

UNIVERSITY OF READING
DEPARTMENT OF MATHEMATICS

NUMERICAL ALGORITHMS
FOR THE SOLUTION OF
CONSERVATION LAWS IN TWO
AND THREE DIMENSIONS

M.J. BAINES

Numerical Analysis Report 4/81

April 1981

SUMMARY

An improved conservative algorithm for the solution of multi-dimensional fluid flow problems is described, based on the ideas of P.L. Roe of the Royal Aircraft Establishment, Bedford. The main features are the weighted allocation of convected mass quantities to nodes and a logical switch to eliminate oscillations.

Contents

§ 1.	Introduction	1
§ 2.	Convection and increments for the one-dimensional method	2
§ 3.	Monotonicity-preservation in one dimension	7
§ 4.	The two-dimensional method	12
§ 5.	Monotonicity-preservation in two dimensions	18
§ 6.	Non-linear aspects in two dimensions	24
§ 7.	A test problem in two dimensions	27
§ 8.	The three-dimensional method	29
§ 9.	Monotonicity-preservation in three dimensions	32
§10.	Conclusion	36
	Acknowledgements	37
	References	38
	Appendix A : Less compact algorithms	39
	Appendix B : Generalised increments	41
	Appendix C : Results for the problem in §7	44

§1. Introduction

This report describes the continuation of work¹ arising from developments in solving the Euler equations for time dependent compressible fluid flows at RAE Bedford. Recently P.L. Roe of RAE Bedford proposed a novel approach to the construction of numerical algorithms for the approximate solution of these equations in one dimension and also devised a switching rule for suppressing unwanted oscillations near discontinuities. Roe's one dimensional algorithm has proved highly successful giving second order accuracy and excellent shock representation in a variety of problems.

The method has been applied in more than one space dimension using operator splitting techniques with mixed results. Here we generalise Roe's one dimensional ideas directly into two or three dimensions.

The one dimensional method has been discussed fully elsewhere^{1,2,3,4}. However, it is convenient to present here a derivation based on convection approximations which is used later when generalising to a higher number of dimensions.

§2. Convection and Increments for the One Dimensional Method

Consider the simple scalar, but possibly nonlinear, equation

$$u_t + f_x = 0 \quad (1)$$

or, equivalently,

$$u_t + f'(u)u_x = 0. \quad (2)$$

Divide the x -axis into equal intervals of length Δx and consider three adjacent nodes x_{k-1} , x_k , x_{k+1} (fig. 1).

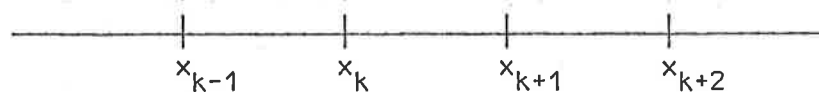


fig. 1

We approximate the physical mechanisation involved in the derivation of (1) or (2) by writing

$$m_{k+\frac{1}{2}} = \int_{x_k}^{x_{k+1}} u dx \quad (3)$$

and supposing that the "mass" $m_{k-\frac{1}{2}}$ in the adjacent cell (x_{k-1}, x_k) is convected into the cell (x_k, x_{k+1}) with speed $(f'(u))_{k+\frac{1}{2}}$, taken to be positive. The cell (x_k, x_{k+1}) then draws in a mass $(f'(u))_{k+\frac{1}{2}} \frac{\Delta t}{\Delta x} m_{k-\frac{1}{2}}$ which replaces $(f'(u))_{k+\frac{1}{2}} \frac{\Delta t}{\Delta x} m_{k+\frac{1}{2}}$ (assuming that each term in brackets lies between 0 and 1). The net mass increment is

$$\begin{aligned} \Delta m_{k+\frac{1}{2}} &= (f'(u))_{k+\frac{1}{2}} \frac{\Delta t}{\Delta x} m_{k-\frac{1}{2}} - (f'(u))_{k+\frac{1}{2}} \frac{\Delta t}{\Delta x} m_{k+\frac{1}{2}} \\ &= (f'(u))_{k+\frac{1}{2}} \frac{\Delta t}{\Delta x} (m_{k-\frac{1}{2}} - m_{k+\frac{1}{2}}). \end{aligned} \quad (4)$$

We approximate $(f'(u))_{k+\frac{1}{2}}$ by

$$\frac{f_{k+1} - f_k}{u_{k+1} - u_k} \quad (5)$$

giving

$$\Delta m_{k+\frac{1}{2}} = \frac{\Delta t}{\Delta x} \left(\frac{f_{k+1} - f_k}{u_{k+1} - u_k} \right) (m_{k-\frac{1}{2}} - m_{k+\frac{1}{2}}) . \quad (6)$$

To approximate $m_{k+\frac{1}{2}}$ return to equation (3). If we use a zero'th order downwind approximation

$$u_{k+1} \Delta x \quad (7)$$

for $m_{k+\frac{1}{2}}$, then (6) becomes

$$\Delta m_{k+\frac{1}{2}} = \Delta t \left(\frac{f_{k+1} - f_k}{u_{k+1} - u_k} \right) (u_k - u_{k+1}) = -\Delta t (f_{k+1} - f_k) . \quad (8)$$

This mass increment is transferred to nodal values of u at the next time step to complete the physical simulation and to ensure conservation. We do this by using the same downwind approximation (7) for $m_{k+\frac{1}{2}}$ at the forward time step and increment u_{k+1} by an amount

$$g_{k+\frac{1}{2}} = \frac{1}{\Delta x} \Delta m_{k+\frac{1}{2}} = -\frac{\Delta t}{\Delta x} (f_{k+1} - f_k) . \quad (9)$$

The expression (5), approximating $\frac{\partial f}{\partial u}$, was assumed to be positive in the foregoing. If (5) is negative so that mass is drawn in from the right in fig. 1, not only is the mass increment altered but so is the downwind approximation (7) for (3). If (7) is replaced by

$$u_k \Delta x \quad (10)$$

and the mass increment is drawn from the cell (x_{k+1}, x_{k+2}) , equation (4) becomes

$$\Delta m_{k+\frac{1}{2}} = -(f'(u))_{k+\frac{1}{2}} \frac{\Delta t}{\Delta x} m_{k+\frac{3}{2}} + (f(u))_{k+\frac{1}{2}} \frac{\Delta t}{\Delta x} m_{k+\frac{1}{2}} \quad (11)$$

(since $f'(u) < 0$), and equation (9) becomes

$$\begin{aligned}
 g_{k+\frac{1}{2}} &= \frac{1}{\Delta x} \Delta m_{k+\frac{1}{2}} = \frac{\Delta t}{\Delta x} \left(\frac{f_{k+1} - f_k}{u_{k+1} - u_k} \right) (-u_{k+1} + u_k) \\
 &= -\frac{\Delta t}{\Delta x} (f_{k+1} - f_k), \quad (12)
 \end{aligned}$$

as before. However, in line with (11), the increment will now be allocated to u_k at the downwind end of the interval. Hence nodal values can receive increments from intervals both to the left and to the right. It is evident from the form of (8) that the total mass is conserved in the sense that

$$\sum m = \text{constant} \quad (13)$$

when summed over all the intervals of the grid.

For the simple advection equation

$$u_t + au_x = 0 \quad (14)$$

(9) or (12) reduces to

$$g_{k+\frac{1}{2}} = a \frac{\Delta t}{\Delta x} (u_k - u_{k+1}), \quad (15)$$

to be assigned to u_{k+1} if $a > 0$ and to u_k if $a < 0$.

It is easy to verify that the scheme described here is exact for specimen functions $u = 1$, $u = x$, but not for $u = x^2$. The method is therefore of first order and it can be identified as a purely upwind scheme. The upwindedness discourages oscillations whilst conservation and the built-in directionality ensure good shock capturing capability⁴.

We wish now to construct a higher order scheme along these lines. For second order accuracy we demand the scheme to be exact also for the specimen function $u = x^2$. We obtain the necessary flexibility by assuming that the increment (15) may be assigned to both the left and right ends of the interval (x_k, x_{k+1}) with weights α and β respectively. Then we can find values of α, β which ensure second order accuracy as

follows.

Writing

$$g_{k+\frac{1}{2}} = v(u_k - u_{k+1}) \quad (16)$$

where

$$v = \frac{a\Delta t}{\Delta x} \quad (17)$$

is the CFL number, and taking $u = x^2$ with the origin at $x = x_k$ (see fig. 1), we find that the total increment to u_k from the intervals (x_k, x_{k+1}) and (x_{k-1}, x_k) is

$$\alpha v\{-(\Delta x)^2\} + \beta v\{(\Delta x)^2\} = (-\alpha + \beta)v(\Delta x)^2, \quad (18)$$

whilst the exact value should be $(a\Delta t)^2$. Hence we require that

$$-\alpha + \beta = v. \quad (19)$$

It is easy to show that the corresponding condition for first order accuracy is the same as that for conservation, namely, that the increment is fully allocated, or

$$\alpha + \beta = 1. \quad (20)$$

Solving (19) and (20) gives

$$\alpha = \frac{1}{2}(1 - v), \quad \beta = \frac{1}{2}(1 + v). \quad (21)$$

With these weights, second order accuracy is achieved in that the specimen polynomials $u = 1$, $u = x$, $u = x^2$ are all exactly convected. The scheme can be identified with the standard Lax-Wendroff scheme⁵. The same weights are obtained for both $a > 0$ and $a < 0$.

If the approximation (10) is used instead of (7), a parallel calculation gives the weights

$$\alpha = \frac{1}{2}(3 - v), \quad \beta = \frac{1}{2}(-1 + v), \quad (22)$$

corresponding to the fully upwinded second order scheme⁶.

Going to second order inevitably results in the loss of monotonicity, as

pointed out by Godunov⁷. However monotonicity can be regained by a switching mechanism, as we shall now show.

§3. Monotonicity-preservation in one dimension

The loss of monotonicity in second order schemes often manifests itself in unwanted oscillations near to rapidly varying data. If part of the mass increment is assigned upstream (by one of the weights) these oscillations can be understood in terms of an excess mass increment.

Suppose that the mass increment in the interval (x_k, x_{k+1}) is less than or equal (in absolute value) to the mass increment in the interval (x_{k-1}, x_k) . Then the contribution sent to x_k from the interval (x_k, x_{k+1}) will be less than or equal to the contribution that would be sent to x_k from the interval (x_{k-1}, x_k) if its increment were wholly extrapolated into the interval (x_k, x_{k+1}) . For this extrapolation of m an overshoot can occur relative to a linear extrapolation of u , for which the scheme is exact. Hence if the absolute value of the mass increment is decreasing in the direction of the stream no such overshoot will occur. If the absolute value of the mass increment is increasing, however, there will in general be an overshoot relative to linear extrapolation.

