ELSEVIER

Applied Numerical Mathematics 16 (1995) 417–438

APPLIED NUMERICAL MATHEMATICS

Positive cell-centered finite volume discretization methods for hyperbolic equations on irregular meshes

M. Berzins *, J.M. Ware

School of Computer Studies, University of Leeds, Leeds LS2 9JT, UK

Abstract

The conditions sufficient to ensure positivity and linearity preservation for a cell-centered finite volume scheme for time-dependent hyperbolic equations using irregular one-dimensional and triangular two-dimensional meshes are derived. The conditions require standard flux limiters to be modified and also involve possible constraints on the meshes. The accuracy of this finite volume scheme is considered and is illustrated by two simple numerical examples.

1. Introduction

An important trend in numerical methods for the spatial discretization of partial differential equations is the move towards using finite element and finite volume methods on unstructured triangular or tetrahedral meshes. The reasoning underlying this trend is that such methods offer one way of solving problems adaptively on general geometries. The finite volume methods used may be split into cell-vertex methods (in which the solution values are positioned at mesh points) and cell-centered methods (in which the solution variables are positioned at the centroids of triangles). Cell-vertex methods have a clear advantage over cell-centered methods in that there are fewer unknowns for a given mesh, but a possible disadvantage is that the area (or volume) of each cell is larger. While both methods have their advocates what is clear is that both classes of methods need to be well-understood. In this respect more work has been done on the analysis and derivation of cell-vertex schemes, e.g. see Barth [1], Struijs et al. [17] and van Leer [18] and the references therein. One of the early papers to make an important advance in this area was that of Cockburn et al. [7] which proves a maximum principle for a discontinuous Galerkin method of order $k + 1$ which may be interpreted as a finite volume type scheme.

In the area of cell-centered schemes on triangles perhaps the first extension of successful one-dimensional schemes to triangles was that of Venkatakrishnan and Barth [19]. Subsequent

* Corresponding author.

SSDI 0168-9274(95)00007-0

modifications (e.g. by Lin et al. [11]) and independent developments (by Berzins et al. [3,6,20]) occurred shortly afterwards. These schemes all attempt to transfer successful regular one-dimensional and quadrilateral mesh two-dimensional schemes (e.g. [16]) to unstructured triangular meshes. The scheme of Durlofsky et al. [8] has similarities with these methods, except that the limited upwind interpolants used are different. More recently Liu [12] showed that a modified form of this method satisfies a maximum principle.

The intention in this paper is to show that the schemes of Ware and Berzins [20] and Venkatakrishnan and Barth [19] satisfy the properties of *linearity preservation* and *positivity*. These properties have been proposed by Struijs et al. [17] as being of importance for multi-space dimensional schemes. The positivity analysis of such methods has often been confined to regular mesh cases (e.g. Spekreijse [16]). The intention in this paper is to extend Spekreijse's analysis to the one-dimensional irregular mesh case and then to the unstructured triangular mesh algorithm of Ware and Berzins [20]. This paper will show that the new scheme has these properties under certain restrictions on the limiter function, the mesh and on the interpolating functions used in the discretization method. The analysis is extended to time integration using the Theta method in a method of lines approach, [2].

An outline of this paper is as follows. Section 2 describes the spatial discretization method analyzed by Spekreijse. The extension of this method to irregular meshes is considered in Section 3. The issue of positive time integration is considered in Section 4. Section 5 extends the approach to unstructured triangular meshes and considers accuracy issues. Section 6 considers the linearity preservation and positivity of the scheme while Section 7 illustrates these results using two simple numerical examples.

2. Spekreijse's discretization method

Spekreijse [16] considers regular square meshes in two-space dimensions by splitting the computation dimensionally. This makes it possible to consider the extension to irregular meshes by looking at the scalar partial differential equation in one space dimension given by

$$u_t + [f(u)]_x = 0, \tag{1}$$

where $f(u)$ is the advective flux function which describes *wave* movements in the solution. Spekreijse [16] assumes that this can be split into positive and negative parts:

$$f(u) = f_\ell(u) + f_r(u), \tag{2}$$

where

$$\frac{\mathrm{d}f_\ell(u)}{\mathrm{d}u} \geqslant 0, \qquad \frac{\mathrm{d}f_r(u)}{\mathrm{d}u} \leqslant 0. \tag{3}$$

In this paper a slightly different set of conditions, due to Cockburn et al. [7], which restricts only the numerical flux function will be used, see below. The analysis undertaken will apply equally to both cases, however.

A spatial mesh, with constant spacing h, is defined by

$$x_{i+1} = x_i + h, \quad i = 1, \ldots, n, \qquad x_1 = a,$$

and the midpoints by $x_{i+1/2} = x_i + \frac{1}{2}h$.

Denote by $U_i(t)$ the solution value $U(x_i, t)$ at the meshpoint x_i at time t. Throughout the paper it will be assumed that all solution values, derivatives and fluxes depend on the time t. The semi-discrete form of (1) is

$$\frac{\partial U_i}{\partial t} + \frac{f_{i+1/2} - f_{i-1/2}}{h} = 0, \tag{4}$$

where $f_{i+1/2}$ and $f_{i-1/2}$ are the fluxes at the midpoints $x_{i+1/2}$ and $x_{i-1/2}$ respectively. Spekreijse's method [16] makes use of an approximate Riemann solver such as the well-known Roe or Osher solvers to calculate these fluxes. The flux calculated by this approximate Riemann solver will be defined as

$$f_{\mathrm{Rm}}\left(U^{\ell}_{i+1/2}, U^{\mathrm{r}}_{i+1/2}\right) \tag{5}$$

and, following Cockburn et al. [7], is assumed to satisfy:

- $f_{\mathrm{Rm}}(u, u) = f(u)$;
- $f_{\mathrm{Rm}}(u, v)$ is nondecreasing in u and nonincreasing in v;
- $f_{\mathrm{Rm}}(\cdot, \cdot)$ is Lipschitz;
- $f_{\mathrm{Rm}}(u, v) = -f_{\mathrm{Rm}}(v, u)$.

In order to use this approach it is necessary to construct left, $U^{\ell}_{i+1/2}$, and right, $U^{\mathrm{r}}_{i+1/2}$, solution values at the midpoints $x_{i+1/2}$. A standard first-order scheme uses $U_i(t)$ as the left value and $U_{i+1}(t)$ as the right value. In Spekreijse's second-order scheme the limited left and right solution values at the cell interface $x_{i+1/2}$ are defined by

$$U^{\ell}_{i+1/2} = U_i + \tfrac{1}{2}(U_i - U_{i-1})\Phi(r_i), \tag{6}$$

$$U^{\mathrm{r}}_{i+1/2} = U_{i+1} - \tfrac{1}{2}(U_{i+2} - U_{i+1})\Phi\left(\frac{1}{r_{i+1}}\right), \tag{7}$$

where $U^{\ell}_{i+1/2}$ and $U^{\mathrm{r}}_{i+1/2}$ are the limited upwind solutions on the left and right respectively. The ratio of gradients, r_i, and the limiter function, $\Phi(\cdot)$, are defined as

$$r_i = \frac{U_{i+1} - U_i}{U_i - U_{i-1}}, \qquad \Phi(R) = \frac{R + |R|}{1 + |R|}, \tag{8}$$

where $\Phi(\cdot)$ is van Leer's harmonic limiter, [16].

The semi-discrete form of (1) now becomes

$$\frac{\partial U_i}{\partial t} = \frac{1}{h}\left[-f_{\mathrm{Rm}}\left(U^{\ell}_{i+1/2}, U^{\mathrm{r}}_{i+1/2}\right) + f_{\mathrm{Rm}}\left(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i-1/2}\right)\right],$$

where $f_{\mathrm{Rm}}(U^{\ell}, U^{\mathrm{r}})$ denotes the flux value calculated by solving the approximate Riemann problem with left and right states U^{ℓ} and U^{r} respectively.

Spekreijse splits the flux function, f, into its positive and negative parts as in (2) and uses the forward Euler method with time step k to get the equations:

$$U_i(t_{n+1}) = U_i(t_n) + \frac{k}{h}\left[f_{\mathrm{r}}\left(U^{\ell}_{i-1/2}\right) - f_{\mathrm{r}}\left(U^{\ell}_{i+1/2}\right) - f_{\ell}\left(U^{\mathrm{r}}_{i-1/2}\right) + f_{\ell}\left(U^{\mathrm{r}}_{i+1/2}\right)\right],$$

where $i = 1, \ldots, n$ and $t_{n+1} = t_n + k$.

