MATHEMATICS OF COMPUTATION
Volume 66, Number 220, October 1997, Pages 1389–1406
S 0025-5718(97)00886-7

ANALYSIS OF A CELL–VERTEX FINITE VOLUME METHOD FOR CONVECTION–DIFFUSION PROBLEMS

K. W. MORTON, MARTIN STYNES, AND ENDRE SÜLI

ABSTRACT. A cell-vertex finite volume approximation of elliptic convection-dominated diffusion equations is considered in two dimensions. The scheme is shown to be stable and second-order convergent in a mesh-dependent L_2-norm.

1. INTRODUCTION

A finite volume formulation is the preferred technique for discretising systems of partial differential equations where conservation is the most important property to be modelled, compressible gas dynamics being the prime example–see Jameson [3] and a large subsequent literature. Of the various formulations that are possible, the cell-vertex scheme is often advocated for its compactness and its accuracy for first-order equations on distorted meshes (see Morton and Paisley [8] and Süli [17]); moreover, Morton et al. [6] and Crumpton et al. [2] have also demonstrated the effectiveness of the cell-vertex scheme for the compressible Navier-Stokes equations (see also Mackenzie and Morton [7]). Numerous practical computations have, indeed, shown this discretisation to be of very general utility, with recent extensions to unstructured three-dimensional meshes on general domains, and applicable to the very high aspect ratio meshes encountered with high Reynolds number, turbulent flows.

However, the resulting system of discrete equations is difficult to solve and its accuracy is hard to analyse. Some of these issues can be studied with simple model problems on rectangular meshes. In the earlier form of the method, for purely hyperbolic problems, when it was referred to as the finite difference box scheme of Thomas [18], Preissmann [13], Wendroff [19], Keller [4] and others, the equations were always solved by marching in a special coordinate direction. This is not possible with the equations for steady inviscid transonic flow and various alternatives have been developed, based on the work of Ni [11]; marching techniques are even less appropriate when second-order viscous terms are present, but Ni's techniques are still effective (see [2] and [7]). The present paper is one in a series devoted to the analysis of the resulting cell-vertex finite volume schemes.

Scalar convection-dominated diffusion problems, with general convective velocity fields, show both the remarkable approximation properties of cell-vertex methods

Received by the editor November 22, 1994 and, in revised form, January 26, 1996 and June 12, 1996.

1991 *Mathematics Subject Classification.* Primary 65N99, 65L10; Secondary 76M25.

Key words and phrases. Finite volume methods, stability, error analysis.

The authors are grateful to the British Council and Forbairt for the generous financial support of this project.

and highlight the challenge posed by their analysis. Mackenzie and Morton [7] presented a two-dimensional cell-vertex finite volume scheme which reduces to a non-standard twelve-point difference scheme on a uniform rectangular mesh; they demonstrated its accuracy in some well-known model problems and analysed its one-dimensional analogue. Morton and Stynes [9] adopted an alternative approach to the one-dimensional problem and analysed the case of pure convection in two dimensions, making use of the techniques of Süli [15], [16], [17]. The present paper is developed from this approach. The key ideas in the analysis are, firstly, to treat the finite volume scheme as a Petrov-Galerkin finite element method with a trial space $\mathcal{U}^h$ consisting of continuous piecewise bilinear functions, a test space $\mathcal{M}^h$ of piecewise constant functions, and an associated discrete bilinear form $B^h(\cdot,\cdot)$; secondly, a mapping $\mathcal{E}$ from $\mathcal{U}^h$ to $\mathcal{M}^h$ is constructed such that $B^h(v,\mathcal{E}v) \geq C\|v\|^2$ for all v in $\mathcal{U}^h$, where C is a fixed positive constant and $\|\cdot\|$ is a suitable norm on $\mathcal{U}^h$, so that B^h is coercive over $\mathcal{U}^h \times \mathcal{M}^h$. The bulk of the effort is in the construction of this mapping.

We consider the model convection-diffusion problem

$$\nabla \cdot (-\varepsilon \nabla u + \vec{a} u) = f \quad \text{on } \Omega, \tag{1.1}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{1.2}$$

where $\Omega = (0,1) \times (0,1) \subset \mathbf{R}^2$, ε is a small positive parameter, and $\vec{a} = (a_1, a_2)$ is a variable convective velocity, $\vec{a} \in (C^1(\bar{\Omega}))^2$. We assume that there exist positive constants α_1 and α_2 such that $a_i \geq \alpha_i$ on $\bar{\Omega}$, $i = 1, 2$, and that $f \in L_2(\Omega)$.

The well-posedness of this problem can be demonstrated by multiplying (1.2) by gu, where g is a bounded non-negative weight-function constructed so that

$$-\varepsilon \nabla^2 g - \vec{a} \cdot \nabla g + (\nabla \cdot \vec{a}) g > 0 \quad \text{in } \overline{\Omega}.$$

Our stability analysis of the cell-vertex finite volume approximation of (1.1), (1.2) makes use of a similar construction, and also requires that the discretisation takes particular forms at inflow and outflow boundaries. To clarify these points we have assumed that both components of $\vec{a}$ are strictly positive; then the construction of g is simplified and the inflow boundaries for the reduced problem (corresponding to $\varepsilon = 0$) are at $x = 0$ and $y = 0$, with the outflow boundaries at $x = 1$ and $y = 1$. In any case, the presence of the zero Dirichlet boundary condition along the outflow boundary of the reduced problem implies that, for $\varepsilon \ll 1$, the analytical solution contains a thin boundary layer in the neighbourhood of this part of $\partial\Omega$.

Wide-ranging comparisons of finite difference, finite element and finite volume methods for this problem are given in [5] and [14]. It is shown in [5] that the distinctive feature of the cell-vertex scheme for convection-diffusion problems is its uniform effectiveness as ε tends to 0, without the use of any adjustable parameters; indeed, as highlighted below, this is also a distinctive feature of the stability analysis. On the other hand, a dominant difficulty arises from the presence of the spurious chequer-board mode, which of course does not appear in one-dimensional problems. In nearly all other methods that suffer from chequer-board oscillations, the spurious mode is damped out by the diffusion term approximation, but not in the cell-vertex scheme where the diffusion term is transparent to the chequer-board mode: in practical computations with the cell-vertex scheme chequer-board oscillations are controlled by a fourth-order artificial dissipation term. However, since the inclusion of such a term complicates the analysis even further, in the present paper we make an assumption on the convective velocity which reduces the generation of the chequer-board mode and so simplifies the argument.

In Section 2 we state the cell-vertex approximation of the convection-diffusion equation (1.1) using the terminology of Petrov-Galerkin finite element methods. In Section 3 we prove that the cell-vertex scheme is stable in a mesh-dependent L_2-norm, uniformly as ε tends to zero. In doing so, we introduce the technical assumption on the velocity field $\vec{a}$ (see (3.1)) which takes the form of a discrete analogue of $\partial_y a_1 + \partial_x a_2 = 0$. A general class of vector functions $\vec{a}$ satisfying condition (3.1) is given by

$$\vec{a}(x, y) = (a_1(x, y), a_2(x, y)),$$

with

$$a_1(x, y) = A_1(x) + By, \qquad a_2(x, y) = A_2(y) - Bx,$$

where A_1 and A_2 are arbitrary functions of x and y, respectively, and B is a real constant. This gives quite a large class of velocity fields which the analysis can handle. However, we believe that condition (3.1) could be overcome by either a slight change in the scheme or a more sophisticated analysis: indeed, in the case of variable-coefficient linear advection, corresponding to $\varepsilon = 0$, the analysis of Balland and Süli [1] establishes the stability of the cell-vertex scheme in the absence of hypothesis (3.1). Unfortunately, the argument in [1] is difficult to extend to the case of $\varepsilon > 0$.

The stability of the scheme is a straightforward consequence of the discrete Gårding inequality stated in Theorem 3.5. Second-order convergence in a mesh-dependent L_2-norm is then deduced from a superconvergence result of Balland and Süli (see Proposition 3.1 in [1]); the resulting error estimate is stated in Theorem 3.7.

