MULTIDIMENSIONAL UPWIND SCHEMES FOR THE SHALLOW WATER EQUATIONS

M.E.HUBBARD
University of Reading, Department of Mathematics,
P.O.Box 220, Whiteknights, Reading, Berkshire, RG6 6AX, U.K.

1 Introduction

Over the past fifteen years multidimensional upwinding techniques have been developed with the intention of superseding the well-established upwind finite volume methods which rely on the solution of one-dimensional Riemann problems. The new methods attempt a more genuinely multidimensional approach to the solution of nonlinear systems of conservation laws by considering a piecewise linear continuous representation of the flow with the data stored at the nodes of the grid (more in line with finite elements). The schemes are constructed from three separate stages: the decomposition of the system of equations into simple, usually scalar, components (see, for example [22, 4, 16, 19]), the construction of a consistent, conservative discrete form of the equations [6] and the subsequent solution of the decomposed system using scalar fluctuation distribution schemes [23, 7]. A detailed description of each of these stages can also be found in [5].

These multidimensional upwind schemes will capture shocks within two or three cells when they are aligned with the mesh [17] and grid adaptation can be used to take advantage of this. On unstructured grids this can be accomplished by refinement [2], which reduces the size of the cells, and by edge swapping, which realigns the grid. However, both selective refinement and edge alignment can, to a large extent, be achieved by a third option, grid movement, which has the added advantage of avoiding the expensive process of changing the number of nodes or the connectivity of the grid.

A general overview of multidimensional upwind techniques on triangular grids will be given here, focusing on the most successful developments to date. In Section 2, fluctuation distribution schemes for solving the scalar advection equation will be described, mainly for steady state problems although recent improvements in the modelling of time-dependent flows will also be described briefly. Section 3 contains details of the discretisation for nonlinear systems of equations, taking as an example the shallow water equations. This includes both the linearisation process and the decomposition stage of the algorithm. Following this, in Section 4 a very simple and cheap algorithm for moving grid nodes is presented, which is then used to improve the accuracy of two-dimensional steady state solutions of these conservation laws on unstructured triangular grids. The effectiveness of the techniques described in each section will be illustrated using a small selection of numerical results, and at the end a brief discussion of the current state of multidimensional upwinding will be given.

2 The Scalar Advection Equation

2.1 One-Dimensional Fluctuation Distribution

In this section a method is constructed for the solution of the one-dimensional scalar advection equation,

$$u_t + f_x \;=\; 0 \qquad \text{or} \qquad u_t + \lambda\, u_x \;=\; 0\;, \tag{2.1}$$

where $f = f(u)$ and $\lambda(u) = \frac{\partial f}{\partial u}$ defines the advection velocity associated with the variable u. Assuming for the moment that the advection velocity is locally constant, the approximation to u is chosen to vary linearly over each cell with continuity at the nodes (see Figure 2.1), in the manner of a finite element scheme rather than the finite volume schemes with which these methods usually compete.

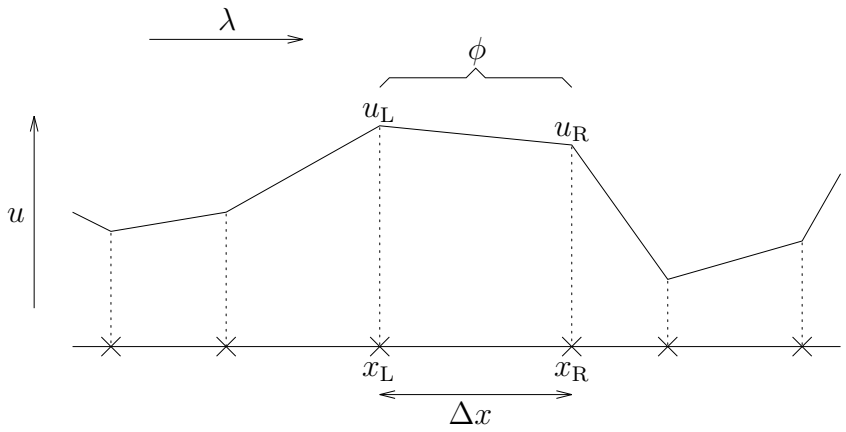

Figure 2.1: The representation of the solution in one dimension.

The **fluctuation** ϕ is now defined. This is a cell-based quantity (closely related to the residual) which, when non-zero, indicates that the system has not yet reached equilibrium [21], and it can be distributed to the nodes of the grid to bring the system closer to equilibrium. In a cell bounded by nodes at $x_{\rm R}$ (right) and $x_{\rm L}$ (left), it is given by

$$\phi \;=\; \int_{x_{\rm L}}^{x_{\rm R}} u_t \,\mathrm{d}x \;=\; -\int_{x_{\rm L}}^{x_{\rm R}} f_x \,\mathrm{d}x \;=\; -\;(f_{\rm R} - f_{\rm L})\;, \tag{2.2}$$

where $f_{\rm R}$ and $f_{\rm L}$ are the values of the flux function f at $x_{\rm R}$ and $x_{\rm L}$ respectively. Since by construction u_x is constant within the cell, a second, equivalent, expression can be obtained for the fluctuation,

$$\phi \;=\; -\int_{x_{\rm L}}^{x_{\rm R}} \lambda(u)\, u_x \,\mathrm{d}x \;=\; -\frac{1}{\Delta x}\left(\int_{x_{\rm L}}^{x_{\rm R}} \lambda(u)\,\mathrm{d}x\right)(u_{\rm R} - u_{\rm L})\;, \tag{2.3}$$

where $\Delta x = x_{\rm R} - x_{\rm L}$ is the length of the cell, *cf.* Figure 2.1.

The integral in (2.3) cannot always be evaluated easily (exact evaluation is important in the construction of a conservative linearisation) but it clearly represents a cell-averaged value

of the advection velocity. However, in this simple case a consistent approximation $\tilde{\lambda}$ of this average can always be obtained locally by comparing (2.2) and (2.3), giving

$$\tilde{\lambda} \;=\; \frac{f_{\mathrm{R}} - f_{\mathrm{L}}}{u_{\mathrm{R}} - u_{\mathrm{L}}}\,, \tag{2.4}$$

where from here onwards $\tilde{\cdot}$ indicates a linearised quantity. This quantity determines the direction of the flow in a cell, which is employed in the construction of any upwind scheme. Note that the discrete fluctuation $\tilde{\phi} = -\tilde{\lambda}(u_{\mathrm{R}} - u_{\mathrm{L}})$ from (2.3), so when $u_{\mathrm{R}} = u_{\mathrm{L}}$, $\tilde{\phi} = 0$ and $\tilde{\lambda}$ is not needed.

When combined with a forward Euler discretisation of the time derivative in (2.1), the **distribution** of the fluctuations can be written in the form of a nodal update scheme,

$$u_i^{n+1} \;=\; u_i^n + \frac{\Delta t}{\Delta x_i} \sum_{j=1}^{N_c} \tilde{\alpha}_i^j \, \tilde{\phi}_j \;, \tag{2.5}$$

where Δx_i is the length of the median dual cell of node i (the distance between the centroids of the two cells adjacent to the node, not necessarily the Δx of (2.3)), $\tilde{\alpha}_i^j$ is the distribution coefficient which indicates the proportion of the fluctuation $\tilde{\phi}_j$ sent from cell j to node i, and N_c is the number of grid cells. Since the resulting iterative scheme is explicit, the superscript n has immediately been dropped from the update for clarity of presentation, and will be assumed from now on.

Summing the fluctuations in equation (2.2) over the whole domain leads to an expression which consists solely of contributions from the boundary. Therefore, as long as each fluctuation is distributed completely, that is

$$\sum_{i=1}^{N_n} \tilde{\alpha}_i^j \;=\; 1 \qquad \text{for} \qquad j \;=\; 1\,,\, 2\,, \ldots,\, N_c\,, \tag{2.6}$$

where $N_n = N_c{+}1$ is the number of grid nodes, a scheme of the form (2.5) will be conservative.

The scheme itself is determined by the choice of the distribution coefficients $\tilde{\alpha}_i^j$, *e.g.* if they are chosen so that the whole fluctuation is sent to the downstream node of the cell according to the sign of $\tilde{\lambda}$, then the resulting scheme is simply first order upwinding. In fact, any one-dimensional finite volume scheme can be rewritten in terms of fluctuation distributions [21]. The advantage of the new form of the scheme is that it generalises naturally to higher dimensions.

2.2 Two-Dimensional Schemes

In two dimensions, the scalar advection equation takes the form

$$u_t + f_x + g_y \;=\; 0 \qquad \text{or} \qquad u_t + \vec{\lambda} \cdot \vec{\nabla} u \;=\; 0\,, \tag{2.7}$$

where $\vec{\lambda} = \left(\frac{\partial f}{\partial u}, \frac{\partial g}{\partial u}\right)^{\mathrm{T}}$ defines the velocity of the advected variable u which, in simple cases, is again approximated by a piecewise linear continuous representation on a triangular grid

(it is slightly less straightforward on quadrilaterals). A scheme can be constructed for the solution of (2.7) by calculating the **fluctuation**,

$$\tilde{\phi} \;=\; -S_{\triangle}\tilde{\vec{\lambda}}\cdot\vec{\nabla}u \;=\; -\int\!\!\int_{\triangle}\vec{\lambda}\cdot\vec{\nabla}u\,\mathrm{d}x\,\mathrm{d}y\;, \tag{2.8}$$

within each cell and then distributing it to the nodes of the grid. The integration in (2.8), when it can be done exactly under the assumption that u varies linearly within each cell (so $\vec{\nabla}u$ is locally constant), introduces the factor of $S_{\triangle}$, the area of the triangle, and a cell-averaged wave speed,

$$\tilde{\vec{\lambda}} \;=\; \frac{1}{S_{\triangle}}\int\!\!\int_{\triangle}\vec{\lambda}\,\mathrm{d}x\,\mathrm{d}y\;. \tag{2.9}$$

Note that given this local approximation and the condition that $\vec{\nabla}\cdot\vec{\lambda}=0$, application of the divergence theorem and a small amount of algebraic manipulation leads to

$$\tilde{\phi} \;=\; -\frac{1}{2}\sum_{l=1}^{N_e}(\tilde{\vec{\lambda}}\cdot\vec{n}_l)\,u_l \quad \oint_{\partial\triangle}(f,g)\cdot\mathrm{d}\vec{n}\;, \tag{2.10}$$

an alternative form for the discrete fluctuation in which $\partial\triangle$ is the boundary of the cell, $N_e=3$ is the number of edges of the cell, and $\vec{n}_l$ is the inward pointing normal to edge l, scaled by the length of that edge.

