MATHEMATICS OF COMPUTATION
VOLUME 60, NUMBER 201
JANUARY 1992, PAGES 189–220

FINITE VOLUME SOLUTIONS OF CONVECTION-DIFFUSION TEST PROBLEMS

J. A. MACKENZIE AND K. W. MORTON

ABSTRACT. The cell-vertex formulation of the finite volume method has been developed and widely used to model inviscid flows in aerodynamics: more recently, one of us has proposed an extension for viscous flows. The purpose of the present paper is two-fold: first we have applied this scheme to a well-known convection-diffusion model problem, involving flow round a 180° bend, which highlights some of the issues concerning the application of the boundary conditions in such cell-based schemes. The results are remarkably good when the boundary conditions are applied in an appropriate manner. In our efforts to explain the high quality of the results we were led to a detailed analysis of the corresponding one-dimensional problem. Our second purpose is thus to gather together various approaches to the analysis of this problem and to draw attention to the supra-convergence phenomena enjoyed by the proposed methods.

1. INTRODUCTION

Since their independent introduction by McDonald [14] and MacCormack and Paullay [11] for the discretization of the transonic Euler equations, finite volume methods have taken a leading role in computational fluid dynamics. The more recent popularization of these methods by Jameson et al. [7], Ni [21] and others has now established them as the dominant discretization schemes in the computation of aeronautical fluid flows. Over the years there have been many variants of the finite volume method, but two main types of formulation have emerged. In the cell-center approach, associated with the name of Jameson, although he has used both successfully, values of the unknowns are held at the centers of the cells over which conservation is imposed. In the cell-vertex scheme they are held at the vertices of these same cells. This presupposes that we are using quadrilateral cells in two dimensions, or hexahedral cells in three dimensions.

Morton and Paisley [18] have given good reasons, stemming mainly from the greater compactness of the stencil, why the cell-vertex formulation should be preferred for modeling inviscid flows. On the other hand, for viscous flows, modelled by the Navier-Stokes equations, one might expect the advantage to lie with the cell-center methods. However, we shall show that our cell-vertex

Received by the editor June 13, 1991.
1991 *Mathematics Subject Classification.* Primary 65N99, 65L10; Secondary 76M25, 76R05.
Key words and phrases. Convection-diffusion, finite volume, cell-vertex.
The work reported here forms a part of the research programme of the Oxford-Reading Institute for Computational Fluid Dynamics.

0025-5718/92 $1.00 + $.25 per page

scheme has some attractive properties in this case too. This paper has arisen from an investigation of the performance of the scheme for two well-known convection-diffusion test problems, proposed at an IAHR workshop, the results of which are summarized by Smith and Hutton [23].

2. The cell-vertex method

We begin by describing the cell-vertex method for the steady convection-diffusion problem

$$\nabla \cdot (\varepsilon \nabla u - \mathbf{a}u) = f \quad \text{in } \Omega, \tag{2.1a}$$

$$u = g \quad \text{on } \Gamma_D, \tag{2.1b}$$

and

$$\partial u/\partial n = 0 \quad \text{on } \Gamma_N, \tag{2.1c}$$

where ε is a positive diffusion coefficient and $\mathbf{a} = (a, b)^T$ is the convective velocity field. The domain Ω is an open bounded region of $\mathbb{R}^2$ with boundary $\Gamma_D \cup \Gamma_N$. We assume that the domain is partitioned by a structured mesh of quadrilaterals and suppose, for simplicity, that its vertices can be labelled $\{(i, j) | i = 0, 1, \ldots, M; j = 0, 1, \ldots, N\}$.

Noting that the left-hand side of problem (2.1a) can be considered as the divergence of a vector flux function $\mathbf{W} = (F, G)$ with $F = \varepsilon u_x - au$ and $G = \varepsilon u_y - bu$, we can obtain an algebraic equation for each interior cell by integrating (2.1a) over the cell, using the divergence theorem to convert this into line integrals of normal fluxes along the cell edges, and approximating these using the trapezoidal rule: for cell C of Figure 1(a),

$$\begin{aligned} \int_C \operatorname{div}(F, G)\, dx dy &= \int_{\partial C} F dy - G dx \\ &\approx \frac{1}{2}[(F_1 - F_3)(y_2 - y_4) + (F_2 - F_4)(y_3 - y_1) \\ &\quad - (G_1 - G_3)(x_2 - x_4) - (G_2 - G_4)(x_3 - x_1)]. \end{aligned} \tag{2.2}$$

With the approximation $U(x, y)$ parametrized by its values $U_{i,j}$ at the vertices, this still leaves ∇u to be approximated at the same points. There are several ways in which this may be done, but we consider mainly that called Method A in Mackenzie [12]. That is, each component of ∇u is also considered as a divergence, and its value at the vertex is obtained as an average over the subsidiary quadrilateral centered at that point and obtained by integrating along the diagonals as in Figure 1(b). Thus we have

$$\left.\frac{\partial u}{\partial x}\right|_1 \approx U_1^{(x)} := \frac{1}{2V_1}[(U_E - U_W)(y_N - y_S) + (U_N - U_S)(y_W - y_E)], \tag{2.3}$$

$$\left.\frac{\partial u}{\partial y}\right|_1 \approx U_1^{(y)} := -\frac{1}{2V_1}[(U_E - U_W)(x_N - x_S) + (U_N - U_S)(x_W - x_E)]. \tag{2.4}$$

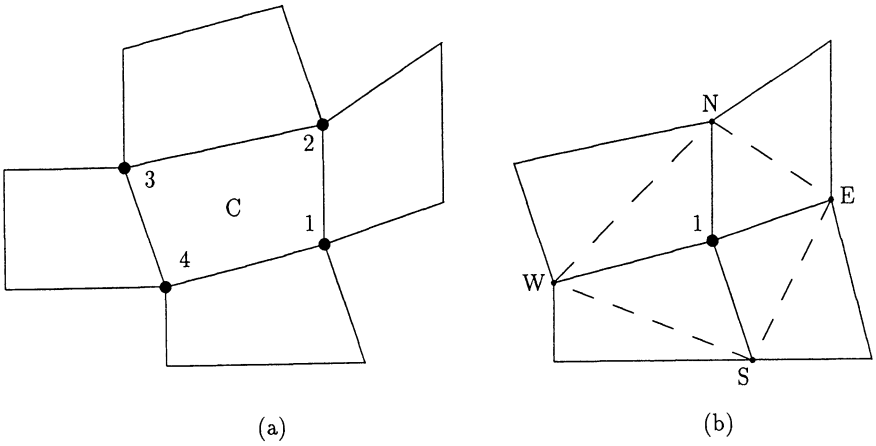


FIGURE 1. Geometric configuration of flow variables: (a) the conservation cell C; (b) the subblock used for a derivative at 1

On a rectangular uniform grid, approximations (2.3) and (2.4) coincide with the standard second-order central difference formulae. To complete the approximation over cell C, the right-hand side of (2.1a) is assumed to be integrated exactly.

It remains to consider how boundary conditions are to be imposed and the set of cell residual equations assembled so as to yield a nonsingular system. We suppose that for the differential system u is prescribed at some points, including all those corresponding to inflow, and otherwise the homogeneous Neumann condition $\partial u/\partial n = 0$ is to be imposed. For the discrete system we shall assume that P boundary vertices have their values prescribed and that $P \geq M + N + 1$, that is, that at least half are prescribed. Then the total number of unknowns is $(M + 1)(N + 1) - P \leq MN$, so that there are sufficient cell equations that may be used to determine them: to obtain an exact match, various algorithms may be used. We prefer one based on upwinded control volumes used in the Moores' method [15], but derived through a Petrov-Galerkin formulation. The derivation starts from the Galerkin equations, which associate each nodal unknown with its test function, which is identical to a piecewise linear trial function. Then, this is replaced by a piecewise constant test function over a quadrilateral—as in a cell-centered finite volume method. Finally, this is shifted upwind to coincide with one of the four cells meeting at the node. The upwinding is based on the convective velocity at the node and results in each nodal unknown being associated with just one cell residual. We shall confine our consideration here to cases where, in turn, each interior cell residual is associated in this way with just one unknown: there may be some boundary cells, where Dirichlet conditions are imposed and the flow is directed outwards, which are not associated with unknowns and their residuals will not be used. A form of the allocation algorithm which will deal with all flow situations is given in Morton [17].

For the boundary edges, the normal flux is approximated as follows: first the derivative along each edge can be approximated by the divided difference in that direction, $(U_P - U_Q)/|\mathbf{r}_P - \mathbf{r}_Q|$ in Figure 2; then the derivative along the adjoining edge at each boundary vertex is extrapolated from the divided

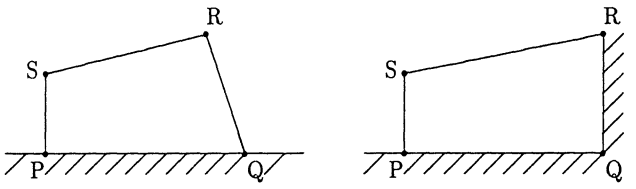


FIGURE 2. Boundary cells

difference along that edge and the derivative at the other vertex of the edge—that is, in Figure 2, we have

$$U_Q^{(QR)} = 2(U_R - U_Q)/|\mathbf{r}_R - \mathbf{r}_Q| - U_R^{(QR)}. \tag{2.5}$$

Finally, these two types of data can be combined to approximate the normal fluxes across each of the boundary edges such as PQ. Clearly $U_Q^{(QR)}$, $U_P^{(PS)}$ and the edge differences can be combined for this purpose whether or not P or Q is a corner point.

The layout of the rest of the paper is as follows: in the next section we present results obtained for a pair of well-known two-dimensional test problems at a wide range of mesh Péclet numbers. Then, in §4, as a first step in attempting to explain these remarkably good results, we analyze a corresponding one-dimensional problem in a number of different ways: we look at monotonicity of solutions, the existence of a maximum principle, an energy identity, and estimation of a discrete Green's function. Finally, in §5 we present numerical evidence to support the analysis of the one-dimensional case.

3. Results for the IAHR/CEGB test problems

The two-dimensional cell-vertex method method has been tested on two steady convection-diffusion problems which were devised by workers at the CEGB for an IAHR workshop in 1981. The first problem involves the convection of a steep inlet temperature profile around a 180° bend. The second, and more difficult, problem involves the calculation of a developing boundary layer. Computational methods for the first problem are reviewed and discussed in Smith and Hutton [23]. For the second problem, a comparison of some finite element solutions can be found in Morton and Scotney [19].