Throughout the paper, C (sometimes subscripted) will denote a generic positive constant, independent of ε and of the mesh-size, and may take different values at different occurrences. We denote by $\|\cdot\|_{H^s(\Omega)}$ and $|\cdot|_{H^s(\Omega)}$ the norm and the seminorm of the hilbertian Sobolev space $H^s(\Omega)$ of index s, and by $\|\cdot\|_{L_p(\Omega)}$ the norm of the Lebesgue space $L_p(\Omega)$, for $1 \leq p \leq \infty$.

2. The cell-vertex discretisation

Consider the uniform square mesh

$$\{(x_i, y_j) \,:\, x_i = ih,\;\; y_j = jh;\;\; i, j = 0, ..., N\}$$

of step-size $h = 1/N$, where N is an integer, $N \geq 3$.

The approximate solution U will be assumed to be continuous and piecewise bilinear on $\bar{\Omega}$, that is, bilinear on each cell

$$K^{ij} \equiv (x_{i-1}, x_i) \times (y_{j-1}, y_j).$$

Following the usual route, we construct the cell-vertex finite volume approximation of problem (1.1), (1.2) by integrating (1.1) over each cell (except for those cells that lie adjacent to the part of the boundary of Ω which is the outflow boundary for the reduced problem corresponding to $\varepsilon = 0$) and using Gauss' Theorem to convert integrals over cells into integrals over cell boundaries; we note that the outflow boundary for the reduced problem is

$$\{(x, y) \in \partial\Omega \,:\, \vec{a}(x, y) \cdot \vec{n}(x, y) > 0\},$$

where $\vec{n}(x, y)$ denotes the unit outward normal to $\partial\Omega$ at $(x, y) \in \partial\Omega$. An approximation of the contour integrals is needed to proceed further: we use the trapezium rule to evaluate integrals of $(\vec{a}u - \varepsilon\nabla u)$, and approximate ∇u by finite differences of u. Motivated by this approximate equality satisfied by the exact solution, we define the cell-vertex approximation of u as a continuous piecewise bilinear function U that satisfies the same relation as u but with approximate equality replaced by the equality sign. The equations resulting from this construction are supplemented with a zero Dirichlet boundary condition.

In order to give a precise definition of the cell-vertex finite volume scheme, we shall employ the terminology of Petrov-Galerkin finite element methods. Thus, we let $\mathcal{U}^h$ denote, for a mesh of size h, the linear space of all continuous piecewise bilinear functions that vanish on $\partial\Omega$, and let $\mathcal{M}^h$ denote the linear space of piecewise constant functions on the mesh which vanish on those K^{ij} for which $i = N$ or $j = N$. Let $I^h : (H_0^1(\Omega) \cap C(\bar{\Omega}))^2 \to (\mathcal{U}^h)^2$ be the interpolation projector onto $(\mathcal{U}^h)^2$. The desired discretisation of the convection term is obtained by defining the bilinear form $B_c : \mathcal{U}^h \times \mathcal{M}^h \to \mathbf{R}$ by

$$B_c(w, p) = (\nabla \cdot I^h(\vec{a}w), p), \tag{2.1}$$

where $(\cdot, \cdot)$ is the inner product in $L_2(\Omega)$. It is easy to see that the use of this bilinear form is equivalent to applying Gauss' Theorem followed by the use of the trapezium rule. Indeed, for $v \in C(\bar{\Omega})$ let v_{ij} denote $v(x_i, y_j)$, and for $q \in \mathcal{M}^h$ let q^{ij} denote the value of q on K^{ij}; then, by choosing p in (2.1) as the characteristic function χ^{ij} of the cell K^{ij}, we have that

$$\begin{aligned} B_c(w, \chi^{ij}) &= \frac{h}{2}\left[(a_1w)_{ij} + (a_1w)_{i,j-1} - (a_1w)_{i-1,j} - (a_1w)_{i-1,j-1}\right] \\ &\qquad + \frac{h}{2}\left[(a_2w)_{ij} + (a_2w)_{i-1,j} - (a_2w)_{i,j-1} - (a_2w)_{i-1,j-1}\right] \\ &= h\mu_y\delta_x(a_1w)_{ij} + h\mu_x\delta_y(a_2w)_{ij}, \end{aligned} \tag{2.2}$$

where we have employed the finite difference operators

$$\begin{aligned} &\delta_x v_{ij} = v_{ij} - v_{i-1,j}, &&\delta_y v_{ij} = v_{ij} - v_{i,j-1}, \\ &\mu_x v_{ij} = (v_{ij} + v_{i-1,j})/2, &&\mu_y v_{ij} = (v_{ij} + v_{i,j-1})/2. \end{aligned}$$

We use the methods of Mackenzie and Morton [7] to discretise the diffusion term in (1.1), together with a simple second-order boundary condition on the inflow boundary. For this purpose we consider the bilinear form $B_d : \mathcal{U}^h \times \mathcal{M}^h \to \mathbf{R}$ defined by

(2.3)

$$B_d(w, p) = -\varepsilon \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} hp^{ij}[\hat{\mu}_y(w_x)_{ij} - \hat{\mu}_y(w_x)_{i-1,j} + \hat{\mu}_x(w_y)_{ij} - \hat{\mu}_x(w_y)_{i,j-1}],$$

where, for $j = 1, \ldots, N-1$, we set

$$\hat{\mu}_y(w_x)_{ij} = \begin{cases} h^{-1}(\mu_x\mu_y w_{i+1,j} - \mu_x\mu_y w_{ij}), & \text{if } i = 1, \ldots, N-1, \\ h^{-1}(2\mu_y w_{1j} - \frac{1}{2}\mu_y w_{2j}), & \text{if } i = 0, \end{cases} \tag{2.4}$$

with $\hat{\mu}_x(w_y)_{ij}$ defined analogously.

Now the cell-vertex finite volume approximation of (1.1), (1.2) is defined as follows: find $U \in \mathcal{U}^h$ such that

$$B^h(U,p) \equiv B_d(U,p) + B_c(U,p) = (f,p) \quad \forall p \in \mathcal{M}^h. \tag{2.5}$$

This is a system of $(N-1)^2$ linear equations in the $(N-1)^2$ unknowns U_{ij}, the nodal values of the continuous piecewise bilinear function $U \in \mathcal{U}^h$, where $i,j = 1,\dots,N-1$. In the next section we show that B^h is coercive over $\mathcal{U}^h \times \mathcal{M}^h$, and therefore U is well-defined.

3. Stability and convergence

The crucial step in the analysis of the cell-vertex scheme is to prove stability via a discrete Gårding inequality that guarantees coercivity in a generalised sense. Let $P^h : L_2(\Omega) \to \mathcal{M}^h$ be the orthogonal projector onto $\mathcal{M}^h$. It is easily seen that $(P^h w)^{ij} = \mu_x\mu_y w_{ij}$ for any w in $\mathcal{U}^h$. We shall consider

$$B^h(w, GP^h w + \lambda P^h \nabla \cdot I^h(\vec{a}w)),$$

where G and λ are suitable elements in $\mathcal{M}^h$ chosen so as to achieve the desired coercivity. We analyse this expression in the following four lemmas.

Let $\Omega_h = (0, x_{N-1}) \times (0, y_{N-1})$. Then, as in Süli [15], [16], [17], Morton and Süli [10], and Morton and Stynes [9], we define the $l_2(\Omega_h)$-seminorm $|v|_{l_2(\Omega_h)}$ of a locally integrable function v by

$$|v|_{l_2(\Omega_h)} = \left\{ \sum_{K^{ij} \subset \Omega_h} h^2 \left| \frac{1}{h^2} \int_{K^{ij}} v\,dx\,dy \right|^2 \right\}^{1/2}.$$

We note that this seminorm is a norm on the linear space $\mathcal{U}^h$.