3. One-dimensional variable mesh formulation

There are two alternative formulations that allow the one-dimensional flux limiter scheme described above to be used on non-uniform meshes. One is a cell-vertex approach, as used in the software of Pennington and Berzins [13], and the other is a cell-centered approach. The cell-centered approach is closer to the two-dimensional case of interest and so will be considered first. In this case the point x_i is assumed to be at the center of a cell of width h_i, and so the spatial mesh is defined by

$$x_{i+1} = x_i + \tfrac{1}{2}(h_i + h_{i+1}), \quad i = 1, \ldots, n, \qquad x_1 = a,$$

and the midpoints by $x_{i+1/2} = x_i + \frac{1}{2}h_i = x_{i+1} - \frac{1}{2}h_{i+1}$.

Three new terms are introduced to cater for the irregular mesh. The first two are the linearly extrapolated upwind values on the left and right of the cell interface: $U^{\mathrm{L}}_{i+1/2}$ and $U^{\mathrm{R}}_{i+1/2}$. The third is the linearly interpolated centered value, $U^{\mathrm{C}}_{i+1/2}$. These terms are defined as follows:

$$U^{\mathrm{L}}_{i+1/2} = U_i + \frac{h_i(U_i - U_{i-1})}{h_{i-1} + h_i}, \tag{9}$$

$$U^{\mathrm{R}}_{i+1/2} = U_{i+1} - \frac{h_{i+1}(U_{i+2} - U_{i+1})}{h_{i+1} + h_{i+2}}, \tag{10}$$

$$U^{\mathrm{C}}_{i+1/2} = U_i + \frac{h_i(U_{i+1} - U_i)}{h_i + h_{i+1}} \tag{11}$$

$$= U_{i+1} - \frac{h_{i+1}(U_{i+1} - U_i)}{h_{i+1} + h_i}, \tag{12}$$

where dependence of the solution values on the time t has been omitted but is understood.

The limited upwind value on the left of the cell interface is given by a modified form of (6), i.e.

$$U^{\ell}_{i+1/2} = U_i + h_i \frac{(U_i - U_{i-1})}{h_i + h_{i-1}} \Phi\left(r^{\ell}_{i+1/2}\right),$$

where the limiter function $\Phi(\cdot)$ may be defined as in (8), and the ratio of gradients with left upwind bias is $r^{\ell}_{i+1/2}$, rather than r_i, and will be defined below. This equation can be rewritten using (9) as

$$U^{\ell}_{i+1/2} = U_i + \Phi\left(r^{\ell}_{i+1/2}\right)\left(U^{\mathrm{L}}_{i+1/2} - U_i\right). \tag{13}$$

A similar process gives the limited upwind value on the right,

$$U^{\mathrm{r}}_{i+1/2} = U_{i+1} + \Phi\left(r^{\mathrm{r}}_{i+1/2}\right)\left(U^{\mathrm{R}}_{i+1/2} - U_{i+1}\right). \tag{14}$$

The irregular mesh equivalent of the ratio of the regular mesh gradients r_i as defined in (8) is

$$r^{\ell}_{i+1/2} = \left[\frac{U_{i+1} - U_i}{\frac{1}{2}(h_i + h_{i+1})}\right] \times \left[\frac{U_i - U_{i-1}}{\frac{1}{2}(h_i + h_{i-1})}\right]^{-1},$$

which may be rewritten using (9)–(12) as

$$r^{\ell}_{i+1/2} = \left[U^{\mathrm{C}}_{i+1/2} - U_i\right] \times \left[U^{\mathrm{L}}_{i+1/2} - U_i\right]^{-1}.$$

Using a similar process on the right, the ratio of gradients is

$$r^{\mathrm{r}}_{i+1/2} = \left[- \frac{U_{i+1} - U_i}{\frac{1}{2}(h_i + h_{i+1})} \right] \times \left[- \frac{U_{i+2} - U_{i+1}}{\frac{1}{2}(h_{i+1} + h_{i+2})} \right]^{-1},$$

which may be again rewritten using (9)–(12) as

$$r^{\mathrm{r}}_{i+1/2} = \left[U^{\mathrm{C}}_{i+1/2} - U_{i+1}\right] \times \left[U^{\mathrm{R}}_{i+1/2} - U_{i+1}\right]^{-1}.$$

The limiter function $\Phi(\cdot)$ is assumed to be unchanged for the moment.

Using the values $U^{\mathrm{L}}_{i+1/2}$, $U^{\mathrm{R}}_{i+1/2}$ and $U^{\mathrm{C}}_{i+1/2}$, the scheme devised by Spekreijse can be extended to the irregular mesh case. Substituting the values defined by (13) and (14) into (4) enables the scheme to be written as

$$\frac{\partial U_i}{\partial t} = \frac{1}{h_i}\left[-f_{\mathrm{Rm}}\left(U^{\ell}_{i+1/2}, U^{\mathrm{r}}_{i+1/2}\right) + f_{\mathrm{Rm}}\left(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i-1/2}\right)\right].$$

Addition and subtraction of the term $f_{\mathrm{Rm}}(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i+1/2})$ gives

$$\begin{aligned} h_i \frac{\partial U_i}{\partial t} = &- \left[f_{\mathrm{Rm}}\left(U^{\ell}_{i+1/2}, U^{\mathrm{r}}_{i+1/2}\right) - f_{\mathrm{Rm}}\left(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i+1/2}\right)\right] \\ &+ \left[f_{\mathrm{Rm}}\left(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i-1/2}\right) - f_{\mathrm{Rm}}\left(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i+1/2}\right)\right]. \end{aligned}$$

At a particular time t_n this can now be written as

$$\frac{\partial U_i}{\partial t} = A^n_{i+1/2}(U_{i+1}(t_n) - U_i(t_n)) - B^n_{i-1/2}(U_i(t_n) - U_{i-1}(t_n)), \tag{15}$$

where

$$A^n_{i+1/2} = -\frac{1}{h_i} \frac{f_{\mathrm{Rm}}\left(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i+1/2}\right) - f_{\mathrm{Rm}}\left(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i-1/2}\right)}{U^{\mathrm{r}}_{i+1/2} - U^{\mathrm{r}}_{i-1/2}} \frac{U^{\mathrm{r}}_{i+1/2} - U^{\mathrm{r}}_{i-1/2}}{U_{i+1}(t_n) - U_i(t_n)},$$

$$B^n_{i-1/2} = \frac{1}{h_i} \frac{f_{\mathrm{Rm}}\left(U^{\ell}_{i+1/2}, U^{\mathrm{r}}_{i+1/2}\right) - f_{\mathrm{Rm}}\left(U^{\ell}_{i-1/2}, U^{\mathrm{r}}_{i+1/2}\right)}{U^{\ell}_{i+1/2} - U^{\ell}_{i-1/2}} \frac{U^{\ell}_{i+1/2} - U^{\ell}_{i-1/2}}{U_i(t_n) - U_{i-1}(t_n)}.$$

Spekreijse's flux splitting approach leads to very similar coefficients:

$$A^n_{i+1/2} = -\frac{1}{h_i} \frac{f_{\ell}\left(U^{\mathrm{r}}_{i+1/2}\right) - f_{\ell}\left(U^{\mathrm{r}}_{i-1/2}\right)}{U^{\mathrm{r}}_{i+1/2} - U^{\mathrm{r}}_{i-1/2}} \frac{U^{\mathrm{r}}_{i+1/2} - U^{\mathrm{r}}_{i-1/2}}{U_{i+1}(t_n) - U_i(t_n)},$$

$$B^n_{i-1/2} = \frac{1}{h_i} \frac{f_{\mathrm{r}}\left(U^{\ell}_{i+1/2}\right) - f_{\mathrm{r}}\left(U^{\ell}_{i-1/2}\right)}{U^{\ell}_{i+1/2} - U^{\ell}_{i-1/2}} \frac{U^{\ell}_{i+1/2} - U^{\ell}_{i-1/2}}{U_i(t_n) - U_{i-1}(t_n)}.$$

Applying the forward Euler method with time step k gives:

$$U_i(t_{n+1}) = U_i(t_n) + kA^n_{i+1/2}(U_{i+1}(t_n) - U_i(t_n)) - kB^n_{i-1/2}(U_i(t_n) - U_{i-1}(t_n)).$$