The domain for both problems is a rectangular region

$$\Omega = \{(x, y) : -1 < x < 1,\ 0 < y < 1\},$$

and the convective velocity field is given analytically by

$$\mathbf{a}(x, y) = (2y(1 - x^2),\ -2x(1 - y^2))^T.$$

3.1. **Problem 1.** The inlet boundary condition along $-1 \le x \le 0$, $y = 0$ is given by

$$U(x, 0) = 1 + \tanh[\alpha(2x + 1)], \tag{3.1}$$

and as in Morton and Scotney [19] we consider only the case $\alpha = 10$. This profile decreases monotonically from $U(0, 0) \approx 2$ down to $U(-1, 0) \approx 0$ with

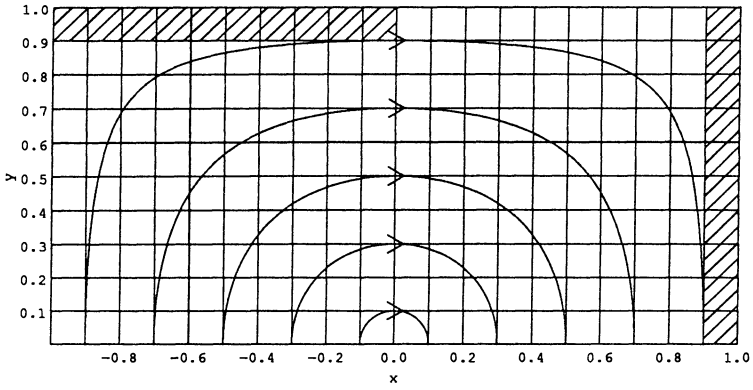
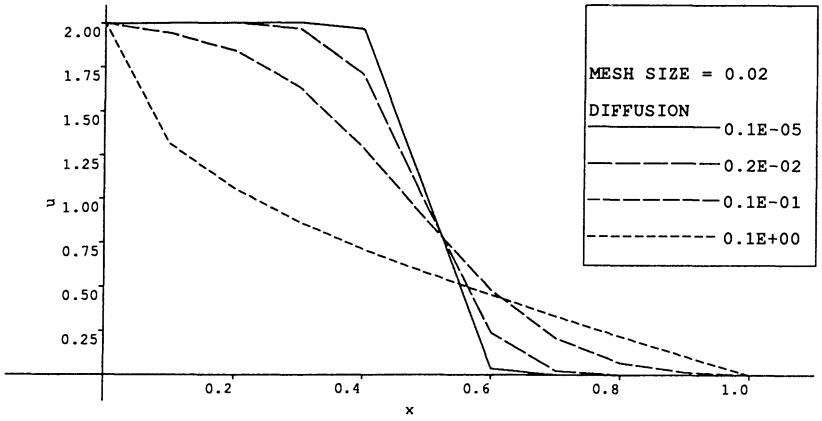


FIGURE 3. Streamlines for the IAHR/CEGB problems: hashing indicates unused cell residuals at outflow Dirichlet boundaries

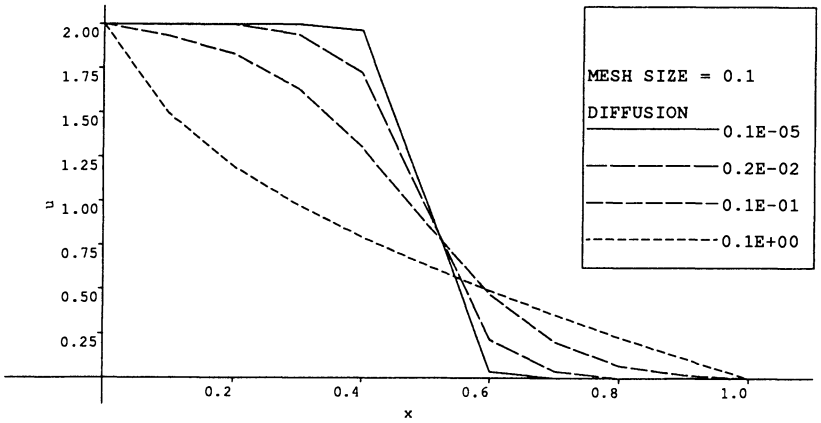
a very steep interior layer centered at $(-1/2, 0)$. The boundary condition on the tangential boundaries, $x = -1$, $y = 1$ and $x = 1$ is given by the compatible Dirichlet condition $U = 1 - \tanh(\alpha)$. Finally, a homogeneous Neumann or natural boundary condition is imposed at the outlet $0 < x \leq 1$, $y = 0$. For comparison with Smith and Hutton [23], the calculations are performed on a uniform grid $\Delta x = \Delta y = 0.1$. The main test of this problem is the calculation of the outflow profile for a wide range of values of ε. Here, we have considered $\varepsilon = 1 \times 10^{-6}$, 2×10^{-3}, 1×10^{-2}, and 1×10^{-1} and with $\max|\mathbf{a}| = 2$ this gives a range of cell Péclet numbers from 2 to 2×10^5. To effectively cover all cases on such a coarse grid is a very severe test for any method: the high curvature of the velocity field near the origin can easily introduce errors due to crosswind diffusion.

For the problem on a $M \times N$ grid there are potentially $(M + 1) \times (N + 1)$ unknowns. Here we have $P = 2(N + 1) + M - 1 + M/2$ Dirichlet boundary conditions, and since $P \geq M + N + 1$, we therefore have sufficient cell equations to determine the unknowns. To obtain an exact match between the unknowns and cell equations, we follow the procedure given in §2. Labelling the (i, j)th cell equation from the bottom left, this results in each equation being associated with just one unknown except for the $M/2$ cell equations $(i = 1, \ldots, M/2;\ j = N)$ and the N cell equations $(i = M;\ j = 1, \ldots, N)$ which are disregarded—see Figure 3; to resolve the ambiguity in the association of the nine nodal unknowns on $x = 0$ one has to appeal to the curvature of the streamlines. To approximate the normal fluxes along the boundary edges, a simplification of the extrapolation procedure described in §2 can be used because of the uniform rectangular mesh.

For comparison, an accurate solution was calculated using a finite difference method on a fine grid where $\Delta x = \Delta y = 0.02$ and the computed output profiles, restricted onto the coarse grid, are shown in Figure 4(a). The results using the cell-vertex method on the standard grid are given in Figure 4(b) and are remarkably good. For the two largest cell Péclet number cases the solutions are sharp and have little or no undershoots or overshoots and are as accurate as one could expect on such a coarse grid. The capability of the cell-vertex finite



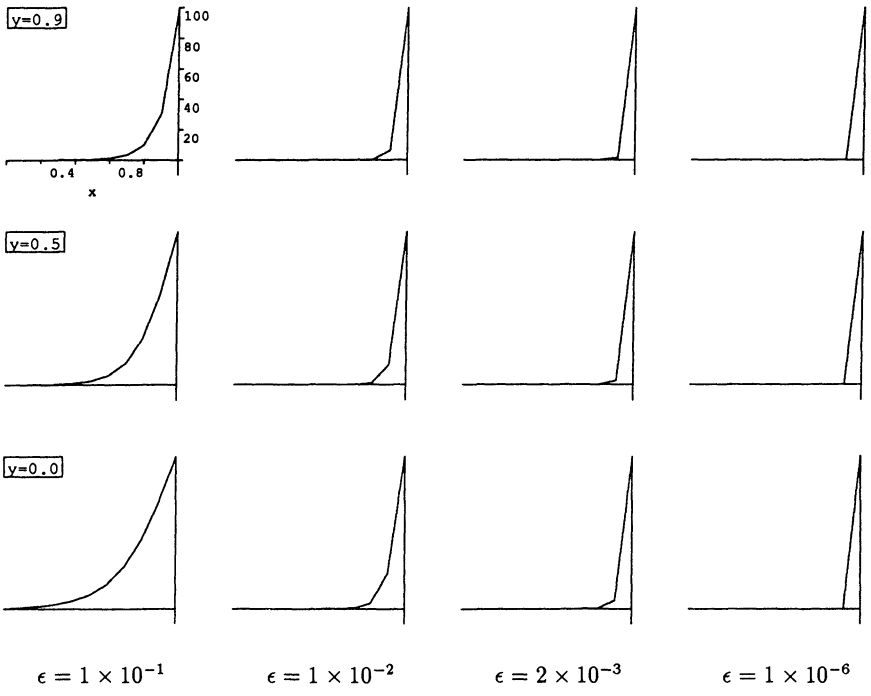
(a)



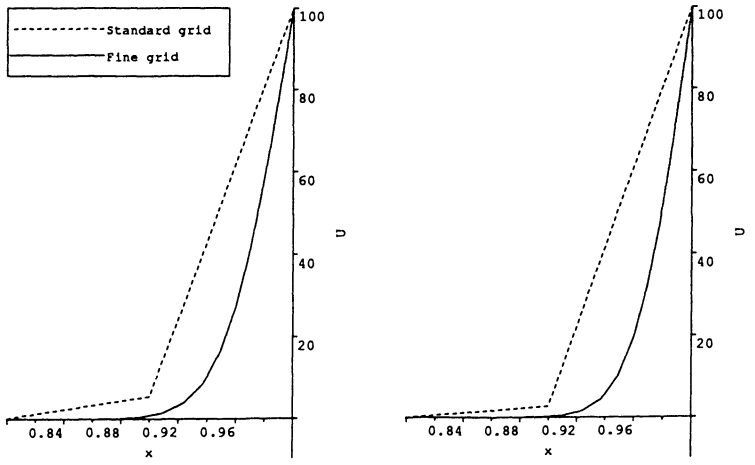
(b)

FIGURE 4. Outlet profiles for first IAHR/CEGB test problem: (a) shows the finite difference fine grid solution restricted to the coarse grid; (b) is the cell-vertex finite volume solution on the standard grid

volume method to cope with viscous dominated problems is demonstrated by the solutions of the remaining two cases, which are at least as accurate as the solutions obtained from other methods. Note that no upwinding parameters, depending on the cell Péclet number, are used in this method—in contrast to the many well-known exponential-fitting methods, Petrov-Galerkin schemes, or upwind difference schemes.



(a)



(b)

FIGURE 5. (a) shows the finite volume boundary layer solutions for different values of ε for the second IAHR/CEGB problem; (b) compares the fine and coarse grid solutions at $y = 0.0$ and $y = 0.5$ with $\varepsilon = 2 \times 10^{-3}$

3.2. **Problem 2.** The second test case considered is a modification of the first problem where the inlet profile is now given by $U(x, 0) = 0$ and on the right-hand tangential boundary, $x = 1$, the Dirichlet condition $U(1, y) = 100$, $0 \leq y \leq 1$, is imposed. The compatible Dirichlet condition $U = 0$ is also set on the remaining two boundaries $x = -1$ and $y = 1$. The main difficulty of this problem lies in the calculation of a developing boundary layer from the corner point $x = 1$, $y = 1$ to the outflow $y = 0$. Figure 5(a) (see p. 195) shows the computed solutions using the finite volume method at the three stations $y = 0.9$, $y = 0.5$ and $y = 0$, for the four values of ε which were considered in the first problem.

The results for the two lower values of the cell Péclet number agree well with streamline-diffusion [6], upwinded [5] and the mixed finite element solutions found in Morton and Scotney [19]. In fact, the finite volume solutions bear a remarkable resemblance to those obtained by the mixed finite element method proposed by Morton and Scotney. For the higher cell Péclet number cases the story is quite different. For instance, when $\varepsilon = 2 \times 10^{-3}$ the thickness of the boundary layer only extends to two cell widths of the standard mesh. On such a coarse grid the finite volume method has performed extremely well and has successfully modelled the thickening of the boundary layer. More detailed pictures of the solution at $y = 0.0$ and $y = 0.5$ are given in Figure 5(b) where the standard grid solutions are compared with a solution on a 40×20 nonuniform grid which has been stretched into the boundary layer. In Morton and Scotney [19] it was found that the streamline-diffusion and upwinded finite element methods, which both aim at giving positive monotone solutions to this problem, completely failed to model the boundary layer for this value of ε. When $\varepsilon = 1 \times 10^{-6}$ the boundary layer is so thin that it cannot be represented on the standard mesh. However, the finite volume method still gives a positive monotone solution which is in stark contrast to the oscillatory behavior of the aforementioned finite element solutions.

The combination of accuracy and monotonicity of the cell-vertex method for both problems makes this method extremely attractive and practicable. In the following sections we attempt to examine the method in more detail by analyzing the one-dimensional version of the scheme.

4. Analysis for a one-dimensional problem

In this section we consider the solution of the following two-point boundary value problem:

$$Lu(x) \equiv \frac{d}{dx}\left[\varepsilon \frac{du}{dx} - au\right] = f, \qquad x \in \Omega = (0, 1),$$

$$u(0) = 0, \qquad u(1) = 1, \tag{4.1}$$

where ε and a are positive constants with $0 < \varepsilon \ll 1$. Although this singular perturbation problem can be solved exactly, we consider it as the simplest model problem for more complicated singular perturbation problems leading to the Navier-Stokes equations at high Reynolds numbers. We shall also consider generalizations of this problem in which a is replaced by a positive function $a(x)$. It turns out that this seemingly innocuous problem is extremely difficult

to analyze to sufficient precision to demonstrate all the attractive properties of the cell-vertex scheme.