Lemma 3.1. *Assume that there exist positive constants α_1 and α_2 such that $a_i \geq \alpha_i$, $i = 1,2$, and that*

$$\mu_x\delta_y(a_1)_{ij} + \delta_x\mu_y(a_2)_{ij} = 0 \tag{3.1}$$

for all i and j. There exist $G \in \mathcal{M}^h$ and positive constants C_i, $i = 1,2,3,4$, such that $C_1 \geq G^{ij} \geq C_2$, $G^{ij} - G^{i+1,j} \geq C_3 h$ and $G^{ij} - G^{i,j+1} \geq C_4 h$ for all i and j, and

$$\begin{aligned} B_c(w, GP^h w) \geq 2C_2|w|^2_{l_2(\Omega_h)} \quad &+ \quad \sum_{j=1}^{N-1} h(\frac{1}{8}\alpha_1 G^{N-1,j} - C_2 h)(\mu_y w_{N-1,j})^2 \\ &+ \quad \sum_{i=1}^{N-1} h(\frac{1}{8}\alpha_2 G^{i,N-1} - C_2 h)(\mu_x w_{i,N-1})^2, \end{aligned}$$

for all $w \in \mathcal{U}^h$ and for all $h \leq h_0(\vec{a})$, where $h_0(\vec{a})$ depends only on $\|\vec{a}\|_{C^2(\overline{\Omega})}$.

Proof. From (2.2) and the definition of $\mathcal{M}^h$ we have that

$$B_c(w, GP^h w) = \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} hG^{ij}(\mu_x\mu_y w_{ij})[\mu_y\delta_x(a_1 w)_{ij} + \mu_x\delta_y(a_2 w)_{ij}]. \tag{3.2}$$

Using the elementary identities

$$\begin{aligned}\delta_x(bc) &= (\mu_x b)(\delta_x c) + (\delta_x b)(\mu_x c),\\ \mu_x(bc) &= (\mu_x b)(\mu_x c) + \frac{1}{4}(\delta_x b)(\delta_x c)\end{aligned}$$

and their δ_x, μ_y analogues, we can rewrite (3.2) as

$$\begin{aligned}B_c(w, GP^h w) = &\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} hG^{ij}(\mu_x\mu_y w_{ij})^2[\delta_x\mu_y(a_1)_{ij} + \mu_x\delta_y(a_2)_{ij}]\\ &+\frac{1}{4}\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} hG^{ij}(\mu_x\mu_y w_{ij})(\delta_x\delta_y w_{ij})[\mu_x\delta_y(a_1)_{ij} + \delta_x\mu_y(a_2)_{ij}]\\ &+\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} hG^{ij}(\mu_x\mu_y w_{ij})(\delta_x\mu_y w_{ij})[\mu_x\mu_y(a_1)_{ij} + \frac{1}{4}\delta_x\delta_y(a_2)_{ij}]\\ &+\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} hG^{ij}(\mu_x\mu_y w_{ij})(\mu_x\delta_y w_{ij})[\frac{1}{4}\delta_x\delta_y(a_1)_{ij} + \mu_x\mu_y(a_2)_{ij}]\\ &\equiv S_1 + S_2 + S_3 + S_4. \qquad (3.3)\end{aligned}$$

We define

$$G(x, y) = \mathrm{e}^{-(\kappa_1 x_{i-1} + \kappa_2 y_{j-1})}, \qquad \text{for } (x, y) \in K^{ij},$$

where κ_l, $l = 1, 2$, are positive constants which will be chosen appropriately in the course of the proof.

First we bound S_1 from below. Observing that

$$(\mu_x\mu_y w_{ij})^2 \le \frac{1}{2}[(\mu_x w_{ij})^2 + (\mu_x w_{i,j-1})^2], \qquad (3.4)$$

it follows that

$$\begin{aligned}S_1 \quad \ge \quad &-\frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} h^2 G^{ij}(\mu_x w_{ij})^2|A_{ij}|\\ &-\frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=1}^{N-2} h^2 G^{i,j+1}(\mu_x w_{ij})^2|A_{i,j+1}|, \qquad (3.5)\end{aligned}$$

with a similar bound in terms of $(\mu_y w_{ij})^2$, where

$$A_{ij} = \frac{1}{h}(\delta_x\mu_y(a_1)_{ij} + \mu_x\delta_y(a_2)_{ij}). \qquad (3.6)$$

Noting that $G^{ij} \ge G^{i,j+1}$ and separating out the term with $j = N-1$ from the first double summation,

$$S_1 \quad \ge \quad -\frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=1}^{N-2} h^2 G^{ij}(\mu_x w_{ij})^2(|A_{ij}| + |A_{i,j+1}|) \qquad (3.7)$$

$$-\frac{1}{2}\sum_{i=1}^{N-1} h^2 G^{i,N-1}(\mu_x w_{i,N-1})^2|A_{i,N-1}|. \qquad (3.8)$$

Now $(\mu_x b_{ij})(\delta_x b_{ij}) = (1/2)\delta_x b_{ij}^2$; we use this identity and sum by parts to get

$$\begin{aligned} S_3 \quad &= \quad \frac{1}{2}\sum_{i=1}^{N-2}\sum_{j=1}^{N-1}(\mu_y w_{ij})^2(G^{ij}B^{ij} - G^{i+1,j}B^{i+1,j}) \\ &\quad + \frac{1}{2}\sum_{j=1}^{N-1}(\mu_y w_{N-1,j})^2 G^{N-1,j}B^{N-1,j}, \end{aligned} \tag{3.9}$$

where

$$B^{ij} = h(\mu_x\mu_y(a_1)_{ij} + \frac{1}{4}\delta_x\delta_y(a_2)_{ij}).$$

Let us write

$$G^{ij}B^{ij} - G^{i+1,j}B^{i+1,j} = (G^{ij} - G^{i+1,j})B^{i+1,j} + G^{ij}(B^{ij} - B^{i+1,j}).$$

Recalling the definition of G^{ij}, it follows that

$$G^{ij} - G^{i+1,j} \geq \frac{1}{2}\kappa_1 h G^{ij},$$

provided $h \leq 1/\kappa_1$. In addition, since $a_1,\ a_2 \in C^1(\bar{\Omega})$,

$$B^{ij} = h(a_1)_{ij} + O(h^2) = h(a_1)_{i-1,j} + O(h^2).$$

Thus, for $0 < h \leq h_0$, where $h_0 = h_0(\vec{a})$, we have

$$B^{ij} \geq \frac{1}{2}h\alpha_1.$$

Similarly, for $0 < h \leq h_0$, where $h_0 = h_0(\vec{a})$ (with a possible adjustment of the previous h_0),

$$|B^{ij} - B^{i+1,j}| \leq 2h^2\|\nabla\vec{a}\|_{L_\infty(\Omega)}.$$

Consequently,

$$G^{ij}B^{ij} - G^{i+1,j}B^{i+1,j} \geq h^2 G^{ij}\left(\frac{1}{4}\kappa_1\alpha_1 - 2\|\nabla\vec{a}\|_{L_\infty(\Omega)}\right).$$

Returning to S_3, we deduce that

$$\begin{aligned} S_3 \quad &\geq \quad \frac{1}{2}\sum_{i=1}^{N-2}\sum_{j=1}^{N-1} h^2 G^{ij}(\mu_y w_{ij})^2\left(\frac{1}{4}\kappa_1\alpha_1 - 2\|\nabla\vec{a}\|_{L_\infty(\Omega)}\right) \\ &\quad + \frac{1}{4}\sum_{j=1}^{N-1} hG^{N-1,j}(\mu_y w_{N-1,j})^2\alpha_1. \end{aligned}$$

Analogously,

$$\begin{aligned} S_4 \quad &\geq \quad \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=1}^{N-2} h^2 G^{ij}(\mu_x w_{ij})^2\left(\frac{1}{4}\kappa_2\alpha_2 - 2\|\nabla\vec{a}\|_{L_\infty(\Omega)}\right) \\ &\quad + \frac{1}{4}\sum_{i=1}^{N-1} hG^{i,N-1}(\mu_x w_{i,N-1})^2\alpha_2. \end{aligned}$$

