

## Coercivity for One-dimensional Cell Vertex Approximations

K. W. Morton

Previous error analysis for the cell vertex scheme has been limited to situations where the cell residuals can be set to zero. However, in practical use for compressible flow computations it is necessary to extend the method by the use of distribution matrices and the careful addition of artificial viscosity terms. In this paper we make a start on the error analysis that is required for this more general method. The chosen example is a one-dimensional convection-diffusion problem with an expansion critical or turning point.

*Key words and phrases:* coercivity, cell vertex methods, finite volume methods, artificial dissipation viscosity, convection-diffusion, expansion critical point, error bounds

The work reported here forms part of the research programme of the Oxford–Reading Institute for Computational Fluid Dynamics.

Oxford University Computing Laboratory  
Numerical Analysis Group  
Wolfson Building  
Parks Road  
Oxford, England OX1 3QD  
*E-mail:* morton@comlab.oxford.ac.uk

February, 1996

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Coercivity for pure convection at an expansion critical point</b>	<b>5</b>
<b>3</b>	<b>Coercivity of diffusion terms on a general mesh</b>	<b>12</b>
<b>4</b>	<b>Coercivity-based error bounds</b>	<b>17</b>
<b>5</b>	<b>References</b>	<b>19</b>

# 1 Introduction

Consider approximating the vector of unknowns  $\mathbf{w}(x, y)$  which satisfy the system of conservation laws

$$\frac{\partial \mathbf{f}}{\partial x} + \frac{\partial \mathbf{g}}{\partial y} = \mathbf{S}, \quad (1.1)$$

together with appropriate boundary conditions. When  $\mathbf{f} \equiv \mathbf{f}(\mathbf{w})$  and  $\mathbf{g} \equiv \mathbf{g}(\mathbf{w})$ , a first order system of equations results which may be of hyperbolic type, of elliptic type or of mixed type; examples are given by the Euler equations for steady, inviscid, compressible gas flow or the unsteady St. Venant equations describing one-dimensional river flow. In both examples one of the most useful and practical schemes of approximation consists of associating the unknowns with the vertices of a mesh (quadrilateral or rectangular in the two cases), applying Gauss' theorem to each cell of the mesh, and using the trapezoidal rule along each edge to approximate the resulting line integrals. For the unsteady, one-dimensional hyperbolic system this is called the box difference scheme and is associated with the names of Wendroff [14], Preissmann [12] and Thomée [13]; for the steady Euler equations it is called the cell vertex finite volume method and has been advocated and developed by Ni [11], Hall [3], Morton and Paisley [8] and many others subsequently.

A second order equation system is obtained when  $\mathbf{f} \equiv \mathbf{f}(\mathbf{w}, \nabla \mathbf{w})$  and  $\mathbf{g} \equiv \mathbf{g}(\mathbf{w}, \nabla \mathbf{w})$ , as would be the case for the Navier–Stokes equations of viscous fluid flow or for any system of convection-diffusion equations. Then two extensions are possible: one can introduce a subsidiary equation system  $\mathbf{Z} = \nabla \mathbf{w}$  and approximate this by the same scheme, as was done by Keller and Cebeci, for the boundary layer equations; or one can recover an approximation to  $\nabla \mathbf{w}$  at each vertex for direct substitution into the flux functions  $\mathbf{f}$  and  $\mathbf{g}$  and the same cell vertex equations as in the first order case, as has been developed in [6], [1] and subsequent papers. The latter is the more generally applicable scheme as it can deal with the complete range of problems, for example, corresponding to all values of the dimensionless Péclet number which parametrises the relative importance of convective and diffusive phenomena. Thus if  $\Omega_\alpha$  is a typical mesh cell and  $\mathbf{R}_\alpha(\cdot)$  denotes the discrete operator on that cell that is derived from (1.1) in the manner described above, the basic cell vertex scheme gives an approximation  $\mathbf{W}$  by setting

$$\mathbf{R}_\alpha(\mathbf{W}) = \mathbf{0} \quad \forall \alpha, \quad (1.2)$$

which are called the *cell residual* equations. If we write  $\mathbf{F}_i$  for the approximation to  $\mathbf{f}(\mathbf{w}, \nabla \mathbf{w})$  obtained at the vertex  $\mathbf{x}_i = (x_i, y_i)$  from  $\mathbf{W}$  and the recovered gradient, then with the vertices of  $\Omega_\alpha$  numbered 1,2,3,4 in anticlockwise order we have

$$\mathbf{R}_\alpha(\mathbf{W}) := \frac{1}{2V_\alpha} \{(\mathbf{F}_1 - \mathbf{F}_3)\delta y_{24} + (\mathbf{F}_2 - \mathbf{F}_4)\delta y_{31}$$

$$-(\mathbf{G}_1 - \mathbf{G}_3)\delta x_{24} - (\mathbf{G}_2 - \mathbf{G}_4)\delta x_{31}\} - \mathbf{S}_\alpha, \quad (1.3)$$

where  $\delta y_{24} := y_2 - y_4$  etc.,  $V_\alpha$  is the measure of the cell  $\Omega_\alpha$  and  $\mathbf{S}_\alpha$  is the average of the source function  $\mathbf{S}$  over the cell.

There is a predominant difficulty with this cell vertex scheme, however, which occurs with both first order and second order equations; namely, the natural association of the unknowns with the vertices and the equations with the cells means that it is often not possible to ensure that there are equal numbers of each after boundary conditions have been applied. Hence, in order to compute approximations to the Euler or Navier–Stokes equations, it is normal practice to form *nodal residuals* at each vertex corresponding to an unknown vector  $\mathbf{W}_i$  by combining the cell residuals from neighbouring cells, and then possibly applying some *artificial viscosity* or *artificial dissipation*; the equations that are actually solved then take the form

$$\mathbf{N}_i(\mathbf{W}) := \frac{\sum_{(\alpha)} V_\alpha [D_{\alpha,i} \mathbf{R}_\alpha + \mathbf{A}_{\alpha,i}]}{\sum_{(\alpha)} V_\alpha} = \mathbf{0} \quad \forall i, \quad (1.4)$$

where the  $D_{\alpha,i}$  are called *distribution matrices* (or distribution factors in the scalar case) and the  $\mathbf{A}_{\alpha,i}$  term represents the artificial viscosity.

Unfortunately, all of the error analysis that has been carried out for the cell vertex scheme applies only to situations where the simple form (1.2) can be used. The purpose of this paper is to initiate the analysis that has to be used when the more general form (1.4) is needed.

Distribution matrices are required when the number of cell equations (1.2) together with the boundary conditions exceed the number of unknowns, so that some of the cell residuals have to be combined. Artificial viscosity terms are needed when the opposite situation occurs and the number of equations is too few. Furthermore, the averaging along a cell edge that leads to the cell residual (1.3) means that the cell equations (1.2) suffer from the spurious chequer-board mode; and the averaging involved in the use of distribution matrices can also introduce further spurious modes. Thus the artificial viscosity is also required to control these modes.

Some of these phenomena can be studied even with one-dimensional problems, and this is the subject of this paper. By adopting the general form (1.4) even for the simple cases when the distribution matrices and artificial viscosity terms are not needed, we avoid the problem of having too few equations; then the well-posedness of the system reduces to establishing a coercivity condition, typically of the form

$$(\mathbf{N}(\mathbf{U}) - \mathbf{N}(\mathbf{V}), \mathbf{U} - \mathbf{V}) \geq \sigma^2 \|\mathbf{U} - \mathbf{V}\|_h^2 \quad (1.5)$$

where  $(\cdot, \cdot)$  is the  $l_2$  inner product and  $\|\cdot\|_h$  is some suitable discrete norm.