4.1. **The finite volume schemes in 1D.** We define a grid, Π_h, as a partition of the unit interval $[0, 1]$, where

$$\Pi_h = \{0 = x_0 < x_1 < \cdots < x_{N-1} < x_N = 1\},$$

which has a variable step size $h_j = x_j - x_{j-1}$ and where we set $h = \max_j h_j$. On this grid we define the usual difference operators

$$\Delta_\pm U_j = \pm(U_{j\pm1} - U_j), \quad D_+ U_j = \frac{\Delta_+ U_j}{h_{j+1}} \quad \text{and} \quad D_- U_j = \frac{\Delta_- U_j}{h_j}. \tag{4.2}$$

The cell-vertex approximation is obtained by integrating (4.1) over the first $N-1$ control volumes to get

$$\varepsilon(U'_j - U'_{j-1}) - a(U_j - U_{j-1}) = \int_{x_{j-1}}^{x_j} f\,dx, \qquad j = 1, \ldots, N-1, \tag{4.3}$$

where U'_j represents an approximation to $u'(x_j)$. If we divide through both sides by h_j we arrive at the discrete equivalent of (4.1); that is $L_h\colon X_h \to Y_h$ is defined by

$$(L_h U)_j \equiv \frac{1}{h_j}[\varepsilon(U'_j - U'_{j-1}) - a_j(U_j - U_{j-1})] = \frac{1}{h_j}\int_{x_{j-1}}^{x_j} f\,dx \equiv (f_h)_j. \tag{4.4}$$

Here, X_h and Y_h are simply $\mathbb{R}^{N-1}$ equipped with suitable norms. Two approximations of the gradient are summarized as follows:

$$U'_j = [\alpha_j D_+ + (1 - \alpha_j) D_-] U_j, \qquad 1 \le j \le N-1, \tag{4.5}$$

where $\alpha_j = h_{j+1}/(h_j + h_{j+1})$ corresponds to Method A of Mackenzie [12] and $\alpha_j = h_j/(h_j + h_{j+1})$ corresponds to Method B: note that in Method A we have $U'_j = (U_{j+1} - U_{j-1})/(h_j + h_{j+1})$. The derivative at $x = 0$ is given by a second-order extrapolation of the gradient from the interior of the domain, that is,

$$U'_0 = 2D_- U_1 - U'_1. \tag{4.6}$$

Note that both of the above schemes involve a four- and a two-point approximation to the second- and first-derivative terms, respectively, and are identical when the grid is uniform. Gushchin and Shchennikov [4] and Lavery [10] have considered this scheme on a uniform mesh, both of them in connection with nonoscillatory solutions of two-point boundary value problems.

4.2. **Approximation of boundary layers and control of spurious solution modes.** The most commonly used first- and second-order schemes reduce to a three-point difference scheme for the 1D model problem (4.1). The cell-vertex scheme, however, uses four points centered on an interval. This results in the scheme having a spurious solution mode for the homogeneous equation, which has to be controlled by the extra boundary condition (4.6) used at the inflow end. So we start our consideration of this scheme with the homogeneous problem on a uniform mesh.

When $f = 0$ the analytical solution of (4.1) is

$$u(x) = \frac{e^{ax/\varepsilon} - 1}{e^{a/\varepsilon} - 1}, \tag{4.7}$$

which increases monotonically and has a steep boundary layer of thickness $O(\varepsilon)$ at $x = 1$.

On a uniform mesh, Methods A and B are identical with $\alpha_j = 1/2$ for all j. The exact solution of the difference equations (4.3) with the boundary condition (4.6) can then be written as

$$U_j = \frac{\mu_1^j - 1 + \frac{\alpha}{\gamma}(\mu_2^j - 1)}{\mu_1^N - 1 + \frac{\alpha}{\gamma}(\mu_2^N - 1)}, \tag{4.8}$$

where

$$\mu_1 = \beta + (1 + \beta^2)^{1/2}, \qquad \mu_2 = \beta - (1 + \beta^2)^{1/2},$$

$$\alpha = \left(\frac{\mu_1 - 1}{1 - \mu_2}\right)^3, \qquad \gamma = \frac{\mu_1}{\mu_2},$$

and $\beta = ah/\varepsilon$ is the cell Péclet number. Note that $\mu_1 > 1$ and is a second-order approximation to e^{β}. However, $-1 < \mu_2 < 0$, and μ_2 is the expected oscillatory solution mode; but this mode decays for increasing j, and an important feature of the scheme is that the resulting solution is monotonic. This is proved in Theorem 4.1.

For the standard three-point central difference scheme,

$$(L_h U)_j = \varepsilon \frac{(U_{j+1} - 2U_j + U_{j-1})}{h^2} - a\frac{(U_{j+1} - U_{j-1})}{2h},$$

the solution of the difference equations is

$$U_j = \frac{\mu_3^j - 1}{(\mu_3^N - 1)}, \qquad \text{where } \mu_3 = \frac{1 + \frac{\beta}{2}}{1 - \frac{\beta}{2}} \tag{4.9}$$

is the $(1, 1)$-Padé approximant of e^{β} which is also second-order accurate. However, unless $\beta \leq 2$, we have $\mu_3 < 0$, and the solution is oscillatory and growing. This condition is very restrictive when ε is small and therefore, although the scheme has no spurious modes, the approximation is poor as $\varepsilon \to 0$ for a fixed h.

For the standard first-order upwind finite difference scheme,

$$(L_h U)_j = \varepsilon \frac{(U_{j+1} - 2U_j + U_{j-1})}{h^2} - a\frac{(U_j - U_{j-1})}{h},$$

we have

$$U_j = \frac{\mu_4^j - 1}{\mu_4^N - 1}, \qquad \text{where } \mu_4 = 1 + \beta \tag{4.10}$$

is the $(0, 1)$-Padé approximant of e^{β}. This scheme has no spurious modes and is monotonic; it is however only first-order accurate.

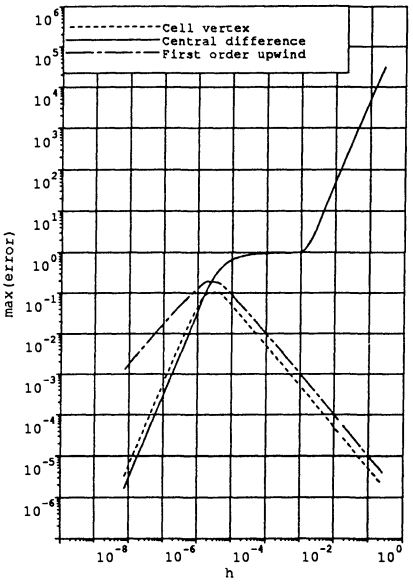


FIGURE 6. $\|E\|_\infty$ for the cell-vertex, central difference and upwind methods with $f = 0$ and $a/\varepsilon = 10^6$

There is a very large literature on three-point schemes, derived from both finite difference and finite element viewpoints, which combine these two approaches in some way—see, for example, O'Riordan and Stynes [22] and Barrett and Morton [1]. However, to compare the cell-vertex scheme with just these two standard finite difference schemes is quite illuminating. In Figure 6 the discrete sup norm of the error $\|E\|_\infty$ in solutions (4.8), (4.9), and (4.10) for $a = 1$ and $\varepsilon = 1.0 \times 10^{-6}$ is plotted against h. It shows that each converges as $h \to 0$, the upwind scheme to first order and both the cell-vertex and central difference methods to second order. However, as the mesh Péclet number increases, the central difference scheme diverges, owing to the growth of its spurious mode, while both the upwind and the cell-vertex schemes tend to the correct solution of the reduced problem. As is well known, only a scheme with exponentially weighted coefficients can give a uniformly accurate error bound (see Doolan et al. [2]), that is, the nodal error is bounded by Ch^p, where C does not depend on h or ε. The cell-vertex scheme, like the fully upwinded scheme, has a peak error where $ah/\varepsilon = O(1)$; but it is consistently better over the whole range and converges as $h \to 0$ with an error little larger than the central difference scheme. We conclude that its spurious mode has no deleterious effects on its performance for this simple problem.

4.3. **Monotonicity of solutions.** We now assume $a(x) \geq a_{\min} > 0$ and introduce $v = u'$, so that (4.1) can be generalized, either to $\varepsilon v' - au' = 0$ or to the conservative form $(\varepsilon v - au)' = 0$. In either case, the homogeneous equation has a monotone solution:

$$\varepsilon u'' = au' \Rightarrow \varepsilon v(x) = \varepsilon v(0) + \int_0^x a(t)v(t)dt \tag{4.11}$$

or

$$\varepsilon u'' = (au)' \Rightarrow \varepsilon v(x) = \varepsilon v(0) + a(x) \int_0^x v(t)dt. \tag{4.12}$$

We deduce such properties of the discrete system in a similar way.

When (4.5) is substituted into (4.3), with a replaced by $a_j = a(x_{j-1/2})$, the full difference equation approximating $\varepsilon u'' = au'$ has the following form for $j = 2, 3, \ldots, N-1$,

$$\begin{aligned} &\varepsilon[\alpha_j D_- U_{j+1} + (1 - \alpha_j - \alpha_{j-1}) D_- U_j - (1 - \alpha_{j-1}) D_- U_{j-1}] \\ &\qquad - a_j h_j D_- U_j = 0. \end{aligned} \tag{4.13}$$

This yields the recurrence relation

$$\begin{aligned} \varepsilon \alpha_j D_- U_{j+1} &= \varepsilon(1 - \alpha_{j-1}) D_- U_{j-1} \\ &\quad + [a_j h_j - \varepsilon(1 - \alpha_j - \alpha_{j-1})] D_- U_j \end{aligned} \tag{4.14}$$

and the boundary condition (4.6) combines with (4.3) and (4.5) to give the starting relation

$$2\varepsilon \alpha_1 D_- U_2 = [a_1 h_1 + 2\varepsilon \alpha_1] D_- U_1. \tag{4.15}$$

The sign of the expression in square brackets in (4.14) is clearly crucial in determining whether the solution is monotone, and we have the following result.

Theorem 4.1. *The approximation given by Method* A, *to the problem* $\varepsilon u'' = au'$ *with* $u(0) = 0$, $u(1) = 1$, *is monotonically increasing if*

$$a_j(h_{j-1} + h_j)(h_j + h_{j+1}) \geq \varepsilon(h_{j-1} - h_{j+1}), \qquad j = 2, \ldots, N-1. \tag{4.16}$$

That for Method B *is monotonically increasing if*

$$a_j(h_{j-1} + h_j)(h_j + h_{j+1}) \geq \varepsilon(h_{j+1} - h_{j-1}), \qquad j = 2, \ldots, N-1. \tag{4.17}$$

Proof. Since (4.15) implies that $D_- U_2$ has the same sign as $D_- U_1$, monotonicity follows if all the coefficients of (4.14) are positive. Moreover, it is clear that α_j for Method A equals $1 - \alpha_j$ for Method B, so that $(1 - \alpha_j - \alpha_{j-1})|_A = -(1 - \alpha_j - \alpha_{j-1})|_B$: calculation of these quantities then gives the quoted conditions. □

The result is not sharp, since it is easy to see that both methods give monotone (indeed, linear) solutions when $a_j \equiv 0$. On a uniform mesh, when the methods are identical, we see that solutions are always monotone for all values of $a(x) > 0$ and ε. It follows that Method B will give a monotone approximation on any decreasing mesh and for all values of a_j and ε, while Method A will not in general do so. Gushchin and Shchennikov [4] were attracted to the four-point scheme on a uniform mesh precisely because of its monotonicity properties. However, they proposed switching the scheme to the standard three-point central difference scheme when $\beta < 1/2$, which is then monotone as shown earlier. However, there is no need to switch from the four-point scheme at this value of β, as the above theorem shows.

When approximating the conservative form of the equation, the last term in (4.13) is replaced by $h_j D_-(a_j U_j)$, for which we write

$$a_j U_j - a_{j-1}U_{j-1} = \frac{a_j + a_{j-1}}{2}(U_j - U_{j-1}) + \frac{U_j + U_{j-1}}{2}(a_j - a_{j-1}). \tag{4.18}$$

Apart from replacing a_j by $\frac{1}{2}(a_j + a_{j-1})$ in (4.14) and(4.15), this also adds an extra term $\frac{1}{2}h_j(U_j + U_{j-1})D_-a_j$ to the equations, and so slightly complicates the conditions guaranteeing monotonicity unless $a(x)$ is nondecreasing.

As a direct consequence of Theorem 4.1 we do, however, have the following result for the nonhomogeneous case.