For this reason we shall allocate to x_k, x_{k+1} with second order weights only that $g_{k+\frac{1}{2}}$ or $g_{k-\frac{1}{2}}$ which has minimum absolute value, and look for some other way to allocate the remainder (if any). It turns out that if the remainder is allocated entirely to the downwind end of the interval (with weight 1 as in the first order method), and that if this procedure is repeated over several intervals in which such a remainder exists, the fully upwinded scheme of (22) is regained. Therefore in smooth regions of the flow this strategy will always produce a second order scheme. At an inflection point, however, some accuracy will be lost although conservation is preserved and, as we shall show, monotonicity is also preserved.

The same result can also be achieved by switching an appropriate amount,

namely $\alpha v(u_k - u_{k+1})$, across the interval. This is Roe's switching procedure³.

We turn now to a proof of the preservation of monotonicity using the above algorithm. The idea for this proof was suggested by P. Sweby⁸ but the form given here is novel. Denote by u^k the value of u_k at the forward time step. Then Roe's version of Lax-Wendroff gives

$$u^k = u_k + \alpha g_{k+\frac{1}{2}} + \beta g_{k-\frac{1}{2}}, \quad (23)$$

where $g_{k+\frac{1}{2}}$ is given by (16) and α, β are given by (21). Then, with a similar definition of $g^{k+\frac{1}{2}}$, we have from (16) that

$$g^{k+\frac{1}{2}} = g_{k+\frac{1}{2}} + v\alpha(g_{k+\frac{1}{2}} - g_{k+\frac{3}{2}}) + v\beta(g_{k-\frac{1}{2}} - g_{k+\frac{1}{2}}). \quad (24)$$

To show monotonicity-preservation we observe from (16) that monotonic data is equivalent to g having the same sign throughout. Provided that the coefficients in the formula for $g^{k+\frac{1}{2}}$ are non-negative, therefore, monotonicity-preservation follows.

This is evidently not the case in (24). But when we adopt the strategy, described earlier, of allocating only $h_{k+\frac{1}{2}}$, where $h_{k+\frac{1}{2}}$ has the same sign as $g_{k+\frac{1}{2}}$ and

$$|h_{k+\frac{1}{2}}| = \min\{|g_{k+\frac{1}{2}}|, |g_{k-\frac{1}{2}}|\}, \quad (25)$$

with second order weights, and the remainder with first order weight, then (24) is replaced by

$$\begin{aligned} g^{k+\frac{1}{2}} = & g_{k+\frac{1}{2}} - v(\alpha h_{k+\frac{3}{2}} + \beta h_{k+\frac{1}{2}}) + v(\alpha h_{k+\frac{1}{2}} + \beta h_{k-\frac{1}{2}}) \\ & - v(g_{k+\frac{1}{2}} - h_{k+\frac{1}{2}}) + v(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}). \end{aligned} \quad (26)$$

Rearranging (26), we obtain

$$\begin{aligned}
g^{k+\frac{1}{2}} &= g_{k+\frac{1}{2}} - v\alpha(g_{k+\frac{1}{2}} - \{g_{k+\frac{1}{2}} - h_{k+\frac{3}{2}}\}) + v(1-\beta)h_{k+\frac{1}{2}} + v\alpha h_{k+\frac{1}{2}} \\
&\quad + v\beta h_{k-\frac{1}{2}} - v g_{k+\frac{1}{2}} + v(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}) \\
&= (1-v\alpha-v)g_{k+\frac{1}{2}} + v\alpha(g_{k+\frac{1}{2}} - h_{k+\frac{3}{2}}) + v(1-\beta+\alpha)h_{k+\frac{1}{2}} \\
&\quad + v\beta h_{k-\frac{1}{2}} + v(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}) \\
&= (1-v)(1-\frac{v}{2})g_{k+\frac{1}{2}} + v\alpha(g_{k+\frac{1}{2}} - h_{k+\frac{3}{2}}) + 2\alpha h_{k+\frac{1}{2}} + v\beta g_{k-\frac{1}{2}} \\
&\quad + v\alpha(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}) \\
&= (1-v)(1-\frac{1}{2}v)g_{k+\frac{1}{2}} + v(1-v)h_{k+\frac{1}{2}} + \frac{1}{2}v(1+v)g_{k-\frac{1}{2}} \\
&\quad + \frac{1}{2}v(1-v)(g_{k+\frac{1}{2}} - h_{k+\frac{3}{2}}) + \frac{1}{2}v(1-v)(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}) \quad (27)
\end{aligned}$$

Provided that $0 \leq v \leq 1$, all the coefficients in the formula (27) for $g^{k+\frac{1}{2}}$ are non-negative. Hence if the data is monotonic such that all the g 's are, say, positive (so that all the h 's are also positive) then, using (25), all the terms on the right hand side of (27) are positive. A similar result holds if they are all negative. Hence monotonicity is preserved.

The use of this strategy admits an even stronger result. If

$|g_{k+\frac{1}{2}}| \geq |g_{k-\frac{1}{2}}|$, $h_{k+\frac{1}{2}} = g_{k-\frac{1}{2}}$ from (25), and (27) becomes

$$\begin{aligned}
g^{k+\frac{1}{2}} &= (1-v)(1-\frac{v}{2})g_{k+\frac{1}{2}} + (1+\alpha)g_{k-\frac{1}{2}} + v\alpha(g_{k+\frac{1}{2}} - h_{k+\frac{3}{2}}) \\
&\quad + v\alpha(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}) \\
&= g_{k-\frac{1}{2}} + (1-v)(1-\frac{v}{2})(g_{k+\frac{1}{2}} - g_{k-\frac{1}{2}}) + v\alpha(g_{k+\frac{1}{2}} - h_{k+\frac{3}{2}}) \\
&\quad + v\alpha(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}) . \quad (28)
\end{aligned}$$

It then follows, by arguments similar to the above, that

$$|g^{k+\frac{1}{2}}| \geq |g_{k-\frac{1}{2}}| . \quad (29)$$

Similarly, if $|g_{k+\frac{1}{2}}| \leq |g_{k-\frac{1}{2}}|$, $h = g_{k+\frac{1}{2}}$ and (27) becomes

$$\begin{aligned}
g^{k+\frac{1}{2}} &= (1 - \nu - \nu\alpha)g_{k+\frac{1}{2}} + \nu\beta g_{k-\frac{1}{2}} + \nu\alpha(g_{k+\frac{1}{2}} - h_{k+\frac{3}{2}}) + \nu\alpha(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}) \\
&= g_{k+\frac{1}{2}} + \nu\beta(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}) + \nu\alpha(g_{k+\frac{1}{2}} - h_{k+\frac{3}{2}}) + \nu\alpha(g_{k-\frac{1}{2}} - h_{k-\frac{1}{2}}),
\end{aligned}
\tag{30}$$

so that

$$|g^{k+\frac{1}{2}}| \geq |g_{k+\frac{1}{2}}| . \tag{31}$$

From (29) and (31) it follows that

$$|g^{k+\frac{1}{2}}| \geq \min\{|g_{k+\frac{1}{2}}|, |g_{k-\frac{1}{2}}|\} , \tag{32}$$

i.e. from (25), that

$$|g^{k+\frac{1}{2}}| \geq |h_{k+\frac{1}{2}}| . \tag{33}$$

Recalling (16) we have the result that the slope of the linear interpolant in an interval at the forward time step cannot fall below (in absolute value) the slope in the interval or its upwind neighbour at the previous time step.

A local monotonicity result follows. If $u_{k-1} \gtrless u_k \gtrless u_{k+1}$, then $u^k \gtrless u^{k+1}$ ($a > 0$). If $a < 0$ the latter inequality is replaced by $u^k \gtrless u^{k-1}$. We can also deduce a regional monotonicity result. If the data is monotonic in a region R then so is the solution at the next time step, with the possible exception of most upwind interval of R . It can then be deduced that piecewise monotonic data leads to a piecewise monotonic solution with at most a one step displacement of the separating nodes in the direction of propagation.

Monotonicity-preservation can be achieved with other definitions of h than (25) but this definition leads to the most compact algorithm (see Appendix A).

Roe's scheme, of which the above is an account, has been widely tested on a

variety of problems in one dimension with impressive results. We now show how the scheme may be generalised directly to two or three dimensions.

§4. The two-dimensional method

Consider the two dimensional scalar equation

$$u_t + f_x + g_y = 0 \quad (34)$$

or
$$u_t + f'(u)u_x + g'(u)u_y = 0 . \quad (35)$$

Subdivide the plane into regular rectangles with cells of side $\Delta x, \Delta y$ and consider a block of four rectangles (see fig. 2).

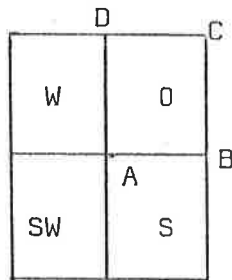


fig. 2

Approximate the physical mechanism involved in the derivation of (34) by defining the "mass" m in the cell as

$$m = \iint_{\text{cell}} u \, dx dy , \quad (36)$$

and supposing that the masses in cells W, SW and S are convected into the cell O with velocity $(f'(u), g'(u))_O$, both components assumed to be positive. Then the cell O draws in masses

$$(f'(u))_O \frac{\Delta t}{\Delta x} \{1 - (g'(u))_O \frac{\Delta t}{\Delta x}\} m_W , \quad (37a)$$

$$(f'(u))_O \frac{\Delta t}{\Delta x} (g'(u))_O \frac{\Delta t}{\Delta y} m_{SW} , \quad (37b)$$

and
$$\{1 - f'(u) \frac{\Delta t}{\Delta x}\} g'(u) \frac{\Delta t}{\Delta y} m_S , \quad (37c)$$

from the cells W, SW and S , respectively (see fig. 3).

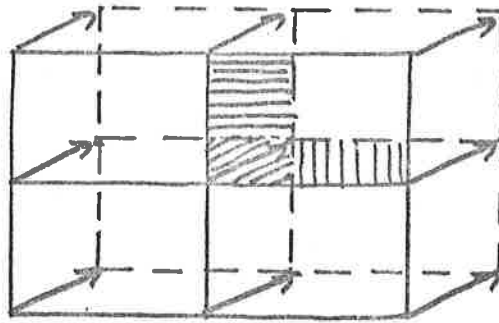


fig. 3

We have assumed that both

$$v_1 = (f'(u))_0 \frac{\Delta t}{\Delta x}, \quad v_2 = (g'(u))_0 \frac{\Delta t}{\Delta y} \quad (38)$$

lie in the interval $[0, 1]$. These masses displace part of the mass m_0 and hence the mass increment is

$$\Delta m_0 = v_1(1 - v_2)(m_W - m_0) + v_1 v_2 (m_{SW} - m_0) + v_2(1 - v_1)(m_S - m_0). \quad (39)$$

To approximate m_0 use the zero'th order downwind approximation

$$u_C \Delta x \Delta y \quad (40)$$

(see fig. 2) with similar approximations for the other masses. Then (39) becomes

$$\Delta m_0 = \Delta x \Delta y [v_1(1 - v_2)(u_D - u_C) + v_1 v_2 (u_A - u_C) + v_2(1 - v_1)(u_B - u_C)]. \quad (41)$$

The mass increment is to be divided by the area $\Delta x \Delta y$ and allocated to nodal values of u at the next time step. Using the same downwind approximation (40) at the forward time step we end up by incrementing u_D by an amount

$$Y_0^C = \frac{1}{\Delta x \Delta y} \Delta m_0 = -v_1(1 - v_2)(u_C - u_D) - v_1 v_2(u_C - u_A) - v_2(1 - v_1)(u_C - u_B) \quad (42)$$

Note that (42) reduces to (16) when $v_1 = v$, $v_2 = 0$.