Now combining the lower bounds for S_1 and S_4 we obtain

$$\begin{aligned}\frac{1}{2}S_1+S_4 \quad \geq \quad & \sum_{i=1}^{N-1}\sum_{j=1}^{N-2} h^2G^{ij}(\mu_x w_{ij})^2\left(\frac{1}{8}\kappa_2\alpha_2 - \|\nabla\vec{a}\|_{L_\infty(\Omega)} - \frac{1}{4}|A_{ij}+A_{i,j+1}|\right)\\ & +\frac{1}{4}\sum_{i=1}^{N-1} hG^{i,N-1}(\mu_x w_{i,N-1})^2\left(\alpha_2 - h|A_{i,N-1}|\right).\end{aligned}$$

Noting that $|A_{i,j(\pm 1)}| \leq 2\|\nabla\vec{a}\|_{L_\infty(\Omega)}$ for $0<h\leq h_0$, where $h_0 = h_0(\vec{a})$ (with a possible adjustment of the previous h_0), it follows that

$$\alpha_2 - h|A_{i,N-1}| \geq \frac{1}{2}\alpha_2.$$

Choosing κ_2 such that

$$\kappa_2 \geq \frac{8}{\alpha_2}\left(1+2\|\nabla\vec{a}\|_{L_\infty(\Omega)}\right),$$

it follows that for $0<h\leq h_0$, where h_0 depends only on $\vec{a}$,

$$\frac{1}{2}S_1+S_4 \geq \sum_{i=1}^{N-1}\sum_{j=1}^{N-2} h^2G^{ij}(\mu_x w_{ij})^2 + \frac{1}{8}\alpha_2\sum_{i=1}^{N-1} hG^{i,N-1}(\mu_x w_{i,N-1})^2.$$

Similarly, choosing κ_1 such that

$$\kappa_1 \geq \frac{8}{\alpha_1}\left(1+2\|\nabla\vec{a}\|_{L_\infty(\Omega)}\right)$$

and using the alternative bound for S_1, we have that

$$\frac{1}{2}S_1+S_3 \geq \sum_{i=1}^{N-2}\sum_{j=1}^{N-1} h^2G^{ij}(\mu_y w_{ij})^2 + \frac{1}{8}\alpha_1\sum_{j=1}^{N-1} hG^{N-1,j}(\mu_y w_{N-1,j})^2.$$

Finally,

$$\begin{aligned}S_1+S_3+S_4 \geq & \sum_{i=1}^{N-2}\sum_{j=1}^{N-1} h^2G^{ij}(\mu_y w_{ij})^2 + \sum_{i=1}^{N-1}\sum_{j=1}^{N-2} h^2G^{ij}(\mu_x w_{ij})^2\\ & +\frac{1}{8}\alpha_1\sum_{j=1}^{N-1} hG^{N-1,j}(\mu_y w_{N-1,j})^2 + \frac{1}{8}\alpha_2\sum_{i=1}^{N-1} hG^{i,N-1}(\mu_x w_{i,N-1})^2,\end{aligned}$$

provided $h\leq h_0(\vec{a})$, and κ_i, $i=1,2$, are chosen as indicated above. Inserting this lower bound into (3.3) and recalling that due to (3.1) the term $S_2=0$, we deduce that

$$\begin{aligned}B_c(w,GP^hw) \geq C_2 & \left(\sum_{i=1}^{N-2}\sum_{j=1}^{N-1} h^2(\mu_y w_{ij})^2 + \sum_{i=1}^{N-1}\sum_{j=1}^{N-2} h^2(\mu_x w_{ij})^2\right.\\ & \left.+\frac{1}{8}\left(\alpha_1\sum_{j=1}^{N-1} hG^{N-1,j}(\mu_y w_{N-1,j})^2 + \alpha_2\sum_{i=1}^{N-1} hG^{i,N-1}(\mu_x w_{i,N-1})^2\right)\right.\end{aligned} \tag{3.10}$$

for all $w \in \mathcal{U}^h$ and for all $h \leq h_0(\vec{a})$. To complete the proof of the lemma we bound from below the right-hand side of this inequality in terms of $\|w\|_{l_2(\Omega_h)}$. This is easily accomplished by defining

$$\mu w_{ij} = \frac{1}{h^2} \int_{K^{ij}} w \, dx \, dy,$$

and noting that, for $w \in \mathcal{U}^h$,

$$\|w\|^2_{l_2(\Omega_h)} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} h^2 (\mu w_{ij})^2,$$

$$\mu w_{ij} = \frac{1}{2}(\mu_x w_{ij} + \mu_x w_{i,j-1}),$$

and

$$\mu w_{ij} = \frac{1}{2}(\mu_y w_{ij} + \mu_y w_{i-1,j}).$$

Since

$$(\mu w_{ij})^2 \leq \frac{1}{2}(\mu_x w_{ij})^2 + \frac{1}{2}(\mu_x w_{i,j-1})^2,$$

and $w = 0$ on $\partial\Omega$, it follows that

$$C_2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} h^2 |\mu w_{ij}|^2 \leq C_2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-2} h^2 (\mu_x w_{ij})^2 + C_2 h \sum_{i=1}^{N-1} h (\mu_x w_{i,N-1})^2.$$

Similarly,

$$C_2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} h^2 |\mu w_{ij}|^2 \leq C_2 \sum_{i=1}^{N-2} \sum_{j=1}^{N-1} h^2 (\mu_y w_{ij})^2 + C_2 h \sum_{j=1}^{N-1} h (\mu_y w_{N-1,j})^2.$$

Substituting the sum of these two inequalities into (3.10), we deduce the desired coercivity of the bilinear form $B_c(\cdot,\cdot)$ for all $w \in \mathcal{U}^h$ and for all $h \leq h_0(\vec{a})$. □

We note that condition (3.1) was necessary in order to remove the term S_2 that contained the second-difference $\delta_x \delta_y$; this term cannot be absorbed into any of the positive terms in the lower bound on $B_c(w, GP^h w)$.

Lemma 3.2. *For all $w \in \mathcal{U}^h$ and all $\lambda \in \mathcal{M}^h$, $\lambda \geq 0$,*

$$B_c(w, \lambda P^h \nabla \cdot I^h(\vec{a}w)) = |\lambda^{1/2} \nabla \cdot I^h(\vec{a}w)|^2_{l_2(\Omega_h)}.$$

Proof. This is immediate from (2.1). □

Lemma 3.3. *Assume that there exist positive constants C_2 and C_5 such that $G^{ij} \geq C_2$, $|G^{ij} - G^{i-1,j}| \leq C_5 h$, and $|G^{ij} - G^{i,j-1}| \leq C_5 h$ for all i and j. Then there* *exist positive constants $C_6 = C_6(C_2, C_5)$ and $h_1 = h_1(C_2, C_5)$, such that*

$$
\begin{aligned}
B_d(w, GP^h w) \geq \frac{1}{8} C_2 \varepsilon & \left[\sum_{i=1}^{N} \sum_{j=1}^{N-1} h^2 (\hat{\mu}_y (w_x)_{i-1,j})^2 \right. \\
& \left. + \sum_{i=1}^{N-1} \sum_{j=1}^{N} h^2 (\hat{\mu}_x (w_y)_{i,j-1})^2 \right] - C_6 \varepsilon |w|^2_{l_2(\Omega_h)} \\
& - \frac{\varepsilon}{8h} \left[\sum_{i=1}^{N-1} h G^{i,N-1} |\mu_x w_{i,N-1}|^2 + \sum_{j=1}^{N-1} h G^{N-1,j} |\mu_y w_{N-1,i}|^2 \right]
\end{aligned}
$$

for all $w \in \mathcal{U}^h$ and all $h \leq h_1$.