For simplicity and compactness, a cell is allowed to contribute its fluctuation only to its own vertices. Since summing the fluctuations over the whole domain reduces to a sum of boundary contributions, a conservative scheme is again assured as long as the whole of each fluctuation is distributed, *cf.* (2.6).

If explicit forward Euler time-stepping is used, this leads to a scheme of the form

$$u_i^{n+1} \;=\; u_i^n + \frac{\Delta t}{S_i}\sum_{j\in\cup\triangle_i}\tilde{\alpha}_i^j\,\tilde{\phi}_j\;, \tag{2.11}$$

where S_i is the area of the median dual cell for node i (one third of the total area of the triangles with a vertex at i), $\tilde{\alpha}_i^j$ is the distribution coefficient which indicates the proportion of the fluctuation $\tilde{\phi}_j$ (a linear function of the data) to be sent from cell j to node i, and $\cup\triangle_i$ represents the set of cells adjacent to node i, as illustrated in Figure 2.2. Note that a slightly more accurate scheme is given by

$$u_i^{n+1} \;=\; u_i^n + \frac{\Delta t}{\sum_{j\in\cup\triangle_i}\tilde{\alpha}_i^j\,S_{\triangle_j}}\sum_{j\in\cup\triangle_i}\tilde{\alpha}_i^j\,\tilde{\phi}_j\;. \tag{2.12}$$

This corresponds to a consistent, mass-lumped, upwind finite element discretisation.

The scheme (2.11) can be written in the form

$$u_i^{n+1} \;=\; \sum_k c_{ik}\,u_k^n\;. \tag{2.13}$$

If the coefficients c_{ik} are allowed to depend on the data, the scheme becomes nonlinear and can be designed to satisfy the following four criteria:

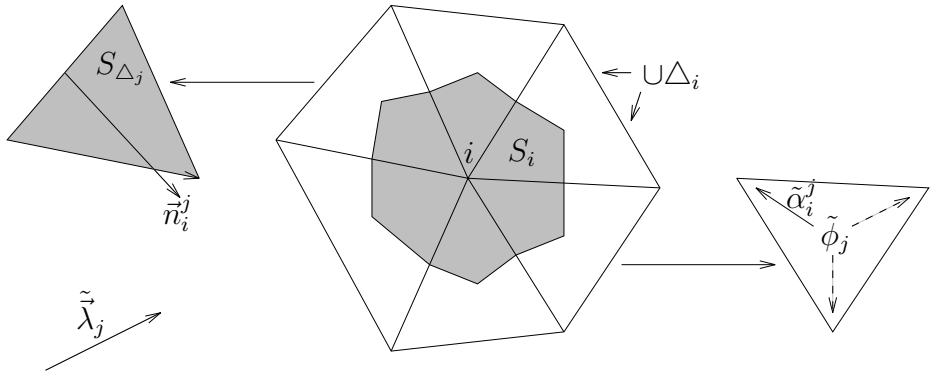

Figure 2.2: The notation used in the representation of the two-dimensional scheme.

- **upwindedness** - the fluctuation within a cell is only sent to the downstream vertices of that cell *i.e.* vertices opposite inflow edges, for which $\tilde{\vec{\lambda}} \cdot \vec{n} > 0$, where $\vec{n}$ is an inward pointing normal to the edge.
- **positivity** - the coefficients c_{ik} are positive, so the scheme cannot produce new extrema in the solution at the new time-step, spurious oscillations will not appear in the solution and the scheme is stable, all for an appropriate time-step restriction.
- **linearity preservation** - no update is sent to the nodes when a cell fluctuation is zero, so linear steady states are preserved and the scheme is second order accurate at the steady state on a regular mesh with a uniform choice of diagonals [7].
- **continuity** - the contributions to the nodes, $\tilde{\alpha}_i^j \tilde{\phi}_j$, depend continuously on the data, avoiding limit cycling as convergence to a steady state is approached.

Note that a linear scheme cannot satisfy both the positivity and the linearity preservation properties simultaneously.

In the search for a scheme which satisfies all of the above properties it is initially advantageous to consider a linear, positive, upwind scheme, the N scheme [7]. By the above definition of upwind, any triangle with only one downstream vertex sends the whole of its fluctuation to that node. This makes these triangles both positive and linearity preserving. For a cell with two inflow sides (choosing, without loss of generality, vertices 1 and 2 to be the downstream nodes), considered in isolation, the N scheme can be written as

$$\begin{aligned} S_1 u_1^{n+1} &= S_1 u_1^n - \Delta t\, \tilde{k}_1 (u_1^n - u_3^n) \\ S_2 u_2^{n+1} &= S_2 u_2^n - \Delta t\, \tilde{k}_2 (u_2^n - u_3^n) \\ S_3 u_3^{n+1} &= S_3 u_3^n \,, \end{aligned} \tag{2.14}$$

where $\tilde{k}_l = \frac{1}{2}\tilde{\vec{\lambda}} \cdot \vec{n}_l$, *cf.* (2.11). Contributions from other triangles are temporarily suppressed. By considering the complete nodal update (2.11), this scheme can be shown to be positive for a restriction on the time-step at a node i, given by

$$\Delta t \leq \frac{S_i}{\sum_{j \in \cup\triangle_i} \max(0, \tilde{k}_i^j)} \,. \tag{2.15}$$

A linearity preserving scheme (which also retains the upwind and continuity properties) can be obtained from a positive upwind scheme [25] such as the N scheme by replacing the overall contributions, $\tilde{\psi}_1$ and $\tilde{\psi}_2$, to the downstream nodes in the two-target case by 'limited' contributions,

$$\begin{aligned} \tilde{\psi}_1^* &= \tilde{\psi}_1 - L(\tilde{\psi}_1\,, -\tilde{\psi}_2) \\ \tilde{\psi}_2^* &= \tilde{\psi}_2 - L(\tilde{\psi}_2\,, -\tilde{\psi}_1)\ . \end{aligned} \tag{2.16}$$

where, in the case of the N scheme presented in (2.14),

$$\tilde{\psi}_1 = -\tilde{k}_1(u_1^n - u_3^n)\ , \qquad \tilde{\psi}_2 = -\tilde{k}_2(u_2^n - u_3^n)\ . \tag{2.17}$$

$L(x, y)$ is any member of the family of symmetric limiter functions, although the minmod limiter [26], given by

$$L(x,y) \;=\; \frac{1}{2}(1+\mathrm{sgn}(xy))\,\frac{1}{2}(\mathrm{sgn}(x)+\mathrm{sgn}(y))\,\min(|x|,|y|)\ , \tag{2.18}$$

is the only one for which the 'limited' scheme remains positive. The resulting scheme is equivalent to the Positive Streamwise Invariant (PSI) scheme [7], satisfies all of the desired properties and is the scheme which is generally used for the distribution of scalar fluctuations for steady state problems.

2.3 Time-Dependent Problems

Whilst the methods described above model steady state flows accurately, none of the positive schemes developed originally [7] is higher than first order accurate in space for time-dependent problems. This problem has recently been addressed successfully:

- by equating the PSI scheme with a mass-lumped finite element scheme, then constructing an appropriate mass matrix (using standard linear trial functions and SUPG-style linear test functions) to give the corresponding full finite element method (which gives up to third order accuracy in some special cases, at the expense of inverting the mass matrix) [15]. flux-corrected transport (FCT) has been applied to enforce a local maximum principle.
- by combining the PSI and Lax-Wendroff schemes in a manner similar to (but more general than) flux-corrected transport, maintaining the local maximum principle satisfied by the PSI scheme but combining it with optimal accuracy [12].

The latter technique will be discussed below in a little more detail. Note first though that the distribution coefficients of the Lax-Wendroff scheme are given by

$$\tilde{\alpha}_i^j \;=\; \frac{1}{3} + \frac{\Delta t}{4S_{\triangle_j}}\tilde{\vec{\lambda}}_j \cdot \vec{n}_i^j\ . \tag{2.19}$$

First some relevant concepts are defined [12], and illustrated in Figure 2.3. Consider a single grid cell in isolation: the **distribution point** is defined to be the point whose local area

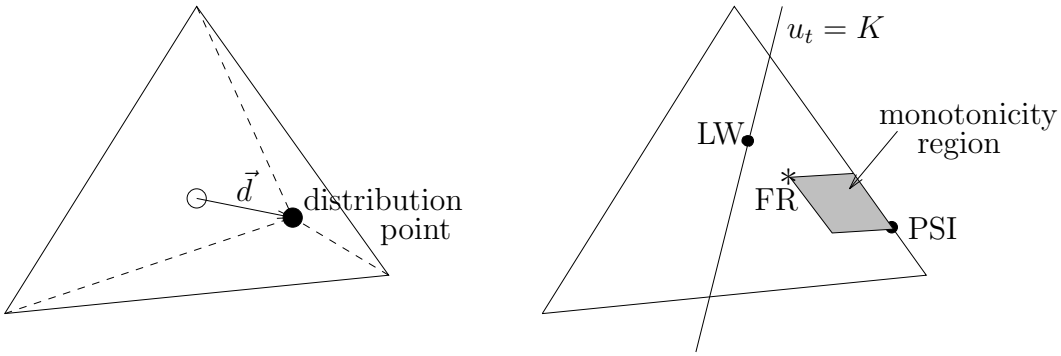

Figure 2.3: The distribution point, diffusion vector and monotonicity region.

coordinates are the distribution coefficients of the scheme for that triangle. The movement of the distribution point is equivalent to the **redistribution** of the fluctuation within the triangle. The **diffusion vector** represents the displacement of the distribution point from the centroid of the triangle (the distribution point of a symmetric central scheme). A scheme with diffusion vector $\vec{d}$ has the second order **equivalent equation**

$$u_t + \vec{\lambda} \cdot \vec{\nabla} u \;=\; \vec{d} \cdot \vec{\nabla}(\vec{\lambda} \cdot \vec{\nabla} u) \;, \tag{2.20}$$

in which the right hand side represents the numerical diffusion of the distribution scheme.