We have assumed above that $v_1, v_2 \geq 0$. When $v_1 < 0$, $v_2 > 0$, (40) is replaced by

$$u_D \Delta x \Delta y, \quad (43)$$

and (42) becomes

$$Y_0^D = v_1(1 - v_2)(u_D - u_C) + v_1 v_2(u_D - u_B) - (1 + v_1)v_2(u_D - u_A) \quad (44)$$

Similarly, if $v_1 > 0$, $v_2 < 0$, we have

$$Y_0^B = -v_1(1 + v_2)(u_B - u_A) + v_1 v_2(u_B - u_D) + v_2(1 - v_1)(u_B - u_C), \quad (45)$$

and, if $v_1, v_2 < 0$,

$$Y_0^A = v_1(1 + v_2)(u_A - u_B) - v_1 v_2(u_A - u_C) + v_2(1 + v_1)(u_A - u_D) \quad (46)$$

When $v_1 = v$, $v_2 = 0$, (44), (45) and (46) also reduce to (16).

Now there are four possible increments instead of one, as in the one-dimensional case. To indicate the essential difference we return to the one-dimensional case and note that $g_{k+\frac{1}{2}}$ as given by (9) is the exact value of the integral

$$- \int_{x_k}^{x_{k+1}} f_x dx \quad (47)$$

which also measures the mass increment (apart from a proportionality constant). In two dimensions, however, Y_0 is only a first order approximation to the corresponding integral

$$- \iint_{\text{cell } 0} (f_x + g_y) dx dy \quad (48)$$

$$= - \int_{\text{boundary of } 0} (f, g) \cdot d\underline{S} \quad (49)$$

Indeed each of the equations (42), (44), (45) and (46) contains a different first order approximation to (48) and they are equivalent to the same order of accuracy.

In each of the four cases allocation to u at the next time level is governed by the appropriate downwind approximation to (36). For example, in the case of (42) we use (40) and allocate to u at the corner C (fig. 2). It is straightforward to show that, in the case of the linear wave equation

$$u_t + au_x + bu_y = 0, \quad (50)$$

for which $v_1 = a \frac{\Delta t}{\Delta x}, \quad v_2 = b \frac{\Delta t}{\Delta y} \quad (51)$

are CFL numbers, this scheme is exact for specimen functions $u = 1, u = x, u = y$ and is therefore of first order.

To construct a second order scheme we suppose that the quantity Y_0^C is allocated to the four corners A, B, C, D (fig. 2) with weights $\alpha, \beta, \gamma, \delta$, respectively. To calculate $\alpha, \beta, \gamma, \delta$ we impose the conditions that the algorithm should be exact for the specimen second degree polynomials $u = x^2, u = y^2, u = xy$. Since $u = (ay - bx)^2$ is an exact solution which is invariant in the direction of the stream, exactness for $u = x^2$ and $u = y^2$ will automatically imply exactness for $u = xy$.

For $u = x^2$, we obtain

$$-\alpha + \beta + \gamma - \delta = v_1, \quad (52)$$

while for $u = y^2$ we have

$$-\alpha - \beta + \gamma + \delta = v_2. \quad (53)$$

The condition for first order accuracy is easily shown to be

$$\alpha + \beta + \gamma + \delta = 1, \quad (54)$$

so that, if we add the equation

$$\alpha - \beta + \gamma - \delta = \lambda, \quad (55)$$

we have the solution

$$\left. \begin{aligned} \alpha &= \frac{1}{4}(1 + \lambda - v_1 - v_2) \\ \beta &= \frac{1}{4}(1 - \lambda + v_1 - v_2) \\ \gamma &= \frac{1}{4}(1 + \lambda + v_1 + v_2) \\ \delta &= \frac{1}{4}(1 - \lambda - v_1 + v_2) \end{aligned} \right\} \quad (56)$$

Second order accuracy is obtained by using the weights (56) with any choice of λ . Moreover the same weights are obtained with all four representations of Y_0 , namely, (42), (44), (45) and (46). This is because all four expressions are identical to the required order.

When $v_1 = v$, $v_2 = 0$, the weights (56) reduce to (21): note that $\alpha + \delta$, $\beta + \gamma$ are the comparative weights here.

If we choose $\lambda = \frac{1}{2}v_1v_2$, the weights (56) can be written in the particularly simple form

$$\left. \begin{aligned} \alpha &= \frac{1}{2}(1 - v_1)\frac{1}{2}(1 - v_2) \\ \beta &= \frac{1}{2}(1 + v_1)\frac{1}{2}(1 - v_2) \\ \gamma &= \frac{1}{2}(1 - v_1)\frac{1}{2}(1 + v_2) \\ \delta &= \frac{1}{2}(1 + v_1)\frac{1}{2}(1 + v_2) \end{aligned} \right\} \quad (57)$$

which are simply products of the weights (21). These are clearly non-negative when $v_1, v_2 \in [0, 1]$.

Upwind weights are easily obtained. If the coordinates of the receiver rectangle with respect to the donor rectangle are $(n\Delta x, m\Delta y)$, the equations for the weights are the same as (52)-(55) with v_1 replaced by $v_1 - 2n$ and v_2 replaced by $v_2 - 2m$. The upwinded weights are therefore

$$\left. \begin{aligned} \alpha' &= \frac{1}{4}(1 + \lambda' + 2n + 2m - v_1 - v_2) \\ \beta' &= \frac{1}{4}(1 - \lambda' - 2n + 2m + v_1 - v_2) \\ \gamma' &= \frac{1}{4}(1 + \lambda' - 2n - 2m + v_1 + v_2) \\ \delta' &= \frac{1}{4}(1 - \lambda' + 2n - 2m - v_1 + v_2) \end{aligned} \right\} \quad (58)$$

These reduce to (56) when $n = m = 0$ and to (22) when $n = 1, m = 0, v_1 = v, v_2 = 0$. Although we refer to (58) as upwind weights, more correctly the weights and allocation are received downwind from the donor cell (which is upwind relative to the receiver cell). The effect is an upwind scheme in the usual terminology.

§5. Monotonicity-preservation in two dimensions

When the increment (42) is allocated with weights (56) there will be loss of monotonicity in general, as expected. However, if we adopt the strategy, described earlier in the one-dimensional case, of allocating with second order weights only that Y with smallest absolute value from the four cells shown in fig. 2, and allocating the remainder with first order weight downwind, we can prove a form of monotonicity-preservation. In fact we shall prove that, under certain conditions, if Y is positive (or negative) for all cells at a certain time level, then it is positive (or negative) at the forward time level.

The value of Y at the forward time level in cell 0 (see fig. 4) is given by

$$\begin{aligned}
 Y^0 = & Y_0 - v_1 v_2 \{ \alpha(Y_{NE} - Y_0) + \beta(Y_N - Y_W) + \gamma(Y_0 - Y_{SW}) + \delta(Y_E - Y_S) \} \\
 & - v_1(1 - v_2) \{ \alpha(Y_{NE} - Y_N) + \beta(Y_N - Y_{NW}) + \gamma(Y_0 - Y_W) + \delta(Y_E - Y_0) \} \\
 & - v_2(1 - v_1) \{ \alpha(Y_{NE} - Y_E) + \beta(Y_N - Y_0) + \gamma(Y_0 - Y_S) + \delta(Y_E - Y_{SE}) \}
 \end{aligned}
 \tag{59}$$

where

$$0 \leq v_1 \leq 1, \quad 0 \leq v_2 \leq 1. \tag{60}$$

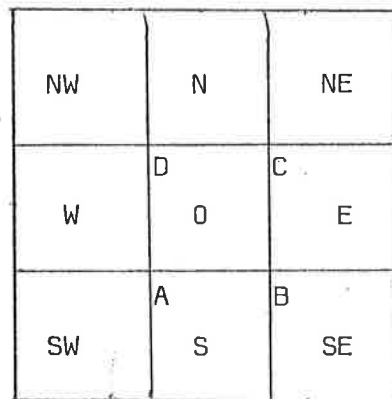


fig. 4

Now define, along the same lines as in one-dimension, the relative quantity

X_0 which has the same sign as Y_0 and is such that

$$|X_0| = \min\{|Y_0|, |Y_W|, |Y_{SW}|, |Y_S|\}, \quad (61)$$

with similar definitions for X_{NE} , etc.

Assign X with second order weights $\alpha, \beta, \gamma, \delta$ and hold $Y-X$ back for the present.

Then Y^0 equals the right hand side of (59) with Y replaced everywhere with X . This may be written

$$\begin{aligned} Y^0 = & Y_0 + v_1 v_2 \{ \alpha(Y_0 - X_{NE}) + \beta(Y_0 - X_N) + \gamma(Y_0 - X_0) + \delta(Y_0 - X_E) \\ & + \alpha(X_0 - Y_0) + \beta(X_W - Y_0) + \gamma(X_{SW} - Y_0) + \delta(X_S - Y_0) \} \\ & + v_1(1 - v_2) \{ \alpha(Y_0 - X_{NE}) + \beta(Y_0 - X_N) + \gamma(Y_0 - X_0) + \delta(Y_0 - X_E) \\ & + \alpha(X_N - Y_0) + \beta(X_{NW} - Y_0) + \gamma(X_W - Y_0) + \delta(X_0 - Y_0) \} \\ & + v_2(1 - v_1) \{ \alpha(Y_0 - X_{NE}) + \beta(Y_0 - X_N) + \gamma(Y_0 - X_0) + \delta(Y_0 - X_E) \\ & + \alpha(X_E - Y_0) + \beta(X_0 - Y_0) + \gamma(X_S - Y_0) + \delta(X_{SE} - Y_0) \} \end{aligned} \quad (62)$$

$$\begin{aligned} = & [1 - (\alpha + \beta + \delta) \{ v_1 v_2 + v_1(1 - v_2) + v_2(1 - v_1) \}] Y_0 \\ & - \gamma \{ v_1 v_2 + v_1(1 - v_2) + v_2(1 - v_1) \} X_0, \end{aligned} \quad (63)$$

together with terms with positive coefficients that take the sign of the X or $Y-X$ that they contain. This is true provided that the weights $\alpha, \beta, \gamma, \delta$ given by (56) are all positive. The condition on λ for this to be true is readily shown to be

$$-1 + |v_1 + v_2| \leq \lambda \leq 1 - |v_1 - v_2|, \quad (64)$$

for $0 \leq v_1, v_2 \leq 1$.

