Journal of Computational Physics **165,** 45–68 (2000)
doi:10.1006/jcph.2000.6595, available online at http://www.idealibrary.com on IDEAL®

Analysis of the Cell-Centred Finite Volume Method for the Diffusion Equation

W. P. Jones* and K. R. Menzies†

**Department of Mechanical Engineering, Imperial College of Science, Technology and Medicine, Exhibition Road, London SW7 2BX, England; and †Rolls Royce plc, P.O. Box 3, Filton, Bristol BS34 7QE, England*
E-mail: w.jones@ic.ac.uk, kevin.menzies@rolls-royce.com

Received March 30, 1999; revised February 10, 2000; published online November 3, 2000

This paper presents a rigorous theoretical analysis of the cell-centred finite volume method for Poisson's equation. We review the traditional Taylor series expansion technique which suggests that the cell-centred method is inconsistent on nonuniform grids, which is not confirmed by numerical experiments. We then present an analysis of the method which confirms that the solution error does indeed reduce as the cell size is reduced. This is supported by numerical calculations.

1. INTRODUCTION

In the numerical solution of fluid flow problems the cell-centred finite volume approach is one of the most commonly used techniques; the user defines the locations of the faces of the computational cells, and the nodes (locations where the variables are calculated) are positioned in the centre of the cells. This is in contrast to the node-based finite volume method where the node locations are defined by the user and the cell face locations are derived as being midway between nodes. The cell-centred finite volume method has the disadvantage that, on a nonuniform grid, conventional Taylor series truncation error analysis shows the method to be inconsistent—i.e., the truncation error contains terms that do not vanish as the grid spacing is progressively reduced. Numerical experiments on model problems indicate that although the truncation error does indeed behave in this way, the solution error (the difference between the numerical and exact solutions) does not. In fact the solution error does decrease as the grid is refined; a Taylor series analysis and numerical examples are given later.

Recently, Süli [6] presented an alternative theoretical analysis of a numerical method for Poisson's equation, using concepts from finite element theory. However, despite the method being called a cell-centred finite volume scheme it is, in fact, a node-based scheme and no numerical results were presented to display the validity of the analysis. This paper extends

0021-9991/00 $35.00

Süli's analysis to a truly cell-centred finite volume scheme and also presents numerical examples to compare to the theoretical predictions. This allows us to show that the analysis, extended to this cell-centred environment, delivers results that can be obtained in practice, which has not previously been demonstrated. The notation used in Süli's paper is largely employed here for consistency.

The model problem to be considered here is the solution of Poisson's equation on the unit square with fixed (Dirichlet) boundary conditions. Let $\Omega = (0, 1) \times (0, 1)$ be our domain of interest and let $\partial\Omega$ denote the boundary of Ω. Then the problem to be considered is to determine u such that

$$-\nabla^2 u = f \qquad \text{in } \Omega \tag{1}$$

$$u = 0 \qquad \text{on } \partial\Omega. \tag{2}$$

The layout of the paper is as follows: We begin by reviewing the Taylor series truncation error approach to both the node-based and cell-centred finite volume schemes and showing how this analysis suggests that the cell-centred method is inconsistent; however, numerical experiments give a similar convergence of the solution error for each of the methods. We then provide a theoretical analysis of the cell-centred method using techniques from finite element theory. We show how the finite volume scheme is equivalent to a finite element scheme and a finite difference scheme. We then demonstrate the stability of the finite difference scheme (which indicates stability of the finite volume scheme) and convergence of the finite element method (implying convergence of the finite volume scheme). These results together imply consistency. The bounds on the error determined by the analysis are then compared with numerical results. The necessary mathematical background for this analysis is found in texts such as [1], [2], and [7].

2. NOTATION

The grid arrangement is illustrated in Fig. 1, which shows the construction of the control volumes. This is common for both grid systems. In the case of the node-based grid the nodes are located first and the control volume faces are derived from these: if a node location is given as (x_i, y_j) then the corresponding cell faces are located at

$$\begin{aligned} x_{i-1/2} &= x_i - \frac{h_i}{2} \\ x_{i+1/2} &= x_i + \frac{h_{i+1}}{2} \\ y_{j-1/2} &= y_j - \frac{k_j}{2} \\ y_{j+1/2} &= y_j + \frac{k_{j+1}}{2}, \end{aligned}$$

where h_i, k_j are as shown in Fig. 1. In the case of the cell-centred grid we locate the cell faces first and derive the locations of the nodes; so for a node located at (x_i, y_j) we have

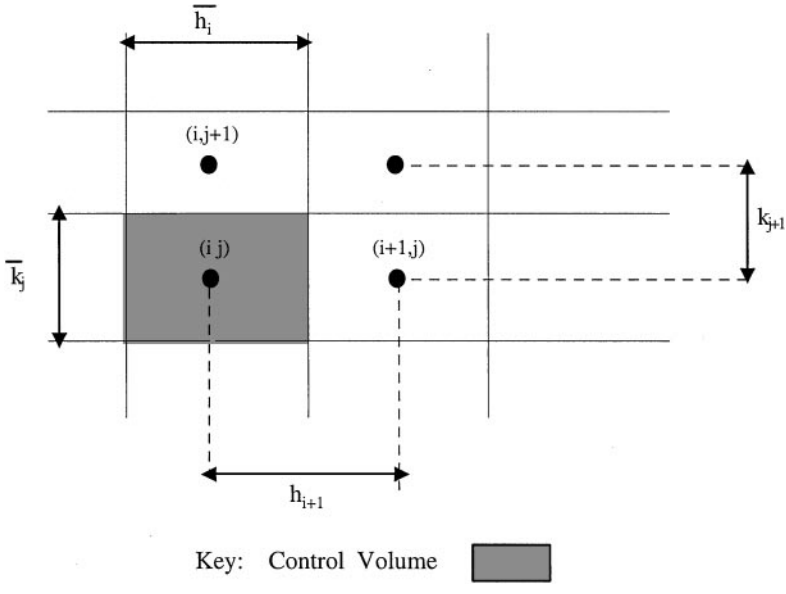

FIG. 1. Grid and control volume construction.

cell face locations defined as

$$x_{i-1/2} = x_i - \frac{\overline{h}_i}{2}$$
$$x_{i+1/2} = x_i + \frac{\overline{h}_i}{2}$$
$$y_{j-1/2} = y_j - \frac{\overline{k}_j}{2}$$
$$y_{j+1/2} = y_j + \frac{\overline{k}_j}{2},$$

where $\overline{h}_i, \overline{k}_j$ are as shown in Fig. 1.

We define the partition of Ω as

$$\overline{\Omega}_x^h = \{x_i, i = 1, \ldots, M : x_0 = 0, x_i - x_{i-1} = h_i, x_M = 1\}$$
$$\overline{\Omega}_y^h = \{y_j, j = 1, \ldots, N : y_0 = 0, y_j - y_{j-1} = k_j, y_N = 1\}$$

and we also define

$$\Omega_x^h = \overline{\Omega}_x^h \cap (0, 1]$$
$$\Omega_y^h = \overline{\Omega}_y^h \cap (0, 1]$$
$$\partial\Omega_x^h = \{0, 1\} \times \Omega_y^h$$
$$\partial\Omega_y^h = \Omega_x^h \times \{0, 1\}$$
$$\Omega^h = \Omega \cap \overline{\Omega}^h$$
$$\partial\Omega^h = \partial\Omega \cap \overline{\Omega}^h,$$

where $\overline{\Omega}^h$ is the Cartesian product of the one-dimensional partitions $\overline{\Omega}_x^h$ and $\overline{\Omega}_y^h$.

With each grid point (x_i, y_j) we associate the finite volume ω_{ij} defined as

$$\omega_{ij} = (x_{i-1/2}, x_{i+1/2}) \times (y_{j-1/2}, y_{j+1/2}). \tag{3}$$

Note that the h_i and $\bar{h}_i$ are related by $h_{i+1} = \frac{1}{2}(\bar{h}_i + \bar{h}_{i+1})$ and similarly $k_{j+1} = \frac{1}{2}(\bar{k}_j + \bar{k}_{j+1})$.

We define the *solution error* and the *truncation error* as follows: if our continuous problem is defined as

$$\mathcal{L}(u) = 0 \tag{4}$$

with solution u and the discrete approximation is denoted by

$$\mathcal{L}_h(u^h) = 0 \tag{5}$$

with solution u^h then we define the solution error as

$$\mathcal{E}_{\text{soln}} = u - u^h \tag{6}$$

and the truncation error as

$$\begin{aligned} \mathcal{E}_{\text{trunc}} &= \mathcal{L}(u) - \mathcal{L}_h(u) \\ &= -\mathcal{L}_h(u). \end{aligned} \tag{7}$$

We therefore see that the truncation error is simply the residual when the exact solution is inserted into the discretised equations. Since we do not usually have the exact solution available we are forced to approximate the truncation error using Taylor series expansions, replacing the derivative terms appearing in the expansion with finite difference approximations.