We begin with the consideration of a pure convection problem with an expansion critical point or turning point. This can be dealt with in the form (1.2) by

splitting the cell residual for the turning point cell, as in [9]; but in section 2 we use the form (1.4) and establish coercivity by appropriate choices of distribution factors and artificial dissipation terms. Then in section 3 we add diffusion to this problem. The coercivity analysis of Morton and Stynes [10] was valid in this case only for a mesh that became more refined in the flow direction, and this is an awkward restriction. However, in [2] it was shown that the method has second order accuracy on any mesh, so in this section we show how a small amount of artificial dissipation could be used to establish coercivity on any mesh. Finally, in section 4 we consider the effect on the accuracy of the scheme of adding these terms to the simple formulation given by (1.2) and (1.3).

## 2 Coercivity for pure convection at an expansion critical point

In conservation law form the convection-diffusion problem in one dimension can be written

$$-\epsilon(au')' + (bu)' = S \quad \text{on } (0, 1). \quad (2.1)$$

We can suppose that  $a(x)$  is strictly positive and interest is focussed on problems which are uniformly well posed as  $\epsilon \rightarrow 0$ . In the absence of turning points, e.g. for  $b(x) > 0$ , such well-posedness is easily established for both the continuous and the discrete problem. We suppose instead that we have a single turning point or critical point, at  $\xi \in (0, 1)$  where  $b(\xi) = 0$ ; then in the case of an expanding convective flow field well-posedness of (2.1) is again simple to show, so in studying the discrete problem we assume that

$$b'(x) \geq \gamma > 0 \quad \text{on } [0, 1]. \quad (2.2)$$

Indeed, in this case the solution of the reduced problem for (2.1), i.e. when  $\epsilon = 0$ , can be given explicitly as

$$u(x) = [1/b(x)] \int_{\xi}^x S(t) dt, \quad x \neq \xi, \quad (2.3)$$

$$u(\xi) = S(\xi)/b'(\xi). \quad (2.4)$$

For  $\epsilon \neq 0$ , a Dirichlet boundary condition at  $x = 0$  or  $x = 1$  will lead to a boundary layer there.

The cell vertex method generally gives an accurate monotone approximation in a boundary layer; and the accuracy of the scheme in the absence of turning points has been thoroughly analysed in [7] and [10]. Here we extend the latter analysis to the turning point problem described above when  $\epsilon = 0$ ; the case of  $\epsilon > 0$  will be dealt with in the next section.

On a nonuniform mesh  $0 = x_0 < x_1 < \dots < x_{J-1} < x_J = 1$ , with interval lengths  $h_j = x_j - x_{j-1}$ , we suppose that the turning point is in the  $k^{\text{th}}$  interval, namely

$$x_{k-1} \leq \xi < x_k, \quad \text{with } k > 1. \quad (2.5)$$

The cell residuals are given by

$$\begin{aligned} R_{j-\frac{1}{2}}(U) := & h_j^{-1}[(b_j U_j - \epsilon a_j U'_j) - (b_{j-1} U_{j-1} - \epsilon a_{j-1} U'_{j-1})] \\ & - \frac{1}{2}(S_{j-1} + S_j), \quad j = 1, 2, \dots, J. \end{aligned} \quad (2.6)$$

Since in this section we consider only the limiting case  $\epsilon \rightarrow 0$ , the form of recovery used to obtain  $U'_j$  is irrelevant. In going to the more general formulation of (1.4) we will use upwind distribution factors for all but the  $k^{\text{th}}$  cell; that is, we set

$$D_{j-\frac{1}{2},j-1} = 1 - s_{j-\frac{1}{2}}, \quad D_{j-\frac{1}{2},j} = 1 + s_{j-\frac{1}{2}}, \quad (2.7)$$

with

$$s_{j-\frac{1}{2}} = \text{sign } b_{j-\frac{1}{2}} \quad \text{for } j \neq k, \quad (2.8)$$

where  $b_{j-\frac{1}{2}}$  is an average value of  $b$  in the cell; but we leave  $s_{k-\frac{1}{2}}$  as a free parameter. Then, as in (1.4) the nodal residuals are defined and the solution is determined by setting for  $i = 0, 1, \dots, J$

$$\begin{aligned} N_i(U) := & (h_i + h_{i+1})^{-1} [h_i (D_{i-\frac{1}{2},i} R_{i-\frac{1}{2}} + A_{i-\frac{1}{2},i}) \\ & + h_{i+1} (D_{i+\frac{1}{2},i} R_{i+\frac{1}{2}} + A_{i+\frac{1}{2},i})] = 0; \end{aligned} \quad (2.9)$$

for the boundary points we need to set  $h_0 = h_1, h_{J+1} = h_J$  and  $R_{-\frac{1}{2}} = R_{J+\frac{1}{2}} = 0$  here. We will leave until later the specification of the artificial viscosity terms, and we will denote by  $N_i^0(U)$  the nodal residuals with these terms set to zero, as well as  $\epsilon = 0$ .

In order to demonstrate coercivity, we introduce the inner product

$$(f, g) := \sum_{j=0}^J \frac{1}{2} (h_j + h_{j+1}) f_j g_j; \quad (2.10)$$

and we use this to define the bilinear form

$$B^h(V, W) := (N(U + V) - N(U), W), \quad (2.11)$$

with  $B^0(\cdot, \cdot)$  resulting when  $N^0$  is used. In this latter case it is clear that

$$\begin{aligned} B^0(V, V) = & \sum_{j=0}^{k-2} V_j \Delta_+(bV)_j + \sum_{k+1}^J V_j \Delta_-(bV)_j \\ & + \frac{1}{2} [(1 - s_{k-\frac{1}{2}}) V_{k-1} + (1 + s_{k-\frac{1}{2}}) V_k] \Delta_-(bV)_k, \end{aligned} \quad (2.12)$$

the coercivity of which can be best studied with the form given in the following lemma. Here  $\Delta_-$  and  $\Delta_+$  are the usual backward and forward undivided difference operators.

**Lemma 2.1** *For the bilinear form without artificial viscosity, we have*

$$\begin{aligned}
2B^0(V, V) = & -b_0 V_0^2 - \sum_1^{k-1} b_j (\Delta_- V_j)^2 + \sum_{k+1}^J b_{j-1} (\Delta_- V_j)^2 + b_J V_J^2 \\
& + \sum_1^{k-1} (\Delta_- b_j) V_{j-1}^2 + (\Delta_- b_k) V_k V_{k-1} + \sum_{k+1}^J (\Delta_- b_j) V_j^2 \\
& + s_{k-\frac{1}{2}} (\Delta_- V_k) \Delta_- (bV)_k.
\end{aligned} \tag{2.13}$$

**Proof** The terms under the summation signs in (2.12) can be broken up to give, for example,

$$\begin{aligned}
V_j \Delta_+ (bV)_j &= (\Delta_+ b_j) V_j^2 + b_{j+1} V_j (V_{j+1} - V_j) \\
&= (\Delta_+ b_j) V_j^2 - \frac{1}{2} b_{j+1} [(V_{j+1} - V_j)^2 - (V_{j+1}^2 - V_j^2)] \\
&= \frac{1}{2} (\Delta_+ b_j) V_j^2 - \frac{1}{2} b_{j+1} (\Delta_- V_{j+1})^2 + \frac{1}{2} \Delta_+ (b_j V_j^2).
\end{aligned}$$