It remains to assign the quantity $Y-X$, to be allocated with weight 1 to the downwind corner C (see fig. 2). Then Y^0 receives additional terms

$$v_1 v_2 \{(Y_{SW} - X_{SW}) - (Y_0 - X_0)\} + v_1 (1 - v_2) \{(Y_W - X_W) - (Y_0 - X_0)\} \\ + v_2 (1 - v_1) \{(Y_S - X_S) - (Y_0 - X_0)\} \quad (65)$$

$$= -\{v_1 v_2 + v_1 (1 - v_2) + v_2 (1 - v_1)\} (Y_0 - X_0), \quad (66)$$

together with terms with positive coefficients that take the sign of the X or Y-X that they contain.

When Y is completely allocated the result is

$$Y^0 = \{1 - (2 - \gamma)v_{12}\}Y_0 + (1 - \gamma)v_{12}X_0, \quad (67)$$

together with terms with positive coefficients that take the sign of the X or Y-X that they contain. Here

$$v_{12} = v_1 v_2 + v_1 (1 - v_2) + v_2 (1 - v_1) \quad (68)$$

and we have used equation (54).

We now seek conditions on λ such that the coefficients in (67) are always positive.

Using (56) these conditions may be written

$$\frac{1}{4}(7 - \lambda - v_1 - v_2)v_{12} \leq 1 \quad (69a)$$

and $\frac{1}{4}(1 + \lambda + v_1 + v_2) \leq 1. \quad (69b)$

Condition (69b) requires $\lambda \leq 1$ and is already covered by (64). Using (64) also we can write the left hand side of (69a) as

$$\frac{1}{4}(7 - \lambda - v_1 - v_2)v_{12} \leq \frac{1}{4}[8 - 2(v_1 + v_2)]v_{12} \quad (70)$$

which is ≤ 1 in a certain region of the v_1, v_2 plane. Since $v_1 + v_2 \geq 0$, it is certainly ≤ 1 if $v_{12} \leq \frac{1}{2}$ which, since

$$v_{12} = 1 - (1 - v_1)(1 - v_2), \quad (71)$$

confines v_1, v_2 to the region shown shaded in fig. 5.

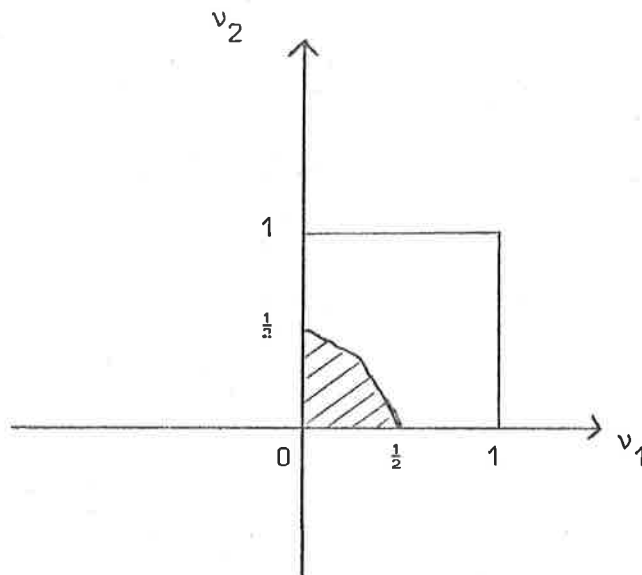


fig. 5

The curve is part of a rectangular hyperbola.

The full condition

$$[2 - \frac{1}{2}(v_1 + v_2)]v_{12} \leq 1 \quad (72)$$

is satisfied in a slightly larger region. These are all sufficient conditions.

In one dimension we were able to identify Y (the corresponding notation was g) as proportional to a difference of u 's, so that the one-signedness of Y was equivalent to the monotonicity of u . In two dimensions this is not the case but we may write Y as a linear combination of three differences, corresponding to the x , y and diagonal directions, namely

$$Y_0 = -v_1(1 - v_2)(u_C - u_D) - v_2(1 - v_1)(u_C - u_B) - v_1v_2(u_C - u_A) .$$

(73)

It is the sign of this quantity that is preserved by the algorithm from one time step to the next and this is the sense in which the algorithm is monotonicity-preserving. It can readily be shown that (73) is an approximation to the component of \underline{v}_u in the direction (a, b) at the point C . Since

the component of ∇u perpendicular to the direction (a, b) is preserved exactly, the local curvature is approximately preserved overall.

We can also obtain the stronger result

$$|Y^0| \geq |X_0| \quad (74)$$

(c.f.(33)). Suppose that the minimum of the four quantities $|Y_0|$, $|Y_W|$, $|Y_{SW}|$, $|Y_S|$ is $|Y_0|$. Then $X_0 = Y_0$ and equation (63) reduces to

$$Y^0 = (1 - v_{12})Y_0, \quad (75)$$

apart from terms with positive coefficients that take the sign of the X or $Y-X$ they contain. Hence $|Y^0| \geq |Y_0|$, since $v_{12} \leq 1$. Now suppose that the minimum in (61) is $|Y_W|$. Then $X_0 = Y_W$ and (63) becomes

$$\begin{aligned} Y^0 &= [1 - (\alpha + \beta + \delta)v_{12}]Y_0 - \gamma v_{12}Y_W \\ &= (1 - v_{12})Y_W + [1 - (\alpha + \beta + \gamma)v_{12}](Y_0 - Y_W), \end{aligned} \quad (76)$$

apart from terms with positive coefficients that take the sign of the X or $Y-X$ that they contain. Hence $|Y^0| \geq |Y_W|$. Similar results hold for $|Y_{SW}|$ being the minimum and for $|Y_S|$ being the minimum. Bringing the four results together gives (75).

A regional monotonicity-preservation result then follows: if Y at one time step is positive (or negative) in a region then it is positive (or negative) at the forward time step in the same region with the possible exception of a rim of boundary cells on the upwind edge of the region.

We have proved that the strategy described at the beginning of this section restores monotonicity in the sense that the sign of Y (as given by Y_0^C in (42) when $0 \leq v_1, v_2 \leq 1$) is preserved on taking a time step.

If the strategy is repeated consistently over a region in which $X_0 = Y_0$, Y_W , Y_{SW} or Y_S , one of the upwind schemes given by (58) (with $n = 0$ or 1

$m = 0$ or 1) is regained, so that second order is maintained in smooth regions for which this is the case.

The switching equivalent of the strategy is to switch the quantities $\alpha(Y_0 - X_0)$, $\beta(Y_0 - X_0)$, $\delta(Y_0 - X_0)$ across the cell 0 from the points A, B, D, respectively, to the point C (fig. 2).

All the results in this section hold in the cases $-1 \leq v_1 \leq 0$, $0 \leq v_2 \leq 1$; $-1 \leq v_1, v_2 \leq 0$; $0 \leq v_1 \leq 1$, $-1 \leq v_2 \leq 0$, with the appropriate choice of Y_0 . Difficulties arise only when v_1 or v_2 change sign in a region as in the non-linear case. We discuss such cases below.

§6. Non-linear aspects in two dimensions

In the non-linear case there are several problems not encountered in the linear case. First there is the question of the evaluation of $\frac{\partial f}{\partial u}$, $\frac{\partial g}{\partial u}$ in order to obtain the quantities v_1, v_2 in (38). Secondly, there is the difficulty of v_1, v_2 varying in sign with the consequent need to vary the downwind corner of a cell and to distinguish between the forms (42), (44), (45), (46).

As far as the calculation of $\frac{\partial f}{\partial u}$, $\frac{\partial g}{\partial u}$ is concerned the obvious choice is to make the approximations

$$\frac{\partial f}{\partial u} = \frac{f_A + f_D - f_B - f_C}{u_A + u_D - u_B - u_C}, \quad \frac{\partial g}{\partial u} = \frac{g_A + g_B - g_C - g_D}{u_A + u_B - u_C - u_D}, \quad (77)$$

which give values for v_1, v_2 and identify the downwind corner of the cell 0 (see fig. 2). It is illuminating to note that the incremental quantity \bar{Y} , calculated from the mass flow integral (49) with trapezium rule approximations to the line integrals on each side, is

$$\bar{Y}_0 = -\frac{\Delta t}{2\Delta x \Delta y} [(f_B + f_C - f_A - f_D)\Delta y + (g_C + g_D - g_A - g_B)\Delta x] \quad (78)$$

$$= -\frac{1}{2} \frac{\Delta t}{\Delta x} (f_B + f_C - f_A - f_D) - \frac{1}{2} \frac{\Delta t}{\Delta y} (g_C + g_D - g_A - g_B) \quad (79)$$

$$= -\frac{1}{2} \bar{v}_1 (u_B + u_C - u_A - u_D) - \frac{1}{2} \bar{v}_2 (u_C + u_D - u_A - u_B) \quad (80)$$

where \bar{v}_1, \bar{v}_2 are the values of v_1, v_2 given by (38) with the approximations (77), namely,

$$\bar{v}_1 = \frac{\Delta t}{\Delta x} \left(\frac{f_A + f_D - f_B - f_C}{u_A + u_D - u_B - u_C} \right), \quad \bar{v}_2 = \frac{\Delta t}{\Delta y} \left(\frac{g_A + g_B - g_C - g_D}{u_A + u_B - u_C - u_D} \right). \quad (81)$$

It is clear from (78) that

$$\sum_{\text{cells}} \bar{Y} = 0, \quad (82)$$

except for boundary values, so that conservation is maintained in the sense of (13).

If we seek weights $\alpha, \beta, \gamma, \delta$ which assign \bar{Y}_0 , given by (80) with constant $\bar{v}_1 = v_1, \bar{v}_2 = v_2$, to the corners A, B, C, D of fig. 2 in such a way that second order accuracy is obtained, then we obtain the familiar weights (56). This is because, to the required order of accuracy, (80) is indistinguishable from (42), (44), (45) and (46). However, the scheme using (80) and the weights (56) cannot be made monotonicity-preserving with any obvious kind of switching; the appropriate downwind form (42), (44), (45) or (46) is needed.

In using a downwind form instead of (80), a third order variation in the scheme is incurred. As long as v_1 and v_2 are of uniform sign the main properties of the scheme are unaffected but in the case of converging flows, where v_1 or v_2 changes sign from one cell to the next, different downwind forms are used and (82) will not hold to the extent of a third order error. This is not true in one dimension where the equivalent of (80) is always used.

In a more sophisticated model of the mass transfer mechanism, (42), (44), (45) or (46) will be replaced by more complicated expressions taking account of convection in all possible directions (see Appendix B). Conservation (in the discretised sense (82)) will then be assured.

If we use (42), (44), (45) and (46) in the non-linear case with v_1, v_2 given by (81) we have to tolerate a small loss of conservation in converging flows. This is of the same order as the loss of accuracy that is tolerated in smooth flows.