The definition of positivity, [17], requires that every new value $U_i(t_{n+1})$ can be written as a convex combination of old values:

$$U_i(t_{n+1}) = \sum_{j=1}^{n} c_j U_j(t_n) \quad \forall c_j \geqslant 0, \tag{16}$$

while $\Sigma c_j = 1$ for consistency. This guarantees, [17], a maximum principle for the discrete steady state solution thus prohibiting the occurrence of new extrema and imposing stability on the explicit scheme. From this definition the requirement on the coefficients $A^n_{i+1/2}$ and $B^n_{i-1/2}$ is that

$$A^n_{i+1/2} \geqslant 0, \qquad B^n_{i-1/2} \geqslant 0, \qquad 1 - kA^n_{i+1/2} - kB^n_{i-1/2} \geqslant 0.$$

Application of the mean value theorem to the definitions of the coefficients $A^n_{i+1/2}$ and $B^n_{i-1/2}$ and use of either Spekreijse's flux function splitting properties defined in (2), or the Riemann solver properties defined in (5), show that this requires that

$$\frac{U^{\mathrm{r}}_{i+1/2} - U^{\mathrm{r}}_{i-1/2}}{U_{i+1}(t_n) - U_i(t_n)} \geqslant 0, \qquad \frac{U^{\ell}_{i+1/2} - U^{\ell}_{i-1/2}}{U_i(t_n) - U_{i-1}(t_n)} \geqslant 0.$$

Consider the right-hand term for example. Substituting from (13) and (9) gives

$$\frac{U^{\ell}_{i+1/2} - U^{\ell}_{i-1/2}}{U_i(t_n) - U_{i-1}(t_n)} = 1 + \frac{h_i}{h_i + h_{i-1}}\Phi\left(r^{\ell}_{i+1/2}\right) - \frac{h_{i-1}}{h_i + h_{i-1}}\frac{\Phi\left(r^{\ell}_{i-1/2}\right)}{r^{\ell}_{i-1/2}}.$$

Following Spekreijse, this is positive if

$$1 + \frac{h_i}{h_i + h_{i-1}}\Phi(R) - \frac{h_{i-1}}{h_i + h_{i-1}}\frac{1}{S}\Phi(S) \geqslant 0 \quad \forall R,S. \tag{17}$$

From this equation and Spekreijse's equation (2.13) in [16] it follows that

$$\alpha \leqslant \frac{2h_i}{h_i + h_{i-1}}\Phi(R) \leqslant M, \qquad -M \leqslant \frac{2h_{i-1}}{h_i + h_{i-1}}\frac{\Phi(R)}{R} \leqslant 2 + \alpha,$$

where $\alpha \in [-2, 0]$ and M is a positive constant. In other words the standard limiter $\Phi(R)$ in Spekreijse's equation (2.13) is replaced by the limiter $\Phi(R)$ multiplied by $2h_i/(h_i + h_{i-1})$. A slight rearrangement of Eq. (17) gives:

$$\frac{h_i}{h_i + h_{i-1}}(1 + \Phi(R)) + \frac{h_{i-1}}{h_i + h_{i-1}}\left(1 - \frac{\Phi(S)}{S}\right) \geqslant 0 \quad \forall R,S.$$

Consideration of extreme mesh ratios in this equation shows that the limiter must satisfy

$$-1 \leqslant \Phi(R) \leqslant M, \qquad -M \leqslant \frac{1}{S}\Phi(S) \leqslant 1 \quad \forall R,S. \tag{18}$$

This shows that standard limiters may need to be modified for the irregular mesh case. For example the van Leer limiter as defined in (8) may be replaced by one which satisfies (18) with $M = 2$, i.e.

$$\Phi(R) = \frac{R + |R|}{1 + \max(1, |R|)}. \tag{19}$$

This new limiter will henceforth be referred to as the modified van Leer limiter in the remainder of this paper.

Remark. In the case when the mesh cells are defined by

$$x_{i+1} = x_i + h_i, \quad i = 1, \ldots, n, \qquad x_1 = a,$$

and the midpoints by $x_{i+1/2} = x_i + \frac{1}{2}h_i$, as in the software of Pennington and Berzins [13], a similar analysis to that above leads to an equivalent equation to (17) given by

$$2 + \frac{h_i}{h_{i-1}}\Phi(R) - \frac{1}{S}\Phi(S) \geqslant 0 \quad \forall R, S.$$

From this it follows that the van Leer limiter may be used without modification in a cell-vertex scheme but other limiters that allow negative values when the mesh ratio h_i/h_{i-1} is large will need to be modified to preserve positivity. For example, if the van Albeda limiter used by Spekreijse and Venkatakrishnan and Barth [19] and defined by

$$\Phi(R) = \frac{R + R^2}{1 + R^2} \tag{20}$$

is used and $R = -0.5$, then $\Phi(-0.5) = -\frac{1}{5}$ and a mesh ratio value of $h_i/h_{i-1} = 10$ will result in the positivity condition being violated.

3.1. Systems of equations

The present proof extends to systems of equations without difficulty providing flux vector splitting is used to decompose the flux function into *positive* and *negative* fluxes (see Roe [14]). The extension to using the Roe and Osher type approximate Riemann solvers is beyond the scope of this paper.

4. Time integration

The above spatial discretization scheme results in a system of differential equations, each of which is of the form of Eq. (4). This system of equations can be written as the initial value problem:

$$\dot{U} = F_N(t, U(t)), \qquad U(0) \text{ given}, \tag{21}$$

where the N-dimensional vector, $\boldsymbol{U}(t)$, is defined by

$$\boldsymbol{U}(t) = \left[U(x_1, t), U(x_2, t), \ldots, U(x_N, t)\right]^{\mathrm{T}}.$$

The point x_i is the center of the ith cell and $U_i(t)$ is a numerical approximation to $u(x_i, t)$. Although Section 3 showed that the discretization scheme is positive when used with the forward Euler method it is necessary to extend this analysis to the method of time integration used by Berzins and Ware [6] and Berzins [2]. Numerical integration of (21) provides the approximation, $V(t)$, to the vector of exact PDE solution values at the mesh points, $\boldsymbol{u}(t)$:

$$V(t) = \left[V(x_1, t), V(x_2, t), \ldots, V(x_N, t)\right]^{\mathrm{T}}.$$

The Theta method code of Berzins and Furzeland [4] used here selects functional iteration automatically for the non-stiff ODEs resulting from convection-dominated problems. The numerical solution at $t_{n+1} = t_n + k$, where k is the time step size, as denoted by $V(t_{n+1})$, is defined by

$$V(t_{n+1}) = V(t_n) + (1-\theta)k\dot{V}(t_n) + \theta k \boldsymbol{F}_N(t_{n+1}, V(t_{n+1})),$$

in which $V(t_n)$ and $\dot{V}(t_n)$ are the numerical solution and its time derivative at the previous time t_n. The value of θ used is bounded by $0.5 \leqslant \theta < 1.0$, and may be chosen by the user or automatically varied to increase the time step, [4]. Values of θ close to 0.5 (e.g. 0.55) give the benefits of almost second-order accuracy plus added stability (see [4] for a detailed discussion of this matter). The time step k is chosen to satisfy a local error control which may be modified to reflect the spatial error present, [2]. The system of equations (4) is solved using functional iteration (see [2]),

$$V^{(m+1)}(t_{n+1}) = V(t_n) + (1-\theta)k\dot{V}(t_n) + \theta k \boldsymbol{F}_N\left(t_{n+1}, V^{(m)}(t_{n+1})\right), \tag{22}$$

where $m = 0, 1, \ldots,$ generally less than 3 and using a second-order predictor or with a predictor based on the forward Euler method:

$$V^{(0)}(t_{n+1}) = V(t_n) + k\boldsymbol{F}_N(t_n, V(t_n)). \tag{23}$$

Berzins [2] shows that one adavantage of using functional iteration is that a Courant number type stability condition is automatically satisfied if functional iteration converges sufficiently fast. The more difficult issue of positivity will be considered below.

Remark. It is possible for the user to select $\theta = 0.5$ and to allow only one corrector iteration to be performed in which case the method is the second-order positivity-preserving Runge–Kutta method used by Shu and Osher [15].

In order to show that the coupling of this time integration scheme with a spatial discretization method is positive, the precise form of the ODE system must be stated, i.e.

$$F_i(t_n, V(t_n)) = -a_i V_i(t_n) + S_N^i(V(t_n)) \tag{24}$$

$$\text{where } S_N^i(V(t_n)) = \sum_{j \neq i} c_{ij} V_j(t_n),$$

and where from Eq. (15) the coefficients $c_{i,j}$ are zero except for

$$c_{i,i+1} = A^n_{i+1/2}, \qquad c_{i,i-1} = B^n_{i-1/2}, \qquad a_i = A^n_{i+1/2} + B^n_{i-1/2}, \tag{25}$$

thus making $S^i_N(V(t_n))$ a positive function for positive values of $V(t_n)$.