Hence we have

$$\sum_0^{k-2} V_j \Delta_+ (bV)_j = \frac{1}{2} \sum_0^{k-2} (\Delta_+ b_j) V_j^2 - \frac{1}{2} \sum_1^{k-1} b_j (\Delta_- V_j)^2 + \frac{1}{2} (b_{k-1} V_{k-1}^2 - b_0 V_0^2),$$

with a similar sum to the right of the turning point. With  $s_{k-\frac{1}{2}} = 0$ , the terms in  $2B^0(V, V)$  associated with the  $k^{\text{th}}$  cell are then just

$$\begin{aligned}
& b_{k-1} V_{k-1}^2 - b_k V_k^2 + (V_{k-1} + V_k)(b_k V_k - b_{k-1} V_{k-1}) \\
&= (b_k - b_{k-1}) V_k V_{k-1},
\end{aligned}$$

which gives all the terms in (2.13), except those dependent on  $s_{k-\frac{1}{2}}$  which are obtained from (2.12).  $\square$

All the terms in (2.13) are positive definite except the product  $V_k V_{k-1}$  and the terms depending on  $s_{k-\frac{1}{2}}$ . In general, the freedom to choose  $s_{k-\frac{1}{2}}$  is inadequate to force the whole expression for  $B^0(V, V)$  to be positive definite and artificial dissipation terms are needed. However, some special cases are worth noting before we consider the general case. If  $b_{k-1} = 0$ , so that  $b_k > 0$ , the two sets of terms give

$$b_k [V_k V_{k-1} + s_{k-\frac{1}{2}} (V_k^2 - V_k V_{k-1})],$$

so that taking  $s_{k-\frac{1}{2}} = 1$  gives the positive definite term  $b_k V_k^2$ ; but then  $B^0(V, V) = 0$  if  $V_i = 0$  for  $i \neq k-1$  while  $V_{k-1} \neq 0$ , which reflects the fact that the system of residual equations contains no equation to determine  $V_{k-1}$ . Similarly, if  $b_k = 0$  and  $b_{k-1} < 0$  then taking  $s_{k-\frac{1}{2}} = -1$  makes  $B^0(V, V) \geq 0$ , but  $V_k$  is not determined by the residual equations. Between these two extremes, the case  $b_k + b_{k-1} = 0$  gives

$$b_k [2V_k V_{k-1} + s_{k-\frac{1}{2}} (V_k^2 - V_{k-1}^2)],$$

which cannot be made positive definite by any choice of  $s_{k-\frac{1}{2}}$ .

Thus we consider the addition of artificial viscosity, both second order and fourth order; note that in the latter case it is more commonly referred to as artificial dissipation. As used for modelling the Navier–Stokes equations in [1], the terms for substituting in (2.9) have the form

$$\begin{aligned} A_{i-\frac{1}{2},i} &= \tau_{i-\frac{1}{2}}^{(2)}(U_i - U_{i-1}) - \tau_{i-\frac{1}{2}}^{(4)}(\delta^2 U_i - \delta^2 U_{i-1}) \\ A_{i-\frac{1}{2},i-1} &= \tau_{i-\frac{1}{2}}^{(2)}(U_{i-1} - U_i) - \tau_{i-\frac{1}{2}}^{(4)}(\delta^2 U_{i-1} - \delta^2 U_i), \end{aligned} \quad (2.14)$$

where  $\delta^2 = \Delta_+ \Delta_- = \Delta_- \Delta_+$ . Note that the sum of these two terms associated with a single cell is zero, so that conservation is unaffected by their inclusion; and the net addition to  $N_i(U)$  at the node  $x_i$  is

$$-(h_i + h_{i+1})^{-1} \Delta_+ [h_i \tau_{i-\frac{1}{2}}^{(2)} \Delta_- U_i - h_i \tau_{i-\frac{1}{2}}^{(4)} \Delta_- \delta^2 U_i]. \quad (2.15)$$

For simplicity, we suppose that  $\tau^{(2)}$  and  $\tau^{(4)}$  are both zero in the intervals near the boundaries; then summing by parts over the contributing interior cells, we have

$$\begin{aligned} 2[B^c(V, V) - B^0(V, V)] &= -\sum V_j \Delta_+ [h_j (\tau_{j-\frac{1}{2}}^{(2)} \Delta_- V_j - \tau_{j-\frac{1}{2}}^{(4)} \Delta_- \delta^2 V_j)] \\ &= \sum h_j [\tau_{j-\frac{1}{2}}^{(2)} (\Delta_- V_j)^2 - \tau_{j-\frac{1}{2}}^{(4)} (\Delta_- V_j) (\Delta_- \delta^2 V_j)] \end{aligned} \quad (2.16)$$

where the notation  $B^c(\cdot, \cdot)$  denotes that only the convection terms are included. In the following lemma we first consider the effect of the second order terms, using a discrete norm which is motivated by the right-hand side of (2.13), namely

$$\begin{aligned} \|V\|_h^2 &:= |b_0| V_0^2 + |b_J| V_J^2 + \sum_0^J (\Delta b)_j V_j^2 \\ &\quad + \sum_1^J |b_{j-\frac{1}{2}}| (\Delta_- V_j)^2, \end{aligned} \quad (2.17)$$

where

$$(\Delta b)_j := \begin{cases} \Delta_+ b_j & \text{for } j \leq k-2 \\ \frac{1}{2} \Delta_- b_k \equiv \frac{1}{2} \Delta_+ b_{k-1} & \text{for } j = k-1, k \\ \Delta_- b_j & \text{for } j \geq k+1 \end{cases} \quad (2.18)$$

and  $|b_{j-\frac{1}{2}}| := \min(|b_{j-1}|, |b_j|)$ . We also suppose in this lemma that  $s_{k-\frac{1}{2}} = 0$ .

**Lemma 2.2** *Addition of second order artificial viscosity of the form (2.14) with*

$$\tau_{k-\frac{1}{2}}^{(2)} \geq D_- b_k \equiv (b_k - b_{k-1})/h_k, \quad (2.19)$$

*and  $\tau_{j-\frac{1}{2}}^{(2)} \geq 0$  for  $j \neq k$ , is sufficient to ensure that with  $s_{k-\frac{1}{2}} = 0$  we have*

$$B^c(V, V) \geq \frac{1}{2} \|V\|_h^2. \quad (2.20)$$



**Proof** The only nonpositive definite term in (2.13), arising from the  $k^{\text{th}}$  interval, can be rewritten by using

$$V_k V_{k-1} = \frac{1}{2}[V_{k-1}^2 + V_k^2 - (\Delta_- V_k)^2]$$

to give, with  $s_{k-\frac{1}{2}} = 0$ ,

$$B^0(V, V) = \frac{1}{2} \|V_h\|^2 - \frac{1}{2} \left[ |b_{k-\frac{1}{2}}| + \frac{1}{2} \Delta_- b_k \right] (\Delta_- V_k)^2. \quad (2.21)$$

Since  $|b_{k-\frac{1}{2}}| \leq \frac{1}{2} \Delta_- b_k$ , the coefficient of  $(\Delta_- V_k)^2$  here can be dominated by the coefficient  $\frac{1}{2} h_k \tau_{k-\frac{1}{2}}^{(2)}$  in the corresponding term provided by (2.16) if (2.19) is satisfied.  $\square$

Now let us consider whether the amount of artificial viscosity can be reduced by choosing  $s_{k-\frac{1}{2}}$  more carefully and also modifying the norm (2.17). Apart from the terms  $(\Delta_- V_{k-1})^2$  and  $(\Delta_+ V_k)^2$ , the only terms involving  $V_{k-1}$  and  $V_k$  in (2.13) can be written, with  $\beta := -b_{k-1}/b_k$  and  $s := s_{k-\frac{1}{2}}$ , as

$$b_k [(1 + \beta) V_k V_{k-1} + s(V_k^2 - \beta V_{k-1}^2 - (1 - \beta) V_k V_{k-1})]; \quad (2.22)$$

and with  $\tau := h_k \tau_{k-\frac{1}{2}}^{(2)}$  the artificial viscosity term that is added is  $\tau(V_k - V_{k-1})^2$ . Choosing  $s$  to minimise the magnitude of  $\tau$  needed to make the result positive definite yields the following result.