Having identified the downwind corner of the cell 0 using (77), there may be some advantage in redefining v_1, v_2 in terms of the local, one-sided, differences at the downwind corner. For example, if the downwind corner is C (fig. 2) and we define

$$v_1 = \frac{\Delta t}{\Delta x} \left(\frac{f_C - f_D}{u_C - u_D} \right), \quad v_2 = \frac{\Delta t}{\Delta y} \left(\frac{g_C - g_B}{u_C - u_B} \right), \quad (83)$$

then (68) becomes

$$Y_0 = -\frac{\Delta t}{\Delta x} (1 - v_2)(f_C - f_D) - \frac{\Delta t}{\Delta y} (1 - v_1)(g_C - g_B) - v_1 v_2 (u_C - u_A) \quad (84)$$

which, for small v_1 and v_2 , is approximately a straight sum of the two corresponding one-dimensional quantities.

§7. A test problem in two dimensions

As a simple test problem for the two-dimensional scheme of §4 and §5 we take the problem considered recently by Dukiewicz and Ramshaw⁹ of the convection of an L-shaped discontinuity across a region. The governing equation is (50) where a and b are constants.

The initial position of the discontinuity is shown in fig. 6: the larger square region shown is a 17×17 grid and the smaller is 12×12 . Δx and Δy are taken to be 0.1, $a = 1.5$, $b = 0.5$, and $\Delta t = 0.2$, so that $v_1 = 0.3$, $v_2 = 0.1$, in tables 1 and 2, which show the values of u after 20 time steps. Table 2 is obtained without switching (i.e. the centred scheme) while in table 1 we have used the switching strategy of §5. Throughout the calculation λ was taken equal to zero. The results were obtained on the NORD-100 system at Reading University and are given in Appendix C.

Boundary conditions were chosen as follows. On upwind boundaries we took $u = 2$ to retain a steady inflow. Had this not been done the oscillations in table 1 would have extended even further upwind. On downwind boundaries u was taken equal to its value on the grid just inside the boundary simulating a transparent boundary condition.

In addition to the tables given, surface plots of u have been obtained. The results are given in Appendix C. Plots 1 and 2 correspond to tables 1 and 2.

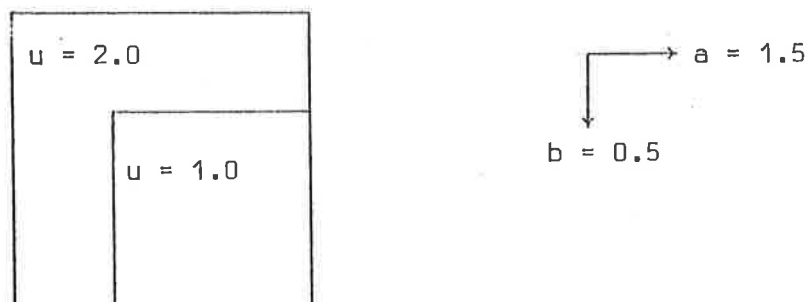


fig. 6

In addition plot 3 shows the results of an unswitched 'Fromm' type scheme (see Appendix A).

§8. The three-dimensional method

We can readily extend the algorithm described in §4, §5 and §6 into three dimensions, for which the corresponding scalar equation is

$$u_t + f_x + g_y + h_z = 0 \quad (85)$$

or
$$u_t + f'(u)u_x + g'(u)u_y + h'(u)u_z = 0 \quad (86)$$

The three dimensional space is subdivided into rectangular parallelepipeds of sides $\Delta x, \Delta y, \Delta z$: a typical cell is shown in fig. 7

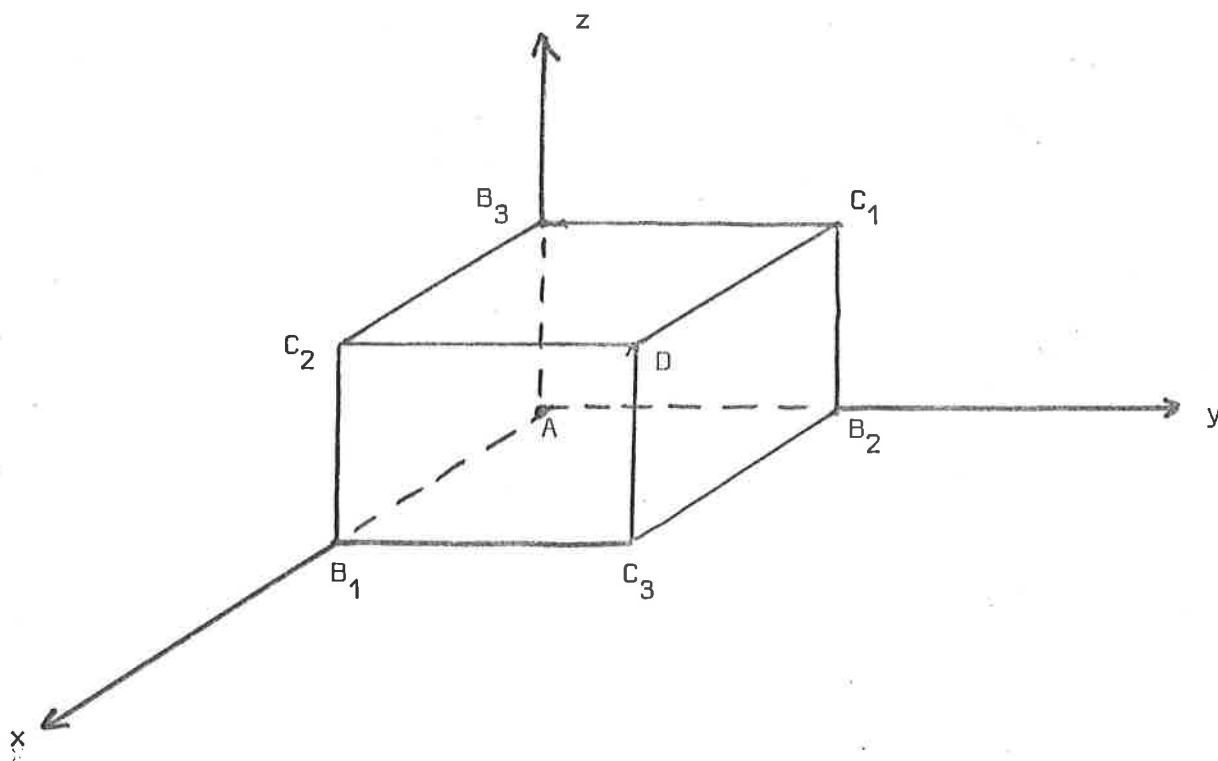


fig. 7

The increment Y_0 corresponding to (42) in two dimensions is

$$Y_0 = -v_1 v_2 v_3 (u_D - u_A) - \sum v_i v_j (1 - v_k) (u_D - u_{B_k}) - \sum (1 - v_i) (1 - v_j) v_k (u_D - u_{C_k}), \quad (87)$$

where the sums here and in what follows are taken over cyclic permutations

of 1, 2, 3. [In the two-dimensional case we could have written (42) as

$$Y_0 = -v_1 v_2 (u_{C_3} - u_A) - \sum_i v_i (1 - v_j) (u_{C_3} - u_{B_j}), \quad (88)$$

where the sum was over permutations of 1, 2.]

The increment (87) is the appropriate one when D is the downwind corner of the cell. Corresponding forms of Y_0 can be written down for the other seven corners of the cell. In (87) v_1, v_2, v_3 are approximately

$$v_1 = \frac{\Delta t}{\Delta x} \frac{\partial f}{\partial u}, \quad v_2 = \frac{\Delta t}{\Delta y} \frac{\partial g}{\partial u}, \quad v_3 = \frac{\Delta t}{\Delta z} \frac{\partial h}{\partial u} \quad (89)$$

where $0 \leq v_1, v_2, v_3 \leq 1$, (90)

and the form of Y_0 corresponds to the approximation

$$u_D \Delta x \Delta y \Delta z \quad (91)$$

for the integral

$$m = \iiint u \, dx dy dz \quad (92)$$

(see fig. 7).

Consider the case of the linear wave equation

$$u_t + a u_x + b u_y + c u_z = 0, \quad (93)$$

where

$$v_1 = a \frac{\Delta t}{\Delta x}, \quad v_2 = b \frac{\Delta t}{\Delta y}, \quad v_3 = c \frac{\Delta t}{\Delta z} \quad (94)$$

are the CFL numbers. In the first order method Y_0 is allocated wholly to the downwind corner D. The resulting scheme is exact for specimen polynomials $u = 1, u = x, u = y, u = z$. To obtain the second order method assume that Y_0 is allocated to the eight corners in fig. 7 with weights $\alpha, \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3$ and δ (in an obvious notation).

For second order we impose the conditions that the algorithm be exact for the

specimen polynomials $u = x^2$, $u = y^2$, $u = z^2$, $u = yz$, $u = zx$, $u = xy$, of second degree. Since $u = (ay - bx)^2$, $u = (bz - cy)^2$, $u = (cx - az)^2$ are all exact solutions of (93) the number of conditions is reduced by three. The remaining three conditions together with the condition for first order accuracy, namely,

$$\alpha + \sum_{i=1}^3 \beta_i + \sum_{i=1}^3 \gamma_i + \delta = 1, \quad (95)$$

give four conditions for the eight unknowns $\alpha, \delta, \beta_i, \gamma_i$ ($i = 1, 2, 3$). There will therefore be a four parameter family of solutions.

The remaining equations for the unknown weights can be written down as

$$\begin{aligned} \text{for } x^2 &: -\alpha - \beta_2 - \beta_3 - \gamma_1 + \beta_1 + \gamma_2 + \gamma_3 + \delta = v_1 \\ \text{for } y^2 &: -\alpha - \beta_3 - \beta_1 - \gamma_2 + \beta_2 + \gamma_3 + \gamma_1 + \delta = v_2 \\ \text{for } z^2 &: -\alpha - \beta_1 - \beta_2 - \gamma_3 + \beta_3 + \gamma_1 + \gamma_2 + \delta = v_3 \end{aligned} \quad (96)$$

and (95), (96) can be solved in terms of four free parameters, say $\lambda_1, \lambda_2, \lambda_3, \lambda_4$.

One possible solution, corresponding to (57) in two dimensions is

$$\left. \begin{aligned} \alpha &= \frac{1}{2}(1 - v_1)^{\frac{1}{2}}(1 - v_2)^{\frac{1}{2}}(1 - v_3) \\ \beta_i &= \frac{1}{2}(1 + v_i)^{\frac{1}{2}}(1 - v_j)^{\frac{1}{2}}(1 - v_k) \\ \gamma_i &= \frac{1}{2}(1 + v_i)^{\frac{1}{2}}(1 + v_j)^{\frac{1}{2}}(1 - v_k) \\ \delta &= \frac{1}{2}(1 + v_i)^{\frac{1}{2}}(1 + v_j)^{\frac{1}{2}}(1 + v_k) \end{aligned} \right\} \quad (97)$$

(i, j, k being cyclic permutations of 1, 2, 3), for which the weights are clearly non-negative since $0 \leq v_1, v_2, v_3 \leq 1$.