Proof. We give details only for the $\hat{\mu}_y$ terms, postponing the analogous contribution from the $\hat{\mu}_x$ terms of (2.3) until later in the proof. Thus we write, for any $w \in \mathcal{U}^h$,

$$
B_d(w, GP^h w) = -\varepsilon \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} h G^{ij} (\mu_x \mu_y w_{ij}) [\hat{\mu}_y (w_x)_{ij} - \hat{\mu}_y (w_x)_{i-1,j}] + (w_y \quad \text{terms})
$$

$$
\begin{aligned}
&= \varepsilon \sum_{j=1}^{N-1} h \left\{ \sum_{i=2}^{N-1} \hat{\mu}_y (w_x)_{i-1,j} [G^{ij} \mu_x \mu_y w_{ij} - G^{i-1,j} \mu_x \mu_y w_{i-1,j}] \right. \\
&\left. -\hat{\mu}_y (w_x)_{N-1,j} G^{N-1,j} \mu_x \mu_y w_{N-1,j} + \hat{\mu}_y (w_x)_{0,j} G^{1,j} \mu_x \mu_y w_{1,j} \right\} + (w_y \quad \text{terms}).
\end{aligned}
\tag{3.11}
$$

Now for $1 \leq i \leq N$ we have that

$$
G^{ij} \mu_x \mu_y w_{ij} - G^{i-1,j} \mu_x \mu_y w_{i-1,j} = h G^{ij} \hat{\mu}_y (w_x)_{i-1,j} + (G^{ij} - G^{i-1,j}) \mu_x \mu_y w_{i-1,j}.
$$

Therefore, using $|G^{ij} - G^{i-1,j}| \leq C_5 h$ together with the arithmetic-geometric mean inequality, we get

$$
\begin{aligned}
& \hat{\mu}_y (w_x)_{i-1,j} (G^{ij} \mu_x \mu_y w_{ij} - G^{i-1,j} \mu_x \mu_y w_{i-1,j}) \\
& \quad \geq h \left[G^{ij} (\hat{\mu}_y (w_x)_{i-1,j})^2 - C_5 |\hat{\mu}_y (w_x)_{i-1,j} \ \mu_x \mu_y w_{i-1,j}| \right] \\
& \quad \geq \frac{1}{2} h \left[G^{ij} (\hat{\mu}_y (w_x)_{i-1,j})^2 - C_5^2 (G^{ij})^{-1} (\mu_x \mu_y w_{i-1,j})^2 \right].
\end{aligned}
\tag{3.12}
$$

On the other hand,

$$
\begin{aligned}
-\hat{\mu}_y (w_x)_{N-1,j} \ \mu_x \mu_y w_{N-1,j} &= h (\hat{\mu}_y (w_x)_{N-1,j})^2 - \hat{\mu}_y (w_x)_{N-1,j} \ \mu_x \mu_y w_{N,j} \\
&\geq \frac{h}{2} (\hat{\mu}_y (w_x)_{N-1,j})^2 - \frac{1}{8h} (\mu_y w_{N-1,j})^2,
\end{aligned}
$$

and therefore

$$
\begin{aligned}
-\hat{\mu}_y (w_x)_{N-1,j} G^{N-1,j} \ \mu_x \mu_y w_{N-1,j} \quad \geq \quad & \frac{h}{2} G^{N-1,j} (\hat{\mu}_y (w_x)_{N-1,j})^2 \\
& - \frac{1}{8h} G^{N-1,j} (\mu_y w_{N-1,j})^2.
\end{aligned}
\tag{3.13}
$$

Analogously, since $\hat{\mu}_y(w_x)_{0,j} + \hat{\mu}_y(w_x)_{1,j} = (4/h)\mu_x\mu_y w_{1,j}$, it follows that

$$\begin{aligned} \hat{\mu}_y(w_x)_{0,j}\ \mu_x\mu_y w_{1,j} &= \frac{h}{4}(\hat{\mu}_y(w_x)_{0,j})^2 + \frac{h}{4}\hat{\mu}_y(w_x)_{0,j}\hat{\mu}_y(w_x)_{1,j} \\ &\geq \frac{h}{8}(\hat{\mu}_y(w_x)_{0,j})^2 - \frac{h}{8}(\hat{\mu}_y(w_x)_{1,j})^2, \end{aligned} \tag{3.14}$$

and therefore

$$\hat{\mu}_y(w_x)_{0,j}G^{1,j}\ \mu_x\mu_y w_{1,j} \geq \frac{h}{8}G^{1,j}(\hat{\mu}_y(w_x)_{0,j})^2 - \frac{h}{8}G^{1,j}(\hat{\mu}_y(w_x)_{1,j})^2. \tag{3.15}$$

Substituting (3.12)–(3.15) into (3.11), absorbing the last term of (3.15) into the corresponding term of (3.12), and noting that $G^{ij} \geq C_2$, we deduce that

$$\begin{aligned} B_d(w, GP^h w) \geq{}& \frac{1}{8}C_2\varepsilon\left\{\sum_{j=1}^{N-1}\sum_{i=1}^{N} h^2|\hat{\mu}_y(w_x)_{i-1,j}|^2 + \sum_{i=1}^{N-1}\sum_{j=1}^{N} h^2|\hat{\mu}_x(w_y)_{i,j-1}|^2\right\} \\ &-\frac{1}{2}C_6\varepsilon\left\{\sum_{j=1}^{N-1}\sum_{i=2}^{N-1} h^2|\mu_x\mu_y w_{i-1,j}|^2 + \sum_{i=1}^{N-1}\sum_{j=2}^{N-1} h^2|\mu_x\mu_y w_{i,j-1}|^2\right\} \\ &-\frac{\varepsilon}{8h}\left\{\sum_{i=1}^{N-1} hG^{i,N-1}|\mu_x w_{i,N-1}|^2 + \sum_{j=1}^{N-1} hG^{N-1,j}|\mu_y w_{N-1,j}|^2\right\} \end{aligned}$$

with $C_6 = C_5^2/C_2$, where we have assumed that h is sufficiently small, namely $h \leq h_1(C_2, C_5)$. Recalling the definition of $|\cdot|_{l_2(\Omega_h)}$, we obtain the desired result. □

We note that with $G^{ij} = \mathrm{e}^{-(\kappa_1 x_{i-1} + \kappa_2 y_{j-1})}$ and κ_1 and κ_2 chosen as in the proof of Lemma 3.1 all hypotheses on G in Lemma 3.3 are satisfied.

Lemma 3.4. *For all $w \in \mathcal{U}^h$ and all $\lambda \in \mathcal{M}^h$, $\lambda \geq 0$,*

$$\begin{aligned} |B_d(w, \lambda P^h\nabla\cdot I^h(\vec{a}w))| \leq{}& (\varepsilon/\sqrt{2})[\sum_{i=1}^{N}\sum_{j=1}^{N-1} h(\lambda^{i-1,j} + \lambda^{ij})(\hat{\mu}_y(w_x)_{i-1,j})^2 \\ &+ \sum_{i=1}^{N-1}\sum_{j=1}^{N} h(\lambda^{i,j-1} + \lambda^{ij})(\hat{\mu}_x(w_y)_{i,j-1})^2 + 2\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} h\lambda^{ij}((P^h\nabla\cdot I^h(\vec{a}w))^{ij})^2], \end{aligned}$$

where we set $\lambda^{0j} = \lambda^{i0} = 0$.

Proof. As in the proof of Lemma 3.3, we write out the details only for the $\hat{\mu}_y$ terms. The Cauchy-Schwarz inequality gives

$$\begin{aligned}
&|B_d(w, \lambda P^h \nabla \cdot I^h(\vec{a}w))| \\
&\quad \le \varepsilon \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \lambda^{ij} h |(P^h \nabla \cdot I^h(\vec{a}w))^{ij}| \\
&\qquad\qquad \times |\hat{\mu}_y(w_x)_{ij} - \hat{\mu}_y(w_x)_{i-1,j}| + (w_y \text{ terms}) \\
&\quad \le \varepsilon \{ \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \lambda^{ij} h [\hat{\mu}_y(w_x)_{ij} - \hat{\mu}_y(w_x)_{i-1,j}]^2 \}^{1/2} \\
&\qquad\qquad \times \{ \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \lambda^{ij} h ((P^h \nabla \cdot I^h(\vec{a}w))^{ij})^2 \}^{1/2} + (w_y \text{ terms}) \\
&\quad \le (\varepsilon/\sqrt{2}) \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \lambda^{ij} h [(\hat{\mu}_y(w_x)_{ij})^2 + (\hat{\mu}_y(w_x)_{i-1,j})^2] \\
&\qquad + (\varepsilon/\sqrt{2}) \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \lambda^{ij} h ((P^h \nabla \cdot I^h(\vec{a}w))^{ij})^2 + (w_y \text{ terms}) \\
&\quad = (\varepsilon/\sqrt{2}) \sum_{i=1}^{N} \sum_{j=1}^{N-1} h (\lambda^{i-1,j} + \lambda^{ij}) (\hat{\mu}_y(w_x)_{i-1,j})^2 \\
&\qquad + (\varepsilon/\sqrt{2}) \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \lambda^{ij} h ((P^h \nabla \cdot I^h(\vec{a}w))^{ij})^2 + (w_y \text{ terms}).
\end{aligned}$$

Including the w_y terms, we obtain the desired result. □

We now combine these four lemmas to reach our coercivity bound.