**Lemma 2.3** *The minimum second order artificial viscosity needed to render  $B^c(V, V)$  positive definite is given by*

$$\tau_{k-\frac{1}{2}}^{(2)} > h_k^{-1} \frac{|b_k b_{k-1}|}{|b_k| + |b_{k-1}|}, \quad (2.23)$$

with  $\tau_{j-\frac{1}{2}}^{(2)} = 0$  for  $j \neq k$ , which occurs when the distribution factors in the turning point cell are given by

$$s_{k-\frac{1}{2}} = \frac{b_k + b_{k-1}}{|b_k| + |b_{k-1}|}. \quad (2.24)$$

**Proof** The coefficient of either  $V_k^2$  or  $V_{k-1}^2$  in (2.22) is negative, according to the sign of  $s \equiv s_{k-\frac{1}{2}}$ . If  $\beta < 1$ , we choose  $s \geq 0$ ; then with the artificial viscosity term added the coefficients of both  $V_k^2$  and  $V_{k-1}^2$  are positive if  $\tau > \beta b_k s = |b_{k-1}| s$ . The whole expression is positive if this condition is strengthened to

$$\tau > \frac{1}{4} b_k [(1 + \beta)(s^2 + 1) - 2(1 - \beta)s], \quad (2.25)$$

and the right-hand side here has its minimum of  $b_k \beta / (1 + \beta)$  at  $s = (1 - \beta) / (1 + \beta)$ . This gives the result of (2.23) and (2.24), which is only changed when  $\beta > 1$  by the change in sign of  $s$ .  $\square$

We have introduced absolute value signs in the denominators of (2.23) and (2.24) to emphasise the fact that the form for  $s$  in (2.24) can be used for all  $s_{j-\frac{1}{2}}$  to generalise (2.8) to the turning point cell. Note that, with this choice, any value of  $\tau_{k-\frac{1}{2}}^{(2)} > 0$  will restore coercivity when  $b_{k-1} = 0$ ; and in the worst case, when  $b_k + b_{k-1} = 0$ , the amount needed is a quarter of that in Lemma 2.2.

However, fourth order dissipation is to be preferred because of its smaller effect on the smooth solution that is generally expected for the present problem. We show in the following lemma how it also can restore coercivity to  $B^c(V, V)$ .

**Lemma 2.4** *Suppose the distribution factors are given by (2.7) with*

$$s_{j-\frac{1}{2}} = \frac{b_j + b_{j-1}}{|b_j| + |b_{j-1}|}. \quad (2.26)$$

*Then  $B^c(V, V)$  is made positive definite by adding fourth order artificial dissipation solely in the turning point cell with*

$$(4 - 2\sqrt{3}) \frac{|b_k b_{k-1}|}{|b_k| + |b_{k-1}|} < h_k \tau_{k-\frac{1}{2}}^{(4)} < (4 + 2\sqrt{3}) \frac{|b_k b_{k-1}|}{|b_k| + |b_{k-1}|}. \quad (2.27)$$

**Proof** Using the same notation as in (2.22) except for  $\tau := h_k \tau_{k-\frac{1}{2}}^{(4)}$ , and adding in the  $(\Delta_- V_{k-1})^2$  and  $(\Delta_+ V_k)^2$  terms, we have from (2.13) and (2.16)

$$\begin{aligned} 2B^c(V, V) &\geq b_k \{ (\Delta_+ V_k)^2 + \beta (\Delta_- V_{k-1})^2 + s(V_k^2 - \beta V_{k-1}^2) \\ &\quad + [(1 + \beta) - s(1 - \beta)] V_k V_{k-1} \} \\ &\quad + \tau (\Delta_- V_k) (2\Delta_- V_k - \Delta_- V_{k-1} - \Delta_+ V_k). \end{aligned} \quad (2.28)$$

From a Cauchy–Schwarz inequality we have

$$|(\Delta_- V_k)(\Delta_+ V_k)| \leq \frac{1}{2} \left[ \left( \frac{\tau}{2b_k} \right) (\Delta_- V_k)^2 + \left( \frac{2b_k}{\tau} \right) (\Delta_+ V_k)^2 \right], \quad (2.29)$$

with a similar bound for  $|(\Delta_- V_k)(\Delta_- V_{k-1})|$ . Hence, by introducing

$$R := \frac{1}{b_k} + \frac{1}{|b_{k-1}|} \equiv \frac{1}{b_k} \left( 1 + \frac{1}{\beta} \right),$$

we obtain, from balancing the first and last pairs of terms in (2.28),

$$\begin{aligned} 2B^c(V, V) &\geq b_k \{ s(V_k^2 - \beta V_{k-1}^2) + [(1 + \beta) - s(1 - \beta)] V_k V_{k-1} \} \\ &\quad + \tau \left( 2 - \frac{1}{4} R \tau \right) (V_k - V_{k-1})^2. \end{aligned} \quad (2.30)$$

Now this is in exactly the same form as was obtained using second order artificial dissipation, and leading to the inequality (2.23). Noting that the right-hand side of (2.23) is just  $(Rh_k)^{-1}$ , we obtain the condition

$$R\tau \left( 2 - \frac{1}{4} R \tau \right) > 1 \quad (2.31)$$

which is satisfied by the inequalities in (2.27).  $\square$

The upper bound on  $\tau_{k-\frac{1}{2}}^{(4)}$  that is required by (2.27) may seem a little restrictive, because of its dependence on  $b$  near the turning point, but it is not so in practice. It comes about through the Cauchy–Schwarz inequality (2.29) when  $b_k$  is small, or the corresponding case when  $b_{k-1}$  is small. However, the fourth order dissipation would normally be applied over a patch of cells and this would overcome this problem. To illustrate the situation, let us write  $\tau_j := h_j \tau_{j-\frac{1}{2}}^{(4)}$  and set  $\tau_j = 0$  except for  $j = k-1, k, k+1$ . Then with simple Cauchy–Schwarz bounds we have from (2.16)

$$\begin{aligned} 2[B^c(V, V) - B^0(V, V)] &= \sum_{k-1}^{k+1} \tau_j (\Delta_- V_j) (2\Delta_- V_j - \Delta_- V_{j-1} - \Delta_- V_{j+1}) \\ &\geq \sum_{k-1}^{k+1} \tau_j \left[ (\Delta_- V_j)^2 - \frac{1}{2} (\Delta_- V_{j-1})^2 - \frac{1}{2} (\Delta_- V_{j+1})^2 \right]. \end{aligned}$$

By choosing

$$\tau_{k-1} = \tau_{k+1} = \frac{1}{2} \tau_k, \quad (2.32)$$

this collapses to

$$2[B^c(V, V) - B^0(V, V)] \geq \tau_k \left[ \frac{1}{2} (\Delta_- V_k)^2 - \frac{1}{4} (\Delta_- V_{k-2})^2 - \frac{1}{4} (\Delta_- V_{k+2})^2 \right]. \quad (2.33)$$

Now from the assumption (2.2), it is clear that

$$|b_{k-2}| \geq h_{k-1} \gamma, \quad b_{k+1} \geq h_{k+1} \gamma;$$

and hence the negative terms in (2.33) can be dominated by the corresponding terms in  $B^0(V, V)$  given by (2.13) if

$$\max \left( \tau_{k-\frac{3}{2}}^{(4)}, \tau_{k+\frac{1}{2}}^{(4)} \right) \leq 2\gamma, \quad \text{i.e.} \quad \tau_{k-\frac{1}{2}}^{(4)} \leq 4\gamma \min \left( \frac{h_{k-1}}{h_k}, \frac{h_{k+1}}{h_k} \right), \quad (2.34)$$

which is a very unrestrictive bound. Comparing (2.33) with (2.30), we have in fact constructed a positive definite  $B^c(V, V)$  so long as

$$\frac{|b_k b_{k-1}|}{|b_k| + |b_{k-1}|} < h_k \tau_{k-\frac{1}{2}}^{(4)} \leq 4 \min(|b_{k-2}|, b_{k+1}), \quad (2.35)$$

rather than the conditions (2.27).