Upwind weights can easily be found, the essential change being the replacement of v_1, v_2, v_3 in (96) by $v_1 - 2n, v_2 - 2m, v_3 - 2l$, respectively (cf. (58)).

§9. Monotonicity-preservation in three dimensions

Once again we propose the variation of the above algorithm which allows 'monotonicity' preservation to be proved. Instead of allocating Y_0 with the above second order weights, we allocate instead X_0 , where X_0 has the same sign as Y_0 and is such that

$$|X_0| = \min |Y_u| \quad (98)$$

where Y_u goes through Y_0 and its adjacent upwind cells, eight cells in all (see fig. 8).

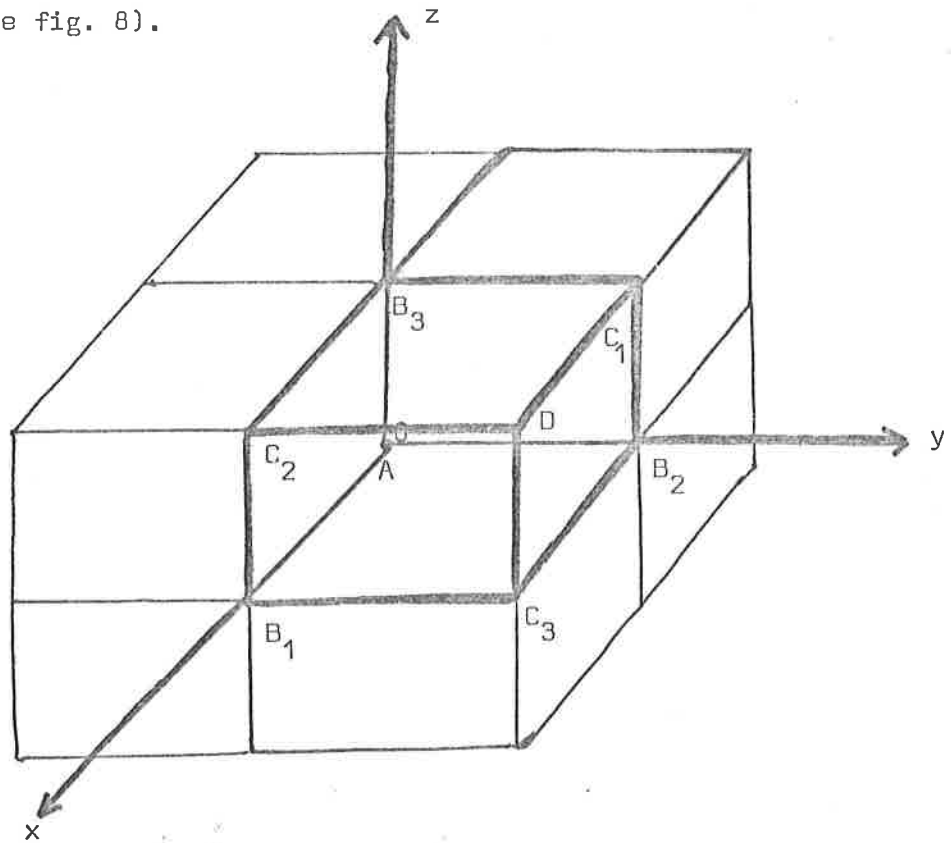


fig. 8

When this has been done the remaining $Y_0 - X_0$ (if any) is allocated to the downwind corner D. Uniform repetition of this procedure can lead to an upwind scheme, as in two dimensions.

'Monotonicity' preservation is proved in the same manner as in §5. For each contribution $X_0 - x$ to Y_0 we replace $X_0 - x$ by $X_0 - Y_0$ together

$Y_0 - X$. Since all the terms $Y_0 - X$ are necessarily of the same sign as Y_0 , by virtue of (98), and all terms X have the same sign as the corresponding Y , we consider only those terms containing Y_0 and X_0 . This leads to an equation which corresponds to (63), namely,

$$Y^0 = [1 - \left(\alpha + \sum_{i=1}^3 \beta_i + \sum_{i=1}^3 \gamma_i \right) v_{123}] Y_0 - \delta v_{123} X_0, \quad (99)$$

together with terms with positive coefficients (using e.g. (97)) that take the sign of the X or $Y-X$ they contain. Here

$$v_{123} = v_1 v_2 v_3 + \sum v_i v_j (1 - v_k) + \sum (1 - v_i)(1 - v_j) v_k. \quad (100)$$

When the remainder $Y - X$ is allocated, additional terms arise, the relevant one (corresponding to (66)) being

$$-v_{123}(Y_0 - X_0). \quad (101)$$

Taking (99) and (101) together gives

$$Y^0 = \{1 - (2 - \delta)v_{123}\} Y_0 + (1 - \delta)v_{123} X_0, \quad (102)$$

apart from terms with positive coefficients that take the sign of the X or $Y - X$ they contain. Here we have used (95).

The coefficients in equation (102) are non-negative if

$$(2 - \delta)v_{123} \leq 1, \quad (103a)$$

$$\delta \leq 1. \quad (103b)$$

With δ given by (97), condition (103b) is satisfied, and condition (103a) can be written

$$\left\{ 2 - \frac{1}{8}(1 + v_1)(1 + v_2)(1 + v_3) \right\} v_{123} \leq 1, \quad (104)$$

which is necessarily satisfied if $v_{123} \leq \frac{1}{2}$. Since

$$v_{123} = 1 - (1 - v_1)(1 - v_2)(1 - v_3) \quad (105)$$

we can illustrate this region in fig. 9.

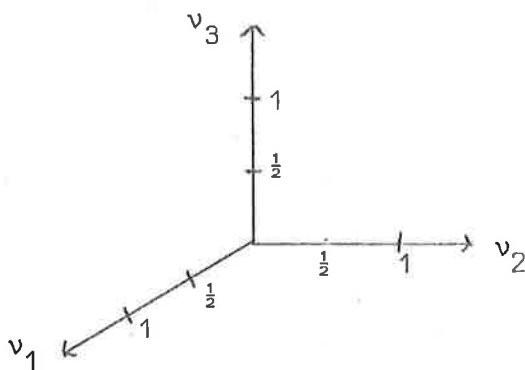


fig. 9

The region corresponding to the full condition (104) is somewhat larger than this. We have thus obtained a sufficient condition for 'monotonicity'-preservation.

As in two dimensions we may identify Y as a linear combination of differences corresponding to various downwind directions (see (87)). It is this quantity whose sign is preserved by the algorithm in a time step, rather than the single difference (16) in one-dimension. It can readily be shown that Y is proportional to an approximation to the component of $\nabla_{\underline{u}}$ in the direction (a, b, c) at C .

We can also show, as in §5, that

$$|Y^0| \geq |X_0|. \quad (106)$$

'Monotonicity'-preservation in a region then follows, except possibly for an upwind skin of boundary cells.

As in two dimensions special problems arise in the approximation of $\frac{\partial f}{\partial u}$, $\frac{\partial g}{\partial u}$, $\frac{\partial h}{\partial u}$ and the identification of the downwind corner. We may define

$$\frac{\partial f}{\partial u} = \frac{f_A + f_{B_2} + f_{C_1} + f_{B_3} - (f_{B_1} + f_{C_3} + f_D + f_{C_2})}{u_A + u_{B_2} + u_{C_1} + u_{B_3} - (u_{B_1} + u_{C_3} + u_D + u_{C_2})} \quad (107)$$

with similar expressions for $\frac{\partial g}{\partial u}$, $\frac{\partial h}{\partial u}$, and use the signs of these quantities

to identify the downwind corner. As in two dimensions these approximations are consistent with a calculation of Y_0 based on the mass flow surface integral

$$- \oint_{\text{boundary of } \Omega} (f, g, h) \cdot d\underline{S} .$$

With the above definitions of $\frac{\partial f}{\partial u}$, etc. we can write this Y_0 either in terms of f, g, h or in terms of u, v_1, v_2, v_3 in an obvious interchangeable way (c.f. (79), (80), (81)), and conservation is self evident (c.f. (82)). The use of a downwind form of Y , e.g. (87) jeopardises exact conservation in converging flows and in such flows should be replaced by a more complex mass transfer expression (c.f. Appendix B). The conservation error is not of lower order than the order of accuracy of the scheme, however.

§10. Conclusion

We have seen that a multi-dimensional extension of P.L. Roe's interval approach can be constructed which leads to new 'monotonicity' preserving algorithms. The principle of convection of mass ensures conservation although in converging flows more sophisticated expressions than are used here are needed (see Appendix B). A simple form of Roe's switching device allows the algorithms to be second order without developing oscillations.

To summarise the algorithm, an estimate is first made of the "mass" increment convected into and out of a cell due to motion in the cell and in adjacent cells with local wavespeeds. This mass is normalised to density and allocated to the nodes of the cell with prescribed weights. The weights are chosen to ensure second order accuracy in smooth parts of the flow and monotonicity-preservation in a discretised sense.

Acknowledgements

Part of this work was carried out while the author held a Vacation Consultancy at RAE Farnborough. It is a pleasure to acknowledge the stimulus and help given by P.L. Roe. Thanks are also due to P.K. Sweby for valuable discussions.

References

1. Baines, M.J. (1980), A numerical algorithm for the solution of systems of conservation laws in two dimensions. Numerical Analysis Report 2/80, Dept. of Maths., Univ. of Reading.
2. Roe, P.L. (1979), An improved version of MacCormack's shock capturing algorithm. RAE Technical Report 79041.
3. Roe, P.L. (1981), Numerical algorithms for the linear wave equations. RAE Technical Report, to appear.
4. Roe, P.L. (1981), The use of the Riemann problem in finite difference schemes. Proc. 7th International Conference on Numerical Methods in Fluid Dynamics, Springer.
5. Lax, P.D. and Wendroff, B. (1960), Systems of conservation laws. Comm. Pure & Applied Maths, 13, 217-237.
6. Warming, R.F. and Beam, R.M. (1976), Upwind second order difference schemes and application in aerodynamic flow. AIAA Journal 14, 1241-1249.
7. Godunov, S.K. (1959), A difference method for the numerical calculation of discontinuous solutions of hydrodynamic equations. Mat. Sbornik 47, 3, 271-306. Translated as JPRS 7225 by U.S. Dept. of Commerce, Nov. 1960.
8. Sweby, P.K. (1981), A high order monotonicity preserving algorithm on an irregular grid for non-linear conservation laws. Numerical Analysis Report 1/81, Dept. of Maths., Univ. of Reading.
9. Dukiewicz, J.K. and Ramshaw, J.D. (1979), Tensor viscosity method in numerical fluid dynamics. J. Comp. Phys. 32, 71-79.
10. Fromm, J.E. (1968), A method for reducing dispersion in convective difference schemes. J. Comp. Phys. 3, 176-189.