The diffusion vector of the Lax-Wendroff scheme is $\vec{d} = \frac{1}{2}\Delta t\, \vec{\lambda}$ and the equivalent equation (2.20) can be rewritten as

$$u_t + \vec{\lambda} \cdot \vec{\nabla} u \;=\; \frac{\vec{\lambda}\Delta t}{2} \cdot \vec{\nabla}(\vec{\lambda} \cdot \vec{\nabla} u) + \left(\vec{d} - \frac{\vec{\lambda}\Delta t}{2} \right) \cdot \vec{\nabla}(\vec{\lambda} \cdot \vec{\nabla} u) \;. \tag{2.21}$$

Hence, any choice of $\vec{d}$ such that

$$\vec{d} - \frac{\vec{\lambda}\Delta t}{2} \;\perp\; \vec{\nabla}(\vec{\lambda} \cdot \vec{\nabla} u) \;=\; -\vec{\nabla} u_t \tag{2.22}$$

will not alter the second order error term in the approximation, so the corresponding distribution scheme should be second order accurate for the given local data. Therefore, moving the distribution point perpendicular to the local value of $\vec{\nabla}(\vec{\lambda} \cdot \vec{\nabla} u)$ should not change the order of accuracy of the local discretisation. It is important to note here that $\vec{\nabla} u_t$ in (2.22) can be approximated locally using the unlimited high order update (which has already been calculated as part of this FCT-type limiting procedure). This allows the overall algorithm to remain compact.

By considering a generalised FCT-like algorithm, in which the monotonic scheme is written in terms of low order (LO) and high order (HO) updates, the distribution coefficients can be expressed locally as

$$\begin{aligned}
\tilde{\alpha}_1 &= \tilde{\alpha}_1^{\mathrm{LO}} + \beta_1 \left(\tilde{\alpha}_1^{\mathrm{HO}} - \tilde{\alpha}_1^{\mathrm{LO}} \right) \;, \\
\tilde{\alpha}_2 &= \tilde{\alpha}_2^{\mathrm{LO}} + \beta_2 \left(\tilde{\alpha}_2^{\mathrm{HO}} - \tilde{\alpha}_2^{\mathrm{LO}} \right) \;, \\
\tilde{\alpha}_3 &= \tilde{\alpha}_3^{\mathrm{LO}} + \beta_3 \left(\tilde{\alpha}_3^{\mathrm{HO}} - \tilde{\alpha}_3^{\mathrm{LO}} \right) \;,
\end{aligned} \tag{2.23}$$

in which the β_k are limiting coefficients. Bounds are constructed on the solution at the new time level which ensure that a local maximum principle is satisfied [12]. These are easily translated into restrictions on the limiting coefficients, generally expressed as

$$\beta_k^{\min} \;\leq\; \beta_k \;\leq\; \beta_k^{\max}\,, \quad k = 1,\, 2,\, 3\,, \tag{2.24}$$

which describes three pairs of 'tramlines' parallel to the edges of the triangle. Placing the distribution point anywhere within the region defined by (2.24), the **monotonicity region**, ensures that the subsequent nodal updates will not create any new local extrema at the next time level and as a result imposes stability on the scheme. Note that the monotonicity region always contains the PSI distribution point so values of β_k always exist.

The calculation of the limited distribution coefficients, given those of the high order (Lax-Wendroff) and low order (PSI) schemes, is as follows:

- find the line passing through the high order distribution point perpendicular to the locally constructed value of $\vec{\nabla} u_t$ (*i.e.* a contour line of u_t).
- calculate the position of the point in the monotonicity region closest to the line defined above (indicated by an asterisk in Figure 2.3) and take this to be the distribution point of the limited scheme. If the line intersects the region then take the point of intersection closest to the high order distribution point.

This process has been called **fluctuation redistribution**. Note that flux-corrected transport, for which

$$\beta_1 \;=\; \beta_2 \;=\; \beta_3 \;=\; \min_{k=1,2,3} \beta_k^{\max}\,, \tag{2.25}$$

is a special case of this.

2.4 Results

The test case used is that of clockwise circular advection, $\vec{\lambda} = (y, -x)^{\mathrm{T}}$, of a given wave profile within the domain $(x, y) \in [-1, 1] \times [0, 1]$. The mesh used is a uniform triangulation created from a regular 65×33 node rectangular grid by inserting diagonals which alternate in direction. Two cases are presented, one being the advection of a square wave profile, for which the inflow boundary conditions are taken to be

$$\begin{aligned} u(x,0) &= 1 \quad \text{for} \quad -0.65 \leq x \leq -0.35 \\ u(x,0) &= 0 \quad \text{elsewhere}\,, \end{aligned} \tag{2.26}$$

and a second which convects a triangular wave profile with the same zero boundary conditions as above but now

$$u(x,0) \;=\; 1.0 - \frac{|x + 0.5|}{0.15} \quad \text{for} \quad -0.65 \leq x \leq -0.35\,. \tag{2.27}$$

In both cases the solution is initially set to zero in the whole of the domain's interior. The results obtained using the N, PSI and Lax-Wendroff schemes are compared with the exact

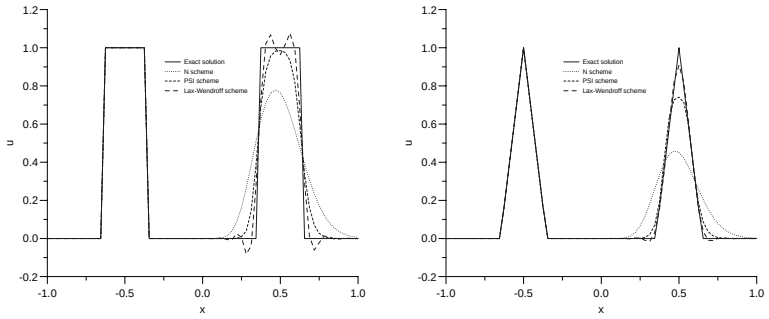

Figure 2.4: Solution on the boundary $y = 0$ for the advection of the square wave (left) and the triangular wave (right).

solution on the given grid in Figure 2.4, illustrating how at the steady state the PSI scheme achieves high accuracy without oscillations.

Unfortunately, this accuracy does not carry over to time-dependent problems, as can be seen in Figure 2.5. This shows that for a simple test case, the advection of a double sine wave profile,

$$u \;=\; \sin(2\pi x)\sin(2\pi y)\,, \tag{2.28}$$

with velocity $\vec{\lambda} = (1, 2)^{\mathrm{T}}$ over the domain $[0, 1] \times [0, 1]$, the PSI scheme is only first order accurate. Surprisingly, the high resolution cell centre finite volume scheme [1] with which it is compared is little better on this type of grid (with alternating diagonals) and the consistent finite element scheme of [15] loses accuracy dramatically as the grid is refined. The fluctuation redistribution approach retains the accuracy achieved by the Lax-Wendroff scheme throughout.

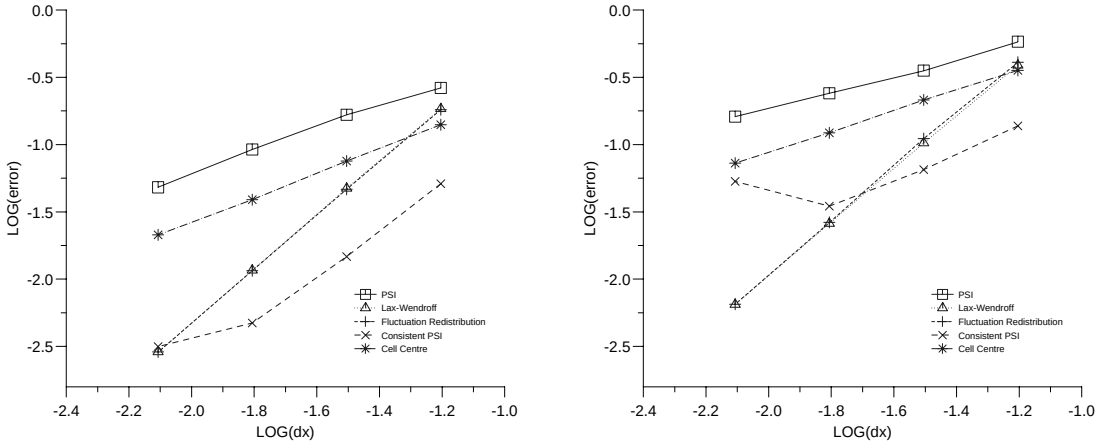

Figure 2.5: L_1 (left) and L_∞ (right) errors for the double sine wave test case on a uniform triangular grid of alternating diagonals for $t = 1.0$.

3 The Shallow Water Equations

3.1 One-Dimensional Flow

A general one-dimensional, homogeneous system of conservation laws can be written

$$\underline{U}_t + \underline{F}_x \;=\; \underline{0} \qquad \text{or} \qquad \underline{U}_t + \mathbf{A}\underline{U}_x \;=\; \underline{0}\,, \tag{3.1}$$

where $\mathbf{A} = \frac{\partial \underline{F}}{\partial \underline{U}}$ is the conservative flux Jacobian matrix. The **flux balance** corresponding to this system is defined in much the same way as the fluctuation in the scalar case, being the cell-based quantity

$$\underline{\Phi} \;=\; \int_{x_\mathrm{L}}^{x_\mathrm{R}} \underline{U}_t \,\mathrm{d}x \;=\; -\int_{x_\mathrm{L}}^{x_\mathrm{R}} \underline{F}_x \,\mathrm{d}x \;=\; -\left(\underline{F}_\mathrm{R} - \underline{F}_\mathrm{L}\right). \tag{3.2}$$

The aim is to linearise this quantity and then decompose it into scalar components to which the fluctuation distribution schemes of Section 2.1 can be applied. The contributions made by the scalar fluctuations to the nodes are then transformed into updates for the original conservative system (3.1). Conservation is assured as long as the system is linearised correctly and the whole of each discrete conservative flux balance $\tilde{\underline{\Phi}}$ is distributed to the grid nodes.