The particular choice of distribution factors given by (2.26) leads to a minor modification to the discrete norm (2.17) if a lower bound to  $B^c(V, V)$  is sought, as in (2.20). If  $\tau_{k-\frac{1}{2}}^{(2)}$  is taken to be double the value on the right of (2.23), then

the quadratic form in (2.22) is modified to  $b_k V_k^2 + |b_{k-1}| V_{k-1}^2$ ; that is, (2.22) with  $s$  given by (2.24) can be written as

$$b_k V_k^2 + |b_{k-1}| V_{k-1}^2 - 2 \frac{|b_k b_{k-1}|}{|b_k| + |b_{k-1}|} (\Delta_- V_k)^2. \quad (2.36)$$

Hence we introduce  $\|V\|_{h*}^2$  as follows,

$$\|V\|_{h*}^2 \equiv \|V\|_h^2, \quad \text{except that} \quad (\Delta b)_{k-1} = |b_{k-1}|, (\Delta b)_k = b_k. \quad (2.37)$$

Also, since  $|b_{k-\frac{1}{2}}|$  is not less than the coefficient of  $2(\Delta_- V_k)^2$  in (2.36), it is easily checked that

$$B^c(V, V) \geq \frac{1}{2} \|V\|_{h*}^2 \quad \text{if} \quad \tau_{k-\frac{1}{2}}^{(2)} \geq 3 |b_{k-\frac{1}{2}}| / h_k. \quad (2.38)$$

Similar results can be obtained using fourth order dissipation, although the coefficients of  $(\Delta_- V_{k-1})^2$  and  $(\Delta_- V_{k+1})^2$  (or those of terms further from the turning point) are reduced by the arguments of (2.28) and (2.29). For example, it is straightforward to show that the choice of coefficients given by (2.32) gives the following result,

$$B^c(V, V) \geq \frac{1}{4} \|V\|_{h*}^2 \quad (2.39)$$

$$\text{if} \quad 5 \min(|b_{k-1}|, b_k) \leq \tau_k \leq 2 \max(|b_{k-2}|, b_{k+1}); \quad (2.40)$$

note that the dissipation can be spread more widely if these conditions cannot be met. We shall use these lower bounds in the error analysis of section 4.

If artificial dissipation is to be applied more generally, as is often needed for more complex problems, its form and the variation of its coefficients needs to be considered more carefully, for example whether it should be based on divided rather than undivided differences. This will not be a major consideration in the present paper. In the next two sections, however, where the effect of artificial dissipation on the coercivity of dissipation terms and on the truncation error are considered, we will need to take account of the wider application of such terms.

### 3 Coercivity of diffusion terms on a general mesh

In [7] an energy analysis of the error obtained with a first order approximation to the diffusive fluxes  $\epsilon a_j U_j'$  in (2.6) was carried out on a general mesh. And in [10] a similar analysis was performed with a second order approximation on a mesh that was restricted to be graded in the flow direction, i.e.  $b_j > 0 \forall j \Rightarrow h_j \geq h_{j+1}$ . However, it has been shown by García-Archilla and Mackenzie [2] that this latter scheme is second order convergent even on a random mesh. In this section we

will therefore show how the addition of artificial viscosity can be used to render this scheme coercive on any mesh.

For simplicity we will apply homogeneous Neumann boundary conditions at both  $x = 0$  and  $x = 1$ , so the nodal residual equations (2.9) are used to determine all the unknown nodal values as in the pure convection case; only the more general definition of the cell residuals given by (2.6) need to be substituted in (2.9), and here we shall assume that  $a_j = 1 \quad \forall j$ . The definitions of (2.10) and (2.11) remain valid as well as the expansion for  $B^0(V, V)$  given by (2.12), but the diffusion terms now need to be added to give the full bilinear form

$$B^h(U, V) = B^c(U, V) + \epsilon B^d(U, V). \quad (3.1)$$

Combining (2.6), (2.7), (2.9) and (2.11) with the boundary conditions  $U'_0 = U'_J = 0$ , extended by  $U'_{-1} = U'_{J+1} = 0$ , gives

$$B^d(U, V) = - \sum_0^J V_j \left[ \frac{1}{2}(1 + s_{j-\frac{1}{2}})\Delta_- U'_j + \frac{1}{2}(1 - s_{j+\frac{1}{2}})\Delta_+ U'_j \right]. \quad (3.2)$$

Summing this by parts and using the identity

$$\Delta_+ \left[ (1 + s_{j-\frac{1}{2}})V_j \right] + \Delta_- \left[ (1 - s_{j+\frac{1}{2}})V_j \right] \equiv (1 + s_{j+\frac{1}{2}})\Delta_+ V_j + (1 - s_{j-\frac{1}{2}})\Delta_- V_j,$$

we then obtain

$$B^d(U, V) = \sum_1^J (\Delta_- V_j) \left[ \frac{1}{2}(1 + s_{j-\frac{1}{2}})U'_{j-1} + \frac{1}{2}(1 - s_{j-\frac{1}{2}})U'_j \right]. \quad (3.3)$$

Now the general form used to give the gradients  $U'_j$  in terms of the nodal values is

$$U'_j = \alpha_j D_+ U_j + (1 - \alpha_j) D_- U_j, \quad (3.4)$$

where  $D_+$  and  $D_-$  denote the divided forward and backward difference operators, respectively. Thus we finally obtain the following quadratic form when  $U = V$ ,

$$B^d(V, V) = \sum_1^J c_{j-\frac{1}{2}} (D_- V_j)^2 + \sum_1^{J-1} d_j (D_- V_j)(D_+ V_j), \quad (3.5)$$

where

$$c_{j-\frac{1}{2}} = h_j \left[ \frac{1}{2}(1 - s_{j-\frac{1}{2}})(1 - \alpha_j) + \frac{1}{2}(1 + s_{j-\frac{1}{2}})\alpha_{j-1} \right] \quad (3.6)$$

$$d_j = h_j \left[ \frac{1}{2}(1 - s_{j-\frac{1}{2}})\alpha_j \right] + h_{j+1} \left[ \frac{1}{2}(1 + s_{j+\frac{1}{2}})(1 - \alpha_j) \right], \quad (3.7)$$

and we set  $\alpha_0 = 0, \alpha_J = 1$ .

