1999

Mixed Upwinding Covolume Methods on Rectangular Grids for Convection-diffusion Problems

So-Hsiang Chou
Bowling Green State University, chou@bgsu.edu

Do Y. Kwak

Panayot S. Vassilevski

SIAM J. SCI. COMPUT.
Vol. 21, No. 1, pp. 145–165

MIXED UPWINDING COVOLUME METHODS ON RECTANGULAR GRIDS FOR CONVECTION-DIFFUSION PROBLEMS*

SO-HSIANG CHOU†, DO Y. KWAK‡, AND PANAYOT S. VASSILEVSKI§

Abstract. We consider an upwinding covolume or control-volume method for a system of first order PDEs resulting from the mixed formulation of a convection-diffusion equation with a variable anisotropic diffusion tensor. The system can be used to model the steady state of the transport of a contaminant carried by a flow. We use the lowest order Raviart–Thomas space and show that the concentration and concentration flux both converge at one-half order provided that the exact flux is in $H^1(\Omega)^2$ and the exact concentration is in $H^1(\Omega)$. Some numerical experiments illustrating the error behavior of the scheme are provided.

Key words. stationary convection-diffusion problems, mixed finite elements, covolume methods, finite volume methods, upwind discretization, error estimates, preconditioning, M-matrices

AMS subject classifications. 65F10, 65N20, 65N30

PII. S1064827597321052

1. Introduction.

The purpose of this paper is to extend the previous results or techniques of covolume conservative schemes developed and analyzed for Stokes and diffusion equations by the authors (see Chou [7]; Chou and Kwak [9], [10]; and Chou, Kwak, and Vassilevski [11]) to convection-diffusion problems, including the practically interesting convection-dominated limit case. The ensuing new scheme combines the covolume methodology from the previous papers and the upwinding strategy to handle the convective term. The present paper deals with the derivation of the scheme, provides error analysis with limiting conditions on the smoothness of the solution, and illustrates the method with numerical experiments. The error analysis is provided economically in the sense that we could have done it the long and direct way (i.e., purely within the finite volume framework); instead, we adapt an existing error analysis in a general mixed finite element setting given in Liu, Wang, and Yan [20]. The bridge of the error analysis lies in the construction of a certain transfer operator (further denoted by γ_h) between the standard Raviart–Thomas (trial) spaces and the piecewise constant test spaces commonly used in the finite volume literature. The covolume scheme simultaneously treats, as in the mixed method, the vector quantity (flux) and the original scalar one (concentration), and provides $O(h^{\frac{1}{2}})$ order approximation to both of them in the L^2 norm. This seems a reasonable approximation for piecewise constant test functions. Also, as for the vector unknown, a one-half order accuracy (at worst) may be acceptable since it is a gradient of the scalar one. In the case of no convection a first order approximation is possible, and even a second

*Received by the editors May 5, 1997; accepted for publication (in revised form) October 8, 1998; published electronically August 26, 1999.
http://www.siam.org/journals/sisc/21-1/32105.html

†Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403-0221 (chou@zeus.bgsu.edu).

‡Department of Mathematics, Korea Advanced Institute of Science and Technology, Taejon, Korea 305-701 (dykwak@math.kaist.ac.kr). The work of this author was partially supported by KOSEF, Korea.

§Center of Informatics and Computing Technology, Bulgarian Academy of Sciences, Acad. G. Bontchev Street, Block 25 A, 1113 Sofia, Bulgaria (panayot@iscbg.acas.bg). The work of this author was partially supported by Bulgarian Ministry for Education, Science, and Technology grant I–95/504 and by Volkswagen Foundation grant I/70480.

order is possible if the mesh subdivision allows for superconvergence. If one departs from the finite volume (covolume) methodology, i.e., allowing higher-order piecewise polynomial spaces similar to the mixed method, then a higher order of convergence is also possible. Those topics, however, are not central to the present paper. The focus here is the extension of the covolume methodology to the convection-diffusion problems that works in a convection-dominated limit. Another topic that we are not concerned with is a local refinement of meshes, an important task we shall pursue in a later paper.

1.1. Problem formulation. Consider the convection-diffusion problem on an axiparallel rectangular (bounded) domain $\Omega \subset \mathbb{R}^2$,

$$\left\{ \begin{array}{rcl} -\nabla \cdot \mathcal{K}\nabla p + \nabla \cdot (\mathbf{b}p) + \alpha p &=& f \text{ in } \Omega, \\ \mathcal{K}\nabla p \cdot \mathbf{n} &=& 0 \text{ on } \Gamma_+ = \{x \in \partial\Omega : \ \mathbf{b}\cdot\mathbf{n} \geq 0\}, \\ p &=& 0 \text{ on } \Gamma_- = \{x \in \partial\Omega : \ \mathbf{b}\cdot\mathbf{n} < 0\}. \end{array} \right. \tag{1.1}$$

One can instead have only Dirichlet boundary condition, i.e., $p = 0$ on $\partial\Omega$. The latter boundary condition simplifies the exposition; hence in most of the technical part of the analysis, only the Dirichlet boundary condition will be assumed. Here $\mathcal{K} = \mathcal{K}(\mathbf{x}) = \ diag\ (\tau_1^{-1}(\mathbf{x}), \tau_2^{-1}(\mathbf{x}))$ is a positive definite diagonal matrix function whose entries are bounded from below and above by positive constants. Furthermore, we shall assume that τ_1, τ_2 are locally Lipschitz. The vector function $\mathbf{b}$ is in the Sobolev space $[W^{1,\infty}(\Omega)]^2$, $f \in L^2(\Omega)$ and $\alpha \in L^\infty(\Omega)$. We impose the following two conditions so that (1.1) is convection-dominated and uniquely solvable.

(H1) There exist two positive constants ϵ_1 and ϵ_2 such that the two constants are small and proportional:

$$\epsilon_2 \ll ||\mathbf{b}||_\infty, \qquad \epsilon_2/\epsilon_1 = \mathcal{O}(1), \tag{1.2}$$

with

$$\epsilon_1 \leq \tau_1^{-1},\ \tau_2^{-1} \leq \epsilon_2 \quad \text{in } \Omega. \tag{1.3}$$

(H2) There exists a positive constant γ_0 such that

$$\alpha + \frac{1}{2}\nabla \cdot \mathbf{b} \geq \gamma_0 \qquad \text{in } \Omega. \tag{1.4}$$

Problem (1.1) can be used to model the steady state of the transport of a contaminant in a porous or anisotropic medium flow. The variable p stands for the concentration of the contaminant, and $\mathbf{b}$ stands for the velocity field of the flow carrying it. In this context, the molecular diffusion tensor $\mathcal{K}$ (assumed to be diagonal for simplicity) has less effect on the physics than the convection term $\mathbf{b}\cdot\nabla p$, but the diffusion term contributes to the smoothness of the solution.