In the first step, a consistent linearisation (indicated again by $\tilde{\cdot}$) is sought for the flux Jacobian $\mathbf{A}$ which ensures that

$$\tilde{\underline{\Phi}} \;=\; -\left(\underline{F}_\mathrm{R} - \underline{F}_\mathrm{L}\right) \;=\; -\tilde{\mathbf{A}}\left(\underline{U}_\mathrm{R} - \underline{U}_\mathrm{L}\right) \qquad \text{and} \qquad \tilde{\mathbf{A}} \;=\; \left.\frac{\partial \underline{F}}{\partial \underline{U}}\right|_{\tilde{\underline{Z}}}\,, \tag{3.3}$$

where $\tilde{\mathbf{A}}$ should have a full set of linearly independent eigenvectors. This is essentially Roe's Property U [20] with, for the purposes of the subsequent decomposition of the discrete system, the additional specification that the discrete flux Jacobian is simply the evaluation of the analytic Jacobian at a given 'Roe-average' state. A number of different possibilities exist for the choice of $\tilde{\underline{Z}}$ in the case of the one-dimensional shallow water equations (see, for example, [9]), all of which are constructed specifically so that (3.3) is satisfied.

In one dimension the shallow water equations can be completely decoupled by diagonalising the system, and this forms the foundation of the decomposition. The linearised equations are transformed into characteristic variables $\underline{W}$, so

$$\underline{W}_t + \tilde{\mathbf{\Lambda}}\underline{W}_x \;=\; \underline{0}\,, \tag{3.4}$$

in which

$$\tilde{\mathbf{\Lambda}} \;=\; \frac{\partial \tilde{\underline{U}}}{\partial \underline{W}}^{-1} \frac{\partial \tilde{\underline{F}}}{\partial \underline{U}} \frac{\partial \tilde{\underline{U}}}{\partial \underline{W}} \tag{3.5}$$

is a diagonal matrix composed of the eigenvalues $\tilde{\lambda}$ of the Jacobian $\tilde{\mathbf{A}}$, and $\frac{\partial \tilde{\underline{U}}}{\partial \underline{W}}$ is the matrix of right eigenvectors of $\tilde{\mathbf{A}}$. This decouples the system into three independent equations, each of the form

$$W_t + \tilde{\lambda} W_x \;=\; 0\,, \tag{3.6}$$

precisely the scalar advection equation solved in Section 2.1. This means that the same numerical schemes can be applied in order to distribute the fluctuation due to (3.6). The original system (3.1), and hence the conservative flux balance, is regained by premultiplying equation (3.4) by $\frac{\partial U}{\partial \underline{W}}$. Thus, the flux balance for the conservative system is

$$\underline{\tilde{\Phi}} \;=\; -\frac{\partial \underline{\tilde{U}}}{\partial \underline{W}}\,\tilde{\mathbf{\Lambda}}(\underline{W}_{\mathrm{R}}-\underline{W}_{\mathrm{L}}) \;=\; -\sum_{k=1}^{N_{eq}}\tilde{\lambda}^k(W^k_{\mathrm{R}}-W^k_{\mathrm{L}})\underline{\tilde{r}}^k \;=\; \sum_{k=1}^{N_{eq}}\tilde{\phi}^k\underline{\tilde{r}}^k\;, \tag{3.7}$$

where $\tilde{\phi}^k$ is the scalar fluctuation corresponding to the k^{th} component (*cf.* (3.6)) of the decomposition, $\underline{\tilde{r}}^k$ is the k^{th} column of $\frac{\partial \underline{\tilde{U}}}{\partial \underline{W}}$ (the k^{th} eigenvector of $\tilde{\mathbf{A}}$) and N_{eq} is the number of decoupled equations (2 for the one-dimensional shallow water equations).

Thus, the overall scheme can be written in a similar form to the nodal update for the scalar scheme (2.5), being

$$\underline{U}_i^{n+1} \;=\; \underline{U}_i^n + \frac{\Delta t}{(\Delta x)_i}\sum_{j=1}^{N_c}\sum_{k=1}^{N_{eq}}(\tilde{\alpha}_i^j)^k\tilde{\phi}_j^k\,\underline{\tilde{r}}_j^k\;, \tag{3.8}$$

where each $(\tilde{\alpha}_i^j)^k$ is calculated from the distribution of the individual components (3.6) and represents the proportion of the scalar fluctuation $\tilde{\phi}_j^k$ due to the k^{th} wave in the j^{th} cell to be sent to node i.

3.2 Conservative Linearisations in Two Dimensions

In conservative form, the two-dimensional shallow water equations are written

$$\underline{U}_t + \underline{F}_x + \underline{G}_y \;=\; \underline{0}\;, \tag{3.9}$$

where

$$\underline{U} \;=\; \begin{pmatrix} d \\ du \\ dv \end{pmatrix}\;, \qquad \underline{F} \;=\; \begin{pmatrix} du \\ du^2+\frac{1}{2}gd^2 \\ duv \end{pmatrix}\;, \qquad \underline{G} \;=\; \begin{pmatrix} dv \\ duv \\ dv^2+\frac{1}{2}gd^2 \end{pmatrix}\;, \tag{3.10}$$

are the vectors of conserved variables and the corresponding conservative fluxes, respectively, in which d is depth, u and v are the x- and y-velocities, and g is the acceleration due to gravity. The associated conservative **flux balance** is given by

$$\underline{\Phi} \;=\; \int\!\!\int_{\triangle}\underline{U}_t\,\mathrm{d}x\,\mathrm{d}y \;=\; -\int\!\!\int_{\triangle}\left(\underline{F}_x+\underline{G}_y\right)\,\mathrm{d}x\,\mathrm{d}y \;=\; \oint_{\partial\triangle}(\underline{F},\underline{G})\cdot\mathrm{d}\vec{n}\;, \tag{3.11}$$

where $\vec{n}$ is an inward pointing normal to the boundary $\partial\triangle$ of the cell.

A linearisation (indicated below by $\tilde{\cdot}$) is sought by which the discrete flux balance can be written

$$\underline{\tilde{\Phi}} \;=\; -S_{\triangle}\left(\mathbf{A}\underline{U}_x+\mathbf{B}\underline{U}_y\right)\Big|_{\underline{\tilde{Z}}}\;, \tag{3.12}$$

where $\mathbf{A}=\frac{\partial F}{\partial \underline{U}}$ and $\mathbf{B}=\frac{\partial G}{\partial \underline{U}}$ are the flux Jacobian matrices, so that the decompositions described below can be applied while retaining conservation [17]. Underlying the construction

of the linearised flux balance is the assumption that if a chosen set of independent variables $\underline{Z}$ varies linearly within each cell (with continuity between cells) then $\underline{\Phi} = \tilde{\underline{\Phi}}$. This is not straightforward for the shallow water equations [18] and it turns out to be more practical to choose

$$\underline{Z} \;=\; \begin{pmatrix} \sqrt{d} \\ \sqrt{d}u \\ \sqrt{d}v \end{pmatrix} \qquad \text{or even} \qquad \underline{Z} \;=\; \begin{pmatrix} d \\ u \\ v \end{pmatrix} \tag{3.13}$$

to average and put into the flux balance (3.12), so the Jacobians are evaluated at $\tilde{\underline{Z}}$ (the arithmetic mean of the vertex values), and the discrete gradients are calculated consistently from the gradient of the linearly varying quantities

$$\tilde{\vec{\nabla}}\underline{Z} \;=\; \frac{1}{2S_\triangle} \sum_{l=1}^{N_{eq}} \underline{Z}_l \, \vec{n}_l \;, \tag{3.14}$$

where $\vec{n}_l$ is the inward pointing normal to the edge opposite vertex l, scaled by the length of the edge. A small correction, in the form of a source term which can be distributed appropriately, may be added to ensure conservation [11].

3.3 Multidimensional Wave Models

The decomposition stage of the algorithm dictates how the linearised conservative flux balance within each triangle of the grid, namely

$$\tilde{\underline{\Phi}} \;=\; -S_\triangle(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}) \cdot \vec{\nabla}\underline{U} \;, \tag{3.15}$$

is divided into simpler components. The fact that $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ are not in general simultaneously diagonalisable makes the two-dimensional situation considerably more complicated than that in one dimension. It is also important to note that $\tilde{\underline{\Phi}}$ is simply the analytic quantity $\underline{\Phi}$ evaluated at a given average state, since this means that decompositions can be applied equivalently to both the original and the linearised equations. For the purposes of simplicity, the notation $\tilde{\cdot}$ can therefore be dropped from the following analysis.