The positive definiteness of  $B^d(V, V)$  is most easily considered when the cell vertex scheme reduces to the three-point fully upwind difference scheme away

from the turning point, for then only  $d_{k-1}$  and  $d_k$  are nonzero in the above expansions. This scheme occurs when the distribution factors given by (2.8), implying that  $s_{j-\frac{1}{2}} = -1$  for  $j \leq k-1$  and  $s_{j-\frac{1}{2}} = 1$  for  $j \geq k+1$ , are coupled with the gradient choice given by

$$\alpha_j = 0 \quad \text{for } j < k-1, \quad \alpha_j = 1 \quad \text{for } j > k. \quad (3.8)$$

Then, writing  $s_{k-\frac{1}{2}} = s$ , we have

$$\begin{aligned} d_{k-1} &= \alpha_{k-1} h_{k-1} + \frac{1}{2}(1+s)(1-\alpha_{k-1})h_k, \quad d_k \\ &= \frac{1}{2}(1-s)\alpha_k h_k + (1-\alpha_k)h_{k+1}, \end{aligned} \quad (3.9)$$

with all other  $d_j$  equal to zero; no choice of  $\alpha_{k-1}$ ,  $\alpha_k$  and  $s$  can make both of these zero. Moreover, from (3.6) we see that the corresponding two cross product terms would have to be dominated by terms given by

$$\begin{aligned} c_{k-\frac{3}{2}} &= (1-\alpha_{k-1})h_{k-1}, \quad c_{k-\frac{1}{2}} = \left[\frac{1}{2}(1-s)(1-\alpha_k) + \frac{1}{2}(1+s)\alpha_{k-1}\right]h_k, \\ c_{k+\frac{1}{2}} &= \alpha_k h_{k+1}. \end{aligned} \quad (3.10)$$

Assuming each  $c_j$  is positive, the necessary and sufficient condition for this domination to hold is that

$$4c_{k-\frac{3}{2}}c_{k-\frac{1}{2}}c_{k+\frac{1}{2}} \geq c_{k+\frac{1}{2}}d_{k-1}^2 + c_{k-\frac{3}{2}}d_k^2. \quad (3.11)$$

This condition is a direct generalisation of the familiar  $4ac \geq b^2$  condition for a quadratic in two variables, and is readily derived from the conditions for the tridiagonal matrix associated with the quadratic form (3.5) to be positive definite; for future reference it is worth noting what these are, namely that the determinants  $\Delta_j$  of the principal minors be positive, where these quantities are given by the recursion  $\Delta_0 = 1$ ,  $\Delta_1 = c_{\frac{1}{2}}$  and

$$\Delta_j = c_{j-\frac{1}{2}}\Delta_{j-1} - \frac{1}{4}d_{j-1}^2\Delta_{j-2}. \quad (3.12)$$

Suppose we take  $s = 1$  and  $\alpha_k = 1$ , thus extending the upwinding on the right down to the turning point cell and making  $d_k = 0$ . Then (3.11) reduces to

$$4\alpha_{k-1}(1-\alpha_{k-1})h_{k-1}h_k \geq [\alpha_{k-1}h_{k-1} + (1-\alpha_{k-1})h_k]^2,$$

which can only be satisfied if

$$\alpha_{k-1}h_{k-1} = (1-\alpha_{k-1})h_k, \quad \text{i.e.} \quad \alpha_{k-1} = h_k/(h_{k-1} + h_k). \quad (3.13)$$

This corresponds to the scheme for calculating the gradient at  $x_{k-1}$  called Method B by Mackenzie and Morton [7].

However, the main practical interest is in the coercivity of the second order accurate scheme that is obtained by using this formula everywhere; that is, by interpolating a quadratic to  $U_{j-1}, U_j, U_{j+1}$  at each triplet of mesh points in order to obtain  $U'_j$ , which gives

$$\alpha_j = h_j/(h_j + h_{j+1}), \quad 1 - \alpha_j = h_{j+1}/(h_j + h_{j+1}). \quad (3.14)$$

This in turn gives for the coefficients (3.6) and (3.7) in the quadratic form (3.5),

$$c_{j-\frac{1}{2}} = \frac{1}{2}(1 - s_{j-\frac{1}{2}}) \frac{h_j h_{j+1}}{h_j + h_{j+1}} + \frac{1}{2}(1 + s_{j-\frac{1}{2}}) \frac{h_{j-1} h_j}{h_{j-1} + h_j} \quad (3.15)$$

$$d_j = \frac{1}{2}(1 - s_{j-\frac{1}{2}}) \frac{h_j^2}{h_j + h_{j+1}} + \frac{1}{2}(1 + s_{j+\frac{1}{2}}) \frac{h_{j+1}^2}{h_j + h_{j+1}}. \quad (3.16)$$

To avoid too much complication, we will use the simple compact conditions

$$d_j^2 \leq c_{j-\frac{1}{2}} c_{j+\frac{1}{2}} \quad \text{for } 2 \leq j \leq J-2, \quad (3.17)$$

$$d_1^2 \leq 2c_{\frac{1}{2}} c_{\frac{3}{2}}, \quad d_{J-1}^2 \leq 2c_{J-\frac{3}{2}} c_{J-\frac{1}{2}} \quad (3.18)$$

which are sufficient to ensure that all the determinants in (3.12) are non-negative and hence that  $B^d(\cdot, \cdot)$  is positive semi-definite.

Now the addition of fourth order artificial dissipation supplements  $\epsilon B^d(V, V)$ , as in (2.16), by

$$\frac{1}{2} \sum h_j \tau_{j-\frac{1}{2}}^{(4)} (\Delta_- V_j) [2\Delta_- V_j - \Delta_- V_{j+1} - \Delta_- V_{j-1}] \quad (3.19)$$

if, as we have so far, we use undivided differences for its definition. The result is that the coefficients in (3.5) are modified to give

$$c_{j-\frac{1}{2}}^* = c_{j-\frac{1}{2}} + \epsilon^{-1} h_j^3 \tau_{j-\frac{1}{2}}^{(4)}, \quad d_j^* = d_j - \frac{1}{2} \epsilon^{-1} h_j h_{j+1} (h_j \tau_{j-\frac{1}{2}}^{(4)} + h_{j+1} \tau_{j+\frac{1}{2}}^{(4)}). \quad (3.20)$$

This strongly suggests replacing (3.5) by the equivalent quadratic form in the undivided differences, and hence introducing the notation

$$\tau_{j-\frac{1}{2}} := \epsilon^{-1} h_j \tau_{j-\frac{1}{2}}^{(4)}, \quad \tilde{c}_{j-\frac{1}{2}} = h_j^{-2} c_{j-\frac{1}{2}}, \quad \tilde{d}_j = (h_j h_{j+1})^{-1} d_j; \quad (3.21)$$

then the condition (3.17) for positive definiteness becomes

$$[\tilde{d}_j - \frac{1}{2}(\tau_{j-\frac{1}{2}} + \tau_{j+\frac{1}{2}})]^2 \leq (\tilde{c}_{j-\frac{1}{2}} + \tau_{j-\frac{1}{2}})(\tilde{c}_{j+\frac{1}{2}} + \tau_{j+\frac{1}{2}}) \quad \forall j. \quad (3.22)$$

It is clear that this can always be satisfied with sufficiently large values of  $\tau^{(4)}$ ; what we now need to do is estimate how large these values have to be.