Let us introduce a new variable $\mathbf{u} = -\mathcal{K}\nabla p$ and write (1.1) as the system of first order partial differential equations (PDEs),

$$\left\{ \begin{array}{rcl} \mathcal{K}^{-1}\mathbf{u} &=& -\nabla p, \\ \mathrm{div}\mathbf{u} + \mathrm{div}(\mathbf{b}p) + \alpha p &=& f, \end{array} \right. \tag{1.5}$$

together with the boundary conditions,

$$\begin{array}{rcl} \mathbf{u}\cdot\mathbf{n} &=& 0 \text{ on } \Gamma_+ = \{x \in \partial\Omega : \ \mathbf{b}\cdot\mathbf{n} \geq 0\}, \\ p &=& 0 \text{ on } \Gamma_- = \{x \in \partial\Omega : \ \mathbf{b}\cdot\mathbf{n} < 0\}. \end{array}$$

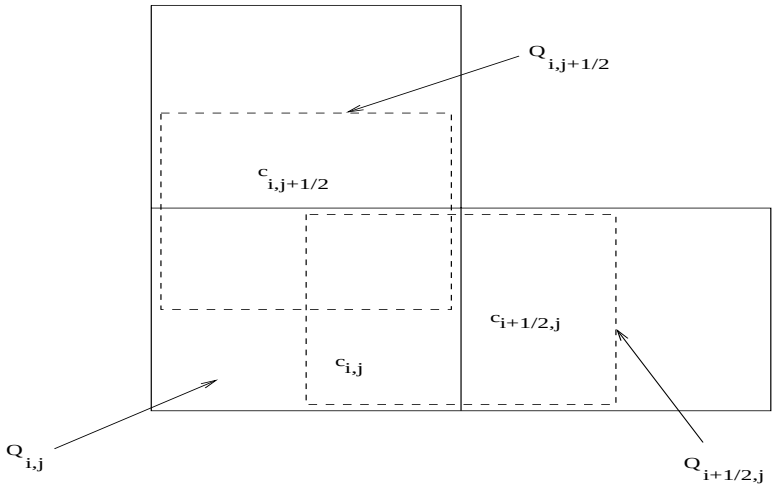

FIG. 1.1. *Primal and dual domains.*

For ease of reference, we shall also refer to the variable $\mathbf{u}$ as the flux or velocity variable and p as the concentration variable. The main purpose of this new system is twofold. In the context of the mixed method we can then approximate both variables to the same order. While in the context of finite volume or difference methods, we can derive conservative schemes from these first order equations because one of them represents a conservation law and the other a constitutive law.

We shall need three domain partitions for the approximation problem. More specifically, let the domain Ω be partitioned (cf. Figure 1) into a union of rectangles $Q_{i,j}$ with centers $c_{i,j}$. This is the primal partition which we shall call $\mathcal{R}_h$. The subindices $\{i+1,j\}, \{i-1,j\}$, $\{i,j+1\}$, and $\{i,j-1\}$ are assigned to the eastern, western, northern, and southern adjacent rectangles, respectively, if they exist. Given $Q_{i,j}$, the two midpoints of its vertical edges are denoted as $c_{i\pm 1/2,j}$, and the two midpoints of horizontal edges as $c_{i,j\pm 1/2}$. Let $c_{i,j} = (x_i, y_j)$ and $c_{i+1/2,j} = (x_{i+1/2}, y_j)$ etc., define

$$Q_{i+1/2,j} := [x_i, x_{i+1}] \times [y_{j-1/2}, y_{j+1/2}] \cap \bar{\Omega},$$
$$Q_{i,j+1/2} := [x_{i-1/2}, x_{i+1/2}] \times [y_j, y_{j+1}] \cap \bar{\Omega}$$

and

$$Q_{i,j} := [x_{i-1/2}, x_{i+1/2}] \times [y_{j-1/2}, y_{j+1/2}].$$

REMARK 1.1. *Note that at the boundary the $Q_{i+1/2,j}$ or $Q_{i,j+1/2}$ is half the size of a typical interior volume.*

Let

$$\mathbf{H}_0 := \mathbf{H}(\mathrm{div}; \Omega) \cap \{\mathbf{u} \cdot \mathbf{n} = 0 \text{ on } \Gamma_+\}, \tag{1.6}$$

where $\mathbf{H}(\mathrm{div};\Omega)$ is the space of all vector-valued functions $\mathbf{w} \in L^2(\Omega)^2$ such that $\mathrm{div}\,\mathbf{w} \in L^2(\Omega)$, and also define

$$\mathcal{L} := \{q \in L^2 : q|_Q \in H^1(Q) \qquad \forall\, Q \in \mathcal{R}_h\}. \tag{1.7}$$

The associated weak formulation of our first order system is as follows.
Find $(\mathbf{u}, p) \in \mathbf{H}_0 \times \mathcal{L}$ such that

$$\begin{array}{rcl} (\mathcal{K}^{-1}\mathbf{u}, \mathbf{v}) & = & (p, \mathrm{div}\,\mathbf{v}) \quad \forall\, \mathbf{v} \text{ in } \mathbf{H}_0, \\ (\mathrm{div}\mathbf{u}, q) + \widetilde{d}_h(p, q) & = & (f, q) \quad \forall\, q \text{ in } \mathcal{L}, \end{array} \tag{1.8}$$

where the bilinear form

$$\widetilde{d}_h(p, q) := \sum_{Q \in \mathcal{R}_h} \left[-\int_Q p\mathbf{b} \cdot \nabla q \, dx \, dy + \int_{\partial Q} \mathbf{b} \cdot \mathbf{n} pq \, d\sigma \right] + (\alpha p, q). \tag{1.9}$$

Note that the first two terms on the right are a modification of the expression $(\mathrm{div}(\mathbf{b}p), q)$, which does not make sense for nonsmooth p.

