iteration requires a very similar expenditure to one iteration of the Newton algorithm. Table 8.7 shows the performance of the linearised method for the Engquist-Osher

N	E-O			E-O- ($p = 1$)		
	Δt	No. iter.	CPU time/s	Δt	No. iter.	CPU time/s
10	35.34	9	0.185×10^{-2}	3.0	13	0.475×10^{-2}
100	13.4	10	0.470×10^{-2}	1.0	16	0.115×10^{-1}
40	5.4	31	0.599×10^{-1}	1.0	36	0.308×10^{-1}
80	5.0	44	0.759×10^{-1}	1.0	56	0.106
160	13.5	60	0.198	1.0	97	0.369
300	16.0	85	0.646	1.0	180	0.151×10^1
640	39.0	141	0.191×10^1	1.0	138	0.335×10^1
1080	84.0	566	0.749×10^1	36.0	560	0.116×10^2

Table 8.7: Linearised implicit algorithm for the E-O and E-O-2 ($p = 1$) schemes for problem 6

scheme and its second order modification ($p = 1$) for problem 6. In each case performance is given for a time step which gives close to the optimum convergence rate. By comparing Tables 8.7 and 8.5 it can be seen that the linearised method and Newton's method are indeed very similar in performance. Of course for the linearised method the performance is dependent on the choice of time step and there is no way to predict the optimal value in advance. However, it is found that good performance is obtained over a much wider range of time steps than for the explicit time stepping method. There also appears to be no limit on the size of the time step

in many cases. In the case of the high order TVD schemes, the linearised method

N	Yee-Roe-Davis			Harten-Yee		
	Δt	No. Ste s	CPU time/s	Δt	No. iter.	CPU time/s
10	4.0	36	0.137×10^{-1}	0.35	55	0.741×10^{-2}
20	1.0	174	0.108	0.51	37	0.538×10^{-1}
40	0.1	56	0.655×10^{-1}	0.45	45	0.547×10^{-1}
80	0.06	68	0.159	0.19	84	0.518
160	0.04968					

8.4 Conclusions

In this chapter the effectiveness of four different meth

Chapter 9

Non-Prismatic Channels

9.1 Scalar Schemes

most natural. Other choices may be more appropriate for reasons of computational efficiency

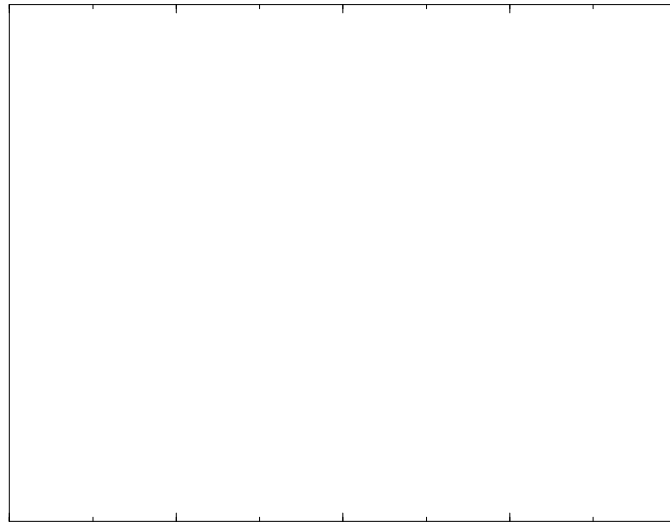
where the flow is subcritical, i.e. $h_{j-1}, h_j > h_c(x_{j+q-1})$ and $h_j, h_{j+1} > h_c(x_{j+q})$, the three schemes all reduce to

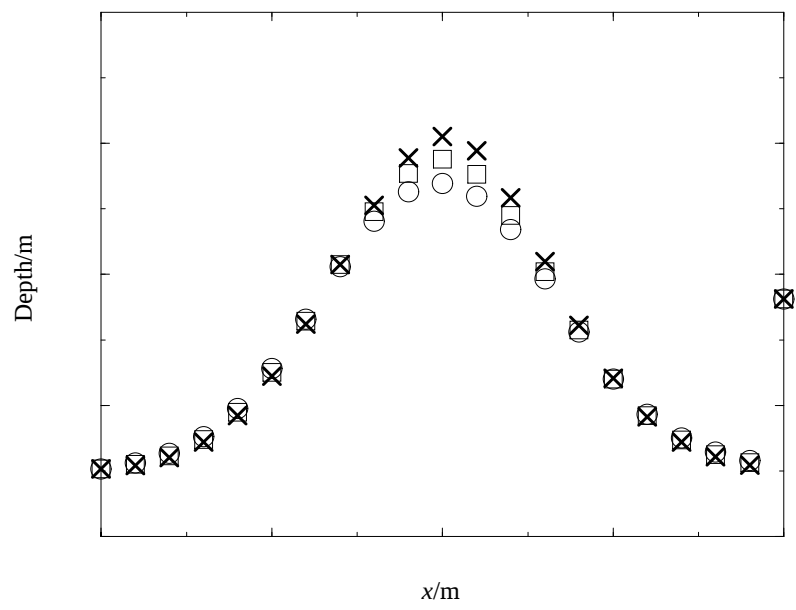
$$\frac{f(x_{j+q}, h_{j+1}) - f(x_{j+q-1}, h_j)}{\Delta x} + D(x_j, h_j) = 0,$$