3. TAYLOR SERIES ANALYSIS

3.1. *Finite Volume Method on a Node-Based Grid*

We first consider the node-based grid as shown in Fig. 1. The node locations are defined followed by the cell face locations, so that the cell faces lie midway between adjacent nodes. The finite volume discretisation to Eq. (1) gives

$$-\frac{1}{h_i}\left(\left.\frac{\partial u}{\partial x}\right|_{i+1/2,j} - \left.\frac{\partial u}{\partial x}\right|_{i-1/2,j}\right) - \frac{1}{k_j}\left(\left.\frac{\partial u}{\partial y}\right|_{i,j+1/2} - \left.\frac{\partial u}{\partial y}\right|_{i,j-1/2}\right) = f_{i,j}. \tag{8}$$

Using two-point central differencing we approximate the derivatives at the cell faces as

$$\left.\frac{\partial u}{\partial x}\right|_{i+1/2,j} \approx \frac{u_{i+1,j} - u_{i,j}}{h_{i+1}} \tag{9}$$

$$\left.\frac{\partial u}{\partial x}\right|_{i-1/2,j} \approx \frac{u_{i,j} - u_{i-1,j}}{h_i}, \tag{10}$$

$$\tag{11}$$

with analogous expressions for the other faces. If we expand the nodal values about the face locations in Taylor series and rewrite the resulting expressions in terms of derivatives at (i, j) then we can easily show that we obtain an expression for the truncation error as

$$\Theta = \frac{1}{12}(h_{i+1} - h_i)\, u_{xxx} + \frac{1}{24}\frac{h_{i+1}^3 + h_i^3}{h_{i+1} + h_i} u_{xxxx} + \frac{1}{12}\left(k_{j+1} - k_j\right) u_{yyy} + \frac{1}{24}\frac{k_{j+1}^3 + k_j^3}{k_{j+1} + k_j} u_{yyy} + \cdots. \tag{12}$$

The leading term in the truncation error is clearly first order on a nonuniform grid, rising to second order in the case of a uniform grid. As the grid spacing goes to zero, the truncation error vanishes and so the approximation is consistent.

3.2. *Finite Volume Method on a Cell-Centred Grid*

We now consider the cell-centred grid arrangement. The control volume surfaces are defined first and the nodes are located in the centre of the control volume. The finite volume discretisation of Poisson's equation is again

$$-\frac{1}{h_i}\left(\left.\frac{\partial u}{\partial x}\right|_{i+1/2,\,j} - \left.\frac{\partial u}{\partial x}\right|_{i-1/2,\,j}\right) - \frac{1}{k_j}\left(\left.\frac{\partial u}{\partial y}\right|_{i,\,j+1/2} - \left.\frac{\partial u}{\partial y}\right|_{i,\,j-1/2}\right) = f_{i,j} \tag{13}$$

and we employ the centred difference approximations to the cell face gradients as before. Proceeding in a similar way to that for the node-based grid we expand the nodal values about the cell faces in Taylor series and express the resulting face derivatives in terms of quantities at the cell centre to obtain the truncation error as

$$\Theta = \frac{1}{4h_i}\left(\frac{\overline{h}_{i+1}^2 - \overline{h}_i^2}{h_{i+1} + h_i} - \frac{\overline{h}_i^2 - \overline{h}_{i-1}^2}{h_i + h_{i-1}}\right) u_{xx} + \frac{1}{4k_j}\left(\frac{\overline{k}_{j+1}^2 - \overline{k}_j^2}{k_{j+1} + k_j} - \frac{\overline{k}_j^2 - \overline{k}_{j-1}^2}{k_j + k_{j-1}}\right) u_{yy} + \frac{1}{24h_i}\left(\frac{\overline{h}_{i+1}^3 + 3\overline{h}_i\overline{h}_{i+1}^2 - 2\overline{h}_i^3}{h_{i+1} + h_i} - \frac{\overline{h}_{i-1}^3 + 3\overline{h}_i\overline{h}_{i-1}^2 - 2\overline{h}_{i-1}^3}{h_i + h_{i-1}}\right) u_{xxx} + \frac{1}{24k_j}\left(\frac{\overline{k}_{j+1}^3 + 3\overline{k}_j\overline{k}_{j+1}^2 - 2\overline{k}_j^3}{k_{j+1} + k_j} - \frac{\overline{k}_{j-1}^3 + 3\overline{k}_j\overline{k}_{j-1}^2 - 2\overline{k}_{j-1}^3}{k_j + k_{j-1}}\right) u_{yyy} + \cdots. \tag{14}$$

Simplifying the coefficient of u_{xx} and that of u_{yy} gives

$$\frac{1}{4h_i}\left(\frac{\overline{h}_{i+1}^2 - \overline{h}_i^2}{h_{i+1} + h_i} - \frac{\overline{h}_i^2 - \overline{h}_{i-1}^2}{h_i + h_{i-1}}\right) = \frac{1}{4}\left(\frac{\overline{h}_{i+1} + \overline{h}_{i-1}}{\overline{h}_i} - 2\right) \tag{15}$$

$$\frac{1}{4k_j}\left(\frac{\overline{k}_{j+1}^2 - \overline{k}_j^2}{k_{j+1} + k_j} - \frac{\overline{k}_j^2 - \overline{k}_{j-1}^2}{k_j + k_{j-1}}\right) = \frac{1}{4}\left(\frac{\overline{k}_{j+1} + \overline{k}_{j-1}}{k_j} - 2\right). \tag{16}$$

These terms vanish on a uniform grid but not on a nonuniform one. Therefore this analysis shows that on a uniform grid the truncation error is again of second order but on a nonuniform grid the approximation is in fact inconsistent, since the leading error term does not vanish as the grid spacing reduces to zero.

4. A SIMPLE NUMERICAL EXAMPLE

We now investigate the validity of the Taylor series expansion using a numerical example. For simplicity we choose a one-dimensional test case. We shall examine the problem

$$-\frac{d^2u}{dx^2} = \pi^2 \sin(\pi x) \qquad \text{for } x \in (0, 1) \tag{17}$$

$$u = 0 \qquad \text{for } x = 0, 1, \tag{18}$$

which has the exact solution

$$u = \sin(\pi x). \tag{19}$$

We discretise this test case on a one-dimensional grid using both the node-based and cell-centred schemes. To investigate the behaviour of the methods with grid nonuniformity, the grids are generated as randomly distributed values in (0, 1) with no imposed constraints on the maximum to minimum cell-size ratio. For the node-based scheme the node locations are randomly generated before being sorted into ascending order; in the case of the cell-centred scheme the cell faces are randomly located. In each case the number of points is chosen as a random variable between 10 and 200. The results are analysed by examining the maximum errors in the approximation on each grid against the maximum cell size. A total of 500 grids have been tried for both the node-based and cell-centred schemes, in order to obtain a reasonable error trend. The behaviour of both schemes with uniform grids has also been checked as a comparison.

Since we have an exact solution for the problem defined in Eqs. (17) and (18) then we may employ the solution error (Eq. (6)) and the truncation error definition (Eq. (7)) in assessing the accuracy of the approximations.

4.1. *Uniform Grid Results*

To check the behaviour of the two methods for the model problem as a baseline, uniform grid calculations were run where the control volume sizes were kept constant; the node-based and cell-centred methods should be identical in this case. The grids were progressively refined and the variation of the errors with grid size was calculated. As expected the results for the node-based and cell-centred grids are identical: analysis of the results yields the solution error varying like $h^{1.9989}$ and the truncation error varying like $h^{1.9976}$ (where h is the cell width) for both the node-based and cell-centred grids. The variations of the solution and truncation errors for the node-based grid are shown in Fig. 2.

4.2. *Nonuniform Grid Results*

We now turn our attention to the nonuniform grid case. Here we have employed randomly distributed grids to remove any bias from systematic grid distributions and to provide a severe test of the methods. The results for the node-based scheme are shown in Fig. 3. We can see that the solution error decreases as the maximum cell size goes to zero but that the truncation error appears to increase; at least there is more scatter in the truncation error results for the smaller cell sizes. The magnitude of the truncation error is always significantly larger than that of the solution error. This is not important if the constant of proportionality does not

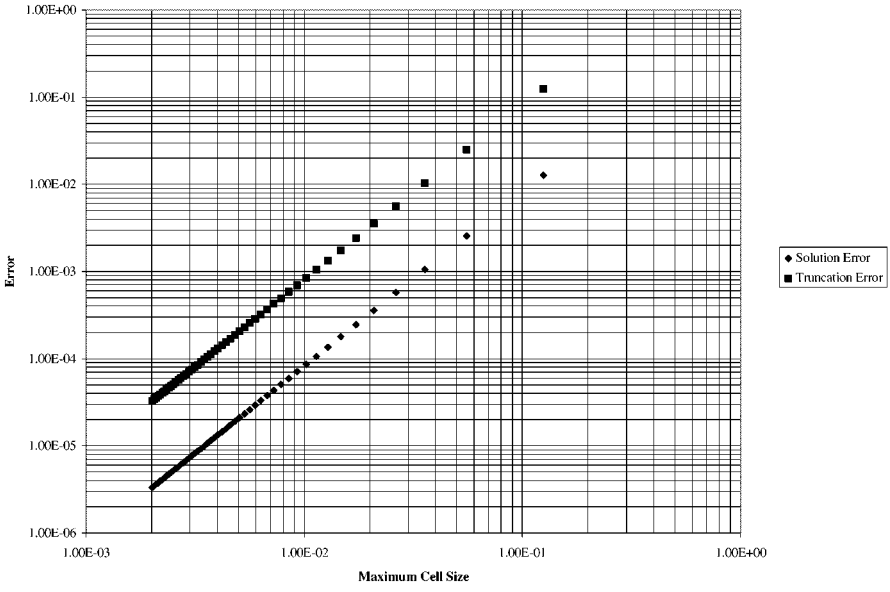

FIG. 2. Error distribution versus cell size for the uniform node-based grid.

vary significantly. If we look at the ratio of the solution error to the truncation error (Fig. 4) we see that it increases as the grid size decreases. The method therefore displays the property of supra-convergence; i.e., the solution error decreases more rapidly than is implied by the truncation error. This suggests that for the node-based scheme, the truncation error is not a reliable estimate of the solution error.