Define the trial space as the lowest order Raviart–Thomas space:

$$\begin{aligned} \mathbf{H}_h := \{&(u_h, v_h) \in \mathbf{H}_0 : u_h(x, y) = a + bx, \\ &v_h(x, y) = c + dy \text{ on } Q_{i,j}\} \end{aligned}$$

and

$$\mathcal{L}_h := \{q_h \in \mathcal{L} : q \text{ is constant over } Q_{ij}\}.$$

Then the natural (upwinding) mixed finite element method [20] corresponding to (1.8) deals with the primal grid only. Find $(\widetilde{\mathbf{u}}_h, \widetilde{p}_h) \in \mathbf{H}_h \times \mathcal{L}_h$ such that

$$\begin{array}{rcl} (\mathcal{K}^{-1}\widetilde{\mathbf{u}}_h, \mathbf{v}_h) - (\mathrm{div}\mathbf{v}_h, \widetilde{p}_h) & = & 0 \quad \forall\, \mathbf{v}_h \text{ in } \mathbf{H}_h, \\ (\mathrm{div}\widetilde{\mathbf{u}}_h, q_h) + d_h(\widetilde{p}_h, q_h) & = & (f, q_h) \quad \forall\, q_h \text{ in } \mathcal{L}_h, \end{array} \tag{1.10}$$

where $d_h(\widetilde{p}_h, q_h)$ is the discretization of the bilinear form $\widetilde{d}_h(p, q)$ of (1.9) involving the upwinding concept

$$\begin{aligned} d_h(p_h, q_h) = &\sum_{Q \in \mathcal{R}_h} \left[\int_{\partial Q} (\mathbf{b} \cdot \mathbf{n})_+ p_h^i q_h^i d\sigma + \int_{\partial Q} (\mathbf{b} \cdot \mathbf{n})_- p_h^o q_h^i d\sigma \right] \\ &+ (\alpha \widetilde{p}_h, q_h), \end{aligned} \tag{1.11}$$

where

$$(\mathbf{b} \cdot \mathbf{n})_+ := \max(\mathbf{b} \cdot \mathbf{n}, 0), \qquad (\mathbf{b} \cdot \mathbf{n})_- := \min(\mathbf{b} \cdot \mathbf{n}, 0),$$

and p_h^i is the trace of p_h on ∂Q taken from the interior of Q and p_h^o is that from the exterior of Q. Therefore, at the inflow boundary Γ_-, $p_h^o = 0$ and hence in (1.11) integration over edges of ∂Q that intersect Γ_- is not actually performed.

We will not use the finite element method (1.10) except in the error analysis.

1.2. The construction of the covolume scheme. Since there are many existing hydrodynamic codes in either finite difference or finite volume methods, we are motivated to design and analyze a finite volume mixed method using the upwinding concept. We will adapt a covolume methodology for the generalized Stokes problem [7] to approximate this system. The basic idea of creating a covolume method is to find a good combination of the finite volume method and the marker and cell (MAC) [17] placements of flow variables. (A balanced survey of the covolume method literature up to 1995 is given by Nicolaides, Porsching, and Hall [21].) For an analysis of the MAC scheme for Stokes problem (see Girault and Lopez [15]). In the MAC scheme, the concentration variable is assigned to the centers of the rectangular volumes and the normal components of the velocity or fluxes are assigned to the edges of the rectangular volumes. The normal approximate velocity is assumed to be constant along any edge. There are several ways [4, 8, 9, 16, 23] to exactly or nearly accomplish this; here we will use the lowest order Raviart–Thomas space $\mathbf{H}_h$ for the approximate velocity field [4, 10]. Note that within each $Q_{i,j}$ the horizontal component of the velocity is linear in x and constant in y, whereas the vertical component is linear in y and constant in x. Thus we have four degrees of freedom assigned at midpoints of edges. For example, on the eastern vertical edge of $Q_{i,j}$, we have only one unknown; the accompanying equation is taken by integrating the first component of the vector equation $(1.5)_1$ over $Q_{i+1/2,j}$. Similarly, to determine the unknown at the northern edge, we integrate $(1.5)_2$ over $Q_{i,j+1/2}$. In other words, if we write the velocity field as $\mathbf{u}_h = (u_h, v_h)$, then $Q_{i,j+1/2}$ is for the determination of v_h and $Q_{i+1/2,j}$ for u_h. We will sometimes call $Q_{i+1/2,j}$ ($Q_{i,j+1/2}$) as a u-volume (v-volume). These volumes are also called the covolumes of $Q_{i,j}$ in the literature.

Throughout this paper the primal partition $\{Q_{ij}\}$ is quasi-regular; i.e., there exists a positive constant C_1 independent of h such that

$$C_1 h^2 \leq \text{ area}\{Q_{i,j}\} \leq h^2 \quad \forall Q_{i,j} \in \mathcal{R}_h,$$

where $h := \max_{i,j}\{h^x_{i,j}, h^y_{i,j}\}$, $h^x_{i,j}, h^y_{i,j}$ are, respectively, the width and height of $Q_{i,j}$.

We now describe the present covolume method. Let $\mathbf{u}_h = (u_h, v_h) \in \mathbf{H}_h$ and $p_h \in \mathcal{L}_h$ be the approximate solution obtained as follows. Integrate the x-component of the first equation of (1.5) over the volume $Q_{i+1/2,j}$ to get

$$\int_{Q_{i+1/2,j}} \tau_1 u_h \, dxdy = -(p_{i+1,j} - p_{i,j}) \int_{y_{j-1/2}}^{y_{j+1/2}} dy, \tag{1.12}$$

where $p_{i,j} = p_h(x_i, y_j), p_{i+1,j} = p_h(x_{i+1}, y_j)$. Integrate the y-component of the first equation of (1.5) over the (control) volume $Q_{i,j+1/2}$ to get

$$\int_{Q_{i,j+1/2}} \tau_2 v_h dxdy = -(p_{i,j+1} - p_{i,j}) \int_{x_{i-1/2}}^{x_{i+1/2}} dx, \tag{1.13}$$

where $p_{i,j+1} = p_h(x_i, y_{j+1})$. Near the boundary of Ω one has to keep in mind that the covolumes are half of the interior covolumes (see Remark 1.1). That is, we then use the Dirichlet boundary value of p at the inflow boundary. Also, if we have boundary conditions of Neumann type (at the outflow boundary) we do not integrate over boundary covolumes; then the Neumann conditions are essential conditions for the flux variable.