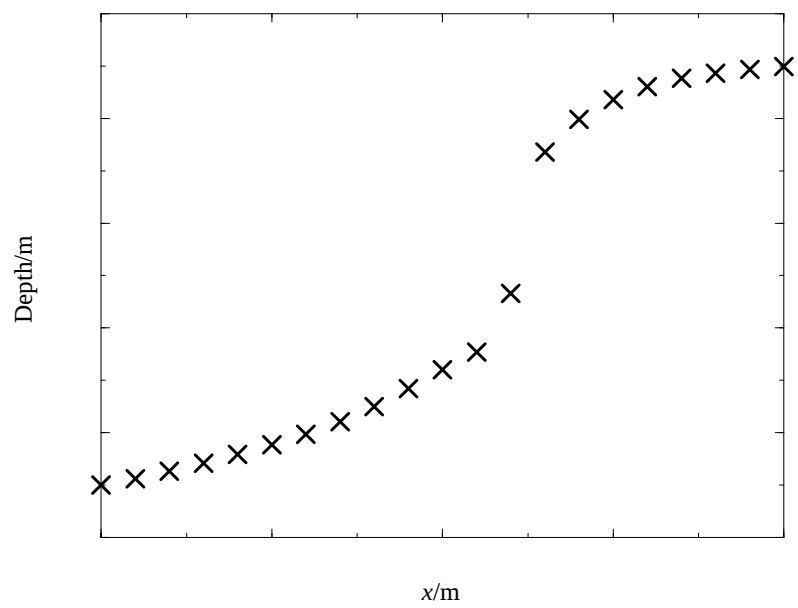
with truncation error

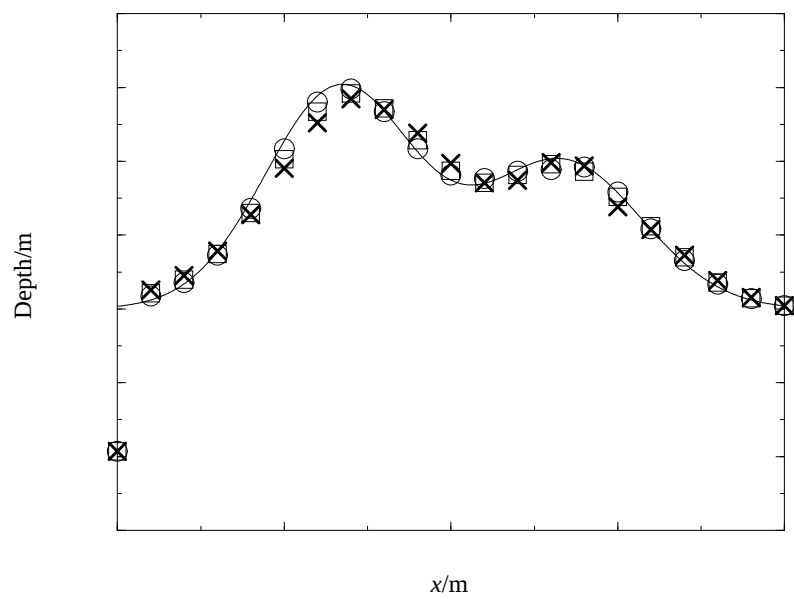
$$\begin{aligned} \text{T.E.} &= \frac{\Delta x}{2} \left(h'' f_h + (h')^2 f_{hh} + 2qh' f_{hx} + (2q-1)f_{xx} \right) + O(\Delta x^2) \\ &= \frac{\Delta x}{2} \frac{d}{dx} (-D + 2(q-1)f_x) + O(\Delta x^2). \end{aligned} \tag{9.5}$$