Further analysis of the solution error data of Fig. 3 indicates that the solution error varies like $h^{2.0667}$; even eliminating the results for the largest maximum cell sizes still returns second-order accuracy of the solution error; i.e., the solution error varies like the square of

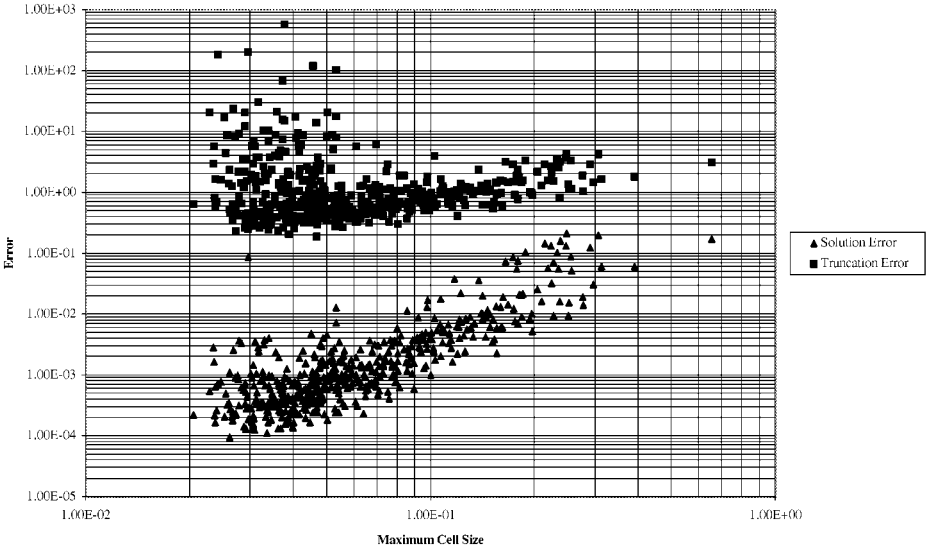

FIG. 3. Error distribution versus maximum cell size for the node-based grid.

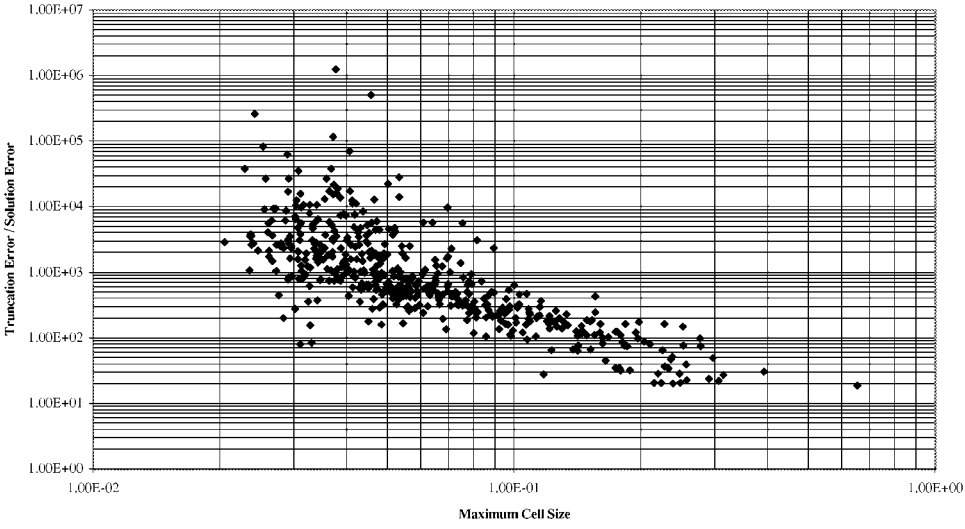

FIG. 4. Ratio of truncation error to solution error for the node-based grid.

the maximum cell size. Therefore, even though the scheme appears to be formally first-order accurate on a nonuniform grid from the Taylor series analysis, the solution error shows that it is in fact second-order accurate.

We now turn to the cell-centred approximation. In Fig. 5 we see the results for the solution error and truncation error again plotted against maximum cell size. The truncation error results do not decrease with reducing cell size, confirming the Taylor series analysis. However, the solution error results do decrease as the maximum cell size reduces, in contradiction to the conclusion drawn from the analysis. As a result the ratio of truncation error to solution error varies significantly as the cell size is reduced (Fig. 6). Analysing the

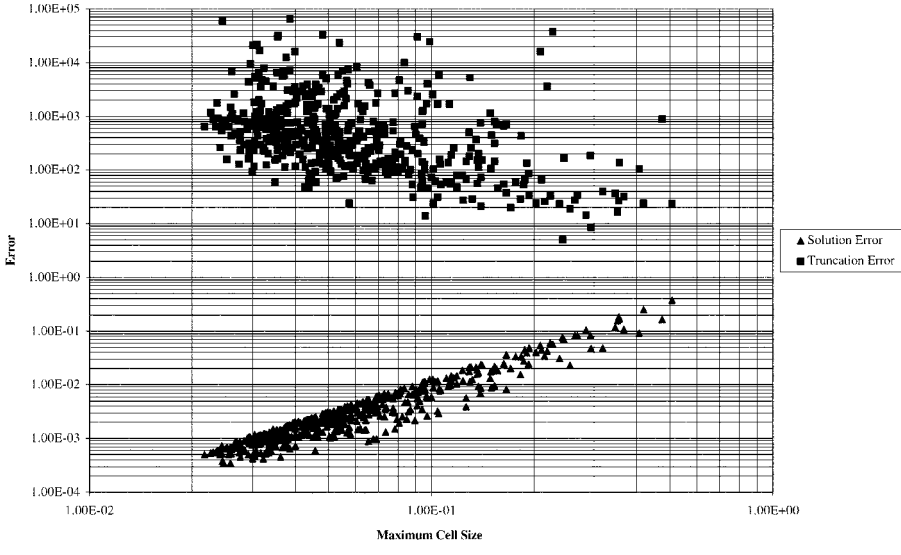

FIG. 5. Error distribution versus maximum cell size for the cell-centred grid.

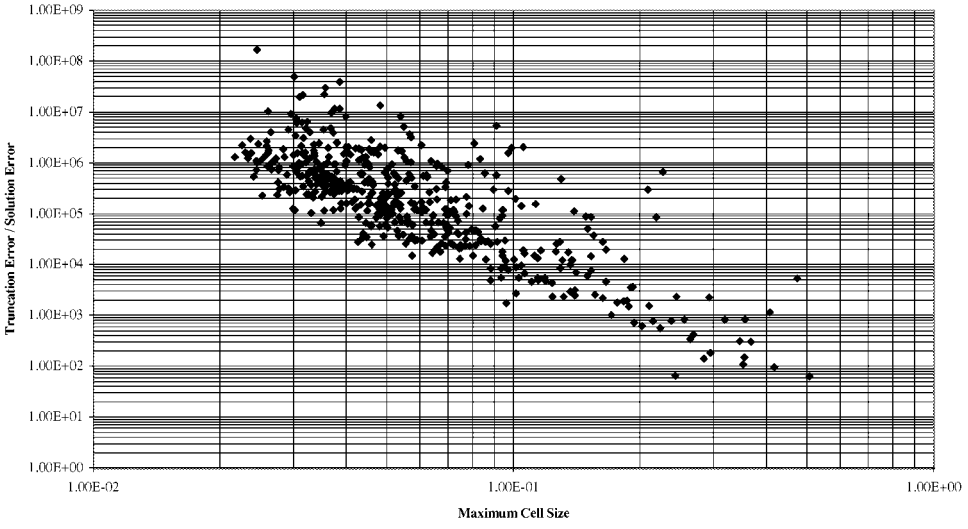

FIG. 6. Ratio of truncation error to solution error for the cell-centred grid.

variation of the solution error as the cell size changes shows that the solution error for the cell-centred scheme also varies with the square of the cell size. Therefore, in contrast to the conclusion drawn from the Taylor series analysis we can say that *the cell-centred finite volume approximation is not only consistent but is second-order accurate even on these nonuniform grids*.

5. AN ALTERNATIVE ANALYSIS OF THE CELL-CENTRED METHOD

We now turn to an alternative analysis of the cell-centred finite volume method. This follows the approach used by Süli [6] for the node-based grid case, extending it to the cell-centred method and showing how it is confirmed by numerical experiments.

We restate the finite volume approximation of u as follows: Let S^h be the set of piecewise bilinear functions defined on the rectangular partition of $\bar{\Omega}$ induced by $\bar{\Omega}^h$. Let $S_0^h \subset S^h$ be the set of those functions vanishing on $\partial\Omega$. The finite volume approximation of u is $u^h \in S_0^h$ satisfying

$$-\frac{1}{h_i k_j}\int_{\partial\omega_{ij}} \frac{\partial u^h}{\partial n}\, ds = \frac{1}{h_i k_j}\int_{x_{i-1/2}}^{x_{i+1/2}}\int_{y_{j-1/2}}^{y_{j+1/2}} f(x,y)\, dx\, dy \tag{20}$$

for $(x, y) \in \Omega^h$.

5.1. *Stability Analysis*

In this section we shall demonstrate the stability of the finite difference analogue of our finite volume approximation, which in turn demonstrates the stability of the finite volume scheme.