Applying the predictor to the ith equation gives

$$V_i^{(0)}(t_{n+1}) = (1 - ka_i)V_i(t_n) + kS^i_N(V(t_n)).$$

Substituting this value in the corrector gives

$$\begin{aligned} V_i^{(1)}(t_{n+1}) = V_i(t_n) - a_i k\theta\big[(1 - ka_i)V_i(t_n) + kS^i_N(V(t_n))\big] \\ + k\theta S^i_N\big(V^{(0)}(t_{n+1})\big) + k(1-\theta)\big[-a_i V_i(t_n) + S^i_N(V(t_n))\big], \end{aligned}$$

which may be written as

$$V_i^{(1)}(t_{n+1}) = V_i(t_n)\big[1 - ka_i + \theta k^2 a_i^2\big] + k\big[1 - 2k\theta a_i\big]S^i_N(V(t_n)) + k^2\theta S^i_N\big(S^i_N(V(t_n))\big).$$

The next corrector iterations may be analyzed by noting that the solution at the mth iteration has the form:

$$V_i^{(m)}(t_{n+1}) = P_0^m V_i(t_n) + k\sum_{l=1}^{m+1} P_l^m \big(S^i_N\big)^l (V(t_n)), \tag{26}$$

where the superscript on (S^i_N) indicates repeated evaluations of the function, e.g. the last term in the previous equation. Substituting this expression into (22) gives rise to the following recurrence relations between the polynomial coefficients, P_l^m,

$$\begin{aligned} P_0^{m+1} &= 1 - a_i k + a_i\theta k(1 - P_0^m), \\ P_1^{m+1} &= k\big(1 - \theta(1 + a_i P_1^m) + \theta P_0^j\big), \\ P_j^{m+1} &= k\theta\big(P_{j-1}^m - a_i P_j^m\big), \quad j = 2, \ldots, m+1, \\ P_0^0 &= 1 - ka_i, \qquad P_1^0 = k. \end{aligned}$$

All these coefficients must be positive for the method to be positive. Evaluation of these coefficients using an algebraic manipulation package shows that the critical condition is that the coefficient P_m^m is positive where

$$P_m^m = k^m\theta^{m-1}(1 - mk\theta a_i). \tag{27}$$

This shows that although the CFL number decreases with increased iterations the magnitude of the terms is multiplied by successive powers of k. From Eq. (24) the predictor will preserve positivity if

$$1 - ka_i \geqslant 0,$$

while for the mth corrector iteration to preserve positivity

$$1 - \theta mka_i \geqslant 0 \quad \text{or} \quad ka_i \leqslant \frac{1}{\theta m}.$$

Combining the last two equations and substituting from (25) gives a CFL-like condition

$$k\left(A^n_{i+1/2}+B^n_{i-1/2}\right) \leqslant \mathrm{Min}\left(1,\frac{1}{\theta m}\right). \tag{28}$$

In practice m is no higher than three and often one or two.

5. Triangular mesh discretization method

Although the two-dimensional method considered below was developed for systems of equations, for ease of exposition, consider the class of scalar PDEs:

$$\frac{\partial u}{\partial t}+\frac{\partial f}{\partial x}+\frac{\partial g}{\partial y}=0, \tag{29}$$

where $f=f(x,\ y,\ u)$ and $g=g(x,\ y,\ u)$ are the flux functions in x and y respectively and with appropriate boundary and initial conditions.

The cell-centered finite volume scheme described here uses triangular elements as the control volumes over which the divergence theorem is applied. The finite volume representation of a solution is formally piecewise constant within each control volume and is not associated with any particular position. To allow the construction of high-order schemes however the centroid of the triangle is defined as the nodal position and the solution value is associated with that point. In Fig. 1 for example, the solution at the centroid of triangle i is U_i,

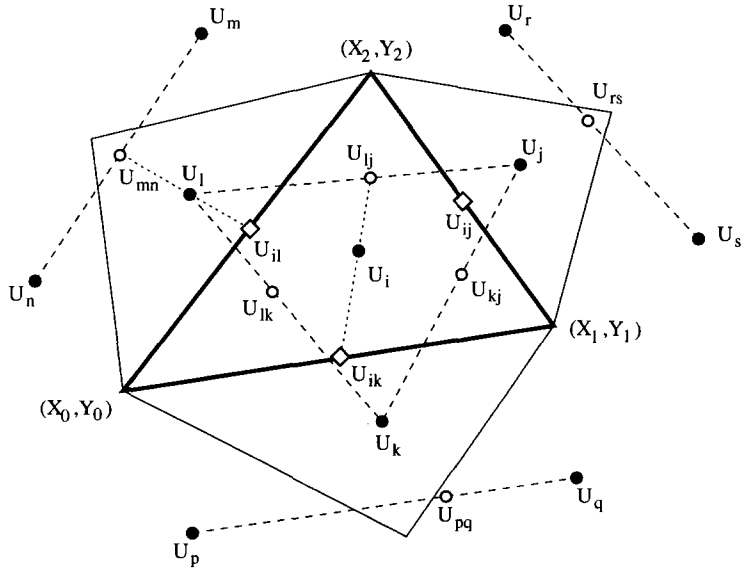

Fig. 1. Construction of interpolants. • centroid solution values; ○ interpolated solution values; ◇ midpoints of edges.

the solutions at the centroids of the triangles surrounding triangle i are U_l, U_j and U_k and the next level of centroid values used by the discretization method on the ith triangle are: U_m, U_n, U_p, U_q, U_r and U_s. The mesh point at which a solution value, say U_s, is defined is denoted by (x_s, y_s).

Integration of (29) on the ith triangle gives:

$$\int_{A_i} \frac{\partial u}{\partial t}\, \mathrm{d}\Omega = -\int_{A_i} \left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} \right) \mathrm{d}\Omega, \tag{30}$$

where A_i is the area of triangle i and Ω is the integration variable defined on A_i. The area integral on the left-hand side of (30) is approximated by a one-point quadrature rule. The quadrature point is the centroid of triangle i. By using the divergence theorem, the area integral on the right-hand side is replaced by a line integral around the triangular element:

$$A_i \frac{\partial U_i}{\partial t} = -\oint_{C_i} (f \cdot \boldsymbol{n}_x + g \cdot \boldsymbol{n}_y)\, \mathrm{d}S,$$

where C_i is the circumference of triangle i and S is the integration variable along that circumference. The line integral along each edge is approximated by using the midpoint quadrature rule. The numerical flux is evaluated at the midpoint of the edge:

$$\frac{\partial u}{\partial t} = -\frac{1}{A_i}\left(f_{ik}\Delta y_{0,1} - g_{ik}\Delta x_{0,1} + f_{ij}\Delta y_{1,2} - g_{ij}\Delta x_{1,2} + f_{il}\Delta y_{2,0} - g_{il}\Delta x_{2,0} \right),$$

where $\Delta x_{i,j} = x_j - x_i$, $\Delta y_{i,j} = y_j - y_i$ and f_{ij} and g_{ij} are the fluxes in the x and y directions respectively evaluated at the midpoint of the triangle edge separating the triangles associated with U_i and U_j.