Theorem 3.5. *Assume that there exist positive constants α_1 and α_2 such that $a_i \ge \alpha_i$, $i = 1, 2$, and that*

$$\mu_x \delta_y (a_1)_{ij} + \delta_x \mu_y (a_2)_{ij} = 0$$

for all i and j. Choose $G \in \mathcal{M}^h$ such that $C_1 \ge G_{ij} \ge C_2 > 0$, $G^{ij} - G^{i+1,j} \ge C_3 h$, $G^{ij} - G^{i,j+1} \ge C_4 h$, $|G^{ij} - G^{i-1,j}| \le C_5 h$ and $|G^{ij} - G^{i,j-1}| \le C_5 h$ for all i and j. Let $\lambda \in \mathcal{M}^h$ be defined as follows:

$$\lambda^{ij} = \begin{cases} \frac{C_2}{8\sqrt{2}} h, & \text{if } h \ge 2\sqrt{2}\varepsilon, \\ 0, & \text{otherwise,} \end{cases}$$

with the convention that $\lambda^{0j} = \lambda^{i0} = 0$.

Then, for all $h \le \min(h_0(\vec{a}), h_1(C_2, C_5))$ and all ε such that

$$\frac{\alpha_i h}{\varepsilon} \ge \left(1 - \frac{8h}{\alpha_i}\right)^{-1}, \qquad i = 1, 2, \tag{3.16}$$

we have that

$$B^h(w, GP^h w + \lambda P^h \nabla \cdot I^h(\vec{a}w)) \ge C_2 |w|^2_{l_2(\Omega_h)} + \frac{1}{2} |\lambda^{1/2} \nabla \cdot I^h(\vec{a}w)|^2_{l_2(\Omega_h)} \tag{3.17}$$

for all $w \in \mathcal{U}^h$. Here $h_0(\vec{a})$ and $h_1(C_2, C_5)$ are as in Lemmas 3.1 and 3.3, respectively.

We note that a function G satisfying the conditions of Theorem 3.5 has been constructed in Lemma 3.1. The hypothesis (3.16) requires the mesh Péclet number to be greater than or equal to $1 + Ch$; this condition is automatically satisfied for the convection-dominated diffusion equations considered here.

Proof. Adding the results of the previous lemmas, we obtain

$$\begin{aligned}
&B^h(w, GP^h w + \lambda P^h \nabla \cdot I^h(\vec{a}w)) \\
&\quad \geq (2C_2 - C_6\varepsilon)|w|^2_{l_2(\Omega_h)} + |\lambda^{1/2}\nabla \cdot I^h(\vec{a}w)|^2_{l_2(\Omega_h)} \\
&\qquad - \frac{2\varepsilon}{\sqrt{2}} \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} h\lambda^{ij}((P^h \nabla \cdot I^h(\vec{a}w))^{ij})^2 \\
&\qquad + \varepsilon \sum_{i=1}^{N} \sum_{j=1}^{N-1} h[\frac{1}{8}hC_2 - \frac{1}{\sqrt{2}}(\lambda^{i-1,j} + \lambda^{ij})](\hat{\mu}_y(w_x)_{i-1,j})^2 \\
&\qquad + \varepsilon \sum_{i=1}^{N-1} \sum_{j=1}^{N} h[\frac{1}{8}hC_2 - \frac{1}{\sqrt{2}}(\lambda^{i,j-1} + \lambda^{ij})](\hat{\mu}_x(w_y)_{i,j-1})^2 \\
&\qquad + \sum_{j=1}^{N-1} h\left(\frac{1}{8}(\alpha_1 - \frac{\varepsilon}{h})G^{N-1,j} - C_2 h\right)(\mu_y w_{N-1,j})^2 \\
&\qquad + \sum_{i=1}^{N-1} h\left(\frac{1}{8}(\alpha_2 - \frac{\varepsilon}{h})G^{i,N-1} - C_2 h\right)(\mu_x w_{i,N-1})^2.
\end{aligned} \tag{3.18}$$

Recalling our assumed lower bound on $\alpha_i h/\varepsilon$, it follows that the last two sums are non-negative. In order to deal with the remaining terms, we need ε so small that $2C_2 - C_6\varepsilon \geq C_2 > 0$; since $C_6 = C_5^2/C_2$, this can be achieved by supposing that $\varepsilon \leq (C_2/C_5)^2$. Next, we claim that

$$\frac{2\varepsilon}{\sqrt{2}}h\lambda^{ij} \leq \frac{1}{2}h^2\lambda^{ij}$$

for each i and j. For if $\lambda^{ij} = 0$, the inequality is trivial. If $\lambda^{ij} \neq 0$, then $\varepsilon \leq h/(2\sqrt{2})$ by hypothesis, as required.

Finally,

$$\frac{1}{\sqrt{2}}(\lambda^{i-1,j} + \lambda^{ij}) \leq \frac{1}{8}C_2 h,$$

and similarly,

$$\frac{1}{\sqrt{2}}(\lambda^{i,j-1} + \lambda^{ij}) \leq \frac{1}{8}C_2 h.$$

Using the above inequalities in (3.18), the result follows. □

We can now derive a bound on $u - U$.

Theorem 3.6. *Suppose that the hypotheses of Theorem 3.5 hold. Then, for all sufficiently small ε and h, such as in (3.16),*

$$\begin{aligned}&|u-U|_{l_2(\Omega_h)}+|\lambda^{1/2}\nabla\cdot I^h(\vec{a}(u-U))|_{l_2(\Omega_h)}\\ &\quad\le C\varepsilon[\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}h^2\{\frac{1}{h}[\hat{\mu}_y(u_x^I)_{ij}-\frac{1}{h}\int_{y_{j-1}}^{y_j}u_x(x_i,y)\,dy\\ &\quad -\hat{\mu}_y(u_x^I)_{i-1,j}+\frac{1}{h}\int_{y_{j-1}}^{y_j}u_x(x_{i-1},y)\,dy]\\ &\quad +\frac{1}{h}[\hat{\mu}_x(u_y^I)_{ij}-\frac{1}{h}\int_{x_{i-1}}^{x_i}u_y(x,y_j)\,dx\\ &\quad -\hat{\mu}_x(u_y^I)_{i,j-1}+\frac{1}{h}\int_{x_{i-1}}^{x_i}u_y(x,y_{j-1})\,dx]\}^2]^{1/2}\\ &\quad +C|\nabla\cdot(I^h(\vec{a}u)-\vec{a}u)|_{l_2(\Omega_h)}+|u-u^I|_{l_2(\Omega_h)},\end{aligned}$$

where u^I is the interpolant of u from $\mathcal{U}^h$.