In a region of the solution where the flo



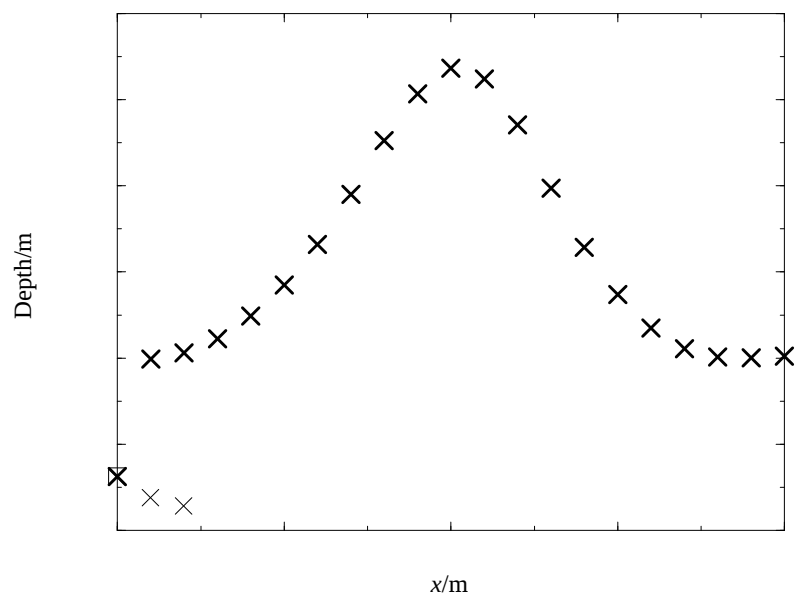


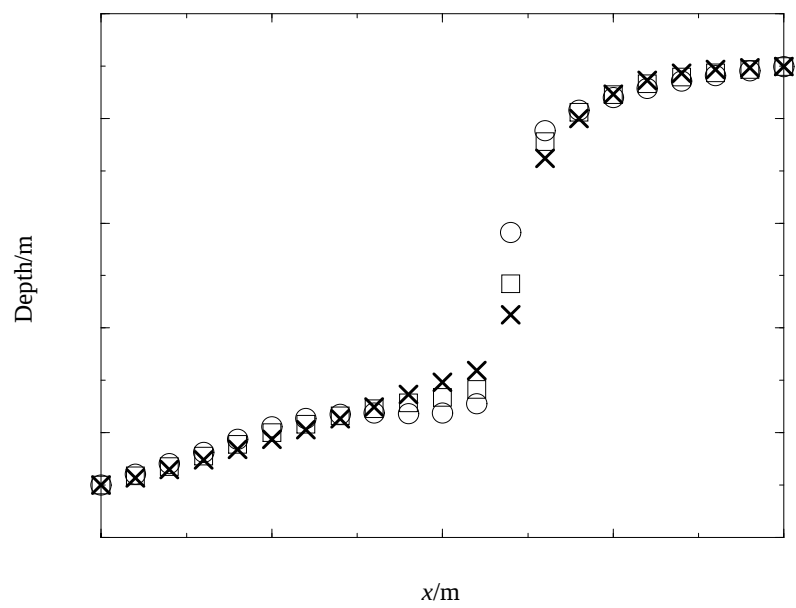




$$\begin{aligned}\chi_j^0 &= 1 - \chi\left(\frac{pf_h(x_j, h_j)}{\sqrt{\Delta x}}\right) - \chi\left(\frac{-pf_h(x_j, h_j)}{\sqrt{\Delta x}}\right) = 1 - \chi_{j+1}^- - \chi_{j-1}^+, \\ \chi_j^+ &= \chi\left(\frac{-pf_h(x_{j+1}, h_{j+1})}{\sqrt{\Delta x}}\right),\end{aligned}$$

and the function χ is given as before. For a prismatic channel the scheme gives second order accuracy. To see whether this remains the case for a non-prismatic channel we again consider the





extremely problematic, since even the explicit time stepping iteration fails to converge more often than not.

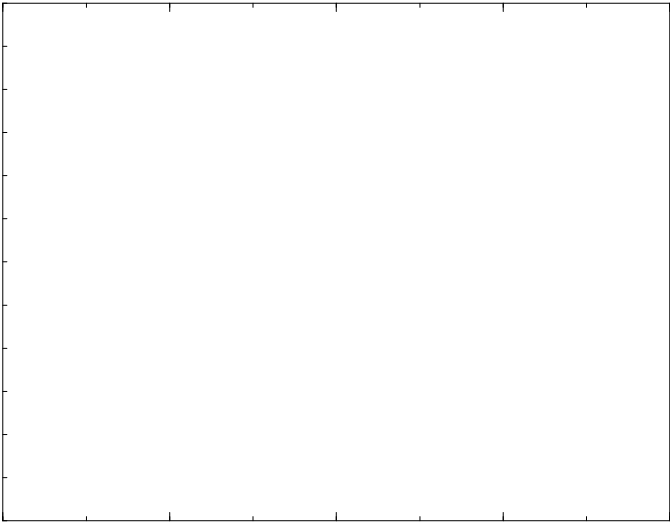
9.2 Roe's Approximate Riemann Solver

We can use the same principle as in the previous section to extend Roe's approximate Riemann solver to the non-prismatic case. The generalised n