5.1.1. Some definitions. Let us define the discrete H^1-norm $\|\cdot\|_{1,h}$ as

$$\|v\|_{1,h} = \left(\|v\|^2 + |v|_{1,h}^2\right)^{1/2}, \tag{21}$$

where $\|\cdot\|$ is the discrete L_2-norm over Ω^h,

$$\|v\| = (v, v)^{1/2} \tag{22}$$

$$(v, w) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \overline{h}_i \overline{k}_j v_{ij} w_{ij}, \tag{23}$$

and $|\cdot|_{1,h}$ is the discrete H^1-seminorm,

$$|v|_{1,h} = \left(\|\Delta_x^- v|]_x^2 + \|\Delta_y^- v|]_y^2 \right)^{1/2}, \tag{24}$$

where we have

$$\|v|]_x^2 = (v, v]_x \tag{25}$$

$$\|v|]_y^2 = (v, v]_y \tag{26}$$

$$(v, w]_x = \sum_{i=1}^{M} \sum_{j=1}^{N-1} h_i \overline{k}_j v_{ij} w_{ij} \tag{27}$$

$$(v, w]_y = \sum_{i=1}^{M-1} \sum_{j=1}^{N} \overline{h}_i k_j v_{ij} w_{ij}. \tag{28}$$

We define the divided difference operators Δ_x^+, Δ_x^-, Δ_y^+, Δ_y^- as

$$\Delta_x^+ u = \frac{u_{i+1,j} - u_{ij}}{h_i} \tag{29}$$

$$\Delta_x^- u = \frac{u_{ij} - u_{i-1,j}}{h_i} \tag{30}$$

$$\Delta_y^+ u = \frac{u_{i,j+1} - u_{ij}}{k_j} \tag{31}$$

$$\Delta_y^- u = \frac{u_{ij} - u_{i,j-1}}{k_j}. \tag{32}$$

We define the discrete H^{-1}-norm as

$$\|v\|_{-1,h} = \sup \frac{|(v, w)|}{\|w\|_{1,h}}, \tag{33}$$

where the supremum is taken over all mesh functions $w \neq 0$ on $\overline{\Omega}^h$ vanishing on $\partial\Omega^h$.

5.2. *Stability Analysis of an Equivalent Finite Difference Method*

Using the definitions of the previous section we may rewrite the finite volume approximation given by Eq. (20) as a finite difference scheme on Ω^h,

$$-(\Delta_x^+ \Delta_x^- + \Delta_y^+ \Delta_y^-) u^h = T_{11} f \qquad \text{in } \Omega^h \tag{34}$$

$$u^h = 0 \qquad \text{on } \partial\Omega^h, \tag{35}$$

where we have

$$(T_{11} f)_{ij} = \frac{1}{h_i k_j} \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{j-1/2}}^{y_{j+1/2}} f(x, y)\, dx\, dy. \tag{36}$$

Let us define the operator $\mathcal{L}^h$ by

$$\mathcal{L}^h v = -(\Delta_x^+ \Delta_x^- + \Delta_y^+ \Delta_y^-) v \tag{37}$$

for mesh functions v defined on $\bar{\Omega}^h$ with $v = 0$ on $\partial\Omega^h$. Then if we take the inner product of $\mathcal{L}^h v$ with v we get

$$\begin{aligned}
(\mathcal{L}^h v, v) &= (-(\Delta_x^+ \Delta_x^-) v, v) + (-(\Delta_y^+ \Delta_y^-) v, v) \\
&= \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \bar{h}_i \bar{k}_j (-\Delta_x^+ \Delta_x^- v_{ij}) v_{ij} + \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \bar{h}_i \bar{k}_j (-\Delta_y^+ \Delta_y^- v_{ij}) v_{ij} \\
&= \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \bar{h}_i \bar{k}_j \left(-\frac{\Delta_x^- v_{i+1,j} - \Delta_x^- v_{ij}}{h_i} \right) v_{ij} \\
&\quad + \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \bar{h}_i \bar{k}_j \left(-\frac{\Delta_y^- v_{i+1,j} - \Delta_y^- v_{ij}}{k_i} \right) v_{ij}.
\end{aligned} \tag{38}$$

Examining the first term on the RHS of Eq. (38) gives

$$\begin{aligned}
&\sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \bar{h}_i \bar{k}_j \left(-\frac{\Delta_x^- v_{i+1,j} - \Delta_x^- v_{ij}}{h_i} \right) v_{ij} \\
&= \sum_{j=1}^{N-1} \bar{k}_j \left\{ -\frac{v_{2,j} - v_{1,j}}{h_2} v_{1,j} + \frac{v_{1,j} - v_{0,j}}{h_1} v_{1,j} - \frac{v_{3,j} - v_{2,j}}{h_3} v_{2,j} + \frac{v_{2,j} - v_{1,j}}{h_2} v_{2,j} + \cdots \right\} \\
&= \sum_{j=1}^{N-1} \bar{k}_j \left\{ -\frac{v_{1,j} - v_{0,j}}{h_2} (v_{1,j} - v_{0,j}) + -\frac{v_{2,j} - v_{1,j}}{h_2} (v_{2,j} - v_{1,j}) + \cdots \right\} \\
&= \sum_{i=1}^{M} \sum_{j=1}^{N-1} h_i \bar{k}_j (\Delta_x^- v_{ij})(\Delta_x^- v_{ij}) \\
&= (\Delta_x^- v, \Delta_x^- v]_x,
\end{aligned} \tag{39}$$

where we have used the fact that $v_{0,j} = 0$ and $v_{M,j} = 0$ by definition. A similar analysis gives

$$\sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \bar{h}_i \bar{k}_j \left(-\frac{\Delta_y^- v_{i+1,j} - \Delta_y^- v_{ij}}{k_j} \right) v_{ij} = (\Delta_y^- v, \Delta_y^- v]_y \tag{40}$$

and so we immediately have

$$\begin{aligned}(\mathcal{L}^h v, v) &= (\Delta_x^- v, \Delta_x^- v]_x + (\Delta_y^- v, \Delta_y^- v]_y \\ &= \|\Delta_x^- v|]_x^2 + \|\Delta_y^- v|]_y^2 \\ &= |v|_{1,h}^2 , \end{aligned} \tag{41}$$

where we have used the definitions of Eqs. (24), (25), and (26).

In addition, since $v = 0$ on $\partial\Omega^h$ then, using Lemma 3.2 in [6], we obtain the relation

$$\|v\|^2 \le \frac{1}{2} |v|_{1,h}^2 , \tag{42}$$

giving

$$\|v\|^2 \le \frac{1}{2}(\mathcal{L}^h v, v), \tag{43}$$

and therefore

$$\|v\|_{1,h} = \left(\|v\|^2 + |v|_{1,h}^2\right)^{1/2} \le \left(\frac{3}{2} |v|_{1,h}^2\right)^{1/2} , \tag{44}$$

immediately yielding

$$\|v\|_{1,h}^2 \le \frac{3}{2}(\mathcal{L}^h v, v). \tag{45}$$

Now the H^{-1}-norm of $\mathcal{L}^h v$ is defined, for some w, by

$$\|\mathcal{L}^h v\|_{-1,h} = \sup \frac{|(\mathcal{L}^h v, w)|}{\|w\|_{1,h}}. \tag{46}$$

Since $v = 0$ on $\partial\Omega^h$ we may say that

$$(\mathcal{L}^h v, v) \le \|v\|_{1,h} \, \|\mathcal{L}^h v\|_{-1,h} \tag{47}$$

and finally

$$\|v\|_{1,h} \le \frac{3}{2} \|\mathcal{L}^h v\|_{-1,h}. \tag{48}$$

This negative norm result indicates that the finite difference method is stable and hence so is the finite volume method by the following analysis.

For the finite difference method to be stable, we require $\|(\mathcal{L}^h)^{-1}\|$ to be uniformly bounded. From Eq. (48) we have demonstrated that $\|v\|_{1,h} \le C \|\mathcal{L}^h v\|_{-1,h}$ for some $C \ge 0$ for mesh functions v and so this implies the uniform boundedness of $\|(\mathcal{L}^h)^{-1}\|$ (see for example [7]).

Uniqueness of the solution follows by application of Theorem 3.2 from [6].

To estimate a bound on u^h, we set $v = u^h$ in Eq. (48), since $u^h = 0$ on $\partial\Omega^h$, to give

$$\|u^h\|_{1,h} \le \frac{3}{2} \|\mathcal{L}^h u^h\|_{-1,h} \tag{49}$$

and so

$$\|u^h\|_{1,h} \le \frac{3}{2} \|T_{11} f\|_{-1,h}. \tag{50}$$

5.3. *Convergence Analysis*

In this section we shall demonstrate the convergence of the finite volume scheme. In this analysis we shall use the concept of anisotropic Sobolev spaces on rectangular subdomains of $\mathbb{R}^2$, where $\mathbb{R}$ is the space of real numbers. Let $\omega = (a, b) \times (c, d)$ and let $r, s \in \mathbb{R}$, with $r, s > 0$. Then $H^{r,s}(\omega)$ is the anisotropic Sobolev space consisting of all functions $u \in L_2(\omega)$ such that

$$|u|_{H^{r,0}(\omega)} = \left\{ \int_c^d |u(\cdot, y)|^2_{H^r(a,b)}\, dy \right\}^{1/2} < \infty \tag{51}$$

$$|u|_{H^{0,s}(\omega)} = \left\{ \int_a^b |u(x, \cdot)|^2_{H^s(c,d)}\, dx \right\}^{1/2} < \infty. \tag{52}$$