Theorem 4.2. *If Method* A *or* B *is used to solve the equation* $\varepsilon u'' - au' = f$ *and either condition* (4.16) *holds for Method* A *or condition* (4.17) *holds for method* B, *then the resulting set of discrete equations is uniquely solvable.*

Proof. If we denote the matrix system of nodal equations by L_h, then all we are required to show is that L_h is nonsingular. This is true if and only if the only solution to $L_h U = 0$ is the trivial solution. By setting $U_0 = U_N = 0$ we know from Theorem 4.1 that $U = 0$, and the result is proved. □

Note that the stability of Methods A and B would ensure that L_h^{-1} would exist for a small enough h and that the equations would then be uniquely solvable. Theorem 4.2 therefore complements the yet to be established stability results ensuring unique solvability.

4.4. **Order of consistency on a nonuniform mesh.** For most practical problems it is necessary to use a graded mesh to capture localized flow features like boundary layers. Therefore, we consider the accuracy of the cell-vertex methods on nonuniform meshes. Methods A and B are now different, and we consider first their truncation errors. If we define the restriction operator R_h such that $(R_h u)_j = u(x_j)$, then the truncation error $\tau \equiv L_h(R_h u) - f_h$ has components

$$\begin{aligned}
\tau_{j-\frac{1}{2}} &= \frac{1}{h_j}\left\{\varepsilon[(u'_j - u'_{j-1})] - a[u(x_j) - u(x_{j-1})] - \int_{x_{j-1}}^{x_j} f\,dx\right\} \\
&= \frac{\varepsilon}{h_j}\{[u'_j - u'_{j-1}] - [u'(x_j) - u'(x_{j-1})]\} \\
&= \frac{\varepsilon}{h_j}(T_j - T_{j-1}), \qquad 1 \le j \le N-1,
\end{aligned} \tag{4.19}$$

where

$$T_j = u'_j - u'(x_j) = [\alpha_j D_+ + (1-\alpha_j)D_-]u(x_j) - u'(x_j). \tag{4.20}$$

For Method A,

$$\begin{aligned}
T_j^{(A)} &= \frac{h_{j+1} - h_j}{2}u''(x_j) + \frac{h_{j+1}^3 + h_j^3}{6(h_{j+1} + h_j)}u'''(x_j) \\
&\quad + \frac{(h_{j+1} - h_j)(h_{j+1}^2 + h_j^2)}{24}u^{(\mathrm{iv})}(\xi_j),
\end{aligned} \tag{4.21}$$

for $j = 1, \ldots, N-1$, where $\xi_j \in (x_{j-1}, x_{j+1})$. With the boundary approximation (4.6) we have

$$T_0^{(A)} = \frac{h_1 - h_2}{2} u_0'' + \frac{h_1^2 - 2h_1h_2 - h_2^2}{6} u'''(x_0) + \frac{h_1^3 - 3h_1^2h_2 - 3h_1h_2^2 - h_2^3}{24} u^{(\mathrm{iv})}(\xi_0), \tag{4.22}$$

where $\xi_0 \in (x_0, x_2)$. Substitution of $T_j^{(A)}$ into (4.19) gives

$$\begin{aligned} \tau_{j-\frac{1}{2}}^{(A)} = \frac{\varepsilon}{2h_j} \Bigg[& (h_{j+1} - 2h_j + h_{j-1}) u''(x_{j-\frac{1}{2}}) \\ & + \frac{1}{6}(2h_{j+1}^2 + h_{j+1}h_j - h_jh_{j-1} - 2h_{j-1}^2) u'''(x_{j-\frac{1}{2}}) \\ & + \frac{1}{12}\left(h_{j+1}\left(h_{j+1}^2 + h_{j+1}h_j + \frac{1}{2}h_j^2 \right) \right. \\ & \left. + h_{j-1}\left(h_{j-1}^2 + h_{j-1}h_j + \frac{1}{2}h_j^2 \right) \right) u^{(\mathrm{iv})}(\xi_{j-\frac{1}{2}}) \Bigg] \end{aligned} \tag{4.23}$$

for $2 \le j \le N-1$, where $\xi_{j-\frac{1}{2}} \in (x_{j-2}, x_{j+1})$ and

$$\tau_{\frac{1}{2}}^{(A)} = \varepsilon \left[\frac{h_2 - h_1}{h_1} u''(x_{\frac{1}{2}}) + \frac{h_2(h_1 + 2h_2)}{6h_1} u'''(x_{\frac{1}{2}}) + \frac{h_2(h_1^2 + 2h_1h_2 + 2h_2^2)}{24h_1} u^{(\mathrm{iv})}(\xi_{\frac{1}{2}}) \right], \tag{4.24}$$

with $\xi_{\frac{1}{2}} \in (x_0, x_2)$. If elements of Y_h are measured in the maximum norm, then in general the truncation error is zeroth-order. If successive mesh lengths are in a ratio of $1 + O(h^2)$, then the truncation error is at least first-order, and if the ratio is $1 + O(h^3)$, then it is second-order. The apparent inaccuracy of method A clearly comes from the central difference approximation of $u'(x_j)$, which is generally not centered at x_j.

In an attempt to rectify the above situation, Method B linearly interpolates the two second-order accurate approximations of the gradient at either side of x_j. If we again replace the true solution in (4.3) and use Taylor expansions, we get

$$T_j^{(B)} = \frac{h_jh_{j+1}}{6}\left[u'''(x_j) + \frac{(h_{j+1} - h_j)}{4} u^{(\mathrm{iv})}(\xi_j) \right], \qquad j = 1, \ldots, N-1, \tag{4.25}$$

where $\xi_j \in (x_{j-1}, x_{j+1})$; and with the boundary approximation (4.6),

$$T_0^{(B)} = -\frac{h_1(h_1 + h_2)}{6}\left[u'''(x_0) + \frac{(2h_1 + h_2)}{4} u^{(\mathrm{iv})}(\xi_0) \right], \tag{4.26}$$

where $\xi_0 \in (x_0, x_2)$. Substitution of $T_j^{(B)}$ into (4.19) gives

$$\tau_{j-\frac{1}{2}}^{(B)} = \varepsilon \left[\frac{(h_{j+1} - h_{j-1})}{6} u'''(x_{j-\frac{1}{2}}) + \frac{(h_{j+1}^2 + h_j^2 + h_{j-1}^2 + h_{j+1}h_j + h_j h_{j-1})}{24} u^{(\mathrm{iv})}(\xi_{j-\frac{1}{2}}) \right] \tag{4.27}$$

for $2 \le j \le N-1$, and

$$\tau_{\frac{1}{2}}^{(B)} = \varepsilon \left[\frac{2h_2 + h_1}{6} u'''(x_{\frac{1}{2}}) + \frac{(h_1^2 + 2h_1h_2 + 2h_2^2)}{24} u^{(\mathrm{iv})}(\xi_{\frac{1}{2}}) \right], \tag{4.28}$$

where $\xi_{\frac{1}{2}} \in (x_0, x_2)$. Therefore, Method B has a truncation error which is at least first-order in the maximum norm. If the ratio of successive cell lengths are in a ratio of $1 + O(h)$, then the truncation error (4.27) is second-order. Clearly, this discretization should be more robust than Method A to distortions of the mesh. In §5 we give some numerical experiments which show that this is the case.

However, it should be noted that the order of convergence of methods on nonuniform grids can be underestimated by a straightforward estimation of the local truncation error. It is possible to obtain different orders of consistency by renorming Y_h and measuring the truncation error accordingly. For example, we may choose the following norm to measure elements in Y_h:

$$\|V\|_{Y_h} = \max_{1 \le j \le N-1} \left| \sum_{k=1}^{j} h_k V_k \right|. \tag{4.29}$$

If we measure τ in this norm and use (4.19), we have

$$\|\tau\|_{Y_h} = \max_{1 \le j \le N-1} \left| \sum_{k=1}^{j} h_k \varepsilon \frac{T_k - T_{k-1}}{h_k} \right| = \max_{1 \le j \le N-1} |\varepsilon(T_j - T_0)|. \tag{4.30}$$

Therefore, in this norm we find that Methods A and B are first- and second-order consistent, respectively. The norm (4.29) is called a Spijker norm and has been used by Spijker in his work on initial value problems [24]. Through the use of this norm it appears that Methods A and B could be more accurate than would be naively expected.

Manteuffel and White [13] have analyzed some well-known finite difference approximations to linear two-point boundary value problems and have shown that many common schemes are second-order accurate although they possess first-order truncation errors on nonuniform grids. This enhancement of truncation error has been called *supra*-convergence by Kreiss et al. [9]. Manteuffel and White rewrite second-order boundary value problems as a system of two first-order equations, each of which are approximated by many methods to second-order consistency in the maximum norm on nonuniform meshes. By a careful elimination of variables they then examine the structure of the local truncation error of the original second-order problem, which is split into a number of parts.

Although not explicitly stated in their paper, the authors similarly renorm Y_h and remeasure the truncation error and show that the order of convergence in the maximum norm for many common schemes is an order greater than the order of consistency.

Implicit in the above discussion on accuracy is that the methods are stable in a way that is defined in the next subsection.

4.5. **Stability and uniform boundedness.** Convergence of consistent difference schemes is usually achieved through the idea of stability. Given the discrete problem

$$L_h U = f_h, \tag{4.31}$$

where U and f_h belong to the finite-dimensional vector spaces X_h and Y_h which have been endowed with norms $\|\cdot\|_{X_h}$ and $\|\cdot\|_{Y_h}$, then we have the following definition.

Definition 4.1. The discretization (4.31) is said to be stable if positive constants h_0 and C exist such that for each $h \leq h_0$, $V \in X_h$

$$\|V\|_{X_h} \leq C \|L_h V\|_{Y_h}. \tag{4.32}$$

From the above definition it is easy to establish the following theorem.

Theorem 4.3. *If for given choices of norms in X_h, Y_h the discretization* (4.31) *is consistent and stable, then* (4.31) *possesses, for h small enough, a unique solution U. Furthermore, these solutions converge and if we set $E = R_h u - U$ and $\tau = L_h(R_h u) - f_h$, then for h small enough*

$$\|E\|_{X_h} \leq C \|\tau\|_{Y_h}, \tag{4.33}$$

so that if the scheme is consistent of order p, then it is convergent of order p.

We may ask if consistency is necessary for convergence and for a certain class of difference schemes this is true.

Definition 4.2. The discretization (4.31) is said to be uniformly bounded if positive constants h_0 and M exist such that for each $h \leq h_0$, $V \in X_h$

$$\|L_h V\|_{Y_h} \leq M \|V\|_{X_h}. \tag{4.34}$$

We then have the following theorem.

Theorem 4.4. *If for given choices of norms in X_h, Y_h the discretization* (4.31) *is convergent of order p and uniformly bounded, then it is consistent of order p. For h sufficiently small, the truncation error, τ, and the error, E, are related by*

$$\|\tau\|_{Y_h} \leq M \|E\|_{X_h}. \tag{4.35}$$

A desirable property of a scheme is that it be stable and uniformly bounded. For such schemes we can deduce the following theorem.

Theorem 4.5. *If for given choices of norms in* X_h, Y_h *the discretization* (4.31) *is stable and uniformly bounded, then it is convergent if and only if it is consistent. For* h *small enough,*

$$M^{-1}\|\tau\|_{Y_h} \leq \|E\|_{X_h} \leq C\|\tau\|_{Y_h}. \tag{4.36}$$

This is a convenient result in that if norms can be chosen such that a method is uniformly bounded and stable, then the optimal order of convergence in the chosen norm is the same as the order of consistency. Schemes which are both stable and uniformly bounded have been called *bistable* by Stummel [25]. Unfortunately, it is often difficult to prove that a method is bistable in a standard norm, which has led to the notion of supra-convergence. If we consider $\|L_h\|_\infty$ for the cell-vertex Methods A and B, we find that it is proportional to h^{-2}, and that it is not uniformly bounded in the maximum norm. Therefore, we are unsure if the method is in fact convergent to a higher order of accuracy than the order of consistency in the maximum norm. In §5 we present some numerical examples which indicate that both Methods A and B are indeed supra-convergent.

4.6. **Maximum principles and error bounds.** In this subsection we attempt to prove stability in the maximum norm by showing that the difference operators generated by Methods A and B both satisfy a maximum principle in mimicry of the maximum principle satisfied by the differential operator. In order to derive a maximum principle and thence an error bound, in the discrete sup norm $\|\cdot\|_\infty$, for the inhomogeneous problem

$$\varepsilon u'' - au' = f \quad \text{with } u(0) = 0,\ u(1) = 1, \tag{4.37}$$

we need to place a *lower bound* on the mesh Péclet number.