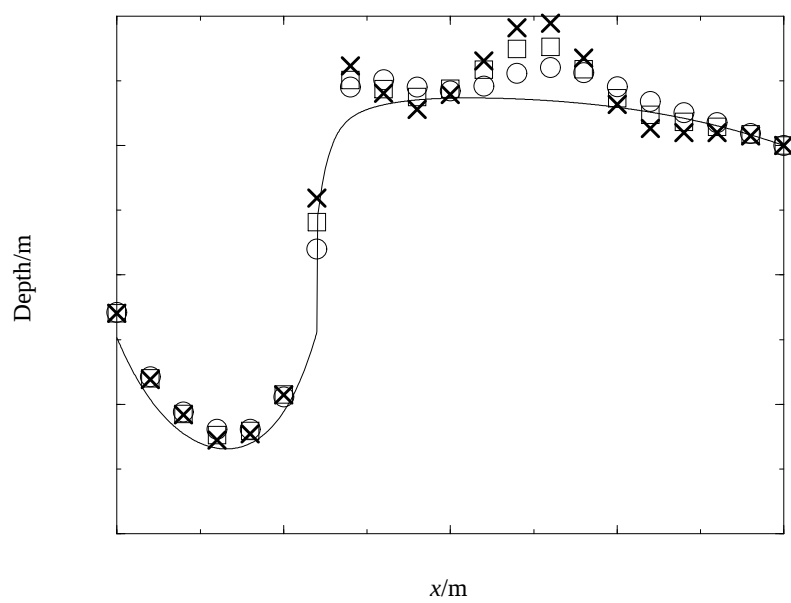
we use the modified formula

$$\left(\tilde{c}_{j+\frac{1}{2}}\right)^2 = \begin{cases} g\left(\frac{I_1(x_{j+q}, A_{j+1}) - I_1(x_{j+q}, A_j)}{A_{j+1} - A_j}\right) & A_j \neq A_{j+1} \\ \frac{gA_j}{T(x_{j+q}, A_j)} & A_j = A_{j+1}, \end{cases}$$

where now the functions T and I_1 give there respective quantities as a function of cross-section and wetted area.



(a) x/m



(a)

channel, because of the additional term (9.7) which has not been decomposed onto the eigenvectors of the Roe matrix. The approach of Priestley[53], which absorbs the additional term into the source term and upwinds the modified source term, does however

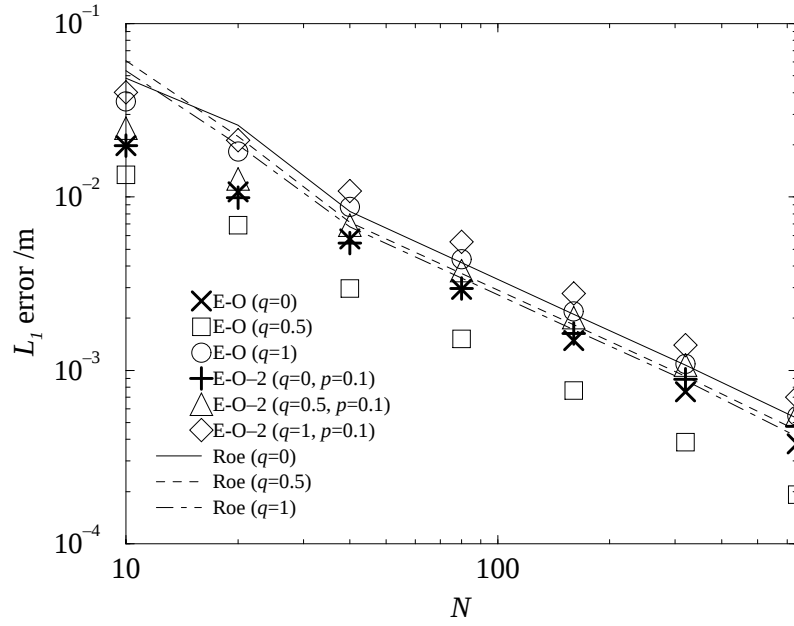


Figure 9.16: L_1 errors for problem 11.

may therefore only be of any benefit for solutions of predominantly one type of flow. The generalisation of the upwind-2 scheme was found to be extremely problematic due non-convergence of the time-stepping iteration, whereas the E-O and E-O-2 schemes were, as for the prismatic case, very amenable to Newton's method. Roe's scheme was also generalised to the non-prismatic case. The scheme has difficulties capturing smooth transitions but this could be cured by addition of an entropy fix. Roe's method was in general found to less accurate than the scalar schemes, due to the large amount of deviation from the expected constant discharge. This cannot be remedied by upwinding the source term without a special treatment of the new term arising from the variation of the cross-section of the channel. This last subject requires further work.

tions of a particular scalar conservation law. Scalar shock capturing schemes could thus be used to compute discontinuous solutions of the steady flow problem. A family of such schemes, which are particularly rich in theoretical results, are the monotone schemes. In Chapter 5 we demonstrated that this richness carries over to the computation of solutions of the steady flow problem. Under the same conditions as the theory in Chapter 4 we demonstrated that the numerical schemes define a solution which converges to the unique physical solution of the steady flow problem (as the grid-spacing vanishes).

adyTriihweecfiTDfiwfieamInCalrthTDfinTfiTDfiemTjfiTDTDfirilefinTfiaTjfiTsTcCrmfiaptTjfiwpaom-TadwTDfiom-

However in general the discharge for Roe's schemes was found to be far from constant at steady state. This was remedied by using an upwind discretisation of the source term. Moreover for a particular form of source term averaging it was found that the scheme gave second order accuracy at steady state. This was explained by showing that, at steady state, the scheme effectively reduces to the trapezium rule. Upwinding of the source term was also used to obtain second order accurate three-point scalar schemes. The resulting generalisations of the Engquist-Osher and first-order upwind schemes were found to give comparable accuracy to Roe's scheme with an upwinded source term. Another more traditional approach to obtaining second order accuracy is through the use of nonlinear limiter functions, leading to five-point TVD schemes. Examples of such schemes were found to be significantly less accurate

than the scheme. This paper is a review of the work of Roe, Osher, and Engquist on this subject.

between grids.