We note that $H^{r,s}(\omega)$ is a Banach space equipped with the norm

$$\|u\|_{H^{r,s}(\omega)} = \left\{ \|u\|^2_{l_2(\omega)} + |u|^2_{H^{r,0}(\omega)} + |u|^2_{H^{0,s}(\omega)} \right\}^{1/2}. \tag{53}$$

Let us define the global error in the approximation as $z = u - u^h$. Then

$$\begin{aligned} -(\Delta_x^+\Delta_x^- + \Delta_y^+\Delta_y^-)z &= -(\Delta_x^+\Delta_x^- + \Delta_y^+\Delta_y^-)(u - u^h) \\ &= -(\Delta_x^+\Delta_x^- + \Delta_y^+\Delta_y^-)u - T_{11} f \\ &= T_{11}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) - (\Delta_x^+\Delta_x^- + \Delta_y^+\Delta_y^-)u \\ &= \left(T_{11}\frac{\partial^2 u}{\partial x^2} - \Delta_x^+\Delta_x^- u \right) + \left(T_{11}\frac{\partial^2 u}{\partial y^2} - \Delta_y^+\Delta_y^- u \right). \end{aligned} \tag{54}$$

Now from the definitions,

$$\begin{aligned} \left(T_{11}\frac{\partial^2 u}{\partial x^2} \right)_{ij} &= \frac{1}{h_i k_j} \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{y_{j-1/2}}^{y_{j+1/2}} \frac{\partial^2 u}{\partial x^2}\, dx\, dy \\ &= \frac{1}{h_i k_j} \int_{y_{j-1/2}}^{y_{j+1/2}} \left\{ \frac{\partial u}{\partial x}\big(x_{i+1/2}, y\big) - \frac{\partial u}{\partial x}\big(x_{i-1/2}, y\big) \right\} dy \\ &= \Delta_x^+ \left(T_{01}^- \frac{\partial u}{\partial x} \right)_{ij}, \end{aligned} \tag{55}$$

where we define T_{01}^- by

$$(T_{01}^- w)_{ij} = \frac{1}{k_j} \int_{y_{j-1/2}}^{y_{j+1/2}} w\big(x_{i-1/2}, y\big)\, dy. \tag{56}$$

We may define T_{10}^- in a similar way from analysis of the $(T_{11}\partial^2 u/\partial y^2)_{ij}$ term.

We may therefore write the equation for the global error as

$$-(\Delta_x^+\Delta_x^- + \Delta_y^+\Delta_y^-)z = \Delta_x^+\eta_1 + \Delta_y^+\eta_2 \qquad \text{in } \Omega^h \tag{57}$$

$$z = 0 \qquad \text{on } \partial\Omega^h, \tag{58}$$

where

$$\eta_1 = T_{01}^- \frac{\partial u}{\partial x} - \Delta_x^- u \tag{59}$$

$$\eta_2 = T_{10}^- \frac{\partial u}{\partial y} - \Delta_y^- u. \tag{60}$$

Applying the estimates derived in the stability analysis gives

$$\|z\|_{1,h} \le \frac{3}{2} \|\Delta_x^+ \eta_1 + \Delta_y^+ \eta_2\|_{-1,h}. \tag{61}$$

Now if w is a mesh function defined on $\bar{\Omega}^h$ such that $w = 0$ on $\partial\Omega^h$ then

$$\begin{aligned}
-(\Delta_x^+ \eta_1 + \Delta_y^+ \eta_2, w) &= -\sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \bar{h}_i \bar{k}_j (\Delta_x^+ \eta_1 + \Delta_y^+ \eta_2)_{ij} w_{ij} \\
&= -\sum_{j=1}^{N-1} \bar{k}_j \{((\eta_1)_{2,j} - (\eta_1)_{1,j}) w_{1,j} + ((\eta_1)_{3,j} - (\eta_1)_{2,j}) w_{2,j} + \cdots\} \\
&\quad - \sum_{i=1}^{M-1} \bar{h}_i \{((\eta_2)_{i,2} - (\eta_2)_{i,1}) w_{i,1} + ((\eta_2)_{i,3} - (\eta_2)_{i,2}) w_{i,2} + \cdots\} \\
&= \sum_{j=1}^{N-1} \bar{k}_j \{(\eta_1)_{1,j} (w_{1,j} - w_{0,j}) + (\eta_1)_{2,j} (w_{2,j} - w_{1,j}) + \cdots\} \\
&\quad + \sum_{i=1}^{M-1} \bar{h}_i \{(\eta_2)_{i,1} (w_{i,1} - w_{i,0}) + (\eta_2)_{i,2} (w_{i,2} - w_{i,1}) + \cdots\} \\
&= \sum_{i=1}^{M} \sum_{j=1}^{N-1} h_i \bar{k}_j (\eta_1)_{ij} \Delta_x^- w_{ij} + \sum_{i=1}^{M-1} \sum_{j=1}^{N} \bar{h}_i k_j (\eta_2)_{ij} \Delta_y^- w_{ij} \\
&= (\eta_1, \Delta_x^- w]_x + (\eta_2, \Delta_y^- w]_y.
\end{aligned} \tag{62}$$

From the definition of the norm $\|\cdot\|_{-1,h}$ we may see that

$$\|\Delta_x^+ \eta_1 + \Delta_y^+ \eta_2\|_{-1,h} \le \|\eta_1|]_x + \|\eta_2|]_y \tag{63}$$

and so

$$\|z\|_{1,h} = \|u - u^h\|_{1,h} \le \frac{3}{2} (\|\eta_1|]_x + \|\eta_2|]_y). \tag{64}$$

We must now estimate the RHS terms in Eq. (64). Let us consider the η_1 term first:

$$\begin{aligned}
(\eta_1)_{ij} &= \frac{1}{k_j} \int_{y_{j-1/2}}^{y_{j+1/2}} \frac{\partial u}{\partial x}\big(x_{i-1/2}, y\big)\, dy - \frac{u_{ij} - u_{i-1,j}}{h_i} \\
&= \frac{1}{k_j} \int_{y_{j-1/2}}^{y_{j+1/2}} \frac{\partial u}{\partial x}\big(x_{i-1/2}, y\big)\, dy - \frac{1}{h_i} \int_{x_{i-1}}^{x_i} \frac{\partial u}{\partial x}(x, y_j)\, dx \\
&= \frac{1}{h_i k_j} \int_{x_{i-1}}^{x_i} \int_{y_{j-1/2}}^{y_{j+1/2}} \left\{ \frac{\partial u}{\partial x}\big(x_{i-1/2}, y\big) - \frac{\partial u}{\partial x}(x, y_j) \right\} dx\, dy.
\end{aligned} \tag{65}$$

If we split η_1 as $\eta_1 = \eta_{11} + \eta_{12}$, where

$$(\eta_{11})_{ij} = \frac{1}{h_i k_j} \int_{x_{i-1/2}}^{x_i} \int_{y_{j-1/2}}^{y_{j+1/2}} \left\{ \frac{\partial u}{\partial x}\left(x_{i-1/2}, y\right) - \frac{\partial u}{\partial x}(x, y_j) \right\} dx\, dy \tag{66}$$

$$(\eta_{12})_{ij} = \frac{1}{h_i k_j} \int_{x_{i-1}}^{x_{i-1/2}} \int_{y_{j-1/2}}^{y_{j+1/2}} \left\{ \frac{\partial u}{\partial x}\left(x_{i-1/2}, y\right) - \frac{\partial u}{\partial x}(x, y_j) \right\} dx\, dy, \tag{67}$$

then we must estimate the terms η_{11} and η_{12}.

Let us first consider η_{11}. We may change variables so that

$$x = x_{i-1/2} + s\bar{h}_i, \quad 0 \le s \le \frac{1}{2} \tag{68}$$

$$y = y_j + t\bar{k}_j, \qquad -\frac{1}{2} \le t \le \frac{1}{2} \tag{69}$$

and define the function $\tilde{v}$ by

$$\tilde{v}(s, t) = h_i \frac{\partial u}{\partial x}(x, y). \tag{70}$$

Then

$$\begin{aligned}(\eta_{11})_{ij} &= \frac{1}{h_i k_j} \int_0^{1/2} \int_{-1/2}^{1/2} \left\{ \frac{1}{h_i} \tilde{v}(0, t) - \frac{1}{h_i} \tilde{v}(s, 0) \right\} \bar{h}_i \bar{k}_j\, ds\, dt \\ &= \frac{\bar{h}_i}{h_i^2} \int_0^{1/2} \int_{-1/2}^{1/2} \{\tilde{v}(0, t) - \tilde{v}(s, 0)\}\, ds\, dt \\ &= \frac{\bar{h}_i}{h_i^2} \tilde{\eta}_{11}, \end{aligned} \tag{71}$$

where

$$\tilde{\eta}_{11} = \int_0^{1/2} \int_{-1/2}^{1/2} \{\tilde{v}(0, t) - \tilde{v}(s, 0)\}\, ds\, dt. \tag{72}$$

Now $\tilde{\eta}_{11}$ is a linear functional (i.e., a linear mapping to the real numbers) with argument $\tilde{v}$, defined on $H^\sigma(\tilde{\omega})$ (with $\sigma > 1/2$) where

$$\tilde{\omega} = \left(0, \frac{1}{2}\right) \times \left(-\frac{1}{2}, \frac{1}{2}\right). \tag{73}$$

Since $\sigma > 1/2$, we may view $\tilde{\eta}_{11}$ as a Trace operator, a continuous linear map, from $H^\sigma(\tilde{\omega}) \to \mathbb{R}$ and so

$$|\tilde{\eta}_{11}| \le C \|\tilde{v}\|_{H^\sigma(\tilde{\omega})}, \quad \sigma > \frac{1}{2} \tag{74}$$

since $|\cdot|$ is the norm on $\mathbb{R}$ (see [7], Section 8 and Theorem 8.1). Therefore $\tilde{\eta}_{11}$ is bounded ($\tilde{v}$ must remain finite) and sublinear (that is, $\tilde{\eta}_{11}$ increases at best as rapidly as its argument $\tilde{v}$).