The fluxes f_{ij} and g_{ij} are evaluated by using approximate Riemann solvers f_{Rm} and g_{Rm} respectively. At the midpoint of each edge one-dimensional Riemann problems are solved in the cartesian directions with the *left* solution value being defined as that internal to triangle i and the *right* solution value being defined as that external to triangle i:

$$\begin{aligned} \frac{\partial u}{\partial t} = -\frac{1}{A_i}\Big(& f_{\mathrm{Rm}}(U_{ik}^{\ell}, U_{ik}^{\mathrm{r}})\Delta y_{0,1} - g_{\mathrm{Rm}}(U_{ik}^{\ell}, U_{ik}^{\mathrm{r}})\Delta x_{0,1} \\ & + f_{\mathrm{Rm}}(U_{ij}^{\ell}, U_{ij}^{\mathrm{r}})\Delta y_{1,2} - g_{\mathrm{Rm}}(U_{ij}^{\ell}, U_{ij}^{\mathrm{r}})\Delta x_{1,2} \\ & + f_{\mathrm{Rm}}(U_{il}^{\ell}, U_{il}^{\mathrm{r}})\Delta y_{2,0} - g_{\mathrm{Rm}}(U_{il}^{\ell}, U_{il}^{\mathrm{r}})\Delta x_{2,0} \Big), \end{aligned} \tag{31}$$

where U_{ij}^{ℓ} is the internal solution, with respect to triangle i, at the midpoint of the edge between U_i and U_j and U_{ij}^{r} is the external solution, with respect to triangle i, on edge j. Note that $U_{i,j}^{\mathrm{r}} = U_{j,i}^{\ell}$ as a consequence of this notation. The approximate Riemann solver satisfies that same conditions as in the one-dimensional case (see Eq. (5)), except that the first condition is replaced by the conditions

$$g_{\mathrm{Rm}}(u, u) = g(u), \qquad f_{\mathrm{Rm}}(u, u) = f(u). \tag{32}$$

Consider for example the two-dimensional advection equation:

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y} = 0,$$

where a and b are positive constants for example. The discrete form (see Eq. (31)) is—assuming that the triangle is aligned to the characteristic directions as in Fig. 1 and given that the solution to the Riemann problem is the product of the upwind value and either a or b—given by

$$\frac{\partial U_i}{\partial t} = -\frac{1}{A_i}\Big[\big(aU_{ik}^{\ell}\big)\Delta y_{0,1} - \big(bU_{ik}^{\mathrm{r}}\big)\Delta x_{0,1} + \big(aU_{ij}^{\ell}\big)\Delta y_{1,2} - \big(bU_{ij}^{\ell}\big)\Delta x_{1,2} + \big(aU_{il}^{\mathrm{r}}\big)\Delta y_{2,0} - \big(bU_{il}^{\ell}\big)\Delta x_{2,0}\Big]. \tag{33}$$

A standard first-order scheme uses the piecewise constant solution on either side of the edge as the upwind values, e.g.

$$U_{ij}^{\ell} = U_i, \qquad U_{ij}^{\mathrm{r}} = U_j.$$

Although this scheme results in numerical solutions with no undershoots or overshoots the amount of numerical diffusion introduced is often not acceptable. Nevertheless Kroner and Rokyta [10] have very recently proved rigorous convergence results which could probably be extended to the method described here.

5.1. Limited interpolants in two dimensions

The approach of using limited linear upwind values to create left and right values for the Riemann solver will now be used on unstructured meshes. In this approach the internal and external values at the cell interface of two triangular elements, U_{ij}^{ℓ} and U_{ij}^{r}, in (31) are replaced with the limited linearly interpolated values defined by

$$U_{ij}^{\ell} = U_i + \Phi\big(r_{ij}^{\ell}\big)\big(U_{ij}^{\mathrm{L}} - U_i\big), \tag{34}$$

$$U_{ij}^{\mathrm{r}} = U_j + \Phi\big(r_{ij}^{\mathrm{r}}\big)\big(U_{ij}^{\mathrm{R}} - U_j\big), \tag{35}$$

where U_{ij}^{L} is the internal linear upwind value, U_{ij}^{R} is the external linear upwind value, r_{ij}^{ℓ} is the internal upwind bias ratio of gradients and r_{ij}^{r} is the external upwind bias ratio of gradients. The internal and external ratio of linear gradients are defined in a similar manner to that in Section 3 by

$$r_{ij}^{\ell} = \frac{U_{ij}^{\mathrm{C}} - U_i}{U_{ij}^{\mathrm{L}} - U_i}, \qquad r_{ij}^{\mathrm{r}} = \frac{U_{ij}^{\mathrm{C}} - U_j}{U_{ij}^{\mathrm{R}} - U_j}, \tag{36}$$

where U_{ij}^{C} is the linear centered value at the cell interface. The choice of limiter function is left open at this point. Eqs. (34), (35) and (36) describe the unstructured flux limiter scheme but in terms of new, and as yet undefined, interpolated and extrapolated values: U_{ij}^{R}, U_{ij}^{R} and U_{ij}^{C}.

In a similar way to Spekreijse, U_{ij}^{L} and U_{ij}^{R} are defined using linear extrapolation but on the unstructured mesh. The value U_{ij}^{L} is constructed by forming a linear interpolant using the

solution values U_i, U_k and U_l at the three centroids. An alternative interpretation is that linear extrapolation is being used based on the solution value U_i and an intermediate solution value (itself calculated by linear interpolation) U_{lk} which lies on the line joining the centroids at which U_l and U_k are defined (see Fig. 1), i.e.

$$U_{ij}^{\mathrm{L}} = U_i + d_{ij,i}\frac{U_i - U_{lk}}{d_{i,lk}}, \tag{37}$$

where the generic term $d_{a,b}$ denotes the positive distance between points a and b, so for example $d_{ij,i}$ denotes the positive distance between points ij and i (see Fig. 1) as defined by

$$d_{i,ij} = \sqrt{(x_i - x_{ij})^2 + (y_i - y_{ij})^2}, \tag{38}$$

where (x_{ij}, y_{ij}) are the coordinates of U_{ij}. The value U_{ij}^{R} is defined in a similar way using the centroid values U_j, U_s and U_r. This also may be viewed as linear extrapolation based on the solution value U_j and an intermediate solution value (itself calculated by linear interpolation) U_{rs} which lies on the line joining the centroids at which U_r and U_s are defined (see Fig. 1), i.e.

$$U_{ij}^{\mathrm{R}} = U_j + d_{ij,j}\frac{U_j - U_{rs}}{d_{j,rs}}. \tag{39}$$

For certain meshes the three centroid points may be collinear in which case it is not possible to define a linear interpolant. In this case the immediate upwind centroid value will be used: internally U_i or externally U_j.

The centered value, U_{ij}^{C}, is constructed from the six values: U_i, U_j, U_k, U_l, U_s and U_r by a series of one-dimensional linear interpolations. Three linear interpolations onto the edge being considered are performed using *opposing* pairs of centroid values (see Fig. 1). U_{lr}, U_{ij} and U_{ks} are found using the pairs U_l and U_r, U_i and U_j and U_k and U_s respectively. If the midpoint of the edge lies between U_{ks} and U_{ij}, then the centered value is found by linear interpolation using these two values. Otherwise the values U_{lr} and U_{ij} are used to compute the centered value at the midpoint by using linear interpolation.

5.2. Interpolation errors

Assuming that all the centroid values are exact, the interpolation errors associated with the linear interpolants defined by (37) and (39) above may be determined by lengthy but straightforward Taylor's series analysis. Denote the interpolation error E_{ij}^{L} by

$$E_{ij}^{\mathrm{L}} = u_{ij}^{\mathrm{L}} - U_{ij}^{\mathrm{L}}, \tag{40}$$

where u_{ij}^{L} is the left exact value (allowing for possible discontinuities in the exact solution) at the midpoint of the edge and it is assumed that the centroid values used to form U_{ij}^{L} are exact. Standard results for linear interpolation then give

$$E_{ij}^{\mathrm{L}} = \frac{1}{2}\left[d_{ij,i}d_{ij,lk}(u_{\eta\eta})_{ij} + \frac{d_{i,ij}}{d_{i,lk}}d_{k,lk}d_{l,lk}(u_{\zeta\zeta})_{lk}\right],$$

where η is a local coordinate along the line through points lk, i and ij and ζ is a local coordinate defined along the line through points l, lk and k.

Hence $(u_{ss})_{ij}$ is the second derivative of u with respect to s evaluated at the point ij. In the same way, denote the interpolation error E_{ij}^{R} by

$$E_{ij}^{\mathrm{R}} = u_{ij}^{\mathrm{R}} - U_{ij}^{\mathrm{R}}, \tag{41}$$

where u_{ij}^{R} is the right exact value at the midpoint of the edge and it is assumed that the centroid values used to form U_{ij}^{R} are exact. Standard results for linear interpolation then give:

$$E_{ij}^{\mathrm{R}} = \frac{1}{2}\left[d_{ij,j} d_{ij,rs} (u_{\mu\mu})_{ij} + \frac{d_{j,ij}}{d_{j,rs}} d_{r,rs} d_{s,sr} (u_{\nu\nu})_{lk} \right],$$

where μ is a local coordinate along the line through points rs, j and ij and ν is a local coordinate defined along the line through points r, rs and s.

Thus from (38) both interpolation errors are second-order in the mesh spacing distances d_{**}.