The schemes discussed thus far are only applicable in the case of a prismatic channel. In Chapter 9 we extended the scalar schemes to the non-prismatic case by allowing the numerical flux functions to depend on the distance along the channel. We investigated different ways of staggering the evaluation points, and for the Engquist-Osher scheme we found that, in the majority of test cases, the most accurate solution was obtained by evaluating the numerical flux at the cell interfaces. The version of the Engquist-Osher scheme with upwinded source term was no longer found to give second order accuracy in all regions of the solution. A particular staggering of the evaluation points may give second order accuracy in one flow regime, but the scheme remains only first order accurate in regions of the opposite type of flow. We concluded that upwinding the source term was only beneficial for solutions of predominantly one type of flow. Further work is required to develop a conservative scheme that is second-order accurate in both flow regimes for a non-prismatic channel.

The same idea as above was used to develop a second order accurate scheme for the

boundaries, however some applications may require an accurate approximation of the solution at the boundaries. Further work is therefore necessary to investigate ways of achieving this.

One approach is to solve the system of difference equations as usual and then, if necessary, extrapolate the solution onto the boundaries. For an upwind scheme, the appropriate one-sided form of the scheme may be used to perform the extrapolation, maintaining the accuracy of the scheme at the boundaries. A difficulty with this technique is to decide when extrapolation is required, for example we must differentiate between a hydraulic jump close to the boundary (for which extrapolation is not appropriate) and a boundary layer, in which the boundary condition is not applicable.

solution without referring to the transient flow. Work is required to investigate how common the existence of multiple solutions is and the behaviour of the scalar schemes in such cases.

Biography

- [1] L Abrahamsson and S Osher. Monotone difference schemes for singular perturbation problems. *SIAM J. Numer. Anal.*, 19(5):979–992, October 1982.
- [2] M J Baines, A Maffio, and A Di Filippo. Unsteady 1-D flows with steep waves in plant channels: The use of Roes’s upwind TVD difference scheme. *Advances in Water Resources*, 15:89–94, 1992.
- [3] P A Burton and P K Sweby. A dynamical approach study of some explicit and implicit TVD schemes and their convergence to steady-state solutions. Numerical Analysis Report 5/95, University of Reading, Department of Mathematics, PO Box 220, Reading RG6 6AF, UK, 1995.
- [4] V T Chow. *Open Channel Hydraulics*. McGraw-Hill Publishing Company, New York, 1959.
- [5] R Courant and K O Friedrichs. *Supersonic Flow and Shock Waves*. Springer-Verlag, New York, 1948.
- [6] R Courant, K O Friedrichs, and H Lewy. Über die partiellen differenzengleichungen der mathematischen physik. *Math. Ann.*, 100:32–74, 1928.
- [7] R Courant, K O Friedrichs, and H Lewy. On the partial differential equations of mathematical physics.

- [10] S F Davis. TVD finite difference schemes and artificial viscosity. ICASE Report 84-20, 1984.
- [11] B de Saint-Venant. Théorie du mouvement non-permanent des eaux avec application aux crues des rivières et à l'introduction des marées dans leur lit. *Acad. Sci. Comptes rendus*, 73:148–154,237–240, 1871.
- [12] .iDDAetj[Standard violation document:- definition and guidelinesoriT tatr dg,5TD5Tc5ci

nSChannelnS

[21] A Harten, B Engquist, S Osh

Notes in Applied Mathematics, volume 942 of *Springer Lecture Notes in Math.*, pages 150–169. Springer Verlag, New York, 1982.

- [33] J Lorenz. Numerical solution of singular perturbation problem with turning points. In H W Knoblock and K Schmitt, editors, *Equadiff 82*, volume 1017 of *Lecture Notes in Applied Mathematics*, pages 432–439. Springer Verlag, New York, 1983.
- [34] J Lorenz. Analysis of difference schemes for a stationary shock problem. *SIAM J. Numer. Anal.*, 21(6):1038–1053, December 1984.
- [35] J Lorenz. Convergence of upwind schemes for a stationary shock. *Mathematics of Computation*, 46(173):45–57, January 1986.
- [36] J Lorenz and R Sanders. On the rate of convergence of viscosity solutions and boundary value problems. *SIAM J. Math. Anal.*, 18(2):306–320, March 1987.
- [37] I MacDonald. Test problems with analytic solutions for steady open channel flow. Numerical Analysis Report 6/94, University of Reading, Department of Mathematics, PO Box 220, Reading RG6 6AF, UK, 1994.
- [38] I MacDonald, M J Baines, N K Nichols, and Samuels P G. Analytic benchmark solutions for open channel flows. To appear in the *Journal of Hydraulic Engineering*, ASCE.
- [39] 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, University of Reading, Department of Mathematics, PO Box 220, Reading RG6 6AF, UK, 1995.
- [40] 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, University of Reading, Department of Mathematics, PO Box 220, Reading RG6 6AF, UK, 1995.
- [41] E M Murman. Analysis of embedded shockwaves calculated by relaxation methods. *AIAA Journal*, 12:636, 1974.