We have now established the properties required to allow us to apply Theorem 4.2 from [6] to obtain

$$|\tilde{\eta}_{11}| \leq C\left(|\tilde{v}|^2_{H^{0,\sigma}(\tilde{\omega})} + |\tilde{v}|^2_{H^{\sigma,0}(\tilde{\omega})}\right)^{1/2}, \quad \frac{1}{2} < \sigma \leq 2. \tag{75}$$

By definition of the Sobolev seminorms we have

$$|\tilde{v}|^2_{H^{0,\sigma}(\tilde{\omega})} = \int_0^{1/2} |\tilde{v}(s,\cdot)|^2_{H^{\sigma}(-1/2,1/2)}\, ds \tag{76}$$

$$|\tilde{v}(s,\cdot)|^2_{H^{\sigma}(-1/2,1/2)} = \int_{-1/2}^{1/2} \left|D_t^{\sigma}\tilde{v}\right|^2 dt, \tag{77}$$

where D_t^{σ} is the generalised derivative of order σ with respect to t. Now changing back to the original variables gives

$$\tilde{v}(s,t) = h_i \frac{\partial u}{\partial x}(x,y) \tag{78}$$

$$dx = \overline{h}_i\, ds \tag{79}$$

$$dy = \overline{k}_j\, dt, \tag{80}$$

and so

$$D_t^{\sigma}\tilde{v} = h_i \overline{k}_j^{\sigma} D_y^{\sigma}\left(\frac{\partial u}{\partial x}\right). \tag{81}$$

This gives us

$$|\tilde{v}(s,\cdot)|^2_{H^{\sigma}(-1/2,1/2)} = \frac{h_i^2 \overline{k}_j^{2\sigma}}{k_j} \int_{y_{j-1/2}}^{y_{j+1/2}} \left|D_y^{\sigma}\left(\frac{\partial u}{\partial x}\right)\right|^2 dy, \tag{82}$$

and so

$$|\tilde{v}|^2_{H^{0,\sigma}(\tilde{\omega})} = \frac{h_i^2 \overline{k}_j^{2\sigma}}{h_i k_j} \int_{x_{i-1/2}}^{x_i} \int_{y_{j-1/2}}^{y_{j+1/2}} \left|D^{\sigma}\left(\frac{\partial u}{\partial x}\right)\right|^2 dx\, dy \tag{83}$$

$$= \frac{h_i^2 \overline{k}_j^{2\sigma-1}}{h_i} \left|\frac{\partial u}{\partial x}\right|_{H^{0,\sigma}(\omega_{ij}^{+})}, \tag{84}$$

where $\omega_{ij}^{+} = (x_{i-1/2}, x_i) \times (y_{j-1/2}, y_{j+1/2})$.

In a similar way, we obtain

$$|\tilde{v}|^2_{H^{\sigma,0}(\tilde{\omega})} = \int_{-1/2}^{1/2} |\tilde{v}(\cdot,t)|^2_{H^{\sigma}(0,1/2)}\, dt \tag{85}$$

$$|\tilde{v}(\cdot,t)|^2_{H^{\sigma}(0,1/2)} = \int_0^{1/2} \left|D_s^{\sigma}\tilde{v}\right|^2 ds. \tag{86}$$

Returning to the original variables again gives

$$D_s^{\sigma}\tilde{v} = h_i \overline{h}_i^{\sigma} D_x^{\sigma}\left(\frac{\partial u}{\partial x}\right), \tag{87}$$

yielding

$$|\tilde{v}(\cdot,t)|^2_{H^\sigma(0,1/2)} = \frac{h_i^2 \overline{h}_i^{2\sigma}}{h_i} \int_{x_{i-1/2}}^{x_i} \left| D_x^\sigma \left(\frac{\partial u}{\partial x} \right) \right|^2 dx, \tag{88}$$

and so

$$|\tilde{v}|^2_{H^{\sigma,0}(\tilde{\omega})} = \frac{h_i^2 \overline{h}_i^{2\sigma}}{h_i k j} \int_{x_{i-1/2}}^{x_i} \int_{y_{j-1/2}}^{y_{j+1/2}} \left| D^\sigma \left(\frac{\partial u}{\partial x} \right) \right|^2 dx\, dy \tag{89}$$

$$= \frac{h_i^2 \overline{h}_i^{2\sigma-1}}{k_j} \left| \frac{\partial u}{\partial x} \right|^2_{H^{\sigma,0}(\omega_{ij}^+)}, \tag{90}$$

where $\omega_{ij}^+ = (x_{i-1/2}, x_i) \times (y_{j-1/2}, y_{j+1/2})$ again. Therefore

$$|\tilde{\eta}_{11}|^2 \le C \left(\frac{h_i^2 \overline{k}_j^{2\sigma-1}}{h_i} \left| \frac{\partial u}{\partial x} \right|^2_{H^{0,\sigma}(\omega_{ij}^+)} + \frac{h_i^2 \overline{h}_i^{2\sigma-1}}{k_j} \left| \frac{\partial u}{\partial x} \right|^2_{H^{\sigma,0}(\omega_{ij}^+)} \right), \tag{91}$$

and so

$$|(\eta_{11})_{ij}|^2 \le C \left(\frac{h_i^2 \overline{k}_j^{2\sigma-1}}{h_i^2} \left| \frac{\partial u}{\partial x} \right|^2_{H^{0,\sigma}(\omega_{ij}^+)} + \frac{\overline{h}_i^{2\sigma+1}}{h_i^2 k_j} \left| \frac{\partial u}{\partial x} \right|^2_{H^{\sigma,0}(\omega_{ij}^+)} \right). \tag{92}$$

If we now turn our attention to $(\eta_{12})_{ij}$ then we may change variables to

$$x = x_{i-1/2} + s\bar{h}_{i-1}, \quad -\frac{1}{2} \le s \le 0 \tag{93}$$

$$y = y_j + t\bar{k}_j, \qquad -\frac{1}{2} \le t \le \frac{1}{2} \tag{94}$$

and again define the function $\tilde{v}$ by

$$\tilde{v}(s,t) = h_i \frac{\partial u}{\partial x}(x,y). \tag{95}$$

We therefore arrive at an analogous estimate for η_{12},

$$|(\eta_{12})_{ij}|^2 \le C \left(\frac{\overline{h}_{i-1}^2 \overline{k}_j^{2\sigma-2}}{h_i^2} \left| \frac{\partial u}{\partial x} \right|^2_{H^{0,\sigma}(\omega_{ij}^-)} + \frac{\overline{h}_{i-1}^{2\sigma+1}}{h_i^2 k_j} \left| \frac{\partial u}{\partial x} \right|^2_{H^{\sigma,0}(\omega_{ij}^-)} \right), \tag{96}$$

where $\omega_{ij}^- = (x_{i-1}, x_{i-1/2}) \times (y_{j-1/2}, y_{j+1/2})$.

If we let $\omega_{ij} = \omega_{ij}^+ \cup \omega_{ij}^-$ then we have $\Omega = \cup_{ij} \omega_{ij}$. Now the square of the Sobolev seminorm displays the property of super-additivity; so if ω_1, ω_2 are mutually disjoint Lebesgue measurable sets then

$$|u|^2_{H^{r,s}(\omega_1)} + |u|^2_{H^{r,s}(\omega_2)} \le |u|^2_{H^{r,s}(\omega_1 \cup \omega_2)}. \tag{97}$$

Therefore

$$\sum_{i,j}\left\{\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\omega_{ij}^+)}+\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\omega_{ij}^-)}\right\}\le\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\Omega)} \tag{98}$$

$$\sum_{i,j}\left\{\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\omega_{ij}^+)}+\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\omega_{ij}^-)}\right\}\le\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\Omega)}. \tag{99}$$

Now since η_{11}, η_{12} represent integrals over disjoint sets whose union represents the domain of integration for η_1, we have

$$|\eta_1|^2=|\eta_{11}|^2+|\eta_{12}|^2, \tag{100}$$

and so from Eqs. (92) and (96) we obtain

$$\begin{aligned}|(\eta_1)_{ij}|^2\le C\frac{\overline{k}_j^{2\sigma-1}}{h_i^2k_j}&\left(\overline{h}_i^2\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\omega_{ij}^+)}+\overline{h}_{i-1}^2\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\omega_{ij}^-)}\right)\\&+C\frac{1}{h_i^2k_j}\left(\overline{h}_i^{2\sigma+1}\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\omega_{ij}^+)}+\overline{h}_{i-1}^{2\sigma+1}\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\omega_{ij}^-)}\right)\end{aligned} \tag{101}$$