Remark. Consider the case of a degenerate triangle in which the three points, say, i, k, l are almost collinear. The distances $d_{k,lk}$ and $d_{l,lk}$ may be as much as a factor of 10 larger than $d_{i,lk}$. Suppose further that, say, $d_{ij,lk} \approx 2 d_{ij,i}$. The expression for E_{ij}^{L} given above then reads:

$$E_{ij}^{\mathrm{L}} = d_{ij,i}^2 \left[(u_{\eta\eta})_{ij} + 50 (u_{\zeta\zeta})_{lk} \right].$$

In experiments we do not appear to have observed a loss of accuracy due to this source of error. Venkatakrishnan and Barth [19] have suggested a modification to the method stencil that overcomes this difficulty.

5.3. Spatial truncation error

The above results on interpolation errors may be combined with standard results for the effect of quadrature errors (see [9]) to show that the underlying method is second-order accurate when the limiter is not used. Consider Eq. (33) and note that the spatial truncation error in the flux derivative approximations for the ith triangle, as denoted by TE_i is, after ignoring the second-order quadrature error, a combination of the interpolation errors defined in Section 5.1, i.e.

$$\begin{aligned} TE_i = -\frac{1}{A_i}\Big[&\left(aE_{ik}^{\mathrm{L}}\right)\Delta y_{0,1} - \left(bE_{ik}^{\mathrm{R}}\right)\Delta x_{0,1} \\ &+ \left(aE_{ij}^{\mathrm{L}}\right)\Delta y_{1,2} - \left(bE_{ij}^{\mathrm{L}}\right)\Delta x_{1,2} + \left(aE_{il}^{\mathrm{R}}\right)\Delta y_{2,0} - \left(bE_{il}^{\mathrm{L}}\right)\Delta x_{2,0} \Big], \end{aligned}$$

where the individual errors are defined in (40) and (41) and where it is assumed that the limiter is set to one. From the results in Section 5.1 it is possible to extract a constant second-order factor, say $d_{\min}^2$, depending on the minimum of the distances, d_{ab}, as defined in (38), from each of the errors in this equation. Assuming that the individual errors all have the form

$$E_{ik}^{\mathrm{L}} = d_{\min}^2 e_{ik}^{\mathrm{L}},$$

the expression for the truncation error may be rewritten as:

$$TE_i = -\frac{d_{\min}^2}{A_i}\left[\left(ae_{ik}^{\mathrm{L}}\right)\Delta y_{0,1} - \left(be_{ik}^{\mathrm{R}}\right)\Delta x_{0,1} + \left(ae_{ij}^{\mathrm{L}}\right)\Delta y_{1,2} - \left(be_{ij}^{\mathrm{L}}\right)\Delta x_{1,2} + \left(ae_{il}^{\mathrm{R}}\right)\Delta y_{2,0} - \left(be_{il}^{\mathrm{L}}\right)\Delta x_{2,0}\right].$$

It is now possible to define two linear functions on the ith triangle $E_f(x, y)$ and $E_g(x, y)$ such that $E_f(x, y)$ has values e_{ik}^{L}, e_{ij}^{L} and e_{il}^{R} at the midpoints ik, ij and il and $E_g(x, y)$ has values e_{ik}^{R}, e_{ij}^{L} and e_{ij}^{L} at the midpoints ik, ij and il. From the linearity of these functions and the divergence theorem it follows that

$$\frac{\partial E_f}{\partial x} = -\frac{1}{A_i}\left[e_{ik}^{\mathrm{L}}\Delta y_{0,1} + e_{ij}^{\mathrm{L}}\Delta y_{1,2} + e_{il}^{\mathrm{R}}\Delta y_{2,0}\right]$$

and

$$\frac{\partial E_g}{\partial y} = \frac{1}{A_i}\left[e_{ik}^{\mathrm{R}}\Delta x_{0,1} + e_{ij}^{\mathrm{L}}\Delta x_{1,2} + e_{il}^{\mathrm{L}}\Delta x_{2,0}\right].$$

Hence the truncation error (ignoring the quadrature error due to the use of the midpoint rule) may be written as

$$TE_i = d_{\min}^2\left[a\frac{\partial E_f}{\partial x} + b\frac{\partial E_g}{\partial y}\right].$$

The error due to the use of the quadrature rule is derived by Jeng and Chen [9]. The extension to handle the case when the limiters are used is as described by Spekreijse [16] and results in observed convergence rates of between one and two (see Section 7 and Durlofsky et al. [8]).

6. Analysis of discretization method

This section will consider whether or not the new scheme has the properties of *linearity preservation* and *positivity*, as proposed in recent work by Struijs et al. [17].

6.1. Linearity-preserving methods

A linearity-preserving spatial discretization method is defined by Struijs et al. [17] as one which preserves the exact steady state solution whenever this is a linear function of the space coordinates x and y, for any arbitrary triangulation of the domain. This is equivalent to second-order accuracy on regular meshes (see [17]). The simplest way to prove a spatial discretization scheme is linearity-preserving is to show that the residual truncation error will be zero when an arbitrary linear solution is substituted.

The following is an outline proof that the unstructured flux limiter scheme is linearity-preserving for a general nonlinear scalar partial differential equation. Consider the discrete form given by (31) with the internal and external values defined by (34), (35) and (36). Consider the

first time step. The centroid values will be point samples of the initial linear profile. Since U_{ij}^{L}, U_{ij}^{R} and U_{ij}^{C} are all created by linear interpolation or extrapolation they will be exact also and $r_{ij}^{\ell} = r_{ij}^{\mathrm{r}} = 1$. Define the limiter function $\Phi(\cdot)$ to have the standard property $\Phi(1) = 1$ (see [16]). The upwind values used in the Riemann solver U_{ij}^{ℓ} and U_{ij}^{r} are now U_{ij}^{L} and U_{ij}^{R} since (34) and (35) simplify. Since U_{ij}^{L} and U_{ij}^{R} are exact they must be the same value, U_{ij}. The discrete equation is now

$$\begin{aligned} A_i \frac{\partial U_i}{\partial t} = & -f_{\mathrm{Rm}}(U_{ik}, U_{ik})\Delta y_{0,1} + g_{\mathrm{Rm}}(U_{ik}, U_{ik})\Delta x_{0,1} \\ & - f_{\mathrm{Rm}}(U_{ij}, U_{ij})\Delta y_{1,2} + g_{\mathrm{Rm}}(U_{ij}, U_{ij})\Delta x_{1,2} \\ & - f_{\mathrm{Rm}}(U_{il}, U_{il})\Delta y_{2,0} + g_{\mathrm{Rm}}(U_{il}, U_{il})\Delta x_{2,0}. \end{aligned}$$

Using the property of the Riemann solver defined by (32) and noting that the midpoint quadrature rule used along the edges is exact for linear data ensures that the discrete approximation for the line integral is exact. The above equation then simplifies to

$$A_i \frac{\partial U_i}{\partial t} = -\oint_{C_i} \left[f(U) \cdot \boldsymbol{n}_x + g(U) \cdot \boldsymbol{n}_y \right] \mathrm{d}S.$$

The one-point area quadrature rule used on the left-hand side is exact for linear data provided the quadrature point is at the centroid. Converting the line integral around the circumference C_i into an area integral using the divergence theorem gives

$$\int_{A_i} \frac{\partial U_i}{\partial t} \, \mathrm{d}\Omega = -\int_{A_i} \frac{\partial}{\partial x} f(U) + \frac{\partial}{\partial y} g(U) \, \mathrm{d}\Omega,$$

and therefore

$$\frac{\partial U_i}{\partial t} + \frac{\partial}{\partial x} f(U) + \frac{\partial}{\partial y} g(U) = 0,$$

which is equivalent to the original differential equation (29). The initial linear solution will thus be preserved providing that sufficient accuracy is used in the time integration method.