- [42] E M Murman and Cole J D. Calculation of plane steady transonic flows. *AIAA Journal*, 9:114, 1971.
- [43] M Nagumo. Über die differentialgleichung $y'' = f(x, y, y')$. *Proc. Phys. Math. Soc. of Japan*, 19:861–866, 1937.
- [44] 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.
- [45] R E O'Malley. *Introduction to Singular Perturbations*. Academic Press, New York, 1974.
- [46] J M Ortega and W C Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.
- [47] S Osher. Nonlinear singular perturbation problems and one sided difference schemes. *SIAM J. Numer. Anal.*, 18(1):129–144, February 1981.
- [48] S Osher. Riemann solvers, the entropy condition and difference approximations. *SIAM J. Numer. Anal.*, 21:217–235, 1984.
- [49] S Osher and F Solomon. Upwind difference schemes for hyperbolic systems of conservation laws. *Mathematics of Computation*, 38(158):339–374, April 1982.
- [50] C E Pearson. On the differential equation of boundary layer type. *J. Maths. Phys.*, 47:134–154, 1968.
- [51] W H Press, S A Teukolsky, W T Vetterling, and B P Fanner. *Numerical Recipes in Fortran*. Cambridge University Press, Cambridge, 2nd edition, 1992.
- [52] R K Price and P G Samuels. A computational hydraulic model for rivers. *Proc. of the Inst. of Civ. Engrs.*, 69(2):87–96, 1980.

- [53] A Priestley. Roe type schemes for super-critical flow in rivers. Numerical Analysis Report 13/89, University of Reading, Department of Mathematics, PO Box 220, Reading RG6 6AF, UK, 1989.
- [54] A Priestley. A quasi-Riemann method for the solution of one-dimensional shallow water flow. *J. Comput. Phys.*, 106(1):139–146, 1993.
- [55] P L Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, 43:357–372, 1981.
- [56] P L Roe. Generalised formulation of TVD Lax-Wendroff schemes. ICASE Report 84-53, 1984.
- [57] 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*, Springer Lecture Notes in Mathematics. Springer Verlag, 1986.
- [58] P G Samuels. *Modelling River and Flood Plain Flow using the Finite Element Method*. PhD thesis, University of Reading, Department of Mathematics, 1985.
- [59] P G Samuels. Backwater lengths in rivers. *Proc. Instn Civ. Engrs.*, 87:571–582, 1989.
- [60] P G Samuels and C P Skeels. Stability limits for Preissmann’s scheme. *Journal of Hydraulic Engineering*, 115(1):1–11, 1989.

- [65] K R Stromberg. *An Introduction to Classical Real Analysis*. Wadsworth, Inc., Belmont, California, 1981.
- [66] P K Sweby. *Shock Capturing Schemes*. PhD thesis, University of Reading, Department of Mathematics, 1982.
- [67] P K Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM J. Numer. Anal.*, 21(5):995–1011, October 1984.
- [68] J P Vila. Simplified Godunov schemes for 2x2 systems of conservation laws. *SIAM J. Numer. Anal.*, 23(6):1173–1192, December 1986.
- [69] S Wolfram. *Mathematica, A System for Doing Mathematics by Computer*. Addison-Wesley, New York, 1988.

Here the operator $\mathcal{T}_j^{\text{E-O-2}}$ is given by

$$\begin{aligned}\mathcal{T}_j^{\text{E-O-2}}h &= \frac{g^{\text{E-O}}(h_{j+1}, h_j) - g^{\text{E-O}}(h_j, h_{j-1})}{\Delta x} \\ &\quad + \chi_j^- D(x_{j-1}, h_{j-1}) + \chi_j^0 D(x_j, h_j) + \chi_j^+ D(x_{j+1}, h_{j+1}),\end{aligned}$$

where

$$\begin{aligned}\chi_j^- &= \chi\left(\frac{pf'(h_{j-1})}{\sqrt{\Delta x}}\right), \\ \chi_j^0 &= 1 - \chi\left(\frac{pf'(h_j)}{\sqrt{\Delta x}}\right) - \chi\left(\frac{-pf'(h_j)}{\sqrt{\Delta x}}\right) = 1 - \chi_{j+1}^- - \chi_{j-1}^+, \\ \chi_j^+ &= \chi\left(\frac{-pf'(h_{j+1})}{\sqrt{\Delta x}}\right),\end{aligned}$$

$p \geq 0$ is a parameter and χ is the smooth increasing function

$$\chi(r) = \begin{cases} 0 & r < 0 \\ r^2 & 0 \leq r \leq \frac{1}{2} \\ \frac{1}{2} - (1-r)^2 & \frac{1}{2} \leq r \leq 1 \\ \frac{1}{2} & r > 1 \end{cases}$$

connecting the values 0 and $\frac{1}{2}$. It is convenient to write

$$\mathcal{T}_j^{\text{E-O-2}}h = \frac{\hat{g}_{j+\frac{1}{2}} - \hat{g}_{j-\frac{1}{2}}}{\Delta x}.$$

for $0 \leq j+1 \leq N$, $h_{j+1} \in [\alpha, \beta]$ and

$$\frac{\partial \hat{g}_{j+\frac{1}{2}}}{\partial h_j} \geq 0,$$

for $0 \leq j \leq N$, $h_j \in [\alpha, \beta]$.