So far, three main types of multidimensional decomposition have been proposed:

- **simple wave models**: $\underline{\Phi}$ is decomposed into contributions due to plane wave solutions of the shallow water equations. These correspond to eigenvalues λ, which give the wave speeds, and eigenvectors $\underline{r}$, which give the transformation of the individual wave fluctuations back to perturbations of the conservative variables, of the matrix $(\mathbf{A}, \mathbf{B}) \cdot \vec{n}_\theta$, θ being the propagation direction. This gives three possible wave types (one shear and two gravity) which can be used.
  The wave model consists of a set of N_w such waves, the number of each type being prespecified (making sure that all types of solution can be modelled), each of which is associated with a propagation direction θ and a strength φ. Some of the values of θ are also specified so as to leave, in this case, 6 (#space dimensions $\times$ #waves) parameters. The remaining values of θ and φ are calculated by solving the gradient decomposition

$$\vec{\nabla}\underline{U} \;=\; \sum_{k=1}^{N_w} \underline{r}^k \, (\vec{\nabla} W^k)^{\mathrm{T}} \;=\; \sum_{k=1}^{N_w} \varphi^k \, \underline{r}^k \, (\vec{n}_{\theta^k})^{\mathrm{T}} \tag{3.16}$$

(a system of 6 equations in 6 unknowns), W^k nominally being a variable associated with the k^{th} plane wave. This leads immediately to the decomposed flux balance

$$\underline{\Phi} \;=\; -S_\triangle(\mathbf{A},\mathbf{B})\cdot\vec{\nabla}\underline{U} \;=\; -S_\triangle\sum_{k=1}^{N_w}\varphi^k\,\lambda^k\,\underline{r}^k \;=\; \sum_{k=1}^{N_w}\phi^k\,\underline{r}^k\;. \tag{3.17}$$

Each simple wave contribution is assigned a velocity $\vec{\lambda}$ and a gradient $\vec{\nabla}W$ so that

$$\phi \;=\; -S_\triangle\,\varphi\,\lambda \;=\; -S_\triangle\,\vec{\lambda}\cdot\vec{\nabla}W \tag{3.18}$$

gives the fluctuation in the usual form, *cf.* (2.8), which can be distributed using the methods of Section 2.2. Compare the models of Roe [22] and Rudgyard [24] for different ways in which (3.18) may be constructed.
Several alternative simple wave models have been suggested [19] but they have so far proved unsatisfactory for two main reasons: (i) the propagation directions depend on solution gradients which vary rapidly from cell to cell, inhibiting convergence to the steady state and robustness in general, (ii) every practical model contains too many waves, leading to a surfeit of numerical dissipation and hindering accuracy. Ideally for the two-dimensional shallow water equations $N_w = 3$ (the number of equations in the system) so that the linearity preserving nature of the scalar scheme is retained.

- **characteristic decompositions**: although the matrices $\mathbf{A}$ and $\mathbf{B}$ cannot generally be diagonalised simultaneously, an approximate diagonalisation can be constructed via a 3-parameter similarity transformation [3], giving a system in 'characteristic' variables $\underline{W}$,

$$\underline{W}_t + \mathbf{A}_W\underline{W}_x + \mathbf{B}_W\underline{W}_y \;=\; \underline{0}\;, \tag{3.19}$$

in which the 3 free parameters are chosen so that the new Jacobians $\mathbf{A}_W$ and $\mathbf{B}_W$ are, in some sense, close to being diagonal. This is treated as a decoupled set of inhomogeneous equations, each with a fluctuation of the form

$$\phi \;=\; -S_\triangle(\vec{\lambda}\cdot\vec{\nabla}W + q)\;, \tag{3.20}$$

(the distribution coefficients can be calculated as they would for the homogeneous fluctuation but then used to distribute the complete ϕ) and a conservative flux balance,

$$\underline{\Phi} \;=\; \sum_{k=1}^{N_{eq}}\phi^k\,\underline{r}^k\;, \tag{3.21}$$

in which $\underline{r}^k$ are the columns of the similarity transformation matrix $\frac{\partial\underline{U}}{\partial\underline{W}}$.
These methods have the correct number of components for linearity preservation ($N_w = N_{eq}$) but the propagation directions, which depend on the parameters which define the similarity transformation, are sometimes chosen to depend on solution gradients. However, their main disadvantage is the presence of the source terms in (3.21) which destroy positivity and hence robustness.

- **preconditioned decompositions**: the effect of the source terms created by the characteristic decomposition can be minimised by attempting to diagonalise a preconditioned form of the equations [19]. The decomposed flux balance once more takes the

form (3.21), but now $\underline{r}^k$ are the columns of the matrix $\frac{\partial \underline{U}}{\partial \underline{Q}} \mathbf{P}^{-1} \frac{\partial \underline{Q}}{\partial \underline{W}}$, $\underline{Q}$ being an intermediate set of (symmetrising) variables, introduced to simplify the algebra, and $\mathbf{P}$ is the preconditioning matrix. Careful choice of the preconditioner gives an optimal decoupling of the system, complete in supercritical flow but unavoidably including a coupled 2×2 elliptic subsystem for subcritical flow.
One of the most successful of these schemes, adapted from the hyperbolic/elliptic decomposition for the Euler equations of Mesaros and Roe [16], is discussed in more detail below.

In all of the above cases, the application of forward Euler time-stepping results in a scheme which, if it is assumed that only scalar schemes will be used in the distribution, even when components are coupled, takes the form

$$\underline{U}_i^{n+1} \;=\; \underline{U}_i^n + \frac{\Delta t}{S_i} \sum_{j \in \cup \triangle_i} \sum_{k=1}^{N_w} (\alpha_i^j)^k \phi_j^k \, \underline{r}_j^k \;, \tag{3.22}$$

cf. (3.8) and (2.11), where ϕ_j^k and $\underline{r}_j^k$ take their forms from the particular wave model used, and $(\alpha_i^j)^k$ are the appropriate distribution coefficients.

3.4 An Optimal Decomposition

In order to simplify the analysis the system (3.9) is first transformed into the symmetrising variables,

$$\partial \underline{Q} \;=\; \begin{pmatrix} \sqrt{\frac{g}{d}}\, \partial d \\ \partial q \\ q \, \partial \theta \end{pmatrix} \;, \tag{3.23}$$

where $q = \sqrt{u^2 + v^2}$ is the flow speed and $\theta = \tan^{-1}\left(\frac{v}{u}\right)$ is the direction of the flow. The shallow water equations therefore become

$$\underline{Q}_t + \mathbf{A}_{\mathrm{Q}} \, \underline{Q}_x + \mathbf{B}_{\mathrm{Q}} \, \underline{Q}_y \;=\; \underline{0} \;, \tag{3.24}$$

in which the flux Jacobians are symmetric matrices given by

$$\mathbf{A}_{\mathrm{Q}} \;=\; \frac{\partial \underline{Q}}{\partial \underline{U}} \, \mathbf{A}_{\mathrm{U}} \, \frac{\partial \underline{U}}{\partial \underline{Q}} \qquad \text{and} \qquad \mathbf{B}_{\mathrm{Q}} \;=\; \frac{\partial \underline{Q}}{\partial \underline{U}} \, \mathbf{B}_{\mathrm{U}} \, \frac{\partial \underline{U}}{\partial \underline{Q}} \;. \tag{3.25}$$

The system (3.24) is simplified even further when it is written in terms of the streamwise coordinates, ξ and η, for which

$$\frac{\partial}{\partial \xi} \;=\; \frac{u}{q} \frac{\partial}{\partial x} + \frac{v}{q} \frac{\partial}{\partial y} \qquad \text{and} \qquad \frac{\partial}{\partial \eta} \;=\; -\frac{v}{q} \frac{\partial}{\partial x} + \frac{u}{q} \frac{\partial}{\partial y} \;, \tag{3.26}$$

leading to the equations

$$\underline{Q}_t + \mathbf{A}_{\mathrm{Q}}^{\mathrm{S}} \, \underline{Q}_\xi + \mathbf{B}_{\mathrm{Q}}^{\mathrm{S}} \, \underline{Q}_\eta \;=\; \underline{0} \;, \tag{3.27}$$

where

$$\mathbf{A}_{\mathrm{Q}}^{\mathrm{S}} \;=\; \frac{u \mathbf{A}_{\mathrm{Q}} + v \mathbf{B}_{\mathrm{Q}}}{q} \qquad \text{and} \qquad \mathbf{B}_{\mathrm{Q}}^{\mathrm{S}} \;=\; \frac{-v \mathbf{A}_{\mathrm{Q}} + u \mathbf{B}_{\mathrm{Q}}}{q} \;. \tag{3.28}$$

These equations are preconditioned by an appropriate matrix $\mathbf{P}$, giving

$$\underline{Q}_t + \mathbf{P}\left(\mathbf{A}^{\mathrm{S}}_{\mathrm{Q}}\,\underline{Q}_\xi + \mathbf{B}^{\mathrm{S}}_{\mathrm{Q}}\,\underline{Q}_\eta\right) \;=\; \underline{0}\,, \tag{3.29}$$

and this system is transformed into 'characteristic' variables,

$$\underline{W}_t + \mathbf{A}^{\mathrm{S}}_{\mathrm{W}}\,\underline{W}_\xi + \mathbf{B}^{\mathrm{S}}_{\mathrm{W}}\,\underline{W}_\eta \;=\; \underline{0}\,, \tag{3.30}$$

where

$$\mathbf{A}^{\mathrm{S}}_{\mathrm{W}} \;=\; \frac{\partial \underline{W}}{\partial \underline{Q}}\,\mathbf{P}\,\mathbf{A}^{\mathrm{S}}_{\mathrm{Q}}\,\frac{\partial \underline{Q}}{\partial \underline{W}} \qquad \text{and} \qquad \mathbf{B}^{\mathrm{S}}_{\mathrm{W}} \;=\; \frac{\partial \underline{W}}{\partial \underline{Q}}\,\mathbf{P}\,\mathbf{B}^{\mathrm{S}}_{\mathrm{Q}}\,\frac{\partial \underline{Q}}{\partial \underline{W}}\,. \tag{3.31}$$

For an appropriate choice of $\mathbf{P}$ the system (3.30) is either fully or partially diagonalised depending on whether the flow is supercritical or subcritical.

The original system (3.9) is recovered by reversing the transformation carried out above. This leads back to the conservative flux balance,

$$\underline{\Phi} \;=\; \frac{\partial \underline{U}}{\partial \underline{Q}}\,\mathbf{P}^{-1}\frac{\partial \underline{Q}}{\partial \underline{W}}\,\underline{\Phi}_W \;=\; \sum_{k=1}^{N_w} \phi^k\,\underline{r}^k\,, \tag{3.32}$$

where $\underline{\Phi}_W$ is the flux balance of (3.30).

The preconditioner of Mesaros and Roe [16] (which in turn was based on that of van Leer, Lee and Roe [28] developed originally for the purpose of convergence acceleration) can be adapted straightforwardly to a form which can be used with the shallow water equations [11], giving

$$\mathbf{P} \;=\; \frac{1}{q}\begin{pmatrix} \frac{\varepsilon F^2}{\beta\kappa} & -\frac{\varepsilon F}{\beta\kappa} & 0 \\ -\frac{\varepsilon F}{\beta\kappa} & \frac{\varepsilon}{\beta\kappa}+\varepsilon & 0 \\ 0 & 0 & \frac{\beta}{\kappa} \end{pmatrix}, \tag{3.33}$$

where $F = q/\sqrt{gd}$ is the local Froude number of the flow,

$$\beta \;=\; \sqrt{|F^2-1|}\quad, \qquad \kappa \;=\; \max(F,1)\,, \tag{3.34}$$

and $\varepsilon = \varepsilon(F)$ is a function which satisfies $\varepsilon(0) = \frac{1}{2}$ and $\varepsilon(F) = 1$ for $F \geq 1$ (giving the correct behaviour in the preconditioned system at stagnation and continuity of the optimal decomposition through the critical point), and is taken here to be

$$\varepsilon(F) \;=\; \begin{cases} -F^3 + \frac{3}{2}F^2 + \frac{1}{2} & \text{for} \quad 0 \;\leq\; F \;\leq\; 1 \\ 1 & \text{for} \quad F \;>\; 1\,. \end{cases} \tag{3.35}$$

Two different forms for the characteristic variables $\underline{W}$ are taken, depending on whether the flow is supercritical or subcritical. These are given by, respectively,

$$\partial\underline{W}_{\mathrm{sup}} \;=\; \begin{pmatrix} \sqrt{\frac{g}{d}}\beta\,\partial d + Fq\,\partial\theta \\ \sqrt{\frac{g}{d}}\beta\,\partial d - Fq\,\partial\theta \\ \sqrt{\frac{g}{d}}\,\partial d + F\,\partial q \end{pmatrix} \qquad \text{and} \qquad \partial\underline{W}_{\mathrm{sub}} \;=\; \begin{pmatrix} \frac{\beta g}{q}\,\partial d \\ q\,\partial\theta \\ \sqrt{\frac{g}{d}}\,\partial d + F\,\partial q \end{pmatrix}, \tag{3.36}$$

and this choice gives a continuous representation through the critical point [16]. Note though that the decomposition is singular when $F = 0$, at stagnation points. In supercritical flow this leads to characteristic Jacobians in (3.30) of the form

$$\mathbf{A}_{\mathrm{W}}^{\mathrm{S}} \;=\; \begin{pmatrix} \frac{\beta}{\kappa} & 0 & 0 \\ 0 & \frac{\beta}{\kappa} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{B}_{\mathrm{W}}^{\mathrm{S}} \;=\; \begin{pmatrix} \frac{1}{\kappa} & 0 & 0 \\ 0 & \frac{1}{\kappa} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{3.37}$$

the system is fully decoupled and each component can be treated as a scalar advection equation. In subcritical flow (and at the critical point since the above decomposition is actually singular at $F = 1$, not a problem in practice),

$$\mathbf{A}_{\mathrm{W}}^{\mathrm{S}} \;=\; \begin{pmatrix} -\varepsilon\beta & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \varepsilon \end{pmatrix}, \qquad \mathbf{B}_{\mathrm{W}}^{\mathrm{S}} \;=\; \begin{pmatrix} 0 & \varepsilon & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.38}$$

which contains a coupled 2×2 subsystem.

In remains slightly unclear as to the best method of distribution for the elliptic subsystem. One possibility is to treat it as a pair of scalar equations with source terms [19], but it seems to be more accurate to use a system Lax-Wendroff scheme with 2×2 matrix coefficients

$$\boldsymbol{\alpha}_{\boldsymbol{i}}^{\boldsymbol{j}} = \frac{1}{3}\,\mathbf{I} + \frac{\Delta t}{4 S_{\triangle_j}}\,(\mathbf{A}, \mathbf{B})_j \cdot \vec{n}_i^{\,j}\,, \tag{3.39}$$

even though this necessarily introduces a discontinuity into the distribution at the critical point, even though the overall representation is continuous. It is the latter approach which will be used here so, rather than (3.22), the scheme actually takes the form

$$\underline{U}_i^{n+1} \;=\; \underline{U}_i^n + \frac{\Delta t}{S_i} \sum_{j \in \cup \triangle_i} \left((\underline{r}_j^1, \underline{r}_j^2)\, \boldsymbol{\alpha}_{\boldsymbol{i}}^{\boldsymbol{j}}\, \boldsymbol{\phi}_{\boldsymbol{j}} + (\alpha_i^j)^3 \phi_j^3\, \underline{r}_j^3 \right), \tag{3.40}$$

containing one scalar component and one subsystem, with corresponding vector fluctuation $\boldsymbol{\phi}_{\boldsymbol{j}}$.

3.5 System Distribution Schemes

A closely related scheme developed recently opts for a generalisation of the scalar schemes to systems of equations [27] rather than employing a wave model along with scalar distribution schemes. The discretisation therefore has the form

$$\underline{U}_i^{n+1} \;=\; \underline{U}_i^n - \frac{\Delta t}{S_i} \sum_{j \in \cup \triangle_i} \tilde{\boldsymbol{\alpha}}_i^j\, \underline{\tilde{\Phi}}_j \tag{3.41}$$

in which the distribution coefficients $\tilde{\boldsymbol{\alpha}}_i^j$ are now matrices.

The distribution coefficients can be constructed so that they satisfy generalised positivity and linearity preservation properties, but they should also be invariant under similarity

transformations, *i.e.* the contribution made from the conservative flux balance to the vertices of the cell is independent of the choice of variables for which the actual distribution is performed. Generalisations of all of the major scalar schemes have been constructed, although it is not yet clear which of the current versions of the system PSI scheme is the best, since the different possible ways of applying the limiter have different advantages, the most obvious being rather expensive.

3.6 Results

The test case presented here is that of flow through a symmetric constricted channel of length 4, whose breadth is given by

$$B(x) \;=\; \begin{cases} 1.0 - (1.0 - B_{\min})\cos^2(\pi(x-2.0)) & \text{for} \quad |x-2.0| \;\leq\; 0.5 \\ 1.0 & \text{otherwise}\,, \end{cases} \tag{3.42}$$

where $B_{\min} = 0.92$ is the minimum channel breadth and x is the distance into the channel (so the throat is positioned at the midpoint of the constriction, $x = 2.0$). The 2114 node, 4054 cell grid on which the numerical results have been obtained is shown in Figure 3.1 along with three steady state solutions distinguished by their freestream Froude numbers: (i) $F_\infty = 0.5$, completely subcritical and hence symmetric about the throat of the channel, (ii) $F_\infty = 0.71$, transcritical with a stationary hydraulic jump in the constriction downstream of the throat, and (iii) $F_\infty = 2.0$, completely supercritical, with a criss-cross pattern of undular jumps downstream of the throat. Simple characteristic boundary conditions are applied in each case. The solutions have been obtained using the hyperbolic/elliptic decomposition described in Section 3.4, applying the PSI scheme to each of the decoupled scalar components and the system Lax-Wendroff scheme to the subcritical elliptic subsystem. The results show that the scheme can accurately model each of these different types of flow.

4 Grid Adaptation

4.1 A Simple Node Movement Algorithm

The two-dimensional grid adaptation algorithm presented here is a very simple form of node movement. It takes the form of an iteration where, at each step, nodes are moved to a weighted average of the positions of the centroids of the neighbouring triangles [8, 10]. The new nodal position can thus be written as

$$\vec{x}_i^{new} \;=\; \frac{\sum_{j\in\cup\triangle_i} w_j \vec{x}_j}{\sum_{j\in\cup\triangle_i} w_j}\,, \tag{4.1}$$

where the $\vec{x}_j$ are the positions of the centroids, w_j are the cell weights and j indicates the cells adjacent to node i.

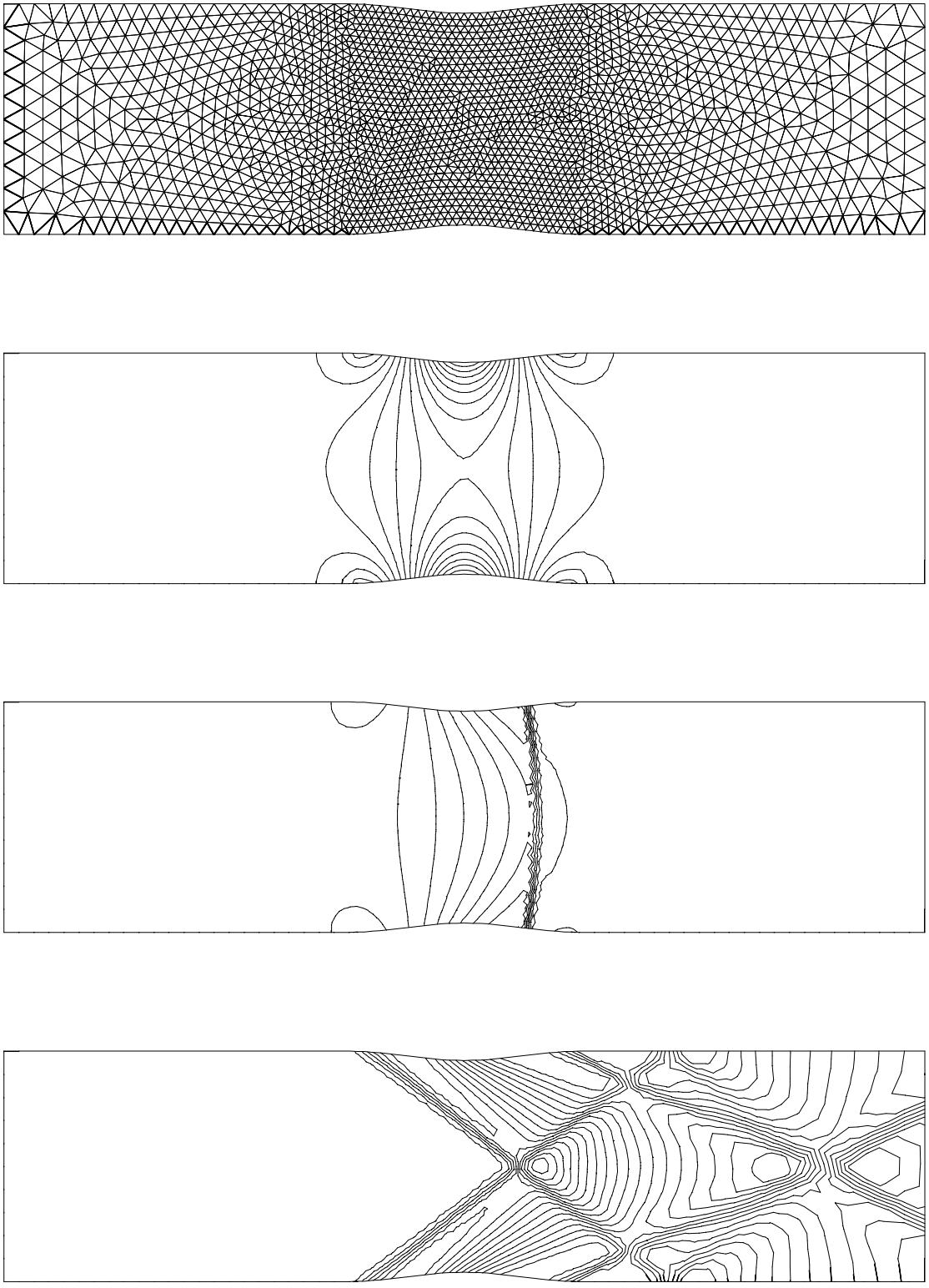

Figure 3.1: The grid and contours of depth for solutions of the subcritical (top), transcritical (middle) and supercritical (bottom) symmetric constricted channel test cases.

The weights here are chosen to depend on local approximations to the first and second derivatives of the solution (u for the scalar advection equation), so

$$w \;=\; \left(1 + \beta_1 |\vec{\nabla} u|^2 + \beta_2 (\vec{\nabla}^2 u)^2\right)^{1/2} \;, \tag{4.2}$$

where β_1 and β_2 are arbitrary parameters and $S_\triangle$ is the area of the triangle (u is replaced by d, the flow depth, in the case of the shallow water equations). The choice of $\beta_1 = 1$ and $\beta_2 = 0$ gives a simple generalisation of the weights which lead to arc length equidistribution in one dimension. Although there is no corresponding equidistribution property in two dimensions, the algorithm will still tend to move nodes towards regions where the weights are high. In the above case this means regions of high first and/or second derivatives, such as those found at shocks, but the weights can be modified so that nodes are attracted towards any detectable feature of the flow. The algorithm can also be easily generalised to three dimensions.

In one dimension mesh tangling can be avoided by ensuring that the chosen weights are always positive. In higher dimensions though, particularly on the highly distorted grids which become common once the mesh is allowed to move, tangling occurs quite readily. Even with positive weights in (4.1), it is possible for a node at the vertex of a triangle to be overtaken by the opposite edge of that triangle, thus causing the cell to 'flip' and acquire a negative area. This can be avoided by artificially limiting the distance which a node can move [10]. A simple but rather restrictive limit is

$$(\Delta x_i)_{\max} \;=\; \min_{j \in \cup \triangle_i} \left(\frac{S_j}{\max_{k=1,3} l_{jk}} \right) \;, \tag{4.3}$$

where S_j is the area of cell j and l_{jk} is the length of edge k of cell j. This expression is equivalent to half the smallest height of the surrounding triangles. A second restriction is also imposed which places a lower limit on the radius of the inscribed circle of each cell. This avoids extremely distorted meshes and the possibility of a prohibitively small limit on the time-step. Using this strategy, a displacement can be found for all nodes, including boundary nodes which must then be projected back on to the nearest point on the boundary, and 'corner' nodes, even though they are then forced to remain fixed.

Once all the displacements have been found, the nodal positions are updated in a block. The solution is then obtained on the new grid using linear interpolation of the solution on the previous grid (although the solution need not be updated at all - both choices are non-conservative but this is subsequently rectified).

4.2 Solution strategy

The method by which node movement is combined with multidimensional upwinding to obtain steady state solutions to the two-dimensional conservation laws can be expressed in three stages:

1) run the time-stepping algorithm on an initial, fixed grid until the solution appears steady (but long before convergence is achieved).

2) run the time-stepping interspersed with the grid movement until the grid has adapted to the steady solution. In this work, each time-step is alternated with a single node movement iteration.

3) fix the grid and run the time-stepping algorithm to convergence using the solution from step 2) as initial conditions.

The grid movement in step 2) can be initiated when the L_2 norm of the residual over the grid drops below a certain level (typically a drop of 2 or 3 orders of magnitude from the initial residual), effectively when the flow has stopped changing.

It is unlikely that the combination of time-stepping and grid movement will lead to a converged solution if allowed to run indefinitely, and it would be impractical to attempt this. Also, this stage of the method is not, as it stands, conservative due to the simple interpolation step of the grid movement. However, since steady state solutions are sought, the grid can be frozen after a fixed number of time-steps (100 for scalar advection, 500 for the shallow water equations) after which the solution strategy returns solely to the conservative time-stepping scheme. Local time-stepping has been used throughout to accelerate convergence, particularly useful on the more distorted meshes.

4.3 Results

The grid adaptation algorithm described above has been combined with the PSI scheme and applied to the two circular advection test cases of Section 2.4. For the square wave profile the adaptation parameters have been taken to be $\beta_1 = 1.0$ and $\beta_2 = 0.01$, while for the triangular profile, for which the curvature of the solution profile has more importance, $\beta_1 = 0.1$ and $\beta_2 = 1.0$. Figure 5.1 shows that in both cases the solution is improved enormously by the application of this simple grid adaptation procedure, as long as the adaptation parameters are chosen with care.

The adaptation has also been applied to the shallow water equations and is illustrated here by supercritical flow over a wedge, which induces an oblique hydraulic jump at its foot. In the case used here, the slope of the wedge is given by $\theta = 8.95^\circ$ and the inflow conditions are given by $d = 1.0\text{m}$, $u = 8.57\text{ms}^{-1}$ and $v = 0.0\text{ms}^{-1}$. The unadapted and adapted grids (taking $\beta_1 = 1.0$ and $\beta_2 = 0.001$) and their respective solutions are depicted in Figure 5.2, showing how much the capturing of the hydraulic jump can be improved by the movement of the nodes. More surprisingly, it has also been noticed that convergence to the steady state is generally faster on the adapted grids than on the more uniform starting grids. The hyperbolic/elliptic scheme described in Section 3.4 has again been used to approximate the shallow water equations.

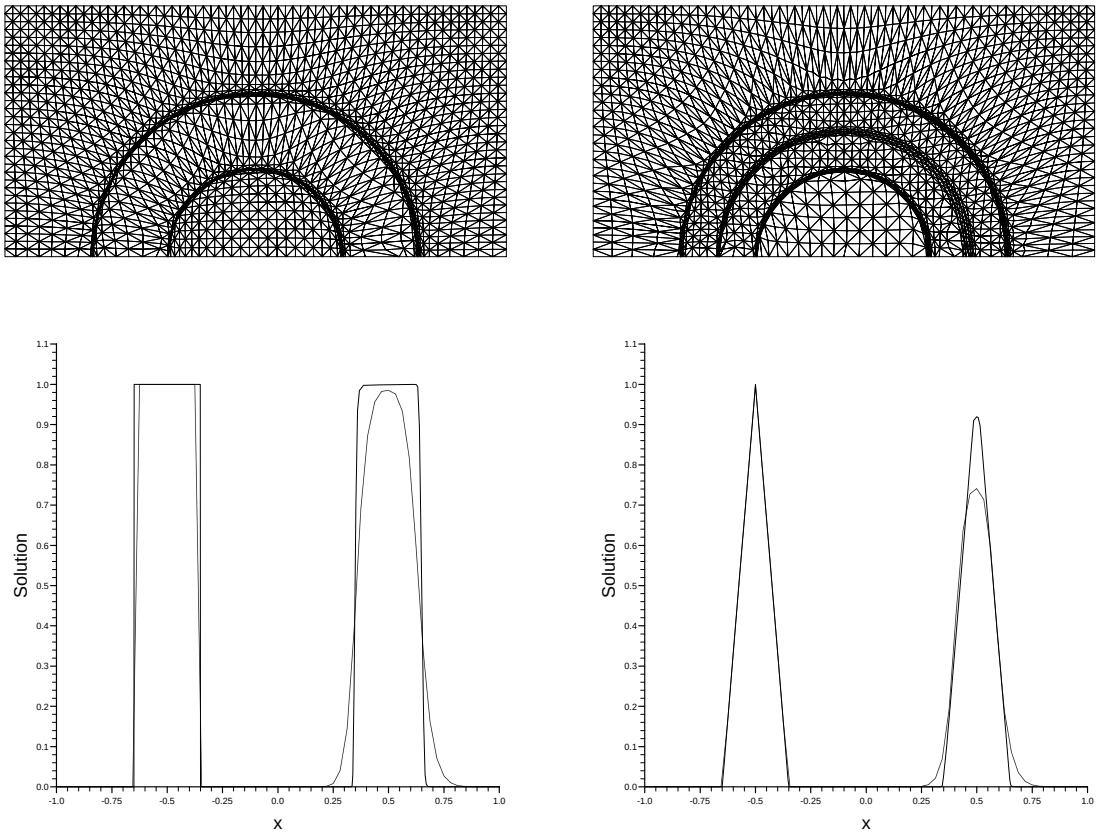

Figure 5.1: Adapted grids and solution on the boundary $y = 0$ for the advection of the square wave (left) and the triangular wave (right). Solid lines indicate the adapted solution, dotted lines the unadapted solution.

5 The Current State

Some of the more well-established multidimensional upwind schemes have been presented here, along with a brief description of a few of the more recent advances. This family of methods has now achieved a degree of success which has allowed them to be applied in practical situations where two-dimensional steady state flows are being approximated (see, for example [14]), although up to now this has been predominantly in the field of aerodynamics. More recently, they have also been applied to problems in hydraulic engineering (the source terms which appear commonly in the modelling can be incorporated simply, but only at the expense of positivity). Note though that all of the applications presented here have been on triangular grids because, although the schemes can be extended to quadrilateral meshes [19], the linearisation procedure is less natural.

Furthermore, the methods have been shown to combine well with the standard techniques for improving accuracy and efficiency, such as implicit time-stepping [13] and grid adaptation through both refinement [2] and movement [10]. Viscous flow models have also been approximated [27] (the Navier-Stokes equations) using these methods but, although the scalar fluctuation distribution technique can be extended to the advection-diffusion equation (by treating the viscous terms as sources), a Galerkin finite element discretisation of the viscous

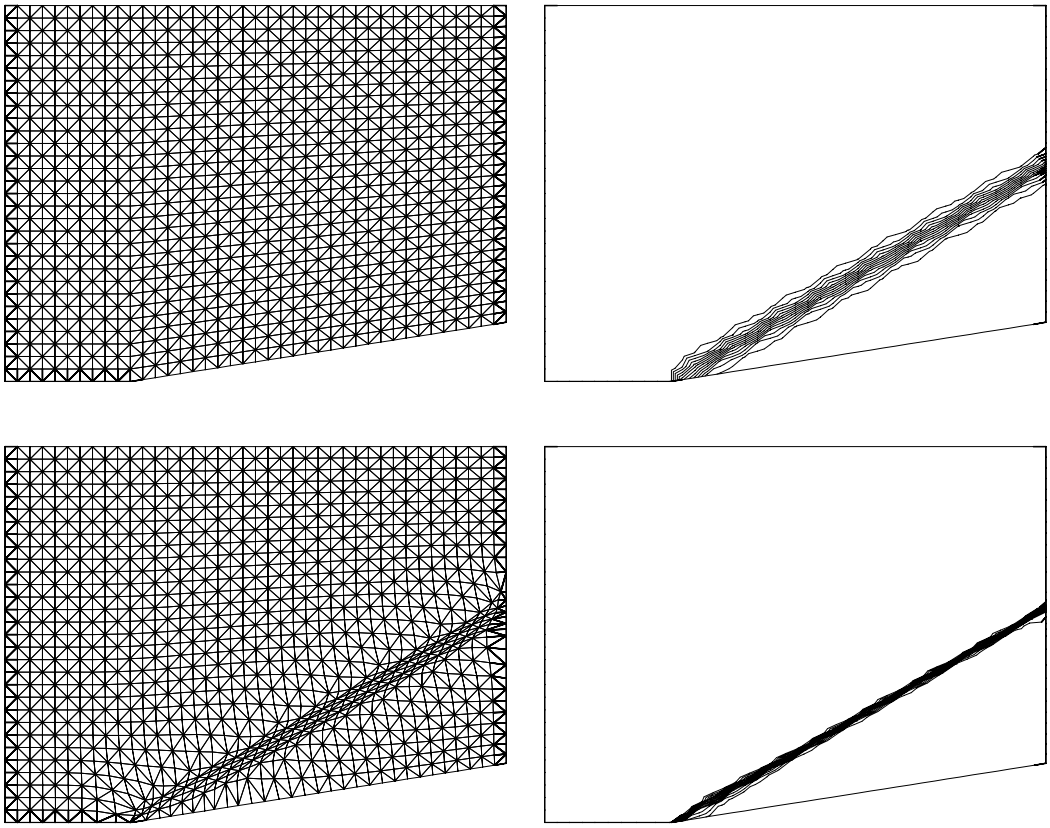

Figure 5.2: The grids and local Froude number contours for the oblique hydraulic jump test case with and without adaptation.

terms is often used [19, 27].

Even now, though, these schemes have their limitations, the most noticeable being that the most accurate of the existing two-dimensional wave models have a singularity at a stagnation point. This can be dealt with satisfactorily in steady state calculations by careful use of standard stabilisation techniques, but remains a problem for time-dependent flows. Because of this, the recent advances in accurate fluctuation distribution schemes for time-dependent problems [12, 15] cannot be taken full advantage of: there is still much work to be done to construct appropriate decompositions for unsteady flows, and this may require an alternative approach to those used so far.

The situation with three-dimensional calculations is less well developed (and of less relevance to shallow water modelling). The system decompositions [27] have been applied with some success and the fluctuation distribution schemes readily generalise to the three-dimensional scalar advection equation (on tetrahedral meshes). It is also fairly simple to construct wave models along similar lines to those described here, but none has yet been proposed which incorporates the additional features apparent in the underlying three-dimensional models, *e.g.* bicharacteristics. Of more relevance to this discussion, the methods can also be adapted to model flow over curved surfaces (flow on a sphere, with meteorological and oceanographical applications), but it is not yet completely clear how the two-dimensional wave models can be applied in such situations.

References

[1] P.Batten, C.Lambert and D.M.Causon, 'Positively conservative high-resolution convection schemes for unstructured elements', *Int. J. for Num. Methods in Engineering*, **39**:1821–1838, 1996.

[2] J.-C.Carette and H.Deconinck, 'Unstructured mesh adaptivity for SUPG formulations based on residual decomposition of the Euler equations', in *VKI LS 1995-02, Computational Fluid Dynamics*, 1995.

[3] H.Deconinck, 'Upwind methods and multidimensional splittings for the Euler equations', in *VKI LS 1991-01, Computational Fluid Dynamics*, 1991.

[4] H.Deconinck, C.Hirsch and J.Peuteman, 'Characteristic decomposition methods for the multidimensional Euler equations', in *Lecture Notes in Physics*, **264**:216–221, 1986.

[5] H.Deconinck and B.Koren (editors), *Euler and Navier-Stokes solvers using multidimensional upwind schemes and multigrid acceleration*, Notes on Numerical Fluid Mechanics, volume 57, Vieweg, 1997.

[6] H.Deconinck, P.L.Roe and R.Struijs, 'A multi-dimensional generalization of Roe's flux difference splitter for the Euler equations', *Computers and Fluids*, **22**:215–222, 1993.

[7] H.Deconinck, R.Struijs, G.Bourgois and P.L.Roe, 'High resolution shock capturing cell vertex advection schemes for unstructured grids', in *VKI LS 1994-05, Computational Fluid Dynamics*, 1994.

[8] G.Erlebacher and P.R.Eiseman, 'Adaptive triangular mesh generation', *AIAA Journal*, **25**:1356–1364, 1987.

[9] P.Glaister, 'Difference schemes for the shallow water equations', Numerical Analysis Report 9/87, Department of Mathematics, University of Reading, 1987.

[10] M.E.Hubbard, 'Multidimensional upwinding and grid adaptation for conservation laws', PhD Thesis, The University of Reading, 1996.

[11] M.E.Hubbard and M.J.Baines, 'Conservative multidimensional upwinding for the steady two dimensional shallow water equations', *J. Comput. Phys.*, **138**:419–448, 1997.

[12] M.E.Hubbard and P.L.Roe, 'Compact high-resolution algorithms for time-dependent advection on unstructured grids', to appear, *Int. J. for Num. Methods in Fluids*, 1999.

[13] E.Issman, G.Degrez and H.Deconinck, 'Implicit iterative methods for a multidimensional upwind Euler/Navier-Stokes solver on unstructured meshes', AIAA Paper 95-1653, 1995.

[14] P.Leyland and B.Khobalatte, 'The kinetic approach to multidimensional upwinding schemes', to appear, *7th Int. Conf. on Hyperbolic Problems*, Zurich, 1998.

[15] J.März, 'Improving time accuracy for residual distribution schemes', VKI PR 1996-17, von Karman Institute for Fluid Dynamics, 1996.

[16] L.M.Mesaros and P.L.Roe, 'Multidimensional fluctuation splitting schemes based on decomposition methods', AIAA Paper 95-1699, 1995.

[17] H.Paillère, 'Multidimensional upwind residual distribution schemes for the Euler and Navier-Stokes equations on unstructured grids', PhD Thesis, Université Libre de Bruxelles, 1995.

[18] H.Paillère, 'It is possible to solve dam-break problems using a multidimensional upwinding approach', Thèse annexe, Université Libre de Bruxelles, 1995.

[19] H.Paillère, E.van der Weide and H.Deconinck, 'Multidimensional upwind methods for inviscid and viscous compressible flows', in *VKI LS 1995-02, Computational Fluid Dynamics*, 1995.

[20] P.L.Roe, 'Approximate Riemann Solvers, parameter vectors, and difference schemes', *J. Comput. Phys.*, **43(2)**:357–372, 1981.

[21] P.L.Roe, 'Fluctuations and signals - A framework for numerical evolution problems', in *Numerical Methods for Fluid Dynamics*, Eds. K.W.Morton and M.J.Baines, Academic Press, 1982.

[22] P.L.Roe, 'Discrete models for the numerical analysis of time-dependent multidimensional gas dynamics', *J. Comput. Phys.*, **63**:458–476, 1986.

[23] P.L.Roe, 'Linear advection schemes on triangular meshes', CoA Report 8720, Cranfield Institute of Technology, 1987.

[24] M.A.Rudgyard, 'Multidimensional wave decompositions for the Euler equations', in *VKI LS 1993-04, Computational Fluid Dynamics*, 1993.

[25] D.Sidilkover and P.L.Roe, 'Unification of some advection schemes in two dimensions', ICASE Report 95-10, 1995.

[26] P.K.Sweby, 'High resolution schemes using flux limiters for hyperbolic conservation laws', *SIAM J. Numer. Anal.*, **21**:995–1011, 1984.

[27] E.van der Weide, 'Compressible flow simulation on unstructured grids using multidimensional upwind schemes', PhD Thesis, Technische Universiteit Delft, 1998.

[28] B.van Leer, W.-T.Lee and P.L.Roe, 'Characteristic time-stepping or local preconditioning of the Euler equations', AIAA Paper 91-1552-CP, 1991.