6.2. Positivity

The definition of positivity, [17], requires that every new value can be written as a convex combination of old values (see Eq. (16)). The approach of Spekreijse, already used in Section 3, uses linearization and the mean value theorem via the definition of the coefficients A and B as in (15), to reduce the nonlinear case to what is effectively a linear advection equation. The same approach is implicitly used here in restricting attention to the linear advection equation as defined by (5) and its discrete form, Eq. (33). Note the $\Delta x_{i,j}$ and $\Delta y_{i,j}$ go anticlockwise around the triangular element so

$$\Delta x_{0,1} + \Delta x_{1,2} + \Delta x_{2,0} = \Delta y_{0,1} + \Delta y_{1,2} + \Delta y_{2,0} = 0.$$

This enables Eq. (33) to be rewritten as

$$-A_i\frac{\partial U_i}{\partial t} = a\left(U_{ik}^{\ell} - U_{il}^{\mathrm{r}}\right)\Delta y_{0,1} - b\left(U_{il}^{\ell} - U_{ik}^{\mathrm{r}}\right)\Delta x_{2,0} + a\left(U_{ij}^{\ell} - U_{il}^{\mathrm{r}}\right)\Delta y_{1,2} - b\left(U_{ij}^{\ell} - U_{ik}^{\mathrm{r}}\right)\Delta x_{1,2}.$$

From Eqs. (34) and (35) it can be seen that these internal and external values at the cell interface are a combination of the centroid values and linear upwind values. Without loss of generality, and by using a similar approach to Section 3 and Spekreijse [16], consider the term $a(U_{ik}^{\ell} - U_{il}^{\mathrm{r}})\Delta y_{0,1}$. For positivity it is sufficient to prove that

$$U_{ik}^{\ell} - U_{il}^{\mathrm{r}} = \gamma_i U_i - \gamma_l U_l - \gamma_j U_j - \gamma_n U_n - \gamma_k U_k, \tag{42}$$

for positive multipliers γ_i, γ_l, γ_j, γ_n and γ_k thus giving an ODE system of the form of Eq. (24). Thus the intention is to show that for the ith ODE all multipliers of solution coefficients other than U_i are positive and the multiplier of U_i is negative. Using the notation of (37) and (39) the left-hand side of (42) may be written as

$$U_i + d_{ik,i}\frac{U_i - U_{lj}}{d_{i,lj}}\Phi\left(\frac{U_{ik}^{\mathrm{C}} - U_i}{U_{ik}^{\mathrm{L}} - U_i}\right) - U_l - d_{il,l}\frac{U_l - U_{mn}}{d_{l,mn}}\Phi\left(\frac{U_{il}^{\mathrm{C}} - U_l}{U_{il}^{\mathrm{R}} - U_l}\right).$$

After noting that

$$\frac{U_l - U_{mn}}{d_{l,mn}} = \frac{(U_{il}^{\mathrm{C}} - U_l)}{d_{il,l}}\left(\frac{U_{il}^{\mathrm{R}} - U_l}{U_{il}^{\mathrm{C}} - U_l}\right),$$

this may be rewritten as

$$U_i - U_l + \delta_{ik,lj}(U_i - U_{lj})\Phi(R) - (U_{il}^{\mathrm{C}} - U_l)\frac{\Phi(S)}{S}, \tag{43}$$

where

$$R = \left(\frac{U_{ik}^{\mathrm{L}} - U_i}{U_{ik}^{\mathrm{C}} - U_i}\right), \qquad S = \left(\frac{U_{il}^{\mathrm{R}} - U_l}{U_{il}^{\mathrm{C}} - U_l}\right), \qquad \delta_{ik,lj} = \frac{d_{ik,i}}{d_{i,lj}}.$$

The centered value U_{il}^{C} is formed by linear interpolation, i.e.

$$U_{il}^{\mathrm{C}} = \beta_{il}(\alpha_{il}U_l + (1-\alpha_{il})U_i) + (1-\beta_{il})(\alpha_{kn}U_n + (1-\alpha_{kn})U_k)$$
$$\text{for } 0 \leqslant \alpha_{il}, \alpha_{kn}, \beta_{il} \leqslant 1.$$

Similarly

$$U_{lj} = \alpha_{lj}U_l + (1-\alpha_{lj})U_j \quad \text{for } 0 \leqslant \alpha_{il} \leqslant 1. \tag{44}$$

It is worth noting that the need to have positive multipliers in these two linear interpolants

effectively restricts the mesh that can be used. A similar restriction is also used by Lin et al. [11]. Using these last two equations to substitute in (43) gives

$$U_i\left(1+\delta_{ik,lj}\Phi(R)-\frac{\Phi(S)}{S}\beta_{il}(1-\alpha_{il})\right)-U_j\delta_{ik,lj}\Phi(R)(1-\alpha_{il})-U_k(1-\beta_{il})(1-\alpha_{kn})$$
$$\times\frac{\Phi(S)}{S}-U_l\left(1+\delta_{ik,lj}\alpha_{lj}\Phi(R)+\frac{\Phi(S)}{S}(1-\beta_{il}\alpha_{il})\right)-U_n(1-\beta_{il})\alpha_{kn}\frac{\Phi(S)}{S}, \tag{45}$$

which is of the form specified by (42).

Inspection of this equation shows that the *Positivity Condition* is that the limiter $\Phi(\cdot)$ must be positive and must satisfy $\Phi(S)/S \leqslant 1$ as in Eq. (18). One such limiter is the modified van Leer limiter defined by Eq. (19).

6.3. Alternative schemes and limiters

The schemes of Venkatakrishnan and Barth [19] and Lin et al. [11] both use the same upwind interpolants as that considered above but different limiters—which may now be assessed in the light of the above results.

In many situations it is reasonable to expect that the edge midpoint value lies almost halfway between the centroids on either side of the edge and consequently that $\beta_{il} \approx 1$ and $\alpha_{il} \approx \frac{1}{2}$. In this case the positivity condition may be relaxed, for example, to $\Phi(S)/S \leqslant 1.2$, as is satisfied by the van Albeda limiter and defined by Eq. (20) used by Venkatakrishnan and Barth [19]. The proof above also applies to the case in which U_{lj} is replaced by a positive combination of two other centroid values and $d_{lj,i}$ is modified appropriately. Thus the method devised by Venkatakrishnan and Barth [19] for dealing with degenerate upwind triangles also fits into the same framework. The limiter used by Lin et al. [11] differs from the Ware and Berzins scheme in that the limited upwind values U_{ij}^{ℓ} and U_{ij}^{r} are defined by

$$U_{il}^{\ell} = U_i + \operatorname{minmod}\left(U_{ij}^{\mathrm{L}} - U_i,\ k\cdot\left(U_j - U_i\right)\right),$$

$$U_{ij}^{\mathrm{r}} = U_j + \operatorname{minmod}\left(U_{ij}^{\mathrm{R}} - U_j,\ k\cdot\left(U_i - U_j\right)\right),$$

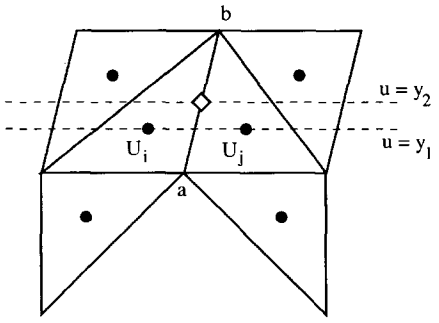

Fig. 2. Demonstration of nonlinearity preserving nature.

where k is some arbitrary constant $k \geqslant 0.5$, the function minmod is defined by

$$\text{minmod}(a, b) = \begin{cases} \min(|a|, |b|) \cdot \text{sign}(a), & \text{if } \text{sign}(a) = \text{sign}(b), \\ 0, & \text{otherwise}, \end{cases}$$

and U_i and U_j are defined as in Section 5. This definition of the limiter function leads to a loss of linearity preservation. Consider the situation in Fig. 2 where the current solution is some linear function of y only, say $u(x, y) = y$. Although the solution is smooth the limiter will not allow the full upwind value to be used at the midpoint of the edge ab as the term $k \cdot (U_j - U_i)$ will be zero. In an attempt to overcome this deficiency other similar limiters are defined by Lin et al. for different triangulation cases in [11]. Lin et al. also proved their scheme satisfies the local maximum principle for certain triangulations.

7. Numerical examples

The following viscous Burgers' equation will be used to illustrate the theoretical results obtained above

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) + \frac{\partial}{\partial y}\left(\frac{u^2}{2}\right) = p\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right),$$

$$(x, y, t) \in [0, 1] \times [0, 1] \times (0, 1.25]$$

with an exact solution of

$$u(x, y, t) = (1 + \exp((x + y - t)/p))^{-1}.$$

The value of p is chosen to be 0.0001 so that the partial differential equation is convection-dominated and the boundary and initial conditions are given by the exact solution. From the exact solution it can be seen that the computed solution should lie in the range [0, 1]. At every time step the computed solution is examined triangle by triangle and the maximum absolute overshoot or undershoot outside the range [0, 1] is noted.