Proof. For brevity, set

$$\theta=GP^h(u^I-U)+\lambda P^h\nabla\cdot I^h(\vec{a}(u^I-U))$$

and

$$\begin{aligned}\chi^{ij}&=\frac{1}{h}[\hat{\mu}_y(u_x^I)_{ij}-\frac{1}{h}\int_{y_{j-1}}^{y_j}u_x(x_i,y)\,dy-\hat{\mu}_y(u_x^I)_{i-1,j}+\frac{1}{h}\int_{y_{j-1}}^{y_j}u_x(x_{i-1},y)\,dy]\\ &+\frac{1}{h}[\hat{\mu}_x(u_y^I)_{ij}-\frac{1}{h}\int_{x_{i-1}}^{x_i}u_y(x,y_j)\,dx-\hat{\mu}_x(u_y^I)_{i,j-1}+\frac{1}{h}\int_{x_{i-1}}^{x_i}u_y(x,y_{j-1})\,dx].\end{aligned}$$

From Theorem 3.5 we have that

$$\begin{aligned}&|u^I-U|^2_{l_2(\Omega_h)}+|\lambda^{1/2}\nabla\cdot I^h(\vec{a}(u^I-U))|^2_{l_2(\Omega_h)}\\ &\le CB^h(u^I-U,\theta)=CB^h(u^I,\theta)-C(f,\theta)\\ &=CB^h(u^I,\theta)-C(\nabla\cdot(-\varepsilon\nabla u+\vec{a}u),\theta)\\ &=-C\varepsilon\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}h^2\theta^{ij}\chi^{ij}+C(\nabla\cdot I^h(\vec{a}u^I)-\nabla\cdot(\vec{a}u),\theta)\\ &\le C|\theta|_{l_2(\Omega_h)}\{\varepsilon[\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}h^2|\chi^{ij}|^2]^{1/2}+|\nabla\cdot(I^h(\vec{a}u^I)-\vec{a}u)|_{l_2(\Omega_h)}\}.\end{aligned}$$

Noting that $I^h(\vec{a}u^I)=I^h(\vec{a}u)$, we deduce that

$$\begin{aligned}&|u^I-U|_{l_2(\Omega_h)}+|\lambda^{1/2}\nabla\cdot I^h(\vec{a}(u-U))|_{l_2(\Omega_h)}\\ &\quad\le C\varepsilon[\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}h^2|\chi^{ij}|^2]^{1/2}+C|\nabla\cdot(I^h(\vec{a}u)-\vec{a}u)|_{l_2(\Omega_h)}.\end{aligned}$$

We combine this with the triangle inequality

$$|u-U|_{l_2(\Omega_h)}\le|u-u^I|_{l_2(\Omega_h)}+|u^I-U|_{l_2(\Omega_h)}$$

and recall the definition of χ^{ij} to complete the argument. □

Hence we easily obtain our final bound on the global error.

Theorem 3.7. *Let the hypotheses of Theorem 3.5 hold. Suppose, further, that $u \in H^s(\Omega) \cap H_0^1(\Omega)$, $s > 2$, and assume that the entries of $\vec{a}$ belong to $C^{\langle s\rangle}(\bar{\Omega})$, where $\langle s\rangle$ denotes the smallest integer greater than or equal to s. Let λ be as in Theorem 3.5.*

There exist positive constants K_1, K_2 and K_3 such that

$$\begin{aligned} |u-U|_{l_2(\Omega_h)} \quad + \quad & |\lambda^{1/2}\nabla\cdot I^h(\vec{a}(u-U))|_{l_2(\Omega_h)} \\ & \leq K_1(\varepsilon,u)h^{r_1-1} + K_2(\varepsilon,u)h^{r_2-1} + K_3(u)h^{r_3}, \end{aligned} \tag{3.19}$$

where

$$\begin{aligned} K_1(\varepsilon,u) &= C_1\varepsilon|u|_{H^{r_1+1}(\Omega)} + C_2|u|_{H^{r_1}(\Omega_h)}, \qquad 1 \leq r_1 \leq \min(s,3), \\ K_2(\varepsilon,u) &= C_1\varepsilon^{1/2}\|u\|_{H^{r_2+1}(\Omega_h)}, \qquad 2 < r_2 \leq \min(s,3), \\ K_3(u) &= C_3|u|_{H^{r_3}(\Omega_h)}, \qquad 1 \leq r_3 \leq 2. \end{aligned}$$

The proof of this theorem relies on the following superconvergence result (see Balland and Süli [1]).

Proposition 3.1. *Given that s is a real number, $s > 1$, there exists a constant C, independent of the mesh-size h, such that*

$$\|P^h\left(\nabla\cdot\vec{d} - \nabla\cdot(I^h\vec{d})\right)\|_{L_2(\Omega)} \leq Ch^{r-1}\,|\vec{d}|_{H^r(\Omega)}, \qquad \text{with } 1 < r \leq \min(s,3),$$

for all $\vec{d} = (d_1,d_2)$ in $(H^s(\Omega))^2$.

We shall also need the following boundary layer estimate.

Proposition 3.2. *Let $D = (0,A)\times(0,B)$, where $A, B > 0$. Suppose that r is a positive real number, and let $D_\tau = (0,\tau)\times(0,B)$ with $0 < \tau < A$. Then*

$$|u|_{H^r(D_\tau)} \leq C\tau^{1/2}\|u\|_{H^{r+1}(D)}.$$

Proof. We shall prove the estimate for $0 \leq r \leq 1$; for $r > 1$, the proof is identical. According to a classical result (see, for example, Chapter 1, Section 4, of Oganesian and Ruhovec [12]):

$$\|u\|_{L_2(D_\tau)} \leq C\tau^{1/2}\|u\|_{H^1(D)}. \tag{3.20}$$

Consequently,

$$|u|_{H^1(D_\tau)} \leq C\tau^{1/2}\|u\|_{H^2(D)}. \tag{3.21}$$

Combining (3.20) and (3.21) we also have that

$$\|u\|_{H^1(D_\tau)} \leq C\tau^{1/2}\|u\|_{H^2(D)}. \tag{3.22}$$

Now inequalities (3.20) and (3.22) imply that $\mathcal{I} : u \mapsto u$ is a bounded linear operator from $H^1(D)$ to $L_2(D_\tau)$ and from $H^2(D)$ to $H^1(D_\tau)$. Using the K-method of function space interpolation it follows that $\mathcal{I}$ is a bounded linear operator from $H^{r+1}(D)$ to $H^r(D_\tau)$, for $0 < r < 1$, and that

$$\|u\|_{H^r(D_\tau)} \leq C\tau^{1/2}\|u\|_{H^{r+1}(D)}.$$

Therefore also,

$$|u|_{H^r(D_\tau)} \leq C\tau^{1/2}\|u\|_{H^{r+1}(D)}, \qquad 0 < r < 1.$$

For $r = 0$ and $r = 1$, the desired inequalities are (3.20) and (3.21), respectively. □

Proof. (of Theorem 3.7) Let us label the three terms on the right-hand side of the inequality in Theorem 3.6 by T_1, T_2 and T_3.

We begin by considering T_1. For the sake of notational simplicity, we define, as in the proof of Theorem 3.6,

$$\begin{aligned}\chi^{ij} &= \frac{1}{h}[\hat{\mu}_y(u_x^I)_{ij} - \frac{1}{h}\int_{y_{j-1}}^{y_j} u_x(x_i,y)\,dy - \hat{\mu}_y(u_x^I)_{i-1,j} + \frac{1}{h}\int_{y_{j-1}}^{y_j} u_x(x_{i-1},y)\,dy] \\ &\quad + \frac{1}{h}[\hat{\mu}_x(u_y^I)_{ij} - \frac{1}{h}\int_{x_{i-1}}^{x_i} u_y(x,y_j)\,dx - \hat{\mu}_x(u_y^I)_{i,j-1} + \frac{1}{h}\int_{x_{i-1}}^{x_i} u_y(x,y_{j-1})\,dx] \\ &\equiv \chi_{(1)}^{ij} + \chi_{(2)}^{ij}, \qquad 1 \le i,j \le N-1.\end{aligned}$$

For $2 \le i \le N-1$ and $1 \le j \le N-1$, a simple application of the Bramble-Hilbert lemma yields

$$|\chi_{(1)}^{ij}| \le Ch^{-2}h^{r-1}|u|_{H^r(T_{ij})}, \qquad 2 < r \le \min(s,4),$$

where $T_{ij} = (x_{i-2}, x_{i+1}) \times (y_{j-1}, y_j)$. Consequently, for $2 \le i \le N-1$ and $1 \le j \le N-1$,

$$\left(\sum_{i=2}^{N-1}\sum_{j=1}^{N-1} h^2|\chi_{(1)}^{ij}|^2\right)^{1/2} \le Ch^{r-2}|u|_{H^r(\Omega)}, \qquad 2 < r \le \min(s,4).$$

Now let us consider the case when $i = 1$ and $1 \le j \le N-1$; recalling the definition of $\hat{\mu}_y(u_x^I)_{0,j}$ and appealing to the Bramble-Hilbert lemma, we deduce that

$$\begin{aligned}\left(\sum_{j=1}^{N-1} h^2|\chi_{(1)}^{1j}|^2\right)^{1/2} &\le C\left(\sum_{j=1}^{N-1} h^2\frac{1}{h^4}h^{2t-2}|u|^2_{H^t((x_0,x_2)\times(y_{j-1},y_j))}\right)^{1/2} \\ &\le Ch^{t-2}|u|_{H^t(\Omega_0)}, \qquad 2 < t \le \min(s,3),\end{aligned}$$

where $\Omega_0 = (x_0, x_2) \times (y_0, y_{N-1})$. Combining these two bounds we get

$$\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} h^2|\chi_{(1)}^{ij}|^2\right)^{1/2} \le C(h^{r-2}|u|_{H^r(\Omega)} + h^{t-2}|u|_{H^t(\Omega_0)}),$$

with $2 < r \le \min(s,4)$ and $2 < t \le \min(t,3)$. Exploiting the boundary layer estimate stated in Proposition 3.1,

$$|u|_{H^t(\Omega_0)} \le Ch^{1/2}\|u\|_{H^{t+1}(\Omega_h)}.$$

Thus,

$$\varepsilon\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} h^2|\chi_{(1)}^{ij}|^2\right)^{1/2} \le C\varepsilon(h^{r-1}|u|_{H^{r+1}(\Omega)} + h^{t-2}h^{1/2}\|u\|_{H^{t+1}(\Omega_h)}),$$

with $1 < r \le \min(s,3)$, $2 < t \le \min(s,3)$. Similarly,

$$\varepsilon\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} h^2|\chi_{(2)}^{ij}|^2\right)^{1/2} \le C\varepsilon(h^{r-1}|u|_{H^{r+1}(\Omega)} + h^{t-2}h^{1/2}\|u\|_{H^{t+1}(\Omega_h)}),$$

with $1 < r \leq \min(s, 3)$, $2 < t \leq \min(s, 3)$. Thus, recalling from the statement of Theorem 3.5 that $\varepsilon \leq Ch$, it follows that

$$T_1 \leq C_1(\varepsilon h^{r-1}|u|_{H^{r+1}(\Omega)} + \varepsilon^{1/2}h^{t-1}\|u\|_{H^{t+1}(\Omega_h)}), \qquad \text{for } 1 < r \leq \min(s, 3),\ 2 < t \leq \min(s, 3).$$

Term T_2 is estimated using Proposition 3.1 with $\vec{d} = \vec{a}u$; we obtain the bound

$$T_2 \leq C_2 h^{r-1}|u|_{H^r(\Omega_h)}, \qquad 1 < r \leq \min(s, 3).$$

Finally, the term T_3 can be bounded using a standard interpolation error estimate to obtain

$$T_3 \leq C_3 h^r |u|_{H^r(\Omega_h)}, \qquad 1 < r \leq \min(s, 2) = 2.$$

Combining the bounds on T_1, T_2 and T_3 yields the desired error estimate. □

4. Conclusions

In this paper we have been concerned with the stability and the convergence of a cell-vertex finite volume method for linear elliptic convection-dominated diffusion equations in the plane. Using a combination of techniques from the theory of finite difference and finite element methods we proved that the scheme is stable, uniformly as the diffusion coefficient tends to zero, and second-order convergent. In addition to the error bound in the mesh-dependent l_2-norm, Theorem 3.7 implies that, provided $u \in H^4(\Omega) \cap H^1_0(\Omega)$, the derivative of the global error in the streamwise direction is $O(h^{3/2})$, as long as $h \geq 2\sqrt{2}\varepsilon$. The results presented here may be extended to tensor-product non-uniform meshes.

References

1. P. Balland and E. Süli, *Analysis of the cell vertex finite volume method for hyperbolic equations with variable coefficients,* SIAM J. Numer. Anal. **34**, No. 3, June 1997.
2. P.I. Crumpton, J.A. Mackenzie and K.W. Morton, *Cell vertex algorithms for the compressible Navier-Stokes equations,* Journal of Computational Physics, **109** (1993), 1–15. MR **94e:**76081
3. A. Jameson, *Acceleration of transonic potential flow calculations on arbitrary meshes by the multiple grid method,* AIAA Paper **79**, p. 1458, 1979.
4. H. Keller, *A new finite difference scheme for parabolic problems,* In: Numerical Solution of Partial Differential Equations II, SYNSPADE 1970 (Ed., B. Hubbard,) Academic Press, 1971, 327–350. MR **43:**2866
5. K. W. Morton, *Numerical Solution of Convection-Diffusion Problems,* Applied Mathematics and Mathematical Computation, **12**, Chapman and Hall, London, 1996.
6. K.W. Morton, P.I. Crumpton and J.A. Mackenzie, *Cell vertex methods for inviscid and viscous flows,* Computers Fluids, **22** (1993), 91–102.
7. J. A. Mackenzie and K. W. Morton, *Finite volume solutions of convection-diffusion test problems,* Mathematics of Computation, **60** (1992), 189–220. MR **93d:**76065
8. K. W. Morton and M. F. Paisley, *A finite volume scheme with shock fitting for the steady Euler equations,* Journal of Computational Physics, **80** (1989), 168–203.
9. K. W. Morton and M. Stynes, *An analysis of the cell vertex method,* Mathematical Modelling and Numerical Analysis, **28** (1994), 699-724. MR **95h:**65072
10. K. W. Morton and E. Süli, *Finite volume methods and their analysis,* IMA Journal of Numerical Analysis, **11** (1991), 241–260. MR **93e:**65145
11. R. H. Ni, *A multiple grid method for solving the Euler equations,* AIAA J. **20** (1982), 1565–1571.
12. L.A. Oganesian and L.A. Ruhovec, *Variational-difference methods for the solution of elliptic equations,* Publ. of the Armenian Academy of Sciences, Yerevan, 1979. (In Russian).

13. A. Preissmann, *Propagation des intumescences dans les canaux et rivieras,* Paper presented at the First Congress of the French Association for Computation, held at Grenoble, France, 1961.
14. H. G. Roos, M. Stynes and L. Tobiska, *Numerical Methods for Singularly Perturbed Differential Equations*, Springer Computational Mathematics, **24**, Springer-Verlag, 1996.
15. E. Süli, *Finite volume methods on distorted partitions: stability, accuracy, adaptivity,* Technical Report NA89/6, Oxford University Computing Laboratory, 1989.
16. E. Süli, *The accuracy of finite volume methods on distorted partitions,* Mathematics of Finite Elements and Applications VII (J.R. Whiteman, ed.) Academic Press, 1991, 253–260. MR **92i:**65171
17. E. Süli, *The accuracy of cell vertex finite volume methods on quadrilateral meshes,* Mathematics of Computation, **59** (1992), 359–382. MR **93a:**65158
18. H. A. Thomas, *Hydraulics of Flood Movements in Rivers,* Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1937.
19. B. Wendroff, *On centered difference equations for hyperbolic systems,* J. Soc. Indust. Appl. Math. **8** (1960), 549–555. MR **22:**7259

Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, United Kingdom
E-mail address: Bill.Morton@comlab.ox.ac.uk

Department of Mathematics, University College, Cork, Ireland
E-mail address: STMT8007@iruccvax.ucc.ie

Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, United Kingdom
E-mail address: Endre.Suli@comlab.ox.ac.uk