Theorem 4.6. *Suppose the problem* (4.37) *is approximated by* (4.4) *and* (4.5) *and the boundary approximation* (4.6) *is applied. Then a maximum principle holds if the conditions*

$$a_j(h_{j+1} + h_j) \geq \varepsilon \tag{4.38}$$

for Method A *and both*

$$\varepsilon(h_{j+1}h_{j-1} + h_j^2 + h_jh_{j-1}) + a_jh_{j+1}(h_{j+1} + h_j)(h_j + h_{j-1}) \geq \varepsilon h_{j+1}^2 \tag{4.39}$$

and

$$a_j(h_{j+1} + h_j)h_{j-1} \geq \varepsilon[h_{j+1} + h_j - h_{j-1}] \tag{4.40}$$

for Method B, *are satisfied for* $j = 2, 3, \ldots, N-1$. *Hence, one obtains the error bound*

$$\|R_hu - U\|_\infty \leq \frac{1}{a_{\min}}\|\tau\|_\infty, \tag{4.41}$$

where the truncation error τ *is defined as in* (4.19).

Proof. The maximum principle takes the form

$$(4.42) \qquad (L_h W)_j \geq 0 \quad \forall j \Rightarrow \max_j W_j \leq \max(W_0, W_N)$$

and follows readily if the coefficient of U_{j-1} in (4.4) is nonnegative; this is because the coefficients of U_{j-2} and U_{j+1} are positive and that of U_j is negative for Method A and if (4.39) holds it is also negative for Method B, and the sum of the coefficients equals zero. This coefficient equals

$$\frac{a_j}{h_j} - \varepsilon \left[\frac{1-\alpha_j-\alpha_{j-1}}{h_j^2} + \frac{1-\alpha_{j-1}}{h_j h_{j-1}} \right]$$

and the conditions (4.38) and (4.40) result from substituting for α_j and α_{j-1}. One merely has to check in addition that U_1 cannot be a maximum, by using the inequality $2\varepsilon\alpha_1 D_- U_2 \geq (a_1 h_1 + 2\varepsilon\alpha_1) D_- U_1$ corresponding to (4.15).

By definition, $L_h(R_h u - U) = \tau$, and the error bound is obtained by a standard argument through construction of a nonnegative mesh function W such that $L_h W_j \geq 1$. We take for this purpose $(1-x)/a_{\min}$ to get (4.41). □

Note that on a uniform mesh both methods require the mesh Péclet number to be at least a half: and on a decreasing mesh, the condition for Method B is less stringent.

Again, the situation with the conservative form of the problem is more complicated: if $a(\cdot)$ is nondecreasing, the theorem holds with $a_j \equiv a(x_{j-1/2})$ replaced by $a(x_{j-1})$; but if $a'(\cdot) < 0$, even the differential equation fails to have a maximum principle.

We end this subsection by noting the effect of using alternative boundary conditions to (4.6). The obvious first-order approximation is $U_0' = D_- U_1$, which leads to (4.15) being replaced by $\varepsilon\alpha_1 D_- U_2 = (a_1 h_1 + \varepsilon\alpha_1) D_- U_1$: that is, it has the effect of halving ε in this equation but leaves Theorems 4.1 and 4.6 unchanged. On the other hand, if the boundary condition is replaced by $U_0' = 0$, (4.15) is replaced by $\varepsilon\alpha_1 D_- U_2 = [a_1 h_1 - (1-\alpha_1)\varepsilon] D_- U_1$ and a lower bound on the mesh Péclet number is required in Theorem 4.1 as well as a possible strengthening of the conditions in Theorem 4.6. Finally, the introduction of a "ghost" cell with $U_{-1} = U_1$ at a point $x = -h_1$, followed by application of (4.5) clearly leads back to the condition $U_0' = 0$.

4.7. **The reduced problem.** Although Theorem 4.6 does not allow us to establish convergence, for a fixed ε, it does however allow us to consider the behavior of the cell-vertex schemes for small values of ε. As is well known, the solution of (4.1) converges as $\varepsilon \to 0$, for $0 \leq x < 1$, to the solution $v(x)$ of the reduced problem

$$(4.43) \qquad a v'(x) = f(x), \qquad v(0) = u_0.$$

What is also well known is that many schemes which are accurate for large values of ε do not behave well as $\varepsilon \to 0$, e.g., central differences. We now consider the cell-vertex schemes using a fixed mesh, Π_h, and examine the solutions of (4.37) as $\varepsilon \to 0$. We find that we first need to bound the truncation errors of Methods A and B, independently of ε, which requires some knowledge of the gradients of the solution. This we do using the following lemma.

Lemma 4.1. *The solution u of* (4.37) *with constant a satisfies*

$$|u^{(i)}| \le c\{1 + \varepsilon^{-i}\exp(-a\varepsilon^{-1}(1-x))\}, \qquad i = 0, 1, \tag{4.44}$$

where c does not depend on ε.

Proof. See Kellogg and Tsan [8]. □

We are now in a position to state the following theorem.

Theorem 4.7. *Suppose* (4.37) *with constant a is approximated on a given mesh Π_h as described in Theorem* 4.6. *Then for either Method* A *or* B *there exist positive constants c_1 and c_2, depending on a and Π_h but not on ε, such that for all $\varepsilon \le c_1$*

$$\|R_h u - U\|_\infty \le c_2 \varepsilon. \tag{4.45}$$

Proof. If we take

$$c_1 \le \min_j a(h_{j+1} + h_j)$$

for Method A, and

$$c_1 \le \min_j a(h_{j+1} + h_j)h_{j-1}/(h_{j+1} + h_j - h_{j-1})$$

for Method B, then the conditions of Theorem 4.6 are satisfied, and we have the error bound

$$\|R_h u - U\|_\infty \le \frac{\varepsilon}{a} \max_j \left| \frac{T_j - T_{j-1}}{h_j} \right|, \tag{4.46}$$

where

$$T_j = u'_j - u'(x_j) = [\alpha_j D_+ + (1 - \alpha_j)D_-]u(x_j) - u'(x_j).$$

From Lemma 4.1 we know that $|u(x)| \le c$ and that the gradient

$$\begin{aligned} |u'(x)| &\le c\{1 + \varepsilon^{-1}\exp(-a\varepsilon^{-1}(1-x))\} \\ &\le c\{1 + \varepsilon^{-1}\exp(-a\varepsilon^{-1}h_{\min})\} \le c_2, \end{aligned}$$

where c_2 is independent of ε. This shows that the truncation error is bounded independently of ε, and the result is proved. □

4.8. **Error bounds from an energy analysis.** Even on a uniform mesh the analysis given above does not establish convergence for a fixed ε, although the error bound (4.41) based on the conventional truncation error (4.19) is then quite good for mesh Péclet numbers greater than a half. Generally, though, one needs an alternative analysis, especially for Method A, in order to obtain an error bound that depends only on the error (4.20) with which the gradient is approximated: this we will now undertake for constant a.

We introduce a notation for the errors in the solution and its gradient at each of the nodes,

$$E_j \equiv U_j - u(x_j), \qquad F_j \equiv U'_j - u'(x_j), \qquad j = 0, \ldots, N. \tag{4.47}$$

The finite volume scheme to approximate (4.37) for a constant a, with L_h given by (4.4), is

$$(L_h U)_j = \frac{1}{h_j}\int_{x_{j-1}}^{x_j} f\,dx = \frac{1}{h_j}\{\varepsilon[u'(x_j) - u'(x_{j-1})] - a[u(x_j) - u(x_{j-1})]\};$$

and by using (4.47), we can write this as

$$\varepsilon(F_j - F_{j-1}) = a(E_j - E_{j-1}), \qquad j = 1, \ldots, N-1. \tag{4.48}$$

Since U_N' is so far undefined, we can use the same relation (4.48) for $j = N$ to give

$$F_N = F_{N-1} + \frac{a}{\varepsilon}(E_N - E_{N-1}). \tag{4.49}$$

Noting that $E_N = E_0 = 0$, we multiply (4.48) and (4.49) by $(E_j + E_{j-1})$ and sum over $j = 1, \ldots, N$ to get

$$\sum_1^N (F_j - F_{j-1})(E_j + E_{j-1}) = 0, \tag{4.50}$$

independent of ε and a. Application of a standard summation-by-parts identity yields

$$\sum_1^N (E_j - E_{j-1})(F_j + F_{j-1}) = 0, \tag{4.51}$$

which can also be written as

$$\begin{aligned} F_0(E_1 - E_0) + F_1(E_2 - E_0) + \cdots + F_{N-1}(E_N - E_{N-2}) \\ + F_N(E_N - E_{N-1}) = 0. \end{aligned} \tag{4.52}$$

This is the desired basic identity.

If $f(x)$ is integrated exactly, as we have assumed, the truncation error results solely from the substitution (4.5) for the gradient, as shown by (4.19) and (4.20), and is therefore proportional to ε. It is now clear that we can write

$$T_j = F_j - [\alpha_j D_+ + (1 - \alpha_j)D_-]E_j, \qquad j = 1, \ldots, N-1. \tag{4.53}$$

An error bound for Method A then results from substituting into (4.52) the expression for F_j given by (4.53). The appropriate inner product and norm for this purpose is given by

$$\langle U, V\rangle_h \equiv \frac{1}{2}h_1 U_0 V_0 + \bar{h}_1 U_1 V_1 + \cdots + \bar{h}_{N-1} U_{N-1} V_{N-1} + \frac{1}{2}h_N U_N V_N, \tag{4.54}$$

where $\bar{h}_j = \frac{1}{2}(h_{j+1} + h_j)$ and $\|U\|_h^2 = \langle U, U\rangle_h$. We also introduce the vector of divided differences, suggested by (4.52),

$$DE \equiv \left\{D_+E_0, \frac{E_2 - E_0}{2\bar{h}_1}, \ldots, \frac{E_N - E_{N-2}}{2\bar{h}_{N-1}}, D_-E_N\right\}, \tag{4.55}$$

in terms of which we obtain the following lemma.

Lemma 4.2. *There are constants C_j, independent of the mesh, such that*

$$|E_j| \le C_j \|DE\|_h , \qquad j = 1, \dots, N-1. \tag{4.56}$$

Proof. Suppose first that j is even. Then

$$\begin{aligned} |E_j|^2 &= |(E_2 - E_0) + \cdots + (E_j - E_{j-2})|^2 \\ &\le \left[\frac{(E_2 - E_0)^2}{2\bar{h}_1} + \cdots + \frac{(E_j - E_{j-2})^2}{2\bar{h}_{j-1}}\right][(h_1 + h_2) + \cdots + (h_{j-1} + h_j)] \\ &\le 2x_j \|DE\|_h^2 . \end{aligned}$$

When j is odd, the same bound is obtained from starting the expansion with $(E_1 - E_0)$. Similarly, we can obtain bounds by starting from the right-hand end. Thus (4.56) follows with

$$C_j = [2\min\{x_j, 1 - x_j\}]^{\frac{1}{2}}. \quad \square \tag{4.57}$$

We are now in a position to give an error bound.