Proof

$$\frac{\partial \hat{g}_{j+\frac{1}{2}}}{\partial h_j} = f'_+(h_j) - \Delta x \chi \left(\frac{pf'(h_j)}{\sqrt{\Delta x}} \right) D_h(x_j, h_j) - \sqrt{\Delta x} p \chi' \left(\frac{pf'(h_j)}{\sqrt{\Delta x}} \right) f''(h_j) D(x_j, h_j).$$

This is zero if $f'(h_j) \leq 0$. In the case $f'(h_j) > 0$

$$\begin{aligned} \frac{\partial \hat{g}_{j+\frac{1}{2}}}{\partial h_j} &\geq f'(h_j) - \Delta x \chi \left(\frac{pf'(h_j)}{\sqrt{\Delta x}} \right) M_1 - \sqrt{\Delta x} p \chi' \left(\frac{pf'(h_j)}{\sqrt{\Delta x}} \right) M_2 \\ &= \frac{\sqrt{\Delta x}}{p} \left(r - p \sqrt{\Delta x} M_1 \chi(r) - p^2 M_2 \chi'(r) \right) \\ &\geq \frac{\sqrt{\Delta x}}{p} \left(r - \chi(r) - \frac{1}{4} \chi'(r) \right), \end{aligned}$$

where

$$r = \frac{pf'(h_j)}{\sqrt{\Delta x}} > 0.$$

Elementary analysis of the function χ shows that

$$r - \chi(r) - \frac{1}{4} \chi'(r) \geq 0,$$

for $r \geq 0$. Next we have

$$\begin{aligned} \frac{\partial \hat{g}_{j+\frac{1}{2}}}{\partial h_{j+1}} &= f'_-(h_{j+1}) + \Delta x \chi \left(\frac{-pf'(h_{j+1})}{\sqrt{\Delta x}} \right) D_h(x_{j+1}, h_{j+1}) \\ &\quad - \sqrt{\Delta x} p \chi' \left(\frac{-pf'(h_{j+1})}{\sqrt{\Delta x}} \right) f''(h_{j+1}) D(x_{j+1}, h_{j+1}). \end{aligned}$$

This is zero if $f'(h_j) \geq 0$. In the case $f'(h_j) < 0$

$$\begin{aligned} \frac{\partial \hat{g}_{j+\frac{1}{2}}}{\partial h_{j+1}} &\leq f'(h_{j+1}) + \Delta x \chi \left(\frac{-pf'(h_{j+1})}{\sqrt{\Delta x}} \right) M_1 + \sqrt{\Delta x} p \chi' \left(\frac{-pf'(h_{j+1})}{\sqrt{\Delta x}} \right) M_2 \\ &= \frac{\sqrt{\Delta x}}{p} \left(-r + p \sqrt{\Delta x} M_1 \chi(r) + p^2 M_2 \chi'(r) \right) \\ &\leq \frac{\sqrt{\Delta x}}{p} \left(-r + \chi(r) + \frac{1}{4} \chi'(r) \right) \\ &\leq 0, \end{aligned}$$

where

$$r = \frac{-pf'(h_j)}{\sqrt{\Delta x}} > 0.$$

The above result leads to the following theorem.

Similarly for $0 < j < N$ we have

$$\mathbf{G}(\cdot)|_j = \beta - \Delta \left(\chi_j^- D(x_{j-1}, \beta) + \chi_j^0 D(x_j, \beta) + \chi_j^+ D(x_{j+1}, \beta) \right),$$

with the non-negative coefficients χ_j^- , χ_j^0 , χ_j^+ being evaluated at $h_{j-1} = h_j = h_{j+1} = \beta$. Since $\beta \geq \bar{h} = \max\{\gamma_0, \gamma_1, m\}$, then $D(x_j, \beta) \geq 0$ for $j = 0, 1, \dots, N$, and we have

$$\mathbf{G}(\cdot) \leq \cdot.$$

As in section 5.2 for $\mathbf{u}, \mathbf{v} \in [\boldsymbol{\alpha}, \cdot]$ we can write

$$\mathbf{G}(\mathbf{h}) - \mathbf{G}(\mathbf{v}) = M(\mathbf{h} - \mathbf{v}),$$

where

$$M = \int_0^1 \mathbf{G}'(\mathbf{h} + s(\mathbf{v} - \mathbf{h})) ds$$

and $\mathbf{G}'$ is of the form (5.15) with

$$p_j = \frac{\Delta}{\Delta x} \frac{\partial \hat{g}_{j-\frac{1}{2}}}{\partial h_{j-1}},$$

$$\begin{aligned} q_j &= 1 - \frac{\Delta}{\Delta x} \left(\frac{\partial \hat{g}_{j+\frac{1}{2}}}{\partial h_j} - \frac{\partial \hat{g}_{j-\frac{1}{2}}}{\partial h_j} \right) - \Delta D_h(x_j, h_j) \\ &= 1 - \Delta D_h(x_j, h_j) - p_{j+1} - r_{j-1}, \end{aligned}$$

$$r_j = -\frac{\Delta}{\Delta x} \frac{\partial \hat{g}_{j+\frac{1}{2}}}{\partial h_{j+1}}.$$