Appendix A. Less Compact Algorithms

One of the reasons why second order methods cause oscillations for rapidly varying data is that the mass increment can be overestimated when the approximations (7), (40) or (91) are used. Only if the modulus of the gradient is not increasing in the direction of the stream is the true mass increment underestimated or correctly estimated. It is also underestimated in the increasing gradient case when the upwind approximation (101) is used.

A more accurate approximation to (3) is

$$\frac{1}{2}(u_{k+1} + u_k)\Delta x \quad (A1)$$

(see fig. 1), which leads to a mass increment which is the average of (8) and the same expression with k replaced by $k - 1$. This gives rise to

$$g_{k+\frac{1}{2}} = -\frac{1}{2}\frac{\Delta t}{\Delta x}(f_{k+1} - f_{k-1}) \quad (A2)$$

(c.f. (9)).

If we follow through the analysis for determining the weights α, β for the allocation of (A2) to the ends of the interval (x_k, x_{k+1}) , we obtain

$$\alpha = 1 - \frac{1}{2}v, \quad \beta = \frac{1}{2}v, \quad (A3)$$

where $v = a\frac{\Delta t}{\Delta x}$ and the equation is $u_t + au_x = 0$, as usual.

This scheme, being the average of the centred scheme and the upwind scheme, is identical to that of Fromm¹⁰. It has the property of calculating exactly the mass increment for the specimen function $u = x^2$, unlike the earlier schemes. Second order accuracy is maintained at both the 'increment calculation' stage and the 'allocation to node' stage of the algorithm. This scheme will not be monotonicity-preserving since it may still overestimate the true mass increment, but by a much smaller amount. Overshoots are therefore likely to be much milder. On the other hand it is less compact than the schemes described earlier, requiring a mass increment calculation over three points.

A corresponding algorithm can be constructed in two or three dimensions. In two dimensions the approximation corresponding to (A1) is

$$\frac{1}{4}(u_A + u_B + u_C + u_D)\Delta x\Delta y \quad (A4)$$

(see fig. 2). A two-dimensional Fromm type scheme can therefore be constructed. Instead of calculating a new mass increment and new weights, we can achieve the same result by allocating one quarter of each increment over the cell 0 and its three upwind neighbours to the corners of 0 with the appropriate central or "upwind" weights (57), (58).

Similarly, in three dimensions one eighth of each increment in eight cells is given to the eight corners of the basic cell with the appropriate weight.

The results of using the two-dimensional scheme on the test problem of §7, with no attempt at switching, are shown in Appendix C. Switching procedures can be devised for these less compact schemes, although this degrades their compactness further.

Appendix B. Generalised increments

In non-linear converging flows it is not a realistic model to assume that the mass increment in a cell comes from one adjacent cell: all such cells must be considered. In the one-dimensional case, returning to fig. 1, we can evaluate three wave speeds in the three intervals (x_{k-1}, x_k) , (x_k, x_{k+1}) , (x_k, x_{k+2}) , namely, $\left(\frac{\partial f}{\partial u}\right)_{k-\frac{1}{2}}$, $\left(\frac{\partial f}{\partial u}\right)_{k+\frac{1}{2}}$, $\left(\frac{\partial f}{\partial u}\right)_{k+\frac{3}{2}}$. These may be estimated by the ratios

$$\frac{f_k - f_{k-1}}{u_k - u_{k-1}}, \quad \frac{f_{k+1} - f_k}{u_{k+1} - u_k}, \quad \frac{f_{k+2} - f_{k+1}}{u_{k+2} - u_{k+1}} \quad (B1)$$

and we denote them by $a_{k-\frac{1}{2}}$, $a_{k+\frac{1}{2}}$, $a_{k+\frac{3}{2}}$. Assuming that each cell draws in a convected mass and has mass drawn out from it by adjacent cells, the interval (x_k, x_{k+1}) will receive a net mass increment

$$\Delta m_{k+\frac{1}{2}} = a_{k+\frac{1}{2}}^+ \frac{\Delta t}{\Delta x} m_{k-\frac{1}{2}} + a_{k+\frac{1}{2}}^- \frac{\Delta t}{\Delta x} m_{k+\frac{3}{2}} - a_{k-\frac{1}{2}}^- \frac{\Delta t}{\Delta x} m_{k+\frac{1}{2}} - a_{k+\frac{3}{2}}^+ \frac{\Delta t}{\Delta x} m_{k+\frac{1}{2}}, \quad (B2)$$

where

$$a_l^+ = \begin{cases} a_l & a_l \geq 0 \\ 0 & a_l < 0 \end{cases} \quad a_l^- = \begin{cases} 0 & a_l > 0 \\ a_l & a_l \leq 0 \end{cases} \quad (B3)$$

Each interval now has a unique downwind end but it will vary from interval to interval. A downwind mass estimate like (7) will vary from interval to interval but there is no danger of losing mass balance because of the construction of (B2). We write

$$m_l = \Delta x \tilde{u}_l \quad (B3)$$

where

$$\tilde{u}_l = \begin{cases} u_{l+\frac{1}{2}} & (a_l \geq 0) \\ u_{l-\frac{1}{2}} & (a_l < 0) \end{cases} \quad (B4)$$

Then (B2) can be written

$$\Delta m_{k+\frac{1}{2}} = a_{k+\frac{1}{2}}^+ \Delta t \tilde{u}_{k-\frac{1}{2}} + a_{k+\frac{1}{2}}^- \Delta t \tilde{u}_{k+\frac{3}{2}} - \bar{a}_{k-\frac{1}{2}} \Delta t \tilde{u}_{k+\frac{1}{2}} - a_{k+\frac{3}{2}}^+ \Delta t \tilde{u}_{k+\frac{1}{2}} \quad (\text{B5})$$

$$= [v_{k+\frac{1}{2}}^+ \tilde{u}_{k-\frac{1}{2}} + v_{k+\frac{1}{2}}^- \tilde{u}_{k+\frac{3}{2}} - v_{k-\frac{1}{2}}^- \tilde{u}_{k+\frac{1}{2}} - v_{k+\frac{3}{2}}^+ \tilde{u}_{k+\frac{1}{2}}] \Delta x \quad (\text{B6})$$

$$\text{where } a_{\ell} \frac{\Delta t}{\Delta x} = v_{\ell}, \quad a_{\ell}^+ \frac{\Delta t}{\Delta x} = v_{\ell}^+, \quad a_{\ell}^- \frac{\Delta t}{\Delta x} = v_{\ell}^- \quad (\text{B7})$$

Normalising to density, we obtain

$$g_{k+\frac{1}{2}} = v_{k+\frac{1}{2}}^+ \tilde{u}_{k-\frac{1}{2}} + v_{k+\frac{1}{2}}^- \tilde{u}_{k+\frac{3}{2}} - v_{k-\frac{1}{2}}^- \tilde{u}_{k+\frac{1}{2}} - v_{k+\frac{3}{2}}^+ \tilde{u}_{k+\frac{1}{2}}, \quad (\text{B8})$$

which is the quantity to be incremented (c.f. (16), (12)).

Weights for second order accuracy will be evaluated just as in §2 since in smooth one way flows (B8) reduces to (16). Indeed in the one-dimensional case there is virtually no change in the method.

When we consider two dimensional flows the difference becomes more crucial because of the difficulties with conservation mentioned in §6. Using fig. 4, suppose that we know wave speeds (a, b) for each cell which may vary in sign. Then

$$\begin{aligned} \Delta m_0 = & v_1^+ (1 - v_2) m_W + v_1^- (1 - v_2) m_E + v_2^+ (1 - v_1) m_S + v_2^- (1 - v_1) m_N \\ & + v_1^+ v_2^+ m_{SW} + v_1^- v_2^+ m_{SE} + v_1^+ v_2^- m_{NW} + v_1^- v_2^- m_{NE} \\ & - v_{1,W}^- (1 - v_{2,W}) m_0 - v_{1,E}^+ (1 - v_{2,E}) m_0 - v_{2,S}^- (1 - v_{1,S}) m_0 \\ & - v_{2,N}^+ (1 - v_{1,N}) m_0 - v_{1,SW}^- v_{2,SW}^- m_0 - v_{1,SE}^+ v_{2,SE}^- m_0 \\ & - v_{1,NW}^- v_{2,NW}^+ m_0 - v_{1,NE}^+ v_{2,NE}^+ m_0, \end{aligned} \quad (\text{B9})$$

$$\text{where } v^+ = v_0^+ = \begin{cases} v & v \geq 0 \\ 0 & v < 0 \end{cases} \quad v^- = v_0^- = \begin{cases} 0 & v \geq 0 \\ v & v < 0 \end{cases} \quad (\text{B10})$$

and similarly for $v_W, v_{SW}, v_S, v_{SE}, v_E, v_{NE}, v_N, v_{NW}$.

Now let

$$m_U = \Delta x \Delta y u_U \quad (\text{B11})$$

where

$$\tilde{u}_0 = \begin{cases} u_C & v_1, v_2 \geq 0 \\ u_D & v_1 < 0, v_2 \geq 0 \\ u_A & v_1, v_2 < 0 \\ u_B & v_1 \geq 0, v_2 < 0 \end{cases} \quad (\text{B12})$$

with similar definitions for $m_W, m_{SW}, m_S, m_{SE}, m_E, m_{NE}, m_N, m_{NW}$. Then, dividing by $\Delta x \Delta y$, we obtain