Theorem 4.8. *Consider the problem and cell-vertex scheme of Theorem* 4.2 *but with constant a. If the mesh is such that*

$$\frac{a\bar{h}_{N-1}}{\varepsilon} > \frac{1}{8}, \tag{4.58}$$

then there is a constant γ such that, for Method A *and the boundary condition* (4.6),

$$|E_j| \le \frac{C_j}{\gamma} \|T\|_h , \qquad j = 1, \dots, N-1, \tag{4.59}$$

where we set $T_N = T_{N-1}$ and define

$$T_0 = 2D_+u(0) - \frac{u(x_2) - u(0)}{2\bar{h}_1} - u'(0). \tag{4.60}$$

Proof. If (4.52) is scaled by one half, the first two terms can be rearranged as follows, by means of (4.6), (4.53), and (4.60):

$$\begin{aligned} &\frac{1}{2}[F_0(E_1 - E_0) + F_1(E_2 - E_0)] \\ &= \frac{1}{2}(E_1 - E_0)[2D_+U_0 - U_1' - u'(0)] + \frac{1}{2}(E_2 - E_0)\left[\frac{E_2 - E_0}{2\bar{h}_1} + T_1\right] \\ &= \frac{1}{2}h_1 D_+E_0\left[2D_+E_0 - \frac{E_2 - E_0}{2\bar{h}_1} + T_0\right] + \bar{h}_1\left(\frac{E_2 - E_0}{2\bar{h}_1}\right)\left[\frac{E_2 - E_0}{2\bar{h}_1} + T_1\right] \\ &\ge \gamma_0\left[\frac{1}{2}h_1(D_+E_0)^2 + \bar{h}_1\left(\frac{E_2 - E_0}{2\bar{h}_1}\right)^2\right] + \frac{1}{2}h_1(D_+E_0)T_0 + \bar{h}_1\left(\frac{E_2 - E_0}{2\bar{h}_1}\right)T_1 , \end{aligned}$$

where

$$\gamma_0 = \max_{c^2}\left[\min\left(2 - \frac{1}{2}c^2, 1 - \frac{h_1}{4\bar{h}_1 c^2}\right)\right];$$

γ_0 attains its minimum value $\frac{1}{2}(3 - \sqrt{2}) \approx 0.7929$ as $h_2/h_1 \to 0$.

At the other end of the sum we use (4.49) to obtain

$$\begin{aligned}
&\frac{1}{2}[F_{N-1}(E_N - E_{N-2}) + F_N(E_N - E_{N-1})] \\
&\quad = \frac{1}{2}F_{N-1}(2E_N - E_{N-1} - E_{N-2}) + \frac{a}{2\varepsilon}(E_N - E_{N-1})^2 \\
&\quad \geq \gamma_1 \left[\bar{h}_{N-1}\left(\frac{E_N - E_{N-2}}{2\bar{h}_{N-1}}\right)^2 + \frac{1}{2}h_N(D_-E_N)^2\right] \\
&\qquad + \left[\bar{h}_{N-1}\left(\frac{E_N - E_{N-2}}{2\bar{h}_{N-1}}\right) + \frac{1}{2}h_N(D_-E_N)\right] T_{N-1},
\end{aligned}$$

where

$$\gamma_1 = \max_{c^2}\left[\min\left(1 - \frac{1}{4}\frac{h_N}{\bar{h}_{N-1}}c^2, \frac{ah_N}{\varepsilon} - \frac{1}{2c^2}\right)\right]. \tag{4.61}$$

Clearly,

$$\gamma_1 > 0 \qquad \text{if } \frac{4\bar{h}_{N-1}}{h_N} > \frac{\varepsilon}{2ah_N},$$

that is, (4.58) is satisfied. We can then take $\gamma = \min(\gamma_0, \gamma_1)$ to obtain, by substituting (4.53) into (4.52),

$$0 \geq \gamma \|DE\|_h^2 + \langle DE, T\rangle_h. \tag{4.62}$$

Hence, $\|DE\|_h \leq \|T\|/\gamma$, and (4.59) follows from Lemma 4.1. □

The proof of the theorem clearly depends heavily on the fact that the centered differences of E occurring in the identity (4.52) are also used in Method A. This does not happen in Method B, and the inner product of E-differences is not positive definite in that case, even for only mildly nonuniform meshes. This is a familiar situation in numerical analysis: the second-order accurate method does not have the stability properties of the first-order method. For this reason, and because condition (4.58) still prevents the proof of convergence, we finally resort to estimating Green's functions.

4.9. **A discrete Green's function estimate.** The key property of the cell-vertex approximation is the constancy of the total flux error expressed in equation (4.48). Good approximation of the gradient at the inflow boundary should therefore be reflected in a good error bound throughout the domain. We denote this constant by K; and we introduce vectors $\mathbf{E} = \{E_j : j = 0, 1, \ldots, N-1\}$, $\mathbf{F} = \{F_j : j = 0, 1, \ldots, N-1\}$ for the function and gradient errors (4.47), a notation that we shall extend to $\mathbf{U}$ and $\mathbf{T}$. Then (4.48) becomes

$$\varepsilon\mathbf{F} - a\mathbf{E} = K\mathbf{1}. \tag{4.63}$$

Also, the relationship (4.5), approximating the gradient by a divided difference, and the boundary condition (4.6) at the inlet end are written

$$RD_+\mathbf{U} = S\mathbf{U}'. \tag{4.64}$$

For Method B and boundary condition (4.6), these matrices are scaled so that

(4.65)

$$R = \begin{bmatrix} \frac{1}{h_1} & & & & \\ \frac{1}{h_1} & \frac{1}{h_2} & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \frac{1}{h_{N-1}} & \frac{1}{h_N} \end{bmatrix}, \qquad S = \begin{bmatrix} \frac{1}{2h_1} & \frac{1}{2h_1} & & & \\ & \frac{1}{h_1}+\frac{1}{h_2} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \frac{1}{h_{N-1}}+\frac{1}{h_N} \end{bmatrix}.$$

For Method A and alternative boundary conditions, R and S have similar forms, with each entry $O(h^{-1})$.

Introducing the truncation error in the gradient approximation given by (4.53), we can combine (4.63) and (4.64) to give

$$\varepsilon R D_+ \mathbf{E} = S(a\mathbf{E} + K\mathbf{1} - \varepsilon \mathbf{T}). \tag{4.66}$$

In principle, this may then be solved for $\mathbf{E}$ and K by means of the boundary conditions $E_0 = E_N = 0$. As a first step, note that the consistency of either method ensures that (4.64) implies $R\mathbf{1} = S\mathbf{1}$; and it is easily seen that R is invertible. Hence, $R^{-1}S\mathbf{1} = \mathbf{1}$, and we introduce the modified truncation error defined by

$$\hat{\mathbf{T}} \equiv R^{-1}S\mathbf{T}. \tag{4.67}$$

This step inverts the linear interpolation operator which calculates the nodal gradients from the first divided differences and modifies (4.56) to

$$-\varepsilon D_+ \mathbf{E} + aR^{-1}S\mathbf{E} = \varepsilon \hat{\mathbf{T}} - K\mathbf{1}, \tag{4.68}$$

with $E_0 = E_N = 0$.

Suppose now we introduce a discrete Green's function

$$\{H_{ij} : i = 1, 2, \ldots, N-1, j = 1, 2, \ldots, N\},$$

by means of which we can write the solution of (4.68) as

$$E_i = \sum_{j=1}^{N} h_j H_{ij} \hat{T}_{j-1}, \qquad i = 1, 2, \ldots, N-1. \tag{4.69}$$

Then our main task is to estimate $\{H_{ij}\}$, in some appropriate norm. It may be worth noting first what this corresponds to for the differential equation. By integrating the original second-order problem from 0 to x, we have

$$[-\varepsilon u' + au]_0^x = \varepsilon g(x) = -\int_0^x f(t)\,dt, \tag{4.70}$$

so that we seek an $H(x, t)$ such that

$$u(x) = \int_0^1 H(x, t) g(t)\,dt. \tag{4.71}$$

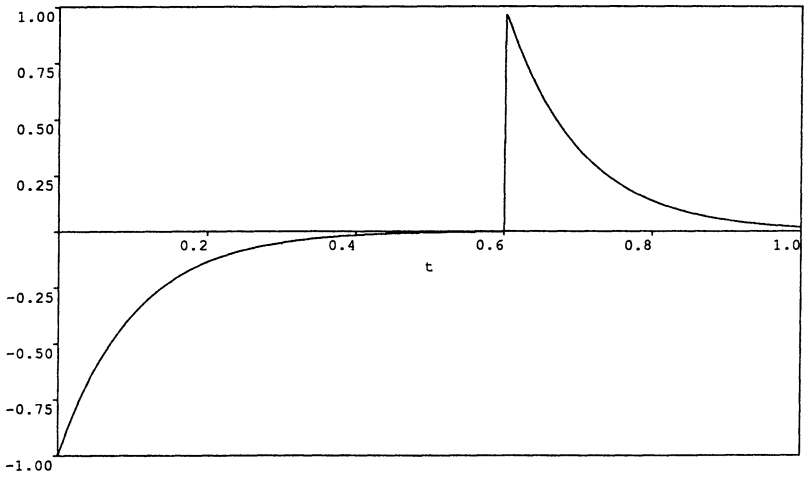


FIGURE 7. Green's function $H(x, t)$ with $x = 0.6$, $a = 1$, and $\varepsilon = 0.1$

It is simple to show that $H(x, t)$ is determined by the properties:

$$\text{(4.72)} \qquad \text{(i)} \int_0^1 H(x, t)\, dt = 0;$$

$$\text{(4.73)} \qquad \text{(ii)} \ \varepsilon \frac{d}{dt} H(x, t) + aH(x, t) = 0 \qquad \text{except at } t = x;$$

$$\text{(4.74)} \qquad \text{(iii)} \ H(x, x_+) - H(x, x_-) = 1.$$

Figure 7 shows a sketch of $H(x, t)$, for typical values of a and ε.

Substituting $\widehat{\mathbf{T}}$ given by (4.67) into (4.68) gives an identity which yields the following defining relations for $\{H_{ij}\}$:

$$\text{(4.75)} \qquad \sum_{j=1}^{N} h_j H_{ij} = 0, \qquad i = 1, 2, \ldots, N-1,$$

$$\text{(4.76)} \qquad H_{ij} - H_{ij-1} + \frac{a}{\varepsilon} \sum_{k=1}^{N} h_k H_{ik} (R^{-1}S)_{k-1, j-1} = \delta_{ij-1}.$$

In particular, it is easily checked that for the purely diffusive case, where $a = 0$, we have

$$\text{(4.77)} \qquad H_{ij} = \begin{cases} -(1 - x_i) & \text{for } j \le i, \\ x_i & \text{for } j \ge i + 1, \end{cases}$$

which corresponds to the exact $H(x, t)$ given by (4.72)–(4.74).

Before embarking on the estimation of H_{ij} when $a \neq 0$, we will first obtain bounds for $\widehat{T}_j$. For Method B, which is our main concern in this this section,

R and S are given by (4.65), and it is readily seen that

$$\widehat{T}_j = h_{j+1}\left[\sum_{k=1}^{j}(-1)^{j-k}\frac{h_k + h_{k+1}}{h_k h_{k+1}}T_k + \frac{(-1)^j}{2h_1}(T_0 + T_1)\right]. \tag{4.78}$$

We now suppose that $u \in C^4(0, 1)$ so that from (4.25)

$$T_j = \frac{h_{j+1}h_j}{6}\left[u'''(x_j) + \frac{1}{4}(h_{j+1} - h_j)\, u^{(\mathrm{iv})}(\xi_j)\right], \qquad j = 1, 2, \cdots, N-1, \tag{4.79}$$

where $\xi_j \in (x_{j-1}, x_{j+1})$; and with boundary condition (4.6), we have from (4.26)

$$T_0 = -\frac{h_1(h_1 + h_2)}{6}\left[u'''(x_0) + \frac{1}{4}(2h_1 + h_2)u^{(\mathrm{iv})}(\xi_0)\right],$$

where $\xi_0 \in (x_0, x_2)$. Hence, rearranging the sum in (4.78) gives, when j is even,

$$\widehat{T}_j = \frac{h_{j+1}}{6}\left[\sum_{k=2}^{j}(-1)^{j-k}\left\{h_k(u'''(x_k) - u'''(x_{k+1})) + \frac{1}{4}(h_{k+1}^2 - h_k^2)u^{(\mathrm{iv})}(\xi_k)\right\}\right.$$
$$\left. + h_{j+1}u'''(x_j) - h_1 u'''(x_1) - \frac{1}{4}(h_2^2 - h_1^2)u^{(\mathrm{iv})}(\xi_1) + \frac{(T_0 + T_1)}{2h_1}\right]; \tag{4.80}$$

and when j is odd,

$$\widehat{T}_j = \frac{h_{j+1}}{6}\left[\sum_{k=3}^{j}(-1)^{j-k}\left\{h_k(u'''(x_k) - u'''(x_{k+1})) + \frac{1}{4}(h_{k+1}^2 - h_k^2)u^{(\mathrm{iv})}(\xi_k)\right\}\right.$$
$$+ h_{j+1}u'''(x_j) - (h_2 u'''(x_2) - (h_1 + h_2)u'''(x_1)$$
$$+ \frac{1}{4}((h_3^2 - h_2^2)u^{(\mathrm{iv})}(\xi_2) - (h_2^2 - h_1^2)u^{(\mathrm{iv})}(\xi_1)))$$
$$\left. + \frac{(T_0 + T_1)}{2h_1}\right]. \tag{4.81}$$

In both cases, since $u \in C^4(0, 1)$,

$$u'''(x_k) - u'''(x_{k-1}) = h_k u^{(\mathrm{iv})}(\eta_k) \tag{4.82}$$

for some $\eta_k \in (x_{k-1}, x_k)$, and therefore $\widehat{T}_j = O(h^2)$ on any mesh.