From the definition of $\|\cdot|]_x$ we have

$$\|\eta_1|]_x^2=\sum_{i=1}^{M}\sum_{j=1}^{N-1}h_i\overline{k}_j\,(\eta_1)_{ij}^2, \tag{102}$$

and so

$$\begin{aligned}\|\eta_1|]_x^2&=C\sum_{i=1}^{M}\sum_{j=1}^{N-1}\frac{\overline{k}_j^{2\sigma-1}}{h_i}\left(\overline{h}_i^2\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\omega_{ij}^+)}+\overline{h}_{i-1}^2\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\omega_{ij}^-)}\right)\\&\quad+C\sum_{i=1}^{M}\sum_{j=1}^{N-1}\frac{1}{h_i}\left(\overline{h}_i^{2\sigma+1}\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\omega_{ij}^+)}+\overline{h}_{i-1}^{2\sigma+1}\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\omega_{ij}^-)}\right)\\&\le C\frac{\overline{h}^{2\sigma+1}}{h}\left(\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\Omega)}+\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\Omega)}\right),\end{aligned} \tag{103}$$

where $\overline{h}=\max_{i,j}(\overline{h}_i,\overline{k}_j)$ and $h=\max_{i,j}(h_i)$. We may, in fact, go further and let $\tilde{h}=\max(h,\overline{h})$ and then set

$$\|\eta_1|]_x^2\le C\tilde{h}^{2\sigma}\left(\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\Omega)}+\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\Omega)}\right). \tag{104}$$

Similarly, we may determine that

$$\|\eta_2|]_y^2\le C\tilde{h}^{2\sigma}\left(\left|\frac{\partial u}{\partial y}\right|^2_{H^{0,\sigma}(\Omega)}+\left|\frac{\partial u}{\partial y}\right|^2_{H^{\sigma,0}(\Omega)}\right) \tag{105}$$

if we define $\tilde{h} = \max_{i,j}(h_i, k_j, \overline{h}_i, \overline{k}_j)$. From the definition of the norm on the anisotropic Sobolev space

$$\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\Omega)} + \left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\Omega)} \leq \left\|\frac{\partial u}{\partial x}\right\|^2_{H^{\sigma,\sigma}(\Omega)} \tag{106}$$

$$\left|\frac{\partial u}{\partial y}\right|^2_{H^{0,\sigma}(\Omega)} + \left|\frac{\partial u}{\partial y}\right|^2_{H^{\sigma,0}(\Omega)} \leq \left\|\frac{\partial u}{\partial y}\right\|^2_{H^{\sigma,\sigma}(\Omega)}. \tag{107}$$

But since $\sigma > 0$ we have $H^{\sigma,\sigma}(\Omega) \equiv H^{\sigma}(\Omega)$ and $\|\cdot\|_{H^{\sigma,\sigma}(\Omega)} \equiv \|\cdot\|_{H^{\sigma}(\Omega)}$; consequently

$$\|\eta_1|]^2_x \leq C\tilde{h}^{2\sigma}\left\|\frac{\partial u}{\partial x}\right\|^2_{H^{\sigma}(\Omega)} \tag{108}$$

$$\|\eta_2|]^2_y \leq C\tilde{h}^{2\sigma}\left\|\frac{\partial u}{\partial y}\right\|^2_{H^{\sigma}(\Omega)}, \tag{109}$$

and thus we have

$$\|u - u^h\|_{1,h} \leq C\tilde{h}^{\sigma}\left(\left\|\frac{\partial u}{\partial x}\right\|_{H^{\sigma}(\Omega)} + \left\|\frac{\partial u}{\partial y}\right\|_{H^{\sigma}(\Omega)}\right), \quad \frac{1}{2} < \sigma \leq 2, \tag{110}$$

where C is some constant, $C > 0$.

We therefore have a bound on the error in terms of the derivatives of u. We would like to express this bound in terms of a norm of u itself. We will now reexpress Eq. (110) in this way.

From the definition of the Sobolev seminorm, recalling that $\Omega = (0, 1) \times (0, 1)$, we have

$$\begin{aligned}
\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma,0}(\Omega)} &= \int_0^1 \left|\frac{\partial u}{\partial x}(\cdot, y)\right|_{H^{\sigma}(0,1)} dy \\
&= \int_0^1 \left\{\sum_{|\alpha|=\sigma}\left\|D^{\alpha}\frac{\partial u}{\partial x}(x, y)\right\|^2_{L^2(0,1)}\right\} dy \\
&= \int_0^1 \left\{\sum_{|\alpha|=\sigma}\int_0^1 \left|D^{\alpha}\frac{\partial u}{\partial x}(x, y)\right|^2 dx\right\} dy \\
&= \sum_{|\alpha|=\sigma}\int_0^1\int_0^1 \left|D^{\alpha}\frac{\partial u}{\partial x}(x, y)\right|^2 dx\, dy \\
&= \sum_{|\alpha|=\sigma}\int_{\Omega} \left|D^{\alpha}\frac{\partial u}{\partial x}(x, y)\right|^2 d\Omega \\
&= \sum_{|\alpha|=\sigma}\left\|D^{\alpha}\frac{\partial u}{\partial x}\right\|^2_{L^2(\Omega)} \\
&= \left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma}(\Omega)}.
\end{aligned} \tag{111}$$

Similarly we obtain

$$\left|\frac{\partial u}{\partial x}\right|^2_{H^{0,\sigma}(\Omega)} = \left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma}(\Omega)} \tag{112}$$

with similar relations for the seminorms of $\frac{\partial u}{\partial y}$. In addition,

$$\begin{aligned}\left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma}(\Omega)} &= \sum_{|\alpha|=\sigma} \int_\Omega \left| D^\alpha \frac{\partial u}{\partial x}\right|^2 d\Omega \\ &= \sum_{|\alpha|=\sigma} \int_\Omega |D^{1+\alpha} u|^2 \, d\Omega \\ &= |u|^2_{H^{1+\sigma}(\Omega)}\end{aligned} \tag{113}$$

and

$$\left|\frac{\partial u}{\partial y}\right|^2_{H^{\sigma}(\Omega)} = |u|^2_{H^{1+\sigma}(\Omega)} . \tag{114}$$

Therefore

$$\begin{aligned}\|\eta_1|]^2_x &\le C\tilde{h}^{2\sigma} \left|\frac{\partial u}{\partial x}\right|^2_{H^{\sigma}(\Omega)} \\ &= C\tilde{h}^{2\sigma} |u|^2_{H^{1+\sigma}(\Omega)}\end{aligned} \tag{115}$$

$$\begin{aligned}\|\eta_2|]^2_y &\le C\tilde{h}^{2\sigma} \left|\frac{\partial u}{\partial y}\right|^2_{H^{\sigma}(\Omega)} \\ &= C\tilde{h}^{2\sigma} |u|^2_{H^{1+\sigma}(\Omega)} ,\end{aligned} \tag{116}$$

and so we finally conclude that

$$\|u - u^h\|_{1,h} \le C\tilde{h}^{\sigma} |u|_{H^{1+\sigma}(\Omega)} , \quad \frac{1}{2} < \sigma \le 2. \tag{117}$$

This result gives us a bound on the error of the finite volume approximation in terms of a norm of the exact solution itself. Since the exponent of the grid-spacing term is bounded away from 0 and the norm of the analytic solution is finite we may conclude that the error will decrease to zero in the limit of zero grid spacing; i.e., the cell-centred approximation is consistent.

We have therefore proved our central result on the consistency of the approximation. In the next section we shall examine some numerical results that follow from this analysis.

6. A TWO-DIMENSIONAL EXAMPLE

6.1. *Analytic Solution*

If we consider the model problem defined by Eq. (1) with Dirichlet boundary conditions as in (2) then taking $f = 2\pi^2 \sin(\pi x)\cos(\pi y)$ and $\Omega = (0, 1) \times (0, 1)$ gives the analytic

solution

$$u(x, y) = \sin(\pi x)\cos(\pi y). \tag{118}$$

Given this solution we may calculate the seminorm $|u|_{H^{1+\sigma}(\Omega)}$ for the cases $\sigma = 1$ and $\sigma = 2$; these are easily shown to be

$$|u|_{H^2(\Omega)} = \pi^2\sqrt{\frac{3}{4}} \tag{119}$$

and

$$|u|_{H^3(\Omega)} = \pi^3. \tag{120}$$

6.2. *Numerical Results*

In the calculations randomly distributed Cartesian grids have been used to solve the model problem. We have allowed a random number of grid lines to be distributed at random locations to avoid any bias of the results from choosing particular nonuniform grid distributions. A total of 1000 cases have been calculated to provide a reasonable sample size. As for the one-dimensional case there were no constraints on the maximum to minimum cell-size ratio. In each case we have calculated the maximum cell size, which corresponds to $\tilde{h}$ in Eq. (117). We have also calculated the maximum solution error $\|u - u^h\|_\infty$ and the H^1-norm of the solution error $\|u - u^h\|_{1,h}$; we wish to demonstrate both that the analysis predicts a behaviour of the H^1-norm that is obtained in practice and that the H^1-norm behaves in a similar manner to the maximum norm. It is after all the maximum norm that is usually calculated in assessing the accuracy of a numerical solution against known values.