Taking a constant value

$$h_j \tau_{j-\frac{1}{2}}^{(4)} \equiv \epsilon \tau_{j-\frac{1}{2}} = \epsilon \tau \quad \forall j, \quad (3.23)$$

ensures, from summing (3.19) by parts, that increasing the artificial viscosity always increases the positive definiteness of  $B^d(\cdot, \cdot)$ ; also the quadratic terms in (3.22) then cancel and the conditions become

$$\tau \geq \frac{\tilde{d}_j^2 - \tilde{c}_{j-\frac{1}{2}} \tilde{c}_{j+\frac{1}{2}}}{2\tilde{d}_j + \tilde{c}_{j-\frac{1}{2}} + \tilde{c}_{j+\frac{1}{2}}} \quad \forall j. \quad (3.24)$$

To bound the right-hand side, we introduce a notation for the mesh ratios and their upper and lower bounds,

$$\gamma \leq \gamma_j := h_{j+1}/h_j \leq \Gamma \quad \forall j. \quad (3.25)$$

Then for  $s_{j-\frac{1}{2}} = s_{j+\frac{1}{2}} = 1$ , we obtain

$$\tilde{c}_{j-\frac{1}{2}} = \frac{1}{h_j(1 + \gamma_{j-1})}, \quad \tilde{d}_j = \frac{\gamma_j}{h_j(1 + \gamma_j)},$$

which leads to the condition, after a little algebra, that we need

$$\tau \geq \frac{1}{h_j(1 + \gamma_j)} \frac{\gamma_j^3 + \gamma_j^3 \gamma_{j-1} - 1 - \gamma_j}{2\gamma_j^2 + 2\gamma_j^2 \gamma_{j-1} + \gamma_j + \gamma_j^2 + 1 + \gamma_{j-1}} \quad \text{for } j \geq k+1.$$

This implies that no artificial dissipation is needed here if the mesh is decreasing towards the boundary, as was shown by Morton and Stynes [10]; a similar result holds for  $j \leq k-2$ . However, for the following lemma we will use the obviously sufficient condition  $\tau \geq \frac{1}{2}\tilde{d}_j$ .

Note that so far we have been imprecise about the boundary conditions for the artificial dissipation, and have previously assumed that it was set to zero near the boundaries. However, for the present purposes it is clearly convenient to apply it to every cell by satisfying (3.23). It is easily checked that this can be achieved by setting  $\delta^2 V_0 = 0 = \delta^2 V_J$  in its definition, and that then all the summing by parts that has been applied is valid. Thus in (3.23), (3.24) and below it is understood that  $j = 1, 2, \dots, J$ .

**Lemma 3.1** *The addition of fourth order dissipation with coefficients given by (3.23) such that*

$$\tau h_j \geq \frac{1}{2} \quad \forall j, \quad (3.26)$$

*is sufficient to make  $B^d(\cdot, \cdot)$  positive definite.*



**Proof** In the worst case, which is when  $s_{j-\frac{1}{2}} = -1$  and  $s_{j+\frac{1}{2}} = 1$  and might occur for  $j = k$ , we have from (3.16) and (3.20) that

$$\tilde{d}_j = \frac{1}{h_j + h_{j+1}} \left( \frac{h_{j+1}}{h_j} + \frac{h_j}{h_{j+1}} \right) \leq \min(h_j^{-1}, h_{j+1}^{-1}).$$

Then the condition  $\tau \geq \frac{1}{2}\tilde{d}_j$  gives (3.26).  $\square$

The condition (3.26) is clearly far from sharp; in particular, the fact that from (3.23) we then have

$$\tau_{j-\frac{1}{2}}^{(4)} \geq \frac{1}{2}\epsilon h_j^{-2}$$

is inconvenient where the mesh is very fine. This strongly suggests that divided differences should be used in defining the artificial dissipation, instead of (2.14).

## 4 Coercivity-based error bounds

We denote the linear interpolant of the exact solution  $u$  of (2.1) by  $u^I$ , the difference  $U - u^I$  by  $E$  and the gradient recovery error  $u'(x_j) - (u^I)'_j$  by  $\eta_j$ . Then since the cell vertex approximation is given by  $N(U) = 0$ , we obtain from (2.11)

$$\begin{aligned} B^h(E, E) &= (N(u^I + E) - N(u^I), E) \\ &= -(N(u^I), E). \end{aligned} \quad (4.1)$$

The term  $N(u^I)$  has the form of a truncation error, in part arising from the cell residual and in part from the artificial dissipation terms. For the former, we assume that we replace the trapezoidal rule in (2.6) by exact integration of the source term and take  $a(x) \equiv 1$ ; then we have

$$\begin{aligned} h_j R_{j-\frac{1}{2}}(u^I) &= \Delta_-(b_j u_j^I - \epsilon(u^I)'_j) - \int_{x_{j-1}}^{x_j} S(x) dx \\ &= \Delta_-(b_j u_j^I - \epsilon(u^I)'_j) - \Delta_-(b_j u(x_j) - \epsilon u'(x_j)) \\ &= \epsilon \Delta_-(u'(x_j) - (u^I)'_j) =: \epsilon \Delta_- \eta_j. \end{aligned} \quad (4.2)$$

If we also use only fourth order artificial dissipation we define, as in (2.14),

$$A_{i-\frac{1}{2},i}^I := -\tau_{i-\frac{1}{2}}^{(4)} \Delta_- \delta^2 u_i^I, \quad \text{etc.} \quad (4.3)$$

in order to obtain the second contribution to the truncation error.

From (2.9), and using the distribution factors given by (2.26) we therefore obtain

$$\begin{aligned} (N(u^I), E) &= \frac{1}{2} \sum_0^J E_j [h_j (D_{j-\frac{1}{2},j} R_{j-\frac{1}{2}}(u^I) + A_{j-\frac{1}{2},j}^I) \\ &\quad + h_{j+1} (D_{j+\frac{1}{2},j} R_{j+\frac{1}{2}}(u^I) + A_{j+\frac{1}{2},j}^I)] \\ &=: S_R + S_A, \end{aligned} \quad (4.4)$$

where, as in (2.12),

$$\begin{aligned}
S_R &= \epsilon \sum_0^J E_j \left[ \frac{1}{2} (1 + s_{j-\frac{1}{2}}) \Delta_- \eta_j + \frac{1}{2} (1 - s_{j+\frac{1}{2}}) \Delta_- \eta_{j+1} \right] \\
&= \epsilon \left\{ \sum_0^{k-2} E_j \Delta_+ \eta_j + \sum_{k+1}^J E_j \Delta_- \eta_j \right. \\
&\quad \left. + \left( \frac{|b_{k-1}| E_{k-1} + b_k E_k}{|b_{k-1}| + b_k} \right) \Delta_- \eta_k \right\}, \tag{4.5}
\end{aligned}$$

and, if artificial dissipation is applied in all cells by setting  $\delta^2 u_0^I$  and  $\delta^2 u_J^I$  to zero,

$$S_A = \frac{1}{2} \sum_0^J E_j \Delta_+ (h_j \tau_{j-\frac{1}{2}}^{(4)} \Delta_- \delta^2 u_j^I), \tag{4.6}$$

with  $\tau_{-\frac{1}{2}}^{(4)} = \tau_{J+\frac{1}{2}}^{(4)} = 0$ . In order to bound the ratio of this inner product with  $\|E\|_{h*}$ , we could sum each of these by parts and make the maximum use of the terms in (2.17) and (2.37). For simplicity, however, we generally will use only the forms (4.5) and (4.6) to obtain

$$|(N(u^I), E)| \leq \|E\|_{h*} \left( \epsilon \|\eta\|_{\Delta} + \frac{1}{2} \|u^I\|_{\tau} \right), \tag{4.7}$$

where, from (2.18) and (2.37),

$$\|\eta\|_{\Delta}^2 = \sum_0^{k-2} \frac{(\Delta_+ \eta_j)^2}{\Delta_+ b_j} + \frac{(\Delta_- \eta_k)^2}{|b_{k-1}| + b_k} + \sum_{k+1}^J \frac{(\Delta_- \eta_j)^2}{\Delta_- b_j}, \tag{4.8}$$

and

$$\|u^I\|_{\tau}^2 = \sum_0^J (\Delta b)_j^{-1} \left| \Delta_+ (h_j \tau_{j-\frac{1}{2}}^{(4)} \Delta_- \delta^2 u_j^I) \right|^2, \tag{4.9}$$

with the notational convention at the boundaries as described above. We conclude by combining these expressions with the bounds derived in sections 3 and 4 to give the following result.