Calculation of the discrete Green's function $\{H_{ij}\}$ starts from the right with an arbitrary value of H_{iN}. Substitution for $(R^{-1}S)$ into (4.76) gives for Method B the successive relations

$$H_{iN} - H_{iN-1} + \frac{a}{\varepsilon}\frac{h_N + h_{N-1}}{h_N h_{N-1}}h_N^2 H_{iN} = 0,$$

$$H_{iN-1} - H_{iN-2} + \frac{a}{\varepsilon}\frac{h_{N-1}+h_{N-2}}{h_{N-1}h_{N-2}}(h_{N-1}^2 H_{iN-1} - h_N^2 H_{iN}) = 0,$$

and generally,

(4.83)

$$H_{ij} - H_{ij-1} + \frac{a}{\varepsilon}\frac{h_j + h_{j-1}}{h_j h_{j-1}}(h_j^2 H_{ij} - h_{j+1}^2 H_{ij+1} + \cdots + (-1)^{N-j} h_N^2 H_{iN}) = \delta_{ij-1}.$$

Then H_{iN} is determined from application of (4.75). The whole procedure is most easily analyzed on a nonincreasing mesh, where we have the following result.

Theorem 4.9. *Suppose* $u \in C^4(0, 1)$ *is the solution of the problem* (4.37) *with constant* a, *and it is approximated by the cell-vertex scheme as in Theorem* 4.2. *If Method* B *is used and the mesh is nonincreasing,* $h_{j+1} \le h_j$, *then the nodal error satisfies*

$$|U_i - u(x_i)| \le 2\|\hat{\mathbf{T}}\|_\infty \tag{4.84}$$

with $\hat{T}_j$ *given by* (4.80), *when* j *is even, and by* (4.81) *when* j *is odd.*

Proof. We can suppose that $H_{iN} > 0$. Then we deduce that

$$0 < H_{iN} < H_{iN-1} < \cdots < H_{ii+1} \tag{4.85}$$

by induction: for, using the induction hypothesis and $h_{j+1} \le h_j$, we obtain

$$\begin{aligned} H_{ij-1} &\ge H_{ij} + \frac{a}{\varepsilon}\frac{h_j + h_{j-1}}{h_j h_{j-1}}[h_j^2(H_{ij} - H_{ij+1}) + h_{j+2}^2(H_{ij+2} - H_{ij+3}) + \cdots] \\ &> H_{ij}. \end{aligned}$$

The δ_{ij-1} in (4.83) initiates a further monotone sequence at H_{ii}, but the combined sequence is no longer monotone and may oscillate. However, we are able to bound its behavior by use of a recurrence obtained by combining $(h_j h_{j-1})/(h_j + h_{j-1})$ times (4.83) with its successor to give

$$\frac{h_{j+1}h_j}{h_{j+1}+h_j}(H_{ij+1} - H_{ij}) + \frac{h_j h_{j-1}}{h_j + h_{j-1}}(H_{ij} - H_{ij-1}) + \frac{a}{\varepsilon}h_j^2 H_{ij} = 0,$$

i.e.,

$$\frac{h_{j-1}}{h_j + h_{j-1}}H_{ij-1} = \left[\frac{(h_{j-1} - h_{j+1})h_j}{(h_{j+1}+h_j)(h_j + h_{j-1})} + \frac{a}{\varepsilon}h_j\right]H_{ij} + \frac{h_{j+1}}{h_{j+1}+h_j}H_{ij+1} \tag{4.86}$$

$$\text{for } j = i-1, i-2, \ldots, 2.$$

The coefficients here are all positive: so if $H_{ii} \le 0$ and $H_{ii-1} \le 0$, then the sequence remains nonpositive; and to have both $H_{ii} \ge 0$ and $H_{ii-1} \ge 0$ will not yield the negative values required to satisfy (4.75). It is also clear from (4.83) that

$$\begin{aligned} H_{ii-1} &\le \left[1 + \frac{a}{\varepsilon}\frac{h_i + h_{i-1}}{h_{i-1}}h_i\right]H_{ii} - \frac{a}{\varepsilon}\frac{h_i + h_{i-1}}{h_i h_{i-1}}\left[h_{i+1}^2(H_{ii+1} - H_{ii+2}) + \cdots\right] \\ &\le \left[1 + \frac{a}{\varepsilon}\frac{h_i + h_{i-1}}{h_{i-1}}h_i\right]H_{ii}. \end{aligned}$$

Hence, if $H_{ii} \leq 0$, then $H_{ii-1} \leq 0$, and all the positive terms in the sum (4.75) result from $j \geq i+1$. It is readily shown that $H_{ii+1} < 1$ so that this part of the sum is bounded by $(1 - x_i)$. If, on the other hand, $H_{ii} > 0$, then it is necessary that $H_{ii-1} < 0$ and the sequence may oscillate before two terms of the same sign cause that sign to hold thereafter. Moreover, it is clear from (4.86) that any oscillation is damped and confined between H_{ii-1} and H_{ii}; for, if we suppose $H_{ij} \leq 0$, then

$$H_{ij-1} \leq \frac{h_{j+1}(h_j + h_{j-1})}{h_{j-1}(h_{j+1} + h_j)} H_{ij+1} \leq H_{ij+1}. \tag{4.87}$$

If oscillation occurs and then the sequence goes negative, the contribution from $j \leq i$ to the positive terms in the sum (4.75) is bounded by $x_i H_{ii}$; but putting $H_{ii-1} \leq 0$ in (4.83) with $j = i$ readily shows that $H_{ii} \leq 1$, so that the sum of all positive terms is seen to be less than unity. If the sequence should go positive after oscillation, we bound the negative terms by $x_i H_{ii-1}$, and by a similar argument we find that $H_{ii-1} \geq -1$.

Thus, by bounding either the positive or negative terms, we establish that $\sum h_j |H_{ij}| \leq 2$, and the desired result follows from (4.69). □

5. Numerical experiments in 1D

We conclude by considering some experiments to validate some of the theoretical issues raised in the previous sections. We also compare the performance of the cell-vertex methods with other finite volume methods.

5.1. **Example 1.** The monotonicity and accuracy of both Methods A and B is demonstrated by applying them to the solution of (4.1) with $f = 0$ where, for simplicity, we take $a = 1$. A nonuniform grid is generated using a smooth mesh function

$$g(s) = 1 - (1 - s)^{\sigma}$$

such that

$$\begin{aligned} x_j &= g(s_j), \qquad j = 0, \ldots, N, \\ s_j &= j/N, \qquad j = 0, \ldots, N, \end{aligned} \tag{5.1}$$

where σ is a positive integer chosen to cluster the mesh points in the boundary layer. Since $g(s) \in C^2(0, 1)$, this ensures that $h_{j+1} - h_j = O(h^2)$. Figure 8 (see next page) shows the computed solutions using Method A for the three cases $\varepsilon = 1 \times 10^{-1}$, 1×10^{-2} and 1×10^{-3} on meshes with $\sigma = 1, 2$ and 3, respectively, and in all cases $N = 64$. These results show very good agreement with the exact solutions and are strictly monotonic despite (4.16) being violated. The results for Method B are identical to those of Method A at plotting accuracy, but in fact are more accurate as can be seen from Table 1, where the l_∞ error of both schemes for the above three cases are given. For comparison, this problem was also solved with two cell-centered finite volume methods (more correctly called vertex-centered methods nowadays). The first cell-centered method gives

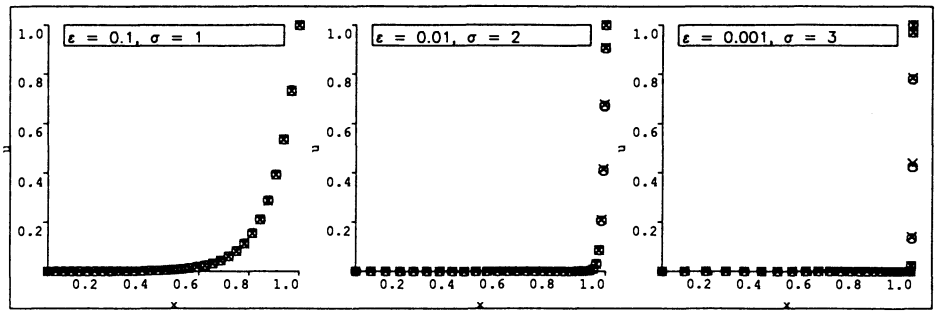


FIGURE 8. Solutions for a 1D singular perturbation problem: o = cell vertex, × = exact

a difference approximation

$$(L_h U)_j = \frac{2}{h_j + h_{j+1}} \left\{ \varepsilon \left(\frac{U_{j+1} - U_j}{h_{j+1}} - \frac{U_j - U_{j-1}}{h_j} \right) - a \left(\frac{U_{j+1} + U_j}{2} - \frac{U_j + U_{j-1}}{2} \right) \right\}, \tag{5.2}$$

which is second-order accurate on uniform meshes. The second cell-centered method tested was the first-order upwind method

$$(L_h U)_j = \frac{2}{h_j + h_{j+1}} \left\{ \varepsilon \left(\frac{U_{j+1} - U_j}{h_{j+1}} - \frac{U_j - U_{j-1}}{h_j} \right) - a(U_j - U_{j-1}) \right\}. \tag{5.3}$$

The results for both methods are also shown in Table 1. As expected, the first-order accurate method is the least accurate of all the methods, owing to numerical diffusion of the first-order approximation of the convective term. The accurate results obtained with the second-order cell-centered method are somewhat surprising, although in each case there are several mesh points in the boundary layer. Moreover, it should be remembered that on a uniform mesh the three-point central difference approximation of the second derivative has a leading coefficient of the truncation error which is 2/5 that of the four-point cell-vertex method. The above results show that on these smoothly varying meshes this increase in accuracy is partially maintained even though the convective terms are less well approximated.