The solution was first computed using Mesh A shown in Fig. 3 but regularly subdivided to contain 2048 triangular elements. The Riemann solver used was the Engquist–Osher solver for the inviscid Burgers' equation. Using the standard van Leer limiter the maximum under/over-

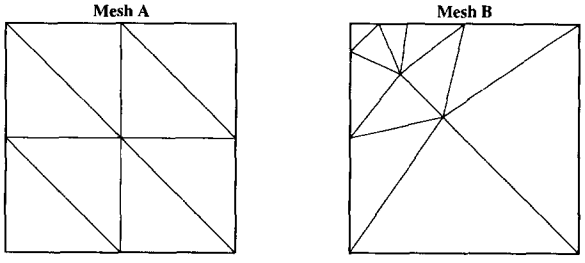

Fig. 3. Meshes used in numerical experiments.

shoot recorded was 0.0. This shows that the unmodified limiter can be used to provide oscillation-free solutions in certain circumstances.

The computation was repeated but now using Mesh B shown in Fig. 3 regularly subdivided to contain 2816 triangular elements. The maximum under/overshoot is now 7.3369e–3 with the van Leer limiter. No overshoot was observed with the new limiter or the van Albeda limiter on either mesh.

The accuracy of the schemes on this problem above is more difficult to assess due to the shock-like behaviour of the solution. In this case Mesh A is used with regular refinement. ONE is the first-order method, VL is the van Leer limiter, MVL is the modified limiter and VA is the van Albeda limiter.

The results in Table 1 show that on a shock problem for which many first-order elements are used (i.e. a flat solution or a zero limiter), all the limiters give only first-order accuracy but that the notionally second-order methods are more accurate by a factor of two. These results are consistent with those obtained by Berzins [2] on regular quadrilateral meshes.

In the light of the above results the accuracy of the method on a problem without shock-like features must be studied. Consider the solution of the linear conservation law

$$u_t + u_x + u_y = 0, \quad (x, y, t) \in [0, 1] \times [0, 1] \times (0, 0.75] \tag{46}$$

with exact solution

$$u(x, y, t) = \sin(2\pi x - t)\sin(2\pi y - t), \tag{47}$$

which is used to specify the initial and boundary conditions. This equation was solved on Mesh A in Fig. 3 using the first-order scheme, original scheme and modified scheme. The L1 error, weighted by element areas, was evaluated at times 0.1 to 1.0 in steps of 0.1 and these then averaged. The smallest mesh used contained 200 elements with a 0.1 mesh spacing and the largest mesh used contained 18200 elements with a 0.0125 mesh spacing. The results of these

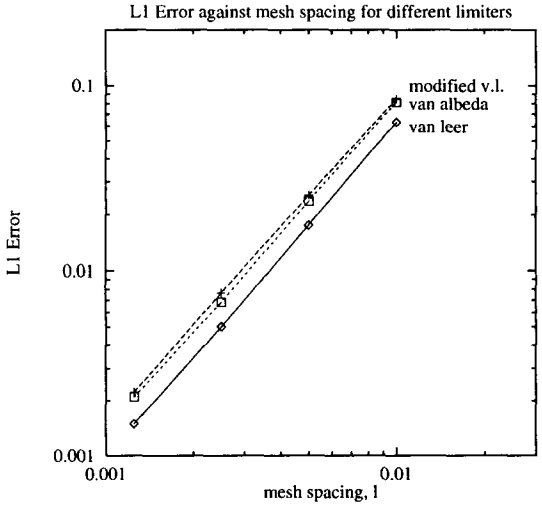

Fig. 4. Log-log graph of error versus mesh spacing.

Table 1
L1 error norms × 1000 for Burgers' equation

Mesh	Time	ONE	VL	MVL	VA
9×9	0.26	0.81	0.69	0.69	0.69
	0.69	4.58	2.80	2.89	2.90
	1.30	5.45	2.85	2.88	3.22
27×27	0.26	0.48	0.40	0.41	0.41
	0.69	1.70	0.86	0.87	0.97
	1.30	1.83	1.13	1.16	1.17
81×81	0.26	0.21	0.17	0.18	0.17
	0.69	0.66	0.39	0.39	0.43
	1.30	0.60	0.31	0.32	0.33

experiments are plotted in a log-log graph shown in Fig. 4. The results show that the scheme with the original limiter has a convergence rate of 1.80 and that with the new limiter has a convergence rate of 1.75. The convergence rate with the van Albeda limiter is 1.76 (see Table 1).

8. Summary

This paper has shown that standard flux limiter schemes may need to be modified when used with cell-centered finite volume schemes on irregular one-dimensional meshes and unstructured triangular meshes in two-space dimensions. A new modified form of the van Leer limiter was introduced together with additional but straightforward conditions on the interpolants in the case of triangular meshes. This combination was shown to ensure both theoretically and experimentally that the new modified scheme of Ware and Berzins [6,20] for unstructured meshes is positive and linearity-preserving for a model problem.

Acknowledgements

The authors would like to thank the referees for their comments and for pointing out references [1] and [19]. Shell Research Ltd and SERC are also thanked for funding.

References

[1] T.J. Barth and D.C. Jespersen, The design and application of upwind schemes on unstructured meshes, AIAA Paper 89-0366, Paper presented at 27th Aerospace Sciences Conference, Reno, NV (1992).

[2] M. Berzins, Temporal error control for convection-dominated equations in two spaced dimensions, *SIAM J. Sci. Comput.*

[3] M. Berzins, P.L. Baehmann, J.E. Flaherty and J. Lawson, Towards an automated finite element solver for time-dependent fluid-flow problems, in: *The Mathematics of Finite Elements and Application* VII (Academic Press, New York, 1991) 181–188.

[4] M. Berzins and R.M. Furzeland, An adaptive theta method for the solution of stiff and non-stiff differential equations, *Appl. Numer. Math.* 8 (1992) 1–19.
[5] M. Berzins, J.M. Ware and J. Lawson, Spatial and temporal error control in the adaptive solution of systems of conservation laws, in: *Advances in Computer Methods for Partial Differential Equations*, IMACS PDE VII (IMACS, New Brunswick, NJ, 1992).
[6] M. Berzins and J.M. Ware, Reliable finite volume methods for time-dependent p.d.e.s, in: J.R. Whiteman, ed., *Mafelap Conference* (Wiley, New York, 1994).
[7] B. Cockburn, Suchung Hou and Chi-Wang Shu, The Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws IV: the multidimensional case, *Math. Comp.* 54 (190) (1990) 545–581.
[8] L.J. Durlofsky, B. Enquist and S. Osher, Triangle based adaptive stencils for the solution of hyperbolic conservation laws, *J. Comput. Phys.* 98 (1992) 64–73.
[9] Y.N. Jeng and J.L. Chen, Truncation error analysis of the finite volume method for a model steady convective equation, *J. Comput. Phys.* 100 (1992) 64–76.
[10] D. Kroner and M. Rokyta, Convergence of upwind finite volume schemes for scalar conservation laws in two dimensions, *SIAM J. Numer. Anal.* 31 (1994) 324–343.
[11] S.Y. Lin, T.M. Wu and Y.S. Chin, Upwind finite-volume method with a triangular mesh for conservation laws, *J. Comput. Phys.* 107 (1993) 324–337.
[12] X.D. Liu, A maximum principle satisfying modification of triangle based adaptive stencils for the solution of scalar hyperbolic conservation laws *SIAM J. Numer. Anal.* 30 (1993) 701–715.
[13] S.V. Pennington and M. Berzins, New NAG Library software for first-order P.D.E.s, *ACM Trans. Math. Software* 20 (1) (1994) 63–99.
[14] P.L. Roe, Characteristic based schemes for the Euler equations, *Ann. Rev. Fluid Mech.* 8 (1986) 337–365.
[15] C.W. Shu and S. Osher, Efficient implementation of E.N.O. shock capturing schemes, *J. Comput. Phys.* 77 (1988) 439–471.
[16] S. Spekreijse, Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws, *Math. Comp.* 49 (179) (1987) 135–155.
[17] R. Struijs, H. Deconinck and P.L. Roe, Fluctuation splitting schemes for the 2D Euler equations, Technical Report, von Karman Institute for Fluid Dynamics, Rhode Saint Genese, Belgium (1991).
[18] B. van Leer, Progress in multi-dimensional upwind differencing, in: M. Napolitano and F. Sabetta, eds., *Proceedings Thirteenth International Conference on Numerical Methods in Fluid Dynamics*, Rome, Italy (1992).
[19] V. Venkatakrishnan and T.J. Barth, Application of a direct solver to unstructured meshes for the Euler and Navier Stokes equations. AIAA Paper 89-0364, Paper presented at 27th Aerospace Sciences Conference, Reno, NV (1992).
[20] J.M. Ware and M. Berzins, Finite volume techniques for time-dependent fluid-flow problems, in: *Advances in Computer Methods for Partial Differential Equations*, IMACS PDE VII (IMACS, New Brunswick, NJ, 1992).