In Fig. 7 we plot the maximum solution error against the maximum cell size. We can see that as the cell size decreases the solution error decreases as well with all of the data

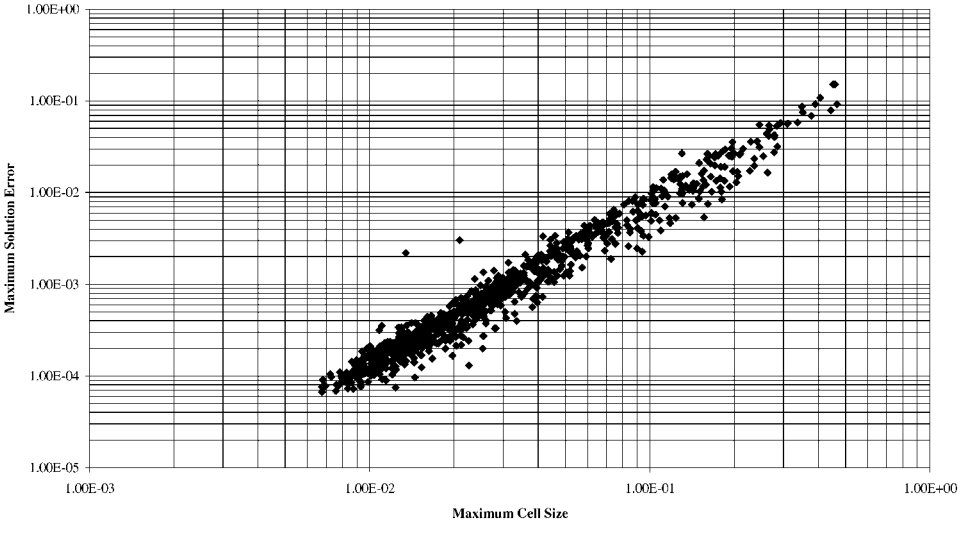

FIG. 7. Two-dimensional example showing variation of maximum solution error with maximum cell size.

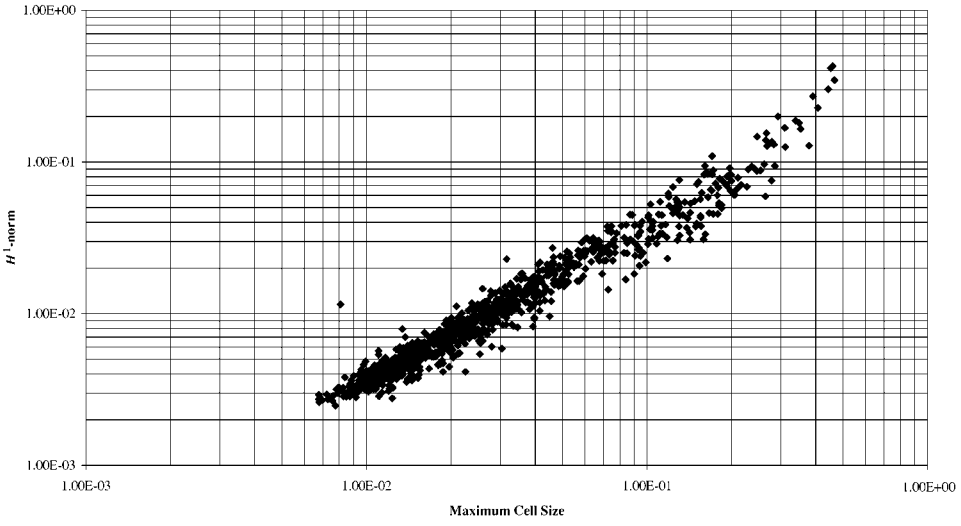

FIG. 8. Two-dimensional example showing variation of H^1-norm with maximum cell size.

points clustered around one line. Strangely this distribution appears smoother than the corresponding result in the one-dimensional case (Fig. 5), although the procedure is the same in both one- and two-dimensional cases. Calculating a best-fit line from the data yields that the solution error varies with $\tilde{h}^{1.728}$, a convergence rate slightly degraded from the one-dimensional case considered above but still close to the $\tilde{h}^2$ convergence expected from the uniform grid case. This is in contrast to the prediction of the truncation error analysis using Taylor series expansions presented in Section 3.2 where the approximation is apparently inconsistent.

Calculating the H^1-norm of the solution error for each example gives the results in Fig. 8. We can see that the H^1-norm tends to zero as the maximum cell size reduces. Applying a best-fit curve again shows that the calculations predict the H^1-norm varying with $\tilde{h}^{1.0306}$. The H^1-norm therefore converges as the grid size decreases. This exponent on the maximum cell size is in the range predicted by the convergence analysis and so confirms the correctness of the approach in Section 4. With the approximation of $\sigma \approx 1$ in Eq. (117) and the value of $|u|_{H^2(\Omega)}$ from Eq. (119) we can also calculate a bound on the constant C. We plot the limiting value of C to give equality in Eq. (117) in Fig. 9. We see that, apart from one outlier point at $C \approx 0.17$, we can estimate a limit of $C \approx 0.12$ for this problem. We can see that C does not change with grid size as expected.

7. DISCUSSION

The results of the numerical example presented in the previous section have confirmed the theoretical analysis. The H^1-norm of the solution error does indeed converge to zero as the maximum cell size decreases and at a rate consistent with the range predicted by the analysis. Although the H^1-norm of the solution error converges less rapidly than the maximum norm shown in Fig. 7 we still have a useful indication of the error behaviour of the approximation. In particular we have presented an analysis of the cell-centred finite volume method which reflects the behaviour seen in numerical experiments such as those in

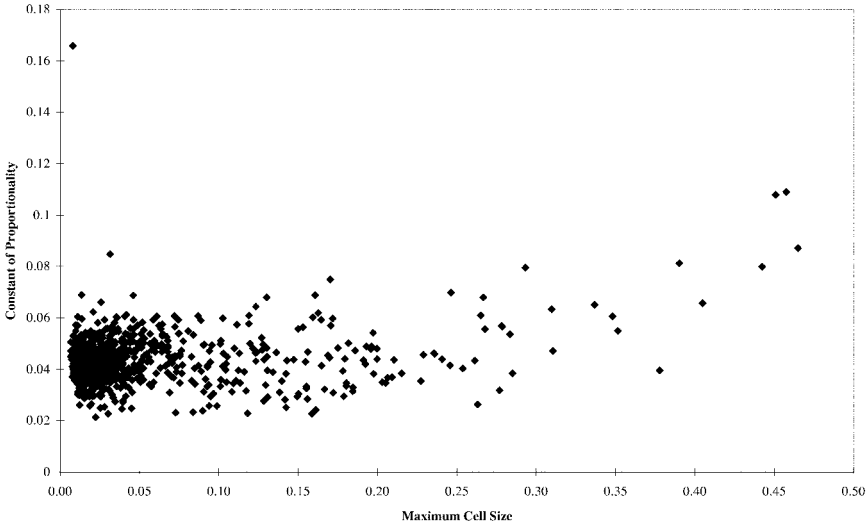

FIG. 9. Two-dimensional example showing value of the constant in Eq. (117) with $\sigma = 1$.

Section 3.2 and have confirmed it by example calculations. This analysis therefore provides us with a firm theoretical foundation for the use of cell-centred grids for diffusion problems. While the familiar Taylor series truncation error analysis suggested that the method was inconsistent, which was not supported by numerical evidence of the convergence of the solution error, we have now been able to demonstrate the convergence of the approximation by more sophisticated means.

The question now arises of whether this analysis can be extended to more complex convection–diffusion equations on more general grids. Some work has been performed in this area already: Süli has analysed the effects of grid distortion on finite volume methods in [5] and has examined the convection–diffusion equation in [4]. However, this work concentrated on the cell vertex finite volume method and was restricted to uniform rectangular grids with restrictions on the velocity field. Another analysis of the convection–diffusion equation, this time with the cell-centred grid formulation and various upwind schemes for the convection terms, was presented in [3]. This analysis was again performed for uniform grids. An analysis of the convection–diffusion equation on nonuniform cell-centred grids will require a more complex and lengthy exposition than that presented here for the diffusion equation.

8. CONCLUSIONS

We have presented a theoretical analysis of the cell-centred finite volume method for the diffusion equation in an attempt to explain the discrepancy between the convergence of the method observed in numerical experiments and the behaviour implied by the Taylor series truncation error analysis. By rewriting the finite volume method as an equivalent finite difference method we were able to prove stability. By invoking concepts from functional analysis we were able to demonstrate convergence of the method in the H^1-norm. Together these results give the consistency of the method. We were able to demonstrate the accuracy of the theoretical analysis by a numerical example which yielded superlinear convergence

of both the maximum norm and the H^1-norm of the solution error; the H^1-norm result confirms the theoretical analysis. While more work is required to extend this approach to the convection–diffusion equation on nonuniform cell-centred grids the present work is valuable because it provides a theoretical underpinning to the evidence of numerical experiments, showing that the cell-centred method is consistent for the diffusion equation in contrast to the results of Taylor series truncation error analysis.

ACKNOWLEDGMENTS

The authors acknowledge support provided by Rolls-Royce plc for this work. The paper is published with the permission of Rolls-Royce plc.

REFERENCES

1. S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods* (Springer-Verlag, Berlin, 1994).
2. M. Feistauer, *Mathematical Methods in Fluid Dynamics*, Pitman Monographs and Surveys in Pure and Applied Mathematics (Longman Scientific and Technical, Harlow, 1993), Vol. 67.
3. R. D. Lazarov, I. D. Mishev, and P. S. Vassilevski, Finite volume methods for convection–diffusion problems, *SIAM J. Numer. Analy.* **33**(1), 31 (1996).
4. K. W. Morton, M. Stynes, and E. Süli, Analysis of a cell-vertex finite volume method for convection–diffusion problems, *Math. Comput.* **66**(220), 1389 (1997).
5. E. Süli, *Finite Volume Methods on Distorted Partitions: Stability, Accuracy, Adaptivity*, Technical Report NA89/6, Oxford University Computing Laboratory, 1989.
6. E. Süli, Convergence of finite volume schemes for Poisson's equation on nonuniform meshes, *SIAM J. Numer. Anal.* **28**(5), 1419 (1991).
7. J. Wloka, *Partial Differential Equations* (Cambridge University Press, Cambridge, 1987).