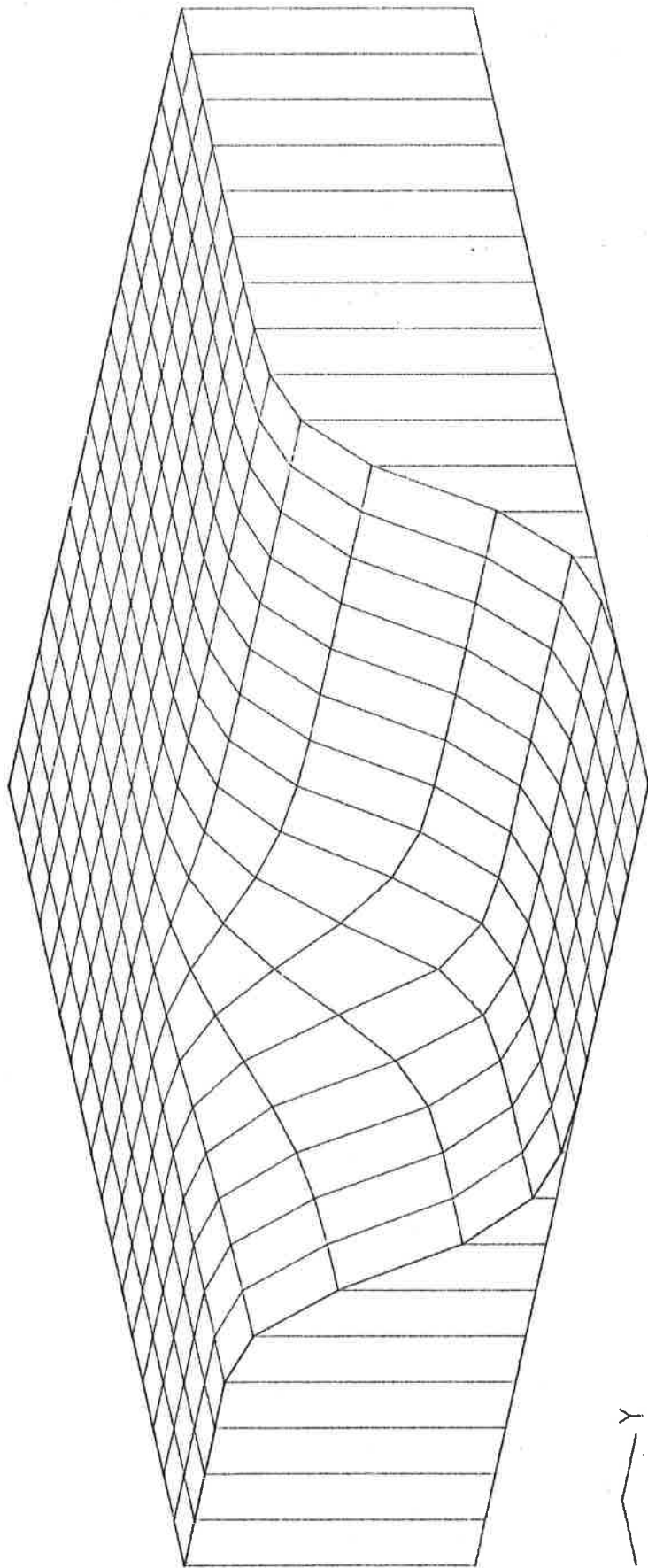
$$\begin{aligned} Y_0 = & v_1^+(1 - v_2)\tilde{u}_W + v_1^-(1 - v_2)\tilde{u}_E + v_2^+(1 - v_1)\tilde{u}_S + v_2^-(1 - v_1)\tilde{u}_N \\ & + v_1^+v_2^+\tilde{u}_{SW} + v_1^-v_2^+\tilde{u}_{SE} + v_1^+v_2^-\tilde{u}_{NW} + v_1^-v_2^-\tilde{u}_{NE} \\ & - v_{1,W}^-(1 - v_{2,W})\tilde{u}_0 - v_{1,SW}^-v_{2,SW}^-\tilde{u}_0 - v_{2,S}^-(1 - v_{1,S})\tilde{u}_0 \\ & - v_{2,N}^+(1 - v_{1,N})\tilde{u}_0 - v_{1,SW}^-v_{2,SW}^-\tilde{u}_0 - v_{1,SE}^+v_{2,SE}^-\tilde{u}_0 \\ & - v_{1,NW}^-v_{2,NW}^+\tilde{u}_0 - v_{1,NE}^+v_{2,NE}^+\tilde{u}_0, \end{aligned} \quad (\text{B13})$$

a single expression for Y_0 (c.f. (42), (44), (45), (46)), which, because of the mass balance, always ensures conservation if completely allocated in the sense of (82).

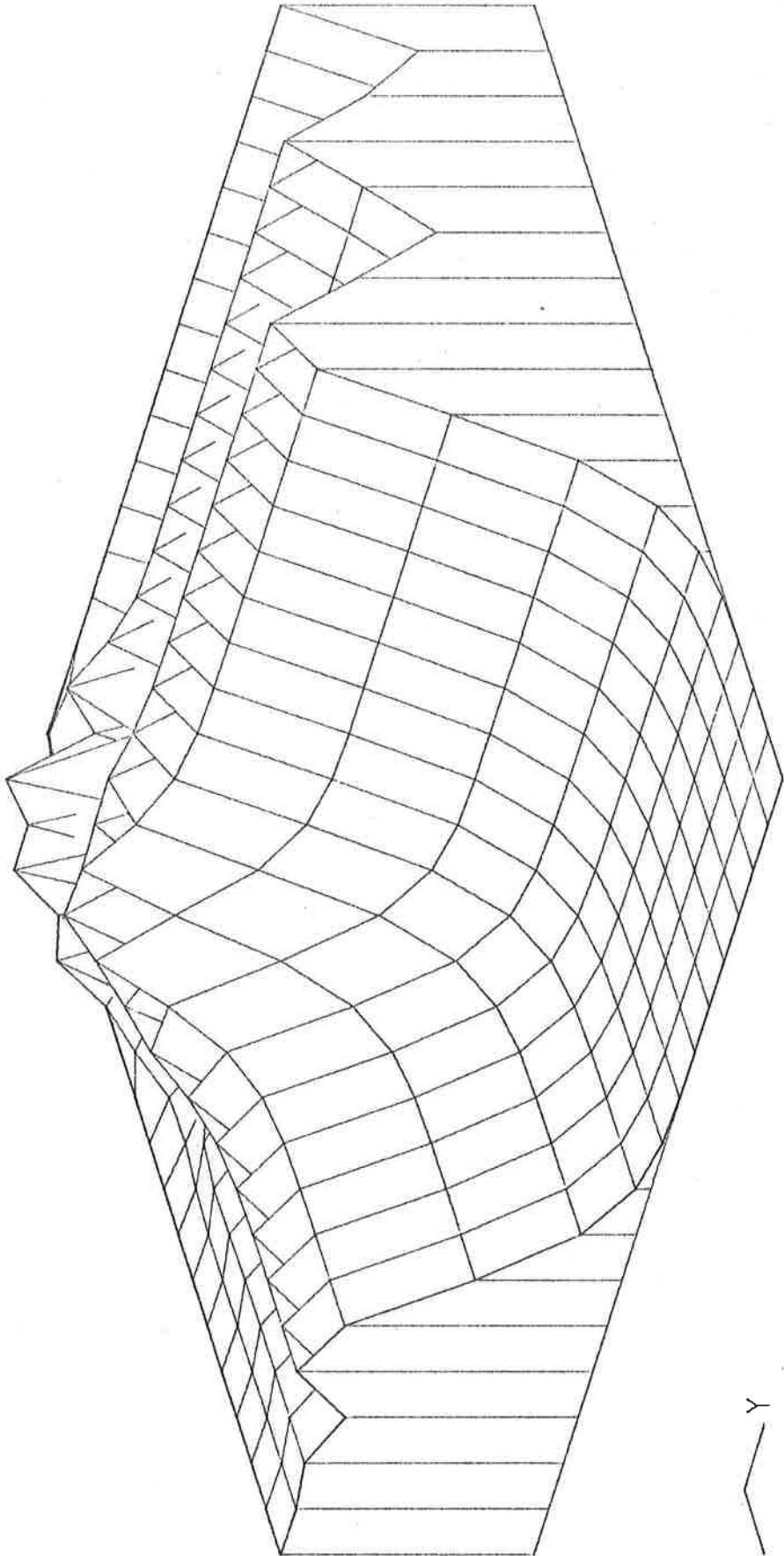
Second order weights (derived from smooth solutions) are as before and switching will ensure smoothing without affecting conservation. A similar approach is available in three dimensions.

Appendix C. Results for the problem in §7

Numerical results as detailed in §7	45
Surface plot of table 1	46
Surface plot of table 2	47
Surface plot of the scheme in Appendix A applied to the problem in §7 (not scaled)	48

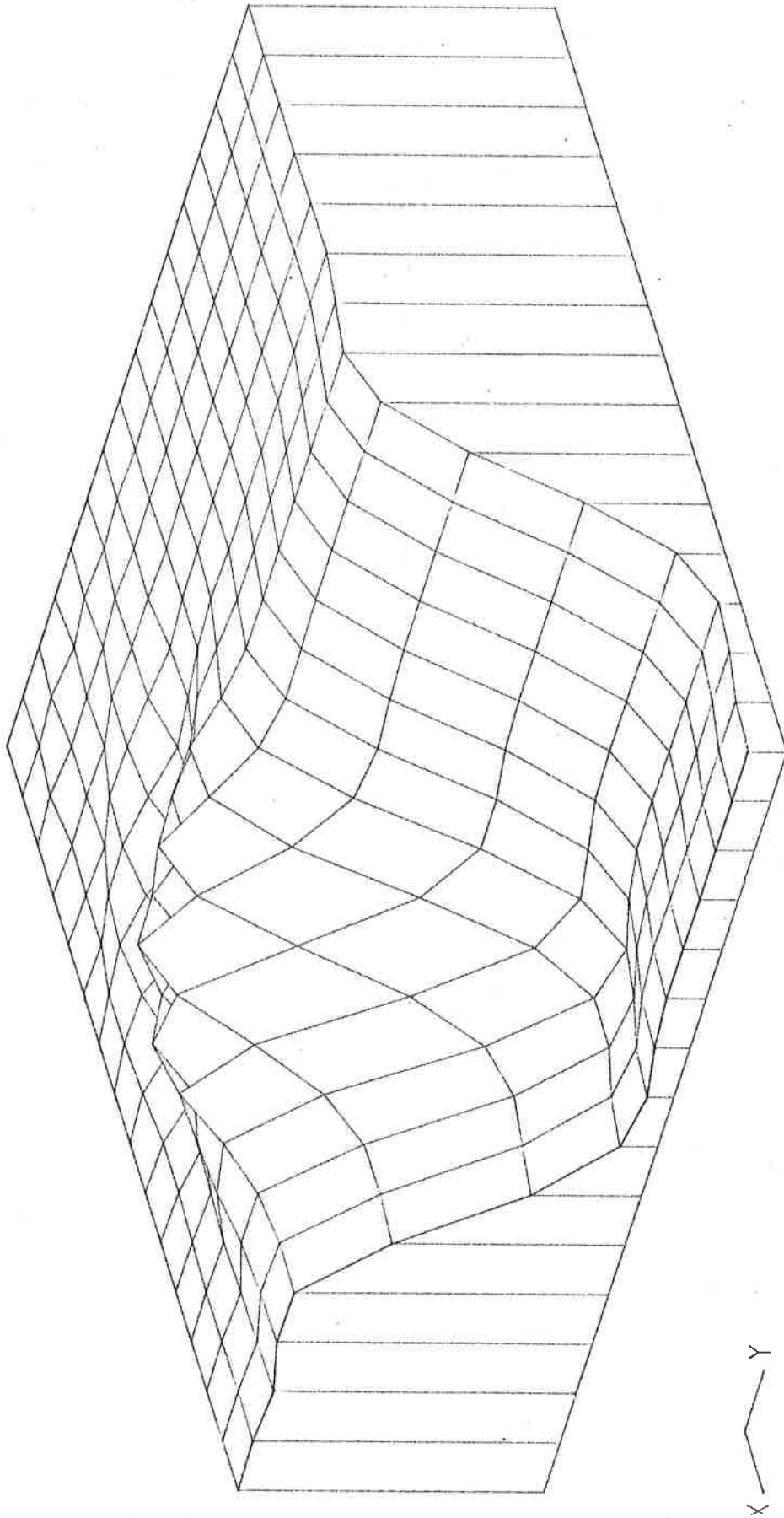


PLOT 1.



X — Y

PLOT 2.



PLOT 3.