The previous lemma shows that the p_j and r_j are non-negative. For $j = 0, 1, 2, \dots, N$ we can write

$$\begin{aligned} q_j &= 1 - \Delta \left(\frac{|f(h_j)|}{\Delta x} + (1 - \chi(r) - \chi(-r)) D_h(x_j, h_j) \right. \\ &\quad \left. + (\chi'(-r) - \chi'(r)) \frac{pf''(h_j)}{\sqrt{\Delta x}} D(x_j, h_j) \right) \quad \left(r = \frac{pf'(h_j)}{\sqrt{\Delta x}} \right) \\ &\geq 1 - \Delta \left(\frac{|f(h_j)|}{\Delta x} + D_h(x_j, h_j) + \frac{p}{\sqrt{\Delta x}} |f''(h_j) D(x_j, h_j)| \right) \\ &\geq 0, \end{aligned}$$

using the fact that $|\chi'(-r) - \chi'(r)| \leq 1$ and condition (A.3).

We estimate the L_1 norm of the matrix $\mathbf{G}'(h)$ by computing the sum of each column. The sum of the first column is p_1 . Consider the expression q_0 where for the sake of argument take $h_{-1} = h_0$. We have shown that $q_0 \geq 0$, and we can write

$$p_1 = 1 - \Delta D_h(x_0, h_0) - r_{-1} - q_0 \leq 1 - \Delta \delta,$$

since $r_{-1} \geq 0$ from Lemma A.1. The sum of the second column is given by

$$q_1 + p_2 = 1 - \Delta D_h(x_1, h_1) - r_0 \leq 1 - \Delta \delta,$$

since $r_0 \geq 0$. For the j^{th} column ($3 \leq j \leq N-2$) the sum is given by

$$r_{j-2} + q_{j-1} + p_j = 1 - \Delta D_h(x_{j-1}, h_{j-1}) \leq 1 - \Delta \delta.$$

The same argument shows that the remaining two column sums satisfy the same bound, hence we conclude that

$$\|\mathbf{G}'(h)\|_1 \leq 1 - \Delta \delta.$$

It follows that

$$\|M\|_1 = \left\| \int_0^1 \mathbf{G}'(\mathbf{h} + s(\mathbf{v} - \mathbf{h})) ds \right\|_1 \leq \int_0^1 \|\mathbf{G}'(\mathbf{h} + s(\mathbf{v} - \mathbf{h}))\|_1 ds \leq 1 - \Delta \delta < 1.$$

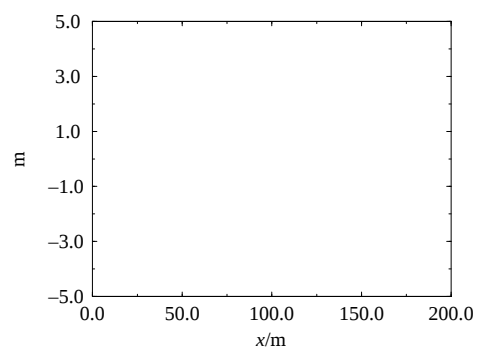
Thus Lemma 5.1 holds with $k = 1 - \Delta \delta$. This completes the proof.

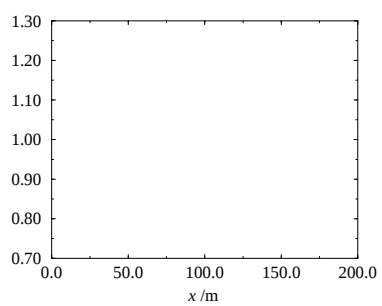
Appendix B

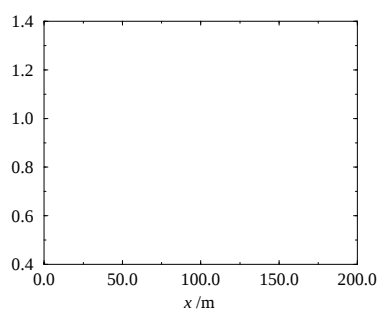
Test Problems for Non-Prismatic Channels

In this appendix we give the details of six test problems with non-prismatic channels, constructed using the method described in Chapter 6. Table B.1 gives the parameters for these problems.

The channel for problems 9, 10, 11 and 12 is shown in Figure B.1. The channel is defined by the following data:







) — J