**Theorem 4.1** *Suppose the cell vertex method defined by (2.9), (2.7) and (2.26) is applied to the expansion critical point problem given by (2.1) and (2.2). Then the addition of fourth order artificial dissipation to satisfy (2.40) and (3.26) enforces coercivity on any mesh and yields the following error bound*

$$\|E\|_{h*} \leq \epsilon [4 \|\eta\|_{\Delta} + \tau_d \|u^I\|_{(4)}] + \tau_c \|u^I\|_k, \tag{4.10}$$

where  $\|\eta\|_\Delta$  is given by (4.8),  $\tau_d = h_{\min}^{-1}$ ,  $\tau_c = 10 |b_{k-\frac{1}{2}}|$ ,

$$\|u^I\|_4^2 = (\Delta b)_0^{-1} (\delta^2 u_1^I)^2 + (\Delta b)_{J-1}^{-1} (\delta^2 u_{J-1}^I)^2 + \sum_1^{J-1} (\Delta b)_j^{-1} (\delta^4 u_j^I)^2, \quad (4.11)$$

with  $\delta^2 u_0^I = 0 = \delta^2 u_J^I$ , and

$$\begin{aligned} \|u^I\|_k^2 &= \frac{1}{4} |b_{k-\frac{3}{2}}|^{-1} (\Delta_- \delta^2 u_{k-1}^I)^2 + |b_{k-\frac{1}{2}}|^{-1} (\Delta_- \delta^2 u_k^I)^2 \\ &\quad + \frac{1}{4} |b_{k+\frac{1}{2}}|^{-1} (\Delta_- \delta^2 u_{k+1}^I)^2. \end{aligned} \quad (4.12)$$

**Proof** Artificial dissipation satisfying (2.40) ensures that  $B^c(E, E) \geq \frac{1}{4} \|E\|_{h*}^2$ , and a further amount satisfying the conditions of Lemma 3.1 ensures that  $B^d(E, E) \geq 0$ . Hence we have  $B^h(E, E) \geq \frac{1}{4} \|E\|_{h*}^2$ . To obtain (4.10) we use (4.1) and (4.4)–(4.6), substituting the two sets of artificial dissipation terms into (4.6) before applying the Cauchy–Schwarz inequality separately. For the diffusion inner product, the bounds are obtained as in (4.7)–(4.9) so as to give (4.11), with  $\tau_d$  obtained from (3.26). However, for the convection inner product it is more convenient to sum by parts to give, from (2.32),

$$-\frac{1}{2} \tau_k [(\Delta_- E_{k-1}) \frac{1}{2} \Delta_- \delta^2 u_{k-1}^I + (\Delta_- E_k) \Delta_- \delta^2 u_k^I + (\Delta_- E_{k+1}) \frac{1}{2} \Delta_- \delta^2 u_{k+1}^I].$$

Then applying a Cauchy–Schwarz inequality to this, with  $\tau_k$  given by (2.40) and  $\tau_c = 2\tau_k$ , yields (4.12).  $\square$

Let us consider briefly the order of accuracy implied by this bound, remembering that we can assume that the solution  $u$  is quite smooth. The coefficients  $(\Delta b)_j^{-1}$  in the norms will normally be  $O(h^{-1})$ ; since we will have  $\Delta \eta = O(h^3)$  on a smooth mesh, we can expect  $\|\eta\|_\Delta$  to be  $O(h^2)$ . Similarly, with  $\delta^4 u^I = O(h^4)$  and  $\tau_d = O(h^{-1})$ , we can expect  $\tau_d \|u^I\|_{(4)}$  to be  $O(h^2)$ , although more care may be needed in the treatment of the  $\delta^2 u_1^I$  and  $\delta^2 u_{J-1}^I$  terms in (4.11), as well as the terms involving  $(\Delta b)_j^{-1}$  for  $j = k-1$  and  $k$ . However, both of these terms are multiplied by the factor  $\epsilon$ . Finally, the locally applied artificial viscosity near the turning point gives a term of the order  $|b_{k-\frac{1}{2}}|^{\frac{1}{2}} \Delta_- \delta^2 u^I$  which is  $O(h^{3.5})$ . Thus in both cases the addition of fourth order artificial viscosity has greatly widened the range of conditions under which the scheme is coercive, without affecting its accuracy.

## References

- [1] P.I. Crumpton, J.A. Mackenzie, and K.W. Morton. Cell vertex algorithms for the compressible Navier-Stokes equations. *J. Comput. Phys.*, 109(1):1–15, 1993.

- [2] B. García-Archilla and J.A. Mackenzie. Analysis of a supraconvergent cell vertex finite-volume method for one-dimensional convection-diffusion problems. *IMA J. Numer. Anal.*, 15:101–115, 1995.
- [3] M.G. Hall. Cell-vertex multigrid schemes for solution of the Euler equations. In K.W. Morton and M.J. Baines, editors, *Proceedings of the Conference on Numerical Methods for Fluid Dynamics, University of Reading*, pages 303–345. Clarendon Press, 1985.
- [4] H.B. Keller and T. Cebeci. Accurate numerical methods for boundary layer flow I: two-dimensional laminar flows. In *Lecture Notes in Physics, Proceedings of Second International Conference on Numerical Methods in Fluid Dynamics*, pages 92–100. Springer-Verlag, 1971.
- [5] H.B. Keller and T. Cebeci. Accurate numerical methods for boundary layer flows II: two-dimensional turbulent flows. *AIAA J.*, 10(9):1193–1199, 1972.
- [6] J.A. Mackenzie. *Cell vertex finite volume methods for the solution of the compressible Navier-Stokes equations*. PhD thesis, Oxford University Computing Laboratory, 11 Keble Road, Oxford, OX1 3QD, 1991.
- [7] J.A. Mackenzie and K.W. Morton. Finite volume solutions of convection-diffusion test problems. *Math. Comp.*, 60(201):189–220, 1992.
- [8] K.W. Morton and M.F. Paisley. A finite volume scheme with shock fitting for the steady Euler equations. *J. Comput. Phys.*, 80:168–203, 1989.
- [9] K.W. Morton, M.A. Rudgyard, and G.J. Shaw. Upwind iteration methods for the cell vertex scheme in one dimension. *J. Comput. Phys.*, 114(2):209–226, 1994.
- [10] K.W. Morton and M. Stynes. An analysis of the cell vertex method. *M<sup>2</sup>AN*, 28(6):699–724, 1994.
- [11] R.-H. Ni. A multiple grid scheme for solving the Euler equations. *AIAA Journal*, 20(11):1565–1571, Nov 1981.
- [12] A. Preissmann. Propagation des intumescences dans les canaux et rivières. In *1st Congrès de l'Assoc. Française de Calc.*, pages 433–442, AFCAL, Grenoble, 1961.
- [13] V. Thomée. A stable difference scheme for the mixed boundary problem for a hyperbolic first order system in two dimensions. *J. Soc. Indust. Appl. Math.*, 10:229–245, 1962.
- [14] B. Wendroff. On centered difference equations for hyperbolic systems. *J. Soc. Indust. Appl. Math.*, 8:549–555, 1960.