TABLE 1. Calculated l_∞ errors for 1D singular perturbation problem for the cell-vertex Methods A and B, a second-order centered finite volume method and a first-order upwind cell-centered method

Scheme	$\varepsilon = 0.1$, $\sigma = 1$	$\varepsilon = 0.01$, $\sigma = 2$	$\varepsilon = 0.001$, $\sigma = 3$
A	1.48E-3	6.63E-3	1.47E-2
B	1.48E-3	4.35E-3	1.04E-2
CC	7.48E-4	2.24E-3	5.41E-3
UP	2.70E-2	3.90E-2	5.37E-2

5.2. **Example 2.** In order to investigate the order of convergence of Methods A and B on general meshes, they were both applied to the solution of (4.1), with

$$a = 1, \qquad \varepsilon = 0.1 \qquad \text{and} \qquad f = \frac{2e^{-x/\varepsilon}}{\varepsilon(e^{-1/\varepsilon} - 1)}.$$

This particular forcing function was chosen in order that the analytical solution

$$u(x) = \frac{e^{-x/\varepsilon} - 1}{e^{-1/\varepsilon} - 1}$$

has a boundary layer at $x = 0$ and so that the accuracy of the boundary approximation (4.6) could be properly tested. A sequence of 600 random meshes ($N - 1$ points placed in (0,1) at random) were generated with N ranging from 100 to 600. The grids were generated with the following algorithm:

```
fix δ > 0
x_0 = a
i = 0
do while x_i < b
        i = i + 1
        h_i = random number in (0, δ) (uniform distribution)
        x_i = x_{i-1} + h_i
end do
J = i
x_J = b
h_J = x_J − x_{J-1}
```

For all of the results shown below the meshes were calculated with $h_{\max}/h_{\min}$ being bounded by 10^7 with the aim of generating very distorted meshes. Figure 9(a) and (b) (see next page) show the error plots for Method A. The scatter diagrams have been fitted by a least squares regression line to give some indication of the slope of the graphs. Figure 9(a) shows the maximum error and has a slope of 1.47. Therefore, the method appears to be at least first-order accurate, despite the method being inconsistent in general. Calculation of the local truncation error shows that $\|\tau\|_\infty \to O(h_{\max}/h_{\min})$ as $h_{\max} \to 0$. As mentioned earlier, for these calculations $h_{\max}/h_{\min} \leq 10^7$, and so the truncation error was very large indeed. Figure 9(b) shows the calculated error in the gradient approximation, $\|F\|_\infty$ for Method A, the slope of the graph being 1.44. As predicted from the earlier analysis, these results indicate that the order of convergence of the method is determined by the accuracy of the calculation of the gradient, which for Method A we have shown to be first-order accurate on arbitrary meshes. Figure 9(c) and (d) show the error plots for Method B. Figure 9(c) shows plots of both the nodal error, which is in the upper portion of the graph, and the local truncation error. The calculated slope of both lines is 2.05 for the nodal error and 0.83 for the truncation error. Therefore, we experimentally observe the supra-convergence property of the method. In addition, Figure 9(d) shows the error in the calculated gradient, the line having a slope of 2.0. As

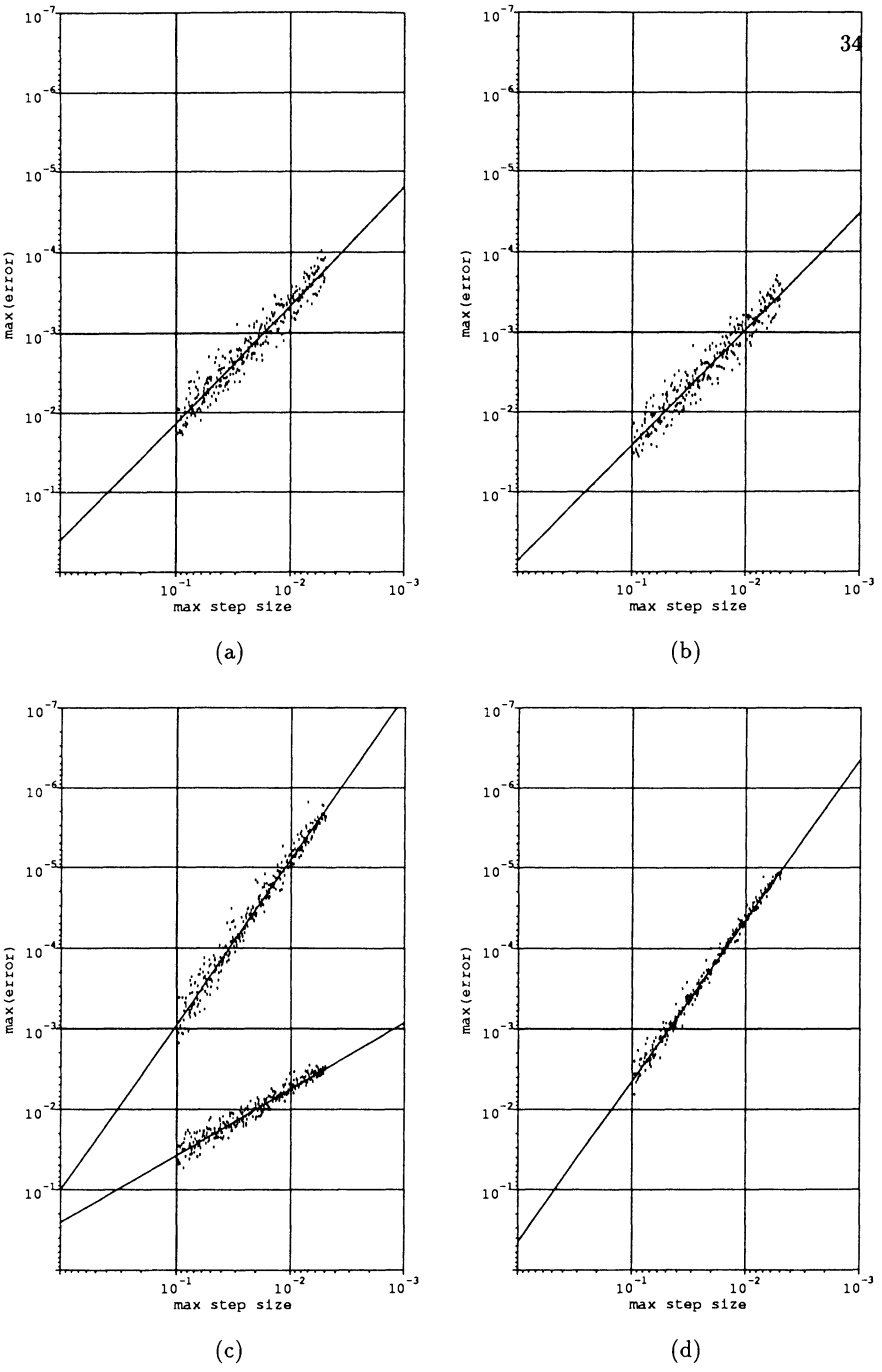


FIGURE 9. Error plots of Methods A and B on random meshes: (a) and (b) show $\|E\|_\infty$ and $\|F\|_\infty$ respectively for Method A; for Method B, (c) shows $\|E\|_\infty$ (upper) and $\|\tau\|_\infty$, and finally (d) shows $\|F\|_\infty$

with Method A, the order of convergence of the gradient approximation is equal to that of the nodal solution.

Finally, a first-order boundary condition $U_0' = U_1'$ was tested on the above problem with Method B; the resulting slope of the regression line through the scatter diagram had a gradient of 0.47, confirming the need to use a second-order boundary approximation in order to retain second-order global accuracy.

6. Conclusions

The final Theorem 4.9 gives us the best and most comprehensive results that we have for Method B. By the same techniques it is possible to prove similar theorems for more general meshes and also for Method A. For completely general meshes one can prove (see [3]) the stability of Methods A and B using compactness arguments developed by Grigorieff. In some cases, error bounds can be established in terms of the local truncation error T rather than $\hat{T}$, as was done in [20].

However, many of the attractive features of these cell-vertex methods are only revealed by the maximum principles and monotonicity results given in §4.4. We also believe that the energy method used in §4.5 is capable of generalization and wider applicability. As was pointed out in the introduction, our purpose in displaying these various techniques of analysis has been to explore those which will be most applicable in 2 or 3 dimensions, where practical interest is focussed and where, as we have seen with the IAHR/CEGB model problem, these methods give such good results without the need of carefully tuned parameters.

Bibliography

1. J. W. Barrett and K. W. Morton, *Approximate symmetrization and Petrov-Galerkin methods for diffusion-convection problems*, Comput. Methods Appl. Mech. Engrg. **45** (1984), 97–122.
2. E. P. Doolan, J. J. H. Miller, and W. H. A. Schilders, *Uniform numerical methods for problems with initial and boundary layers*, Boole Press, Dublin, 1980.
3. B. García-Archilla and J. A. Mackenzie, *Analysis of a supraconvergent cell vertex finite volume method for one-dimensional convection-diffusion problems*, Technical Report NA91/13, Oxford University Computing Laboratory, 11 Keble Road, Oxford, OX1 3QD, 1991. (Submitted for publication)
4. V. A. Gushchin and V. V. Shchennikov, *A monotonic difference scheme of second order accuracy*, U.S.S.R. Comput. Math. and Math. Phys. **14** (1974), 252–256.
5. J. C. Heinrich, P. S. Huyakorn, A. R. Mitchell, and O. C. Zienkiewicz, *An upwind finite element scheme for two-dimensional convective transport equations*, Internat. J. Numer. Methods Engrg. **11** (1977), 131–143.
6. T. J. R. Hughes and A. N. Brooks, *A multi-dimensional upwind scheme with no crosswind diffusion*, Finite Element Methods for Convection Dominated Flows (T. J. R. Hughes, ed.), ASME, New York, 1985, pp. 19–35.
7. A. Jameson, W. Schmidt, and E. Turkel, *Numerical solutions of the Euler equations by finite volume methods using Runge-Kutta time stepping*, AIAA Paper No. 81-1259, 1981.
8. R. B. Kellogg and A. Tsan, *Analysis of some difference approximations for a singular perturbation problem without turning points*, Math. Comp. **32** (1978), 1025–1039.
9. H. O. Kreiss, T. A. Manteuffel, B. Swartz, B. Wendroff, and A. B. White, *Supra-convergent schemes on irregular grids*, Math. Comp. **47** (1986), 537–554.
10. J. E. Lavery, *Nonoscillatory solution of the steady inviscid Burgers' equation by mathematical programming*, J. Comput. Phys. **79** (1988), 436–448.

11. R. W. MacCormack and A. J. Paullay, *Computational efficiency achieved by time splitting of finite difference operators*, AIAA Paper No. 72-154, 1972.
12. J. A. Mackenzie, *The cell vertex method for viscous transport problems*, Technical Report NA89/4, Oxford University Computing Laboratory, 11 Keble Road, Oxford, OX1 3QD, 1989.
13. T. A. Manteuffel and A. B. White, Jr., *The numerical solution of second-order boundary value problems on nonuniform meshes*, Math. Comp. **47** (1986), 511–535.
14. P. W. McDonald, *The computation of transonic flow through two-dimensional gas turbine cascades*, Paper 71-GT-89, ASME, New York, 1971.
15. J. Moore and J. Moore, *Calculation of horseshoe vortex flow without numerical mixing*, Technical Report JM/83–11, Virginia Polytechnic Inst. and State University, Blacksburg, Virginia 24061, 1983. Prepared for presentation at the 1984 Gas Turbine Conference, Amsterdam.
16. K. W. Morton, *Generalised Galerkin methods for hyperbolic problems*, Comput. Methods Appl. Mech. Engrg. **52** (1985), 847–871. Presented at FENOMECH '84, Part III, IV, Stuttgart, 1984.
17. ———, *Finite volume methods and their analysis*, The Mathematics of Finite Elements and Applications, VII MAFELAP 1990 (J. R. Whiteman, ed.), Academic Press, London and New York, 1991, pp. 189–214.
18. K. W. Morton and M. F. Paisley, *A finite volume scheme with shock fitting for the steady Euler equations*, J. Comput. Phys. **80** (1989), 168–203.
19. K. W. Morton and B. W. Scotney, *Petrov-Galerkin methods and diffusion-convection problems in 2D*, The Mathematics of Finite Elements and Applications, V MAFELAP 1984 (J. R. Whiteman, ed.), Academic Press, London and New York, 1985, pp. 343–366.
20. K. W. Morton and E. Süli, *Finite volume methods and their analysis*, Technical Report NA90/14, Oxford University Computing Laboratory, 11 Keble Road, Oxford, OX1 3QD, 1989.
21. R. H. Ni, *A multiple grid method for solving the Euler equations*, AIAA J. **20** (1982), 1565–1571.
22. E. O'Riordan and M. Stynes, *An analysis of a superconvergence result for a singularly perturbed boundary value problem*, Math. Comp. **46** (1986), 81–92.
23. R. M. Smith and A. G. Hutton, *The numerical treatment of convection—a performance/comparison of current methods*, Numer. Heat Transfer **5** (1982), 439–461.
24. M. N. Spijker, *Stability and convergence of finite-difference methods*, PhD thesis, Leiden, Rijksuniversiteit, 1968.
25. F. Stummel, *Biconvergence, bistability and consistency of one-step methods for the numerical solution of initial value problems in ordinary differential equations*, Topics in Numerical Analysis II (J. J. H. Miller, ed.), Academic Press, London, 1975, pp. 197–211.

OXFORD UNIVERSITY COMPUTING LABORATORY, NUMERICAL ANALYSIS GROUP, 11 KEBLE ROAD, OXFORD OX1 3QD, ENGLAND
E-mail address: john.mackenzie@comlab.ox.ac.uk
E-mail address: bill.morton@comlab.ox.ac.uk