Conservative Multidimensional Upwinding for the Shallow Water Equations [1]

M.E.Hubbard, P.Garcia-Navarro [2] and M.J.Baines

Numerical Analysis Report 4/95

Department of Mathematics
P.O.Box 220
University of Reading
Whiteknights
Reading
RG6 6AX
United Kingdom

[1]The first author was funded by DRA Farnborough and the second by the EC Program of Human Capital and Mobility.

[2]Department of Materials and Fluids, University of Zaragoza, Spain.

Abstract

Multidimensional upwinding techniques [**?**, **?**] have been developed with the object of the solution of the Euler equations. However, they can equally well be used to solve other hyperbolic systems of equations. Recently, the method has been adapted for the solution of the shallow water equations [**?**], but due to the subtly different nature of these equations the linearisation of the system used implied that the scheme was not quite conservative.

This report describes a method by which the shallow water equations can be linearised in a truly conservative manner, enabling the use of wave models and fluctuation distribution schemes to give a conservative multidimensional upwinding scheme for the shallow water equations.

Contents

1 Introduction 1

2 A Conservative Linearisation 2

3 Results 7

4 Conclusions 8

1 Introduction

This report is intended as a supplement to reference [?] in which the multidimensional upwinding techniques developed by Roe, Deconinck and Rudgyard amongst others [?, ?, ?], mainly within the context of the Euler equations, were extended to solve the two-dimensional shallow water equations on unstructured triangular grids. In that report two of the wave models developed for the Euler equations, namely Roe's model D and Rudgyard's Mach angle splitting [?], were adapted to decompose the shallow water equations into scalar components (in effect by removing the entropy wave from the decompositions for the Euler equations). The scalar multidimensional PSI scheme [?] was then used to distribute the resulting scalar fluctuations over the grid.

The generalisation was deliberately kept as simple as possible. This had little effect on the derivation of the new wave models and none on the scalar distribution schemes but, due to subtle differences between the systems of equations, the analysis for the creation of a conservative linearisation [?] was no longer valid, and the method therefore lost its conservation property.

In the next section, a simple method is described by which the multidimensional upwinding methods described in [?, ?] can be modified to make them conservative. It essentially follows the construction of a conservative linearisation, carried out in [?] for the Euler equations, but differs in that a source term is introduced in order to ensure that the integration is carried out exactly throughout the analysis and the 'telescopic' property of the fluxes, required for conservation, is satisfied. Brief results, for a test case where the exact solution is known, are then shown using a simple implementation of the conservative algorithm.

2 A Conservative Linearisation

The shallow water equations without friction, like the Euler equations, constitute a nonlinear hyperbolic system of equations and can be written in the form

$$\underline{\mathbf{u}}_t + \underline{\mathbf{F}}_x + \underline{\mathbf{G}}_y = \underline{\mathbf{0}}, \tag{2.1}$$

where

$$\underline{\mathbf{u}} = \begin{pmatrix} h \\ uh \\ vh \end{pmatrix}, \qquad \underline{\mathbf{F}} = \begin{pmatrix} uh \\ u^2h + \frac{gh^2}{2} \\ uvh \end{pmatrix}, \qquad \underline{\mathbf{G}} = \begin{pmatrix} vh \\ uvh \\ v^2h + \frac{gh^2}{2} \end{pmatrix}. \tag{2.2}$$

Here h is the depth of the water, u and v are the x- and y-velocities respectively and g is the acceleration due to gravity.

The object of this analysis is to write this hyperbolic system of equations in the linearised form

$$\underline{\mathbf{u}}_t + \tilde{A}\underline{\mathbf{u}}_x + \tilde{B}\underline{\mathbf{u}}_y = \underline{\mathbf{0}}, \tag{2.3}$$

where $\tilde{A}$ and $\tilde{B}$ are approximations to the flux Jacobians

$$A = \frac{\partial \underline{\mathbf{F}}}{\partial \underline{\mathbf{u}}} = \begin{pmatrix} 0 & 1 & 0 \\ -u^2 + gh & 2u & 0 \\ -uv & v & u \end{pmatrix}, \qquad B = \frac{\partial \underline{\mathbf{G}}}{\partial \underline{\mathbf{u}}} = \begin{pmatrix} 0 & 0 & 1 \\ -uv & v & u \\ -v^2 + gh & 0 & 2v \end{pmatrix}, \tag{2.4}$$

in such a way that the resulting fluctuation can be calculated, decomposed and distributed in a conservative manner. That is to say, the approximation should be consistent and the sum of the fluctuations over the whole domain should reduce to contributions from the boundary only, the 'telescopic' property.

For the Euler equations, Roe *et al* [**?**] insisted that the approximate matrices, which are functions of the three sets of variables $\underline{\mathbf{u}}_1$, $\underline{\mathbf{u}}_2$, $\underline{\mathbf{u}}_3$ (the conservative variables evaluated at the vertices of the cell), satisfy the following three criteria, a two-dimensional version of Roe's Property U [**?**]:

1. The linearisation is consistent, in the sense that

$$\tilde{A}(\underline{\mathbf{u}}, \underline{\mathbf{u}}, \underline{\mathbf{u}}) = A(\underline{\mathbf{u}}), \qquad \tilde{B}(\underline{\mathbf{u}}, \underline{\mathbf{u}}, \underline{\mathbf{u}}) = B(\underline{\mathbf{u}}). \tag{2.5}$$

2. For all angles θ, the matrix $\tilde{A}(\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \underline{\mathbf{u}}_3)\cos\theta + \tilde{B}(\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \underline{\mathbf{u}}_3)\sin\theta$ has real eigenvalues and a complete set of linearly independent eigenvectors.

3. The identities

$$\widehat{\underline{\mathbf{F}}_x} \equiv \tilde{A}(\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \underline{\mathbf{u}}_3)\,\widehat{\underline{\mathbf{u}}_x}, \qquad \widehat{\underline{\mathbf{G}}_y} \equiv \tilde{B}(\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \underline{\mathbf{u}}_3)\,\widehat{\underline{\mathbf{u}}_y}, \tag{2.6}$$

are satisfied for any $\underline{\mathbf{u}}_1$, $\underline{\mathbf{u}}_2$, $\underline{\mathbf{u}}_3$. $\widehat{\underline{\mathbf{F}}_x}$ and $\widehat{\underline{\mathbf{G}}_y}$ are the approximations to the flux derivatives resulting from the linearisation.

However, it is also implicit in the paper that all integrations carried out to evaluate these averages have to be done exactly, otherwise the 'telescopic' property required for conservation is not satisfied precisely. This fourth criterion can be simply expressed by requiring that

$$\sum_{\triangle} V_{\triangle}(\widehat{\underline{\mathbf{F}}_x} + \widehat{\underline{\mathbf{G}}_y}) = \oint_{\partial\Omega} (\underline{\mathbf{F}}, \underline{\mathbf{G}}) \cdot d\vec{n}, \tag{2.7}$$

where $V_{\triangle}$ is the area of the cell, $\partial\Omega$ is the outer boundary of the domain and $\vec{n}$ is the inward pointing normal to the boundary. It was never necessary to impose this condition explicitly in the treatment of the Euler equations because it is automatically satisfied by the approximation, but the condition becomes important when the procedure for creating a conservative linearisation is followed for the shallow water equations. In [**?**] an approximation was used which satisfied the three criteria specified above but used inexact integration and so was not conservative.

As with the Euler equations, assuming linearity of the conservative variables would involve the evaluation of unnecessarily complicated integrals before the linearised Jacobians could be constructed. Furthermore, the derivations of the wave models are based on the matrix $A(\underline{\mathbf{u}})\cos\theta + B(\underline{\mathbf{u}})\sin\theta$. This linearisation does not lead to an approximation to this matrix of the form $A(\overline{\underline{\mathbf{u}}})\cos\theta + B(\overline{\underline{\mathbf{u}}})\sin\theta$, where $\overline{\underline{\mathbf{u}}}$ is some average of the conserved variables, so it is not immediately obvious how the wave decomposition could be applied under these circumstances. Therefore, as in the case of the Euler equations, it is assumed that the parameter vector,

$$\underline{\mathbf{w}} = h^{\frac{1}{2}}(1, u, v)^T, \tag{2.8}$$

is the quantity that varies linearly. However, the consequences of this assumption differ slightly from those for the Euler equations because some of the components of $\underline{\mathbf{u}}$, $\underline{\mathbf{F}}$ and $\underline{\mathbf{G}}$ are polynomials in the components of $\underline{\mathbf{w}}$ of order higher than two (although they are far simpler than those resulting from assuming linearity of the conservative variables).

Multidimensional upwinding methods on triangular grids in two dimensions rely on the evaluation within each triangular cell of the so-called fluctuation

$$\begin{aligned} \Phi &= \int\int_{\triangle} \underline{\mathbf{u}}_t \, dx\, dy \\ &= -\int\int_{\triangle} (\underline{\mathbf{F}}_x + \underline{\mathbf{G}}_y) \, dx\, dy \\ &= -\int\int_{\triangle} (\tilde{A}\underline{\mathbf{u}}_x + \tilde{B}\underline{\mathbf{u}}_y) \, dx\, dy, \end{aligned} \tag{2.9}$$

its decomposition into scalar components and their distribution to the vertices of that cell. The rest of this section is concerned with the calculation of this fluctuation and follows closely that of Roe *et al* [**?**].

Define now the Jacobian matrices,

$$Q = \frac{\partial \underline{\mathbf{u}}}{\partial \underline{\mathbf{w}}} = \begin{pmatrix} 2w_1 & 0 & 0 \\ w_2 & w_1 & 0 \\ w_3 & 0 & w_1 \end{pmatrix}, \tag{2.10}$$

and

$$R = \frac{\partial \underline{\mathbf{F}}}{\partial \underline{\mathbf{w}}} = \begin{pmatrix} w_2 & w_1 & 0 \\ 2g{w_1}^3 & 2w_2 & 0 \\ 0 & w_3 & w_2 \end{pmatrix}, \quad S = \frac{\partial \underline{\mathbf{G}}}{\partial \underline{\mathbf{w}}} = \begin{pmatrix} w_3 & 0 & w_1 \\ 0 & w_3 & w_2 \\ 2g{w_1}^3 & 0 & 2w_3 \end{pmatrix}, \tag{2.11}$$

in terms of w_i, the components of the parameter vector (2.8). Note here that all the components of these matrices are linear functions of $\underline{\mathbf{w}}$ except for a single cubic term which appears in both R and S.

Thus the fluctuation can be written

$$\Phi = -\int\int_{\triangle} (R\underline{\mathbf{w}}_x + S\underline{\mathbf{w}}_y)\, dx\, dy \tag{2.12}$$

$$= -\left(\int\int_{\triangle} R\, dx\, dy\right) \underline{\mathbf{w}}_x - \left(\int\int_{\triangle} S\, dx\, dy\right) \underline{\mathbf{w}}_y, \tag{2.13}$$

since the gradients of the parameter vector variables are constant within each cell.

At this point the theory deviates slightly from that associated with the Euler equations because the assumption of linear variation of $\underline{\mathbf{w}}$ no longer implies that

$$\int\int_{\triangle} R\, dx\, dy = V_{\triangle} R(\overline{\underline{\mathbf{w}}}), \qquad \int\int_{\triangle} S\, dx\, dy = V_{\triangle} S(\overline{\underline{\mathbf{w}}}), \tag{2.14}$$

where $V_{\triangle}$ is the area of the cell and $\overline{\underline{\mathbf{w}}}$ is the value of the parameter vector at the centroid of the triangle.

However, it is still possible to do the integration exactly, simply by using a higher order quadrature for the two cubic terms. Note that this only involves the evaluation of one integral and so is not prohibitively expensive.

Now, defining

$$\tilde{R} = \frac{1}{V_{\triangle}} \int\int_{\triangle} R\, dx\, dy, \qquad \tilde{S} = \frac{1}{V_{\triangle}} \int\int_{\triangle} S\, dx\, dy, \tag{2.15}$$

leads to

$$\Phi = -V_{\triangle}(\tilde{R}\underline{\mathbf{w}}_x + \tilde{S}\underline{\mathbf{w}}_y)$$

$$= -V_{\triangle}(\widehat{\underline{\mathbf{F}}_x} + \widehat{\underline{\mathbf{G}}_y}). \tag{2.16}$$

In fact, $\tilde{R}$ and $\tilde{S}$ only differ from $R(\overline{\underline{\mathbf{w}}})$ and $S(\overline{\underline{\mathbf{w}}})$ in one component each.

All the elements of the matrix Q are linear in $\underline{\mathbf{w}}$ so a similar argument to the one used with the Euler equations gives

$$\widehat{\underline{\mathbf{u}}_x} = Q(\overline{\underline{\mathbf{w}}})\underline{\mathbf{w}}_x, \qquad \widehat{\underline{\mathbf{u}}_y} = Q(\overline{\underline{\mathbf{w}}})\underline{\mathbf{w}}_y. \tag{2.17}$$

Since Q is invertible the fluctuation can now be written

$$\Phi = -V_\triangle(\tilde{R}Q^{-1}(\overline{\underline{\mathbf{w}}})\widehat{\underline{\mathbf{u}}_x} + \tilde{S}Q^{-1}(\overline{\underline{\mathbf{w}}})\widehat{\underline{\mathbf{u}}_y}) \tag{2.18}$$

which gives expressions for the Jacobian matrices

$$\tilde{A} = \tilde{R}Q^{-1}(\overline{\underline{\mathbf{w}}}), \qquad \tilde{B} = \tilde{S}Q^{-1}(\overline{\underline{\mathbf{w}}}). \tag{2.19}$$

These approximations to A and B are consistent and satisfy the 'telescopic' property when substituted back into the expression for the fluctuation (2.9). Unfortunately, since the matrices are not in the form

$$\tilde{A} = A(\overline{\underline{\mathbf{w}}}), \qquad \tilde{B} = B(\overline{\underline{\mathbf{w}}}), \tag{2.20}$$

it is not immediately possible to decompose the fluctuation into scalar components using any of the wave models suggested in [**?**].

To overcome this problem, write

$$R = R(\overline{\underline{\mathbf{w}}}) + \xi_R, \qquad S = S(\overline{\underline{\mathbf{w}}}) + \xi_S, \tag{2.21}$$

where ξ_R and ξ_S are matrix residuals. Then

$$\tilde{A} = A(\overline{\underline{\mathbf{w}}}) + \xi_R Q^{-1}(\overline{\underline{\mathbf{w}}}), \qquad \tilde{B} = B(\overline{\underline{\mathbf{w}}}) + \xi_S Q^{-1}(\overline{\underline{\mathbf{w}}}), \tag{2.22}$$

and the fluctuation can be written

$$\Phi = -V_\triangle(A(\overline{\underline{\mathbf{w}}})\widehat{\underline{\mathbf{u}}_x} + B(\overline{\underline{\mathbf{w}}})\widehat{\underline{\mathbf{u}}_y}) - V_\triangle(\xi_R\underline{\mathbf{w}}_x + \xi_S\underline{\mathbf{w}}_y). \tag{2.23}$$

It has been split into two parts. The first part is analogous to the fluctuation resulting from the corresponding analysis for the Euler equations, and can be decomposed and distributed using a wave model. The second can be considered as a form of source term which must be treated separately. The exact form of this additional term is actually very simple since

$$\xi_R = \begin{pmatrix} 0 & 0 & 0 \\ 2g\zeta & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \xi_S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2g\zeta & 0 & 0 \end{pmatrix}. \tag{2.24}$$

where

$$\zeta = \frac{1}{V_\triangle}\int\!\!\int_\triangle w_1^3 \, dx\, dy - \overline{w_1}^3 \tag{2.25}$$

The scalar ζ represents the difference between integrating a cubic exactly (using say a four point quadrature rule) and integrating it approximately (using a one point quadrature rule). This term will usually be very small and could feasibly be ignored were it not for the fact that it is multiplied by a solution gradient in the expression for the fluctuation (2.23).

If the source term were completely ignored, the approximation would still satisfy the three criteria of the two-dimensional version of Property U, but because the integration is inexact the scheme is no longer conservative, even though the error is likely to be very small. This is effectively the simplification which was made when the concepts of multidimensional upwinding were originally transferred from the Euler equations to the shallow water equations in [**?**]. In fact, in [**?**] it was assumed that the primitive variables varied linearly. This assumption is equally valid and the preceding analysis can still be carried out, resulting in a different and slightly more complicated source term.

All that remains is the decomposition and distribution of the fluctuation which, with the aid of a wave model, can be written

$$\Phi = -\int\int_{\triangle} (\underline{\mathbf{F}}_x + \underline{\mathbf{G}}_y)\, dx\, dy = \sum_{k=1}^{N_e} \phi^k \overline{\underline{\mathbf{r}}}^k + \underline{\xi}, \tag{2.26}$$

where N_e is the number of effective waves, ϕ^k is the fluctuation of the k^{th} wave, $\overline{\underline{\mathbf{r}}}^k$ is the vector corresponding to the projection of ϕ^k on to the conservative variables and ξ is the 'source' term in (2.23). It must be remembered that average values of the primitive variables $\underline{\mathbf{q}}$ and their gradients, which are used to calculate all other cell-averaged quantities necessary for the decomposition [**?**], such as the conservative variables and the flux balance, must be computed exactly from the parameter vector variables, giving

$$\overline{\underline{\mathbf{q}}} = \begin{pmatrix} \overline{w_1}^2 \\ \overline{w_2}/\overline{w_1} \\ \overline{w_3}/\overline{w_1} \end{pmatrix}, \qquad \widehat{\vec{\nabla}\underline{\mathbf{q}}} = \begin{pmatrix} 2\overline{w_1}\vec{\nabla}w_1 \\ (\overline{w_1}\vec{\nabla}w_2 - \overline{w_2}\vec{\nabla}w_1)/\overline{w_1}^2 \\ (\overline{w_1}\vec{\nabla}w_3 - \overline{w_3}\vec{\nabla}w_1)/\overline{w_1}^2 \end{pmatrix}. \tag{2.27}$$

If linearity of the primitive variables is assumed, as it was in [**?**], then these quantities must be calculated directly.

The results in this report were obtained using the 'Froude angle splitting' wave model [**?**] and the PSI scalar distribution scheme [**?**], while the source term was distributed very simply, sending one third to each of the vertices of the triangular cell.

3 Results

Brief results are presented here for a single test case, that of an oblique hydraulic jump, induced by means of the interaction between a supercritical flow and a wall at an angle $\theta = 8.95°$ to the flow. The initial conditions (and the upstream conditions) were $h = 1m$, $u = 8.57ms^{-1}$ and $v = 0ms^{-1}$, *i.e.* a uniform supercritical flow with Froude number, $Fr = 2.74$. The boundary conditions are precisely the same as those used in [**?**].

The exact solution downstream of the jump can be calculated analytically for problems of this type [**?**]. In this case, the predicted values are $h = 1.5m$ and $|\vec{u}| = 7.9556ms^{-1}$ with $Fr = 2.075$, and the jump is at an angle of 30° to the boundary walls at inflow. Figure 3.1 shows the contours of h obtained using the conservative numerical scheme described in this report on a regular 2400 element triangular grid. The result is virtually indistinguishable from that obtained using the previous, non-conservative scheme [**?**]. The discontinuity is captured sharply and is at the correct angle, and the discontinuous water surface is devoid of oscillations. The solution obtained behind the jump agrees closely with the analytic prediction: $h = 1.4993m$ and $|\vec{u}| = 7.9527ms^{-1}$, giving $Fr = 2.075$, which is an improvement on the non-conservative scheme. Unfortunately the solution has not converged to machine accuracy, but this may well be due to the rather crude treatment of the source terms, and could be rectified by using a more sophisticated means of distribution.

Figure 3.1: Contours of water height h.

4 Conclusions

In this report, a method has been described by which the concepts of multidimensional upwinding can be used in the solution of the shallow water equations whilst retaining conservation - a property not satisfied by the first algorithms devised [?, ?]. The resulting scheme has been tested on one test case and gives good results in comparison with the exact analytic solution.

It is suggested that in future, the 'source' terms should be distributed in a more sophisticated manner than has been used here. This may well help to reduce the magnitude of the residuals when the numerical solutions have reached a steady state, improving the convergence properties of the scheme.