Denote the four edges of the boundary of $Q_{i,j}$ as $e_{i+1/2,j}, e_{i,j+1/2}, e_{i-1/2,j}$, and $e_{i,j-1/2}$. Now integrate the second equation of (1.5) over the volume $Q_{i,j}$ and use the upwinding to get

$$
\begin{aligned}
(1.14)\qquad \int_{Q_{i,j}} & \nabla\cdot\mathbf{u}_h\,dxdy+\int_{Q_{i,j}}\alpha p\,dxdy\\
&+p_{i,j}\int_{e_{i+1/2,j}}(\mathbf{b}\cdot\mathbf{n})_+d\sigma+p_{i+1,j}\int_{e_{i+1/2,j}}(\mathbf{b}\cdot\mathbf{n})_-\,d\sigma\\
&+p_{i,j}\int_{e_{i-1/2,j}}(\mathbf{b}\cdot\mathbf{n})_+d\sigma+p_{i-1,j}\int_{e_{i-1/2,j}}(\mathbf{b}\cdot\mathbf{n})_-\,d\sigma\\
&+p_{i,j}\int_{e_{i,j+1/2}}(\mathbf{b}\cdot\mathbf{n})_+d\sigma+p_{i,j+1}\int_{e_{i,j+1/2}}(\mathbf{b}\cdot\mathbf{n})_-\,d\sigma\\
&+p_{i,j}\int_{e_{i,j-1/2}}(\mathbf{b}\cdot\mathbf{n})_+d\sigma+p_{i,j-1}\int_{e_{i,j-1/2}}(\mathbf{b}\cdot\mathbf{n})_-\,d\sigma\\
&=\int_{Q_{i,j}}f\,dxdy.
\end{aligned}
$$

Temporarily, we shall assume all integrals are evaluated exactly. See the last section for final discretization of them using quadratures.

Equations (1.12)–(1.14) are quite intuitive from a physical point of view, but for the error analysis it is more convenient if we state them in terms of bilinear forms. Define the *test* space

$$
\begin{aligned}
(1.15)\qquad \mathbf{Y}_h := \{\mathbf{w}=(u_h,v_h) &: u_h \text{ piecewise constant on } u\text{-volumes},\\
&v_h \text{ piecewise constant on } v\text{-volumes},\\
&\mathbf{w}\cdot\mathbf{n}=0 \quad \text{on covolumes near the outflow boundary } \Gamma_+\}
\end{aligned}
$$

which is used to pick out the control volumes. Here, $\mathbf{n}$ is a normal vector to Γ_+. Furthermore, define the following bilinear forms:

$$
(1.16)\qquad a(\mathbf{u}_h,\mathbf{v}_h)=\int_\Omega \mathcal{K}^{-1}\mathbf{u}_h\cdot\mathbf{v}_h dx,\quad \mathbf{u}_h\in\mathbf{H}_h,\mathbf{v}_h\in\mathbf{Y}_h;
$$

$$
\begin{aligned}
(1.17)\qquad b(\mathbf{v}_h,p_h) := &\textstyle\sum(v_h^1(c_{i+1/2,j}),0)^t\cdot\int_{\partial Q_{i+1/2,j}}p_h\mathbf{n}d\sigma\\
&\textstyle+\sum(0,v_h^2(c_{i,j+1/2}))^t\cdot\int_{\partial Q_{i,j+1/2}}p_h\mathbf{n}d\sigma,\quad \mathbf{v}_h\in\mathbf{Y}_h,\ p_h\in\mathcal{L}_h;
\end{aligned}
$$

$$
(1.18)\qquad c(\mathbf{u}_h,q_h):=\sum q_h(c_{i,j})\int_{Q_{i,j}}\mathrm{div}\mathbf{u}_h dx,\quad \mathbf{u}_h\in\mathbf{H}_h,\ q_h\in\mathcal{L}_h;
$$

$$
\begin{aligned}
(1.19)\qquad d_h(p_h,q_h) \;:=\;& \textstyle\sum_{Q_{ij}\in\mathcal{R}_h}\Big[\int_{\partial Q_{ij}}(\mathbf{b}\cdot\mathbf{n})_+p_h^iq_h^id\sigma+\int_{\partial Q_{ij}}(\mathbf{b}\cdot\mathbf{n})_-p_h^oq_h^id\sigma\Big]\\
&+(\alpha p_h,q_h),\quad p_h,\ q_h\in\mathcal{L}_h,
\end{aligned}
$$

where p_h^i is the trace of p_h on ∂Q_{ij} taken from the interior of Q_{ij} and p_h^o is that from the exterior of Q_{ij}. Note that there is no integration over edges of $Q_{i,j}$ near the inflow boundary (or we may formally let $p_h^o=0$) and at the outflow boundary $(\mathbf{b}\cdot\mathbf{n})_-=0$, hence we do not have to specify p_h^o there.

Define the *transfer* operator $\gamma_h : \mathbf{H}_h \to \mathbf{Y}_h$ connecting the trial space to the test function space in the following equation:

$$\begin{aligned}\gamma_h \mathbf{w}_h &:= (\gamma_h u_h, \gamma_h v_h) \qquad \mathbf{w}_h = (u_h, v_h)\\ &:= \left(\sum u_h(c_{i+1/2,j})\chi_{i+1/2,j}, \sum v_h(c_{i,j+1/2})\chi_{i,j+1/2}\right),\end{aligned}$$

where $\chi_{i+1/2,j}$ and $\chi_{i,j+1/2}$ are the characteristic function of $Q_{i+1/2,j}$ and $Q_{i,j+1/2}$, respectively. Note that we used the same notation γ_h in the component-wise definition and that γ_h is one-to-one and onto.

1.3. Relation to the mixed finite element discretization. The covolume method (1.12)–(1.14) is equivalent to the problem of finding $\{\mathbf{u}_h, p_h\} \in \mathbf{H}_h \times \mathcal{L}_h$ such that

$$\begin{aligned} a(\mathbf{u}_h, \gamma_h \mathbf{v}_h) + b(\gamma_h \mathbf{v}_h, p_h) &= 0 \quad \forall\, \mathbf{v}_h \text{ in } \mathbf{H}_h,\\ c(\mathbf{u}_h, q_h) + d_h(p_h, q_h) &= (f, q_h) \quad \forall q_h \text{ in } \mathcal{L}_h. \end{aligned} \tag{1.20}$$

Here the substitution of $\gamma_h \mathbf{v}_h$ for a test function $w_h \in \mathbf{Y}_h$ is due to the surjectivity of the operator γ_h. This simple observation turns the original Petrov–Galerkin statement into a standard Galerkin one.

We can reformulate (1.20) into a saddle-point problem by further introducing

$$A(\mathbf{u}, \mathbf{v}) := a(\mathbf{u}, \gamma_h \mathbf{v}) = (\mathcal{K}^{-1}\mathbf{u}, \gamma_h \mathbf{v}), \quad \mathbf{u}, \mathbf{v} \in \mathbf{H}_h,$$

$$B(\mathbf{w}_h, q_h) := b(\gamma_h \mathbf{w}_h, q_h), \quad \mathbf{w}_h \in \mathbf{H}_h,\ q_h \in \mathcal{L}_h$$

and noting that $B = -c$ (cf. Lemma 2.3) so that problem (1.20) becomes

$$\begin{aligned} A(\mathbf{u}_h, \mathbf{v}_h) + B(\mathbf{v}_h, p_h) &= 0 \quad \forall\, \mathbf{v}_h \text{ in } \mathbf{H}_h,\\ B(\mathbf{u}_h, q_h) - d_h(p_h, q_h) &= -(f, q_h) \quad \forall q_h \text{ in } \mathcal{L}_h. \end{aligned} \tag{1.21}$$

It is interesting to note that based on the transfer operator γ_h, the forms A and B are bilinear, and hence the above system is in standard form. Nevertheless, the standard (conforming in $\mathbf{H}(\mathrm{div}; \Omega)$) mixed method analysis cannot be used here. This is so because the original PDE cannot be put into the same form: the transfer operator γ_h in the definition of the bilinear form A cannot be extended to the space $\mathbf{H}(\mathrm{div}; \Omega)$. However, upon closer examination we see that the standard mixed method (1.10) for the convection-dominated problem (1.5) differs from the mixed covolume method (1.21) only in the bilinear form A. Thus we can treat the covolume method as one resulting from a "variational crime" of the standard mixed method. A careful analysis of the transfer operator γ_h in connection to this deviation then leads to our error estimate in Theorem 3.1 which demonstrates the one-half order error estimate in the flux variable, as well as in the concentration variable, under the minimal regularity assumption that $\mathbf{u} \in H^1(\Omega)^2$ and $p \in H^1(\Omega)$. The starting point of the proof is a good error equation (3.10) that plays the role of Cea's lemma in the standard finite element analysis. This methodology was initiated in Chou [7] for the generalized Stokes problem on triangular grids, in Chou and Kwak [8, 9] for the same problem on rectangular grids, and in Chou and Li [12] for the "point-centered" or vertex-centered schemes for the variable-coefficient Poisson equation. The diffusion problem (i.e., $\mathbf{b} = 0$ and $\alpha = 0$ in (1.5)) was treated in Chou and Kwak [10] by a mixed covolume method.

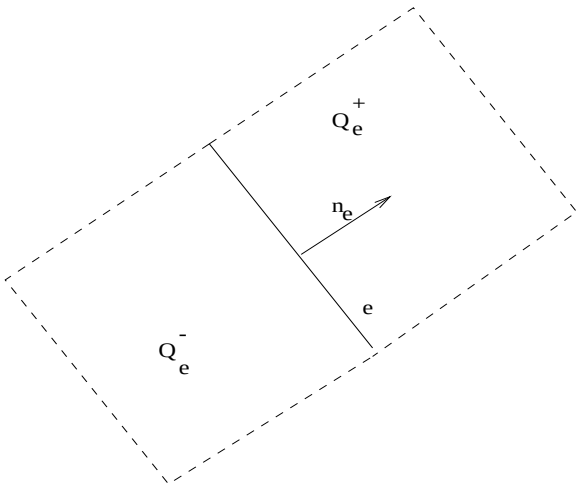

FIG. 2.1. *Neighboring primal volumes Q_e^+ and Q_e^- associated with a pair of edge and a normal vector,* $(e,\ \mathbf{n}_e)$.

1.4. Summary of results in the remaining part of the paper. We summarize now the results of the present paper. In section 2, we study the properties of our discretization scheme. In particular, we show that the saddle-point formulation (1.20) is uniquely solvable for sufficiently small h. The constraint on h is determined by the local Lipschitz constants of the coefficient matrix $\mathcal{K}$. In section 3 we modify the analysis from Liu, Wang, and Yan [20] in the present context to derive our main error estimates. Finally, in section 4 we present extensive numerical experiments that illustrate the error behavior of our discrete solutions. We comment that for the particular case of diagonal diffusion coefficient $\mathcal{K}$, the saddle-point problem can be reduced (implicitly) to a problem only for the concentration variable p_h, which can be solved by a preconditioned iterative method. In fact, due to the upwinding, the reduced problem can be well approximated by an M-matrix, for which good approximate ILU factorization preconditioners are available. Details are found in the last section.

2. Properties of the saddle point formulation. In this section we prove some preliminary lemmas. Let $||\cdot||_j, j = 0, 1$ denote the usual L^2 and H^1 norms, respectively, and we also use $||\cdot||$ for the L^2 norm when there is no confusion. The symbol C will be used as a generic positive constant independent of h and the ϵ_1 of (1.3), and may have different values at different places.

The following two lemmas are slightly more general than the original ones in Liu, Wang, and Yan [20]. To clearly indicate the influence of the boundary conditions we include our proof. Also, the two lemmas below hold in a general setting (2-D and 3-D rectangular or triangular elements; in 3-D, edges should be accordingly replaced by faces of volumes). That is why we have purposely rotated the elements in Figure 2.1.

We adopt now the following notation; for each edge e of a primal volume Q we assign a unit normal vector $\mathbf{n}_e$. Then, given the pair $(e,\ \mathbf{n}_e)$ with e an interior edge, one can uniquely define the neighboring primal volumes Q_e^+ and Q_e^- with the common edge e so that $\mathbf{n}_e$ points towards Q_e^+, see, e.g., Figure 2.1. A vector $\mathbf{n}$ (without subscript) will be considered "outward" to a underlying domain (or volume).

LEMMA 2.1. *Let $\mathcal{E}_0$ be the collection of interior edges of elements in $\mathcal{R}_h$. Then*

the bilinear form $d_h(\cdot,\cdot)$ of (1.19) *can be rewritten as follows:*

$$
\begin{aligned}
d_h(p,q) \;=\;& \frac{1}{2}\sum_{e\in\mathcal{E}}\int_e [p]\,[q]\,|\mathbf{b}\cdot\mathbf{n}|\,d\sigma \\
&+\frac{1}{2}\sum_{e\in\mathcal{E}_0}\int_e \mathbf{b}\cdot\mathbf{n}_e(p_{Q_e^+}+p_{Q_e^-})(q_{Q_e^-}-q_{Q_e^+})\,d\sigma \\
&+(\alpha p,\ q)+\int_{\Gamma_+}\mathbf{b}\cdot\mathbf{n}p^iq^i\,d\sigma,
\end{aligned}
\tag{2.1}
$$

where for each interior edge e, i.e., $e\in\mathcal{E}_0$, $[p]$ denotes the jump of the discontinuous function p.

Proof. For completeness we provide a proof of the lemma. We start from the definition of $d_h(p,q)$; see (1.19). The idea is to rewrite the corresponding sums over the edges $\mathcal{E}$. For a given interior edge e and a pre-assigned unit normal vector $\mathbf{n}_e$, one has two neighboring primal volumes Q_e^+ and Q_e^- that contribute in the rearranged sum. Also, Q_e^+ will contribute with the vector $\mathbf{n}\equiv-\mathbf{n}_e$ and Q_e^- will contribute with the vector $\mathbf{n}=\mathbf{n}_e$. Hence, denoting by $\mathcal{E}_0$ the set of interior edges, one gets

$$
\begin{aligned}
d_h(p,q) \;&=\sum_{Q\in\mathcal{R}_h}\int_{\partial Q}\left[(\mathbf{b}\cdot\mathbf{n})_+p^iq^i+(\mathbf{b}\cdot\mathbf{n})_-p^oq^i\right]\,d\sigma+(\alpha p,\ q) \\
&=\sum_{e\in\mathcal{E}_0}\int_e\Big[(\mathbf{b}\cdot\mathbf{n}_e)_+p_{Q_e^-}q_{Q_e^-}+(\mathbf{b}\cdot\mathbf{n}_e)_-p_{Q_e^+}q_{Q_e^-} \\
&\qquad+(\mathbf{b}\cdot-\mathbf{n}_e)_+p_{Q_e^+}q_{Q_e^+}+(\mathbf{b}\cdot-\mathbf{n}_e)_-p_{Q_e^-}q_{Q_e^+}\Big] \\
&\qquad+(\alpha p,\ q)+\int_{\Gamma_+}\mathbf{b}\cdot\mathbf{n}p^iq^i\,d\sigma.
\end{aligned}
$$

Using now the relations $(\mathbf{b}\cdot-\mathbf{n}_e)_+=-(\mathbf{b}\cdot\mathbf{n}_e)_-$ and $(\mathbf{b}\cdot-\mathbf{n}_e)_-=-(\mathbf{b}\cdot\mathbf{n}_e)_+$ and the abbreviations $w_{Q_e^-}=p_l, w_{Q_e^+}=w_r$ with $w=p,q$, one arrives at

$$
\begin{aligned}
d_h(p,q)- \quad&(\alpha p,\ q)-\int_{\Gamma_+}\mathbf{b}\cdot\mathbf{n}p^iq^i\,d\sigma \\
&=\sum_{e\in\mathcal{E}_0}\int_e\left[(\mathbf{b}\cdot\mathbf{n}_e)_+p_l(q_l-q_r)+(\mathbf{b}\cdot\mathbf{n}_e)_-p_r(q_l-q_r)\right]\,d\sigma \\
&=\sum_{e\in\mathcal{E}_0}\int_e\left[\left((\mathbf{b}\cdot\mathbf{n}_e)_+-(\mathbf{b}\cdot\mathbf{n}_e)_-\right)p_l(q_l-q_r)\right. \\
&\qquad\left.+(\mathbf{b}\cdot\mathbf{n}_e)_-(p_r+p_l)(q_l-q_r)\right]\,d\sigma \\
&=\sum_{e\in\mathcal{E}_0}\int_e\left[|\mathbf{b}\cdot\mathbf{n}_e|(p_l-p_r)(q_l-q_r)+|\mathbf{b}\cdot\mathbf{n}_e|p_r(q_l-q_r)\right. \\
&\qquad\left.+(\mathbf{b}\cdot\mathbf{n}_e)_-(p_r+p_l)(q_l-q_r)\right]\,d\sigma \\
&=\sum_{e\in\mathcal{E}_0}\int_e\left[\tfrac{1}{2}|\mathbf{b}\cdot\mathbf{n}_e|(p_l-p_r)(q_l-q_r)\right. \\
&\qquad\left.+(q_l-q_r)\left(\tfrac{1}{2}|\mathbf{b}\cdot\mathbf{n}_e|(p_l-p_r)+|\mathbf{b}\cdot\mathbf{n}_e|p_r+(\mathbf{b}\cdot\mathbf{n}_e)_-(p_r+p_l)\right)\right]\,d\sigma \\
&=\sum_{e\in\mathcal{E}_0}\int_e\left[\tfrac{1}{2}|\mathbf{b}\cdot\mathbf{n}_e|(p_l-p_r)(q_l-q_r)\right. \\
&\qquad\left.+(q_l-q_r)\left(\tfrac{1}{2}|\mathbf{b}\cdot\mathbf{n}_e|(p_l+p_r)+(\mathbf{b}\cdot\mathbf{n}_e)_-(p_r+p_l)\right)\right]\,d\sigma \\
&=\sum_{e\in\mathcal{E}_0}\int_e\left[\tfrac{1}{2}|\mathbf{b}\cdot\mathbf{n}_e|(p_l-p_r)(q_l-q_r)+\mathbf{b}\cdot\mathbf{n}_e\tfrac{1}{2}(p_r+p_l)(q_l-q_r)\right]\,d\sigma. \quad \square
\end{aligned}
$$

Substituting $p = q$ in the formula (2.1), noticing that $-\mathbf{n}_e$ is an outward normal to Q_e^+ and $\mathbf{n}_e$ is an outward normal to Q_e^-, and that $\int_Q \operatorname{div}(\mathbf{b}p^2)\, dx\, dy = \int_{\partial Q} \mathbf{b}\cdot\mathbf{n}p^2\, d\sigma$ (since $p_Q = \text{const}$), one easily gets the next result.

LEMMA 2.2. *For any $q \in \mathcal{L}_h$ the bilinear form of* (1.19) *satisfies*

$$d_h(q,q) = \left(\left(\alpha + \frac{1}{2}\nabla\cdot\mathbf{b}\right)q, q\right) + \frac{1}{2}|||q|||^2, \tag{2.2}$$

where

$$|||q|||^2 = \sum_{e\in\mathcal{E}_0}\int_e [q]^2 |\,\mathbf{b}\cdot\mathbf{n}|\, d\sigma. \tag{2.3}$$

The next lemma can be proved by direct evaluation [13].

LEMMA 2.3. *The following holds:*

$$B(\mathbf{w}_h, q_h) = b(\gamma_h\mathbf{w}_h, q_h) = -c(\mathbf{w}_h, q_h) \qquad \forall\, \mathbf{w}_h \in \mathbf{H}_h, q_h \in \mathcal{L}_h.$$

Now by Lemma 2.3 problem (1.20) becomes

$$A(\mathbf{u}_h, \mathbf{w}_h) + B(\mathbf{w}_h, p_h) \;=\; 0 \quad \forall\, \mathbf{w}_h \text{ in } \mathbf{H}_h, \tag{2.4}$$

$$B(\mathbf{u}_h, q_h) - d_h(p_h, q_h) \;=\; -(f, q_h) \quad \forall\, q_h \text{ in } \mathcal{L}_h. \tag{2.5}$$

The fact that γ_h is a bounded self-adjoint operator with respect to the L^2 inner product can be found in [13].

LEMMA 2.4. *The following relations hold:*

$$(\gamma_h\mathbf{u}_h, \mathbf{w}_h) = (\mathbf{u}_h, \gamma_h\mathbf{w}_h) \quad \forall\, \mathbf{u}_h, \mathbf{w}_h \in \mathbf{H}_h. \tag{2.6}$$

There also exists a positive constant C independent of h such that

$$||\gamma_h\mathbf{u}_h||_0 \le C||\mathbf{u}_h||_0 \quad \forall\, \mathbf{u}_h \in \mathbf{H}_h. \tag{2.7}$$

We next show that $A(\mathbf{w}_h, \mathbf{w}_h)$ is coercive for sufficiently small h.

LEMMA 2.5. *There exists a constant h_0 such that for all $h \le h_0$*

$$A(\mathbf{w}_h, \mathbf{w}_h) \ge C\epsilon_1^{-1}\|\mathbf{w}_h\|_0^2 \qquad \mathbf{w}_h \in \mathbf{H}_h,$$

where the constant C is independent of h and ϵ_1.

Proof. Write $\mathbf{w}_h = (u_h, v_h) \in \mathbf{H}_h$. Then

$$a(\mathbf{w}_h, \gamma\mathbf{w}_h) = \sum u_h(c_{i+1/2,j})\int_{Q_{i+1/2,j}} \tau_1(x,y)u_h(x,y)dxdy \tag{2.8}$$

$$+\sum v_h(c_{i,j+1/2})\int_{Q_{i,j+1/2}} \tau_2(x,y)v_h(x,y)dxdy \tag{2.9}$$

$$= I + II. \tag{2.10}$$

It suffices to show that $I \ge C\epsilon_1^{-1}\|u_h\|_0^2$. Let $Q_{ij}^- := Q_{i-1/2,j} \cap Q_{ij}$ and $Q_{ij}^+ := Q_{i+1/2,j} \cap Q_{ij}$. Then

$$\begin{aligned} I &= \sum u_h(c_{i-1/2,j})\int_{Q_{ij}^-} \tau_1(x,y)u_h(x,y)dxdy \\ &\quad + \sum u_h(c_{i+1/2,j})\int_{Q_{ij}^+} \tau_1(x,y)u_h(x,y)dxdy \\ &= III + IV, \end{aligned}$$

where

$$\begin{aligned} III &= \sum u_h(c_{i-1/2,j}) \int_{Q_{ij}^-} (\tau_1(x,y) - \tau_1(c_{ij})) u_h(x,y) dxdy \\ &\quad + \sum u_h(c_{i-1/2,j}) \tau_1(c_{ij}) \int_{Q_{ij}^-} u_h(x,y) dxdy \\ &= V + VI \\ IV &= \sum u_h(c_{i+1/2,j}) \int_{Q_{ij}^+} (\tau_1(x,y) - \tau_1(c_{ij})) u_h(x,y) dxdy \\ &\quad + \sum u_h(c_{i+1/2,j}) \tau_1(c_{ij}) \int_{Q_{ij}^+} u_h(x,y) dxdy \\ &= VII + VIII. \end{aligned}$$

Using the linearity of u_h in x and constant in y, we can easily derive by direct computation that

$$VI + VIII \geq C\epsilon_1^{-1} \|u_h\|_0^2,$$

while by Lipschitz continuity of τ_1 and the Simpson's rule that

$$\begin{aligned} V + VII &\leq M\epsilon_1^{-1} h \left(\sum \int_{Q_{ij}^-} |u_h(c_{i-1/2,j}) u_h(x,y)| dxdy \right. \\ &\quad \left. + \int_{Q_{ij}^+} |u_h(c_{i+1/2,j}) u_h(x,y)| dxdy \right) \\ &= M\epsilon_1^{-1} h \left(\sum \|u_h(c_{i-1/2,j})\|_{Q_{ij}^-} \|u_h\|_{Q_{ij}^-} + \|u_h(c_{i+1/2,j})\|_{Q_{ij}^+} \|u_h\|_{Q_{ij}^+} \right) \\ &\leq 6M\epsilon_1^{-1} h \sum \|u_h\|_{Q_{ij}}^2 \\ &= 6M\epsilon_1^{-1} h \|u_h\|_0^2. \end{aligned}$$

Thus we have

$$I \geq C\epsilon_1^{-1} \|u_h\|_0^2 - C\epsilon_1^{-1} h \|u_h\|_0^2$$

and so

$$I \geq C\epsilon_1^{-1} ||u_h||_0^2$$

for h sufficiently small. □

Now note that the bilinear form a of (1.16) is also well defined over $\mathbf{H}_h \times \mathbf{H}_h$. With this in mind we state the following approximation property of γ_h, whose simple proof can be found in [10] or directly verified.

There exists a constant C independent of h and ϵ_1 such that

$$\|(I - \gamma_h)\mathbf{w}_h\|_0 \leq Ch ||\mathbf{w}_h||_{1,h}, \tag{2.11}$$

where

$$||\mathbf{w}_h||_{1,h}^2 = ||\mathbf{w}_h||_0^2 + |\mathbf{w}_h|_{1,h}^2$$

and the discrete seminorm $\mathbf{w}_h = (w_h, x_h) \in \mathbf{H}_h$ is defined as

$$|\mathbf{w}_h|^2_{1,h} := \sum_Q ||\nabla w_h||^2_{0,Q} + ||\nabla x_h||^2_{0,Q}.$$

LEMMA 2.6. There exists a constant C independent of h and ϵ_1 such that

$$a(\mathbf{u}, (I - \gamma_h)\mathbf{w}_h) \le C\epsilon_1^{-1} h ||\mathbf{u}||_1 ||\mathbf{w}_h||_0 \quad \forall\, \mathbf{w}_h \in \mathbf{H}_h \tag{2.12}$$

and $\forall\, \mathbf{u} \in \mathbf{H}^1$.

Proof. Let $\mathcal{E}_h$ be the familiar interpolation operator from $\mathbf{H}^1(\Omega)$ to $\mathbf{H}_h$ with $\int_e \mathbf{q} \cdot \mathbf{n} de$, flux across edge, as its degrees of freedom ([22], pp. 550–554). Then

$$||\mathbf{q} - \mathcal{E}_h \mathbf{q}||_0 \le Ch|\mathbf{q}|_1 \quad \forall\, \mathbf{q} \in \mathbf{H}^1(\Omega). \tag{2.13}$$

Denoting by $\mathcal{K}_0$ the piecewise constant average of $\mathcal{K}$, one has

$$\begin{aligned} a(\mathbf{u}, (I - \gamma_h)\mathbf{w}_h) &= ((\mathcal{K}^{-1} - \mathcal{K}_0^{-1})\mathbf{u}, (I - \gamma_h)\mathbf{w}_h) + (\mathcal{K}_0^{-1}\mathbf{u}, (I - \gamma_h)\mathbf{w}_h) \\ &\le C\epsilon_1^{-1} h \|\mathbf{u}\|_0 \|\mathbf{w}_h\|_0 + (\mathbf{u}, \mathcal{K}_0^{-1}(I - \gamma_h)\mathbf{w}_h) \\ &= C\epsilon_1^{-1} h \|\mathbf{u}\|_0 \|\mathbf{w}_h\|_0 + (\mathbf{u}, (I - \gamma_h)\mathcal{K}_0^{-1}\mathbf{w}_h) \\ &= C\epsilon_1^{-1} h \|\mathbf{u}\|_0 \|\mathbf{w}_h\|_0 + (\mathbf{u} - \mathcal{E}_h\mathbf{u}, (I - \gamma_h)\mathcal{K}_0^{-1}\mathbf{w}_h) \\ &\quad + ((I - \gamma_h)\mathcal{E}_h\mathbf{u}, \mathcal{K}_0^{-1}\mathbf{w}_h), \end{aligned}$$

where we have used the symmetry and boundedness of γ_h. Also, we used the fact that the coefficient $\mathcal{K}^{-1}$ is locally Lipschitz. Now the second term on the right side of the last equation can be bounded by $C\epsilon_1^{-1} h \|\mathbf{u}\|_1 \|\mathbf{w}_h\|_0$ using (2.13) and boundedness of $I - \gamma_h$, and the third term can be bounded using (2.11) and the fact that

$$||\mathcal{E}_h \mathbf{u}||_{1,h} \le C||\mathbf{u}||_1.$$

This completes the proof. ∎

We next show that our covolume method (1.21) has a unique solution.

LEMMA 2.7. For h sufficiently small, there is a unique $(\mathbf{u}_h, p_h) \in \mathbf{H}_h \times \mathcal{L}_h$ for the system:

$$\begin{aligned} A(\mathbf{u}_h, \mathbf{w}_h) + B(\mathbf{w}_h, p_h) &= 0 \quad \forall\, \mathbf{w}_h \text{ in } \mathbf{H}_h, \\ B(\mathbf{u}_h, q_h) - d_h(p_h, q_h) &= -(f, q_h) \quad \forall q_h \text{ in } \mathcal{L}_h. \end{aligned}$$

Proof. Define the bilinear form on $\mathbf{H}_h \times \mathcal{L}_h$:

$$\mathcal{A}(\mathbf{z_h}, s; \mathbf{w_h}, t) := A(\mathbf{z_h}, \mathbf{w_h}) + B(\mathbf{w_h}, s) - B(\mathbf{z_h}, t) + d_h(s, t).$$

Obviously the above system is equivalent to

$$\mathcal{A}(\mathbf{u}_h, p_h; \mathbf{w_h}, q_h) = \phi(\mathbf{w_h}, q_h) \qquad \forall (\mathbf{w_h}, q_h) \in \mathbf{H}_h \times \mathcal{L}_h,$$

where

$$\phi(\mathbf{w_h}, q_h) := (f, q_h)$$

is a linear functional on $\mathbf{H}_h \times \mathcal{L}_h$. By Lemma 2.2

$$\mathcal{A}(\mathbf{w_h}, q_h; \mathbf{w_h}, q_h) = a(\mathbf{w_h}, \gamma_h \mathbf{w_h}) + \Big(\Big(\alpha + \frac{1}{2}\nabla \cdot \mathbf{b}\Big) q_h, q_h\Big) + \frac{1}{2}|||q_h|||^2. \tag{2.14}$$

It suffices to show that $\mathcal{A}(\mathbf{w_h}, q_h; \mathbf{w_h}, q_h) = 0$ admits only a zero solution, which can be inferred by the coercivity of $\mathcal{A}(\mathbf{w_h}, q_h; \mathbf{w_h}, q_h)$ implied by Lemma 2.5 and (1.4). ∎

3. Error estimates. We now prove our main convergence result.

THEOREM 3.1. *Let the rectangular partition family $\{Q_{ij}\}$ of the domain Ω be quasi-regular, and let $\{\mathbf{u}_h, p_h\}$ be the solution of the problem* (1.21) *and $\{\mathbf{u}, p\}$ of the problem* (1.8). *Then there exists a positive constant C independent of h and ϵ_1,*

$$\|\mathbf{u} - \mathbf{u}_h\|_0 + \epsilon_1^{1/2}\|p - p_h\|_0 \le Ch(\|\mathbf{u}\|_1 + h^{-1/2}\epsilon_1^{1/2}||p||_1), \tag{3.1}$$

provided that $\mathbf{u} \in \mathbf{H}^1$ and $p \in H^1$.

Proof. We will use, as a bridge of error analysis, the mixed method (1.10). Find $(\widetilde{\mathbf{u}}_h, \widetilde{p}_h) \in \mathbf{H}_h \times \mathcal{L}_h$ such that

$$a(\widetilde{\mathbf{u}}_h, \mathbf{w}_h) + B(\mathbf{w}_h, \widetilde{p}_h) = 0 \quad \forall\, \mathbf{w}_h \text{ in } \mathbf{H}_h, \tag{3.2}$$

$$B(\widetilde{\mathbf{u}}_h, q_h) - d_h(\widetilde{p}_h, q_h) = -(f, q_h) \quad \forall q_h \text{ in } \mathcal{L}_h. \tag{3.3}$$

Let

$$\widetilde{\mathcal{A}}(\mathbf{z_h}, s; \mathbf{w_h}, t) := a(\mathbf{z_h}, \mathbf{w_h}) + B(\mathbf{w_h}, s) - B(\mathbf{z_h}, t) + d_h(s, t) \tag{3.4}$$

be a bilinear form on $\mathbf{H}_h \times \mathcal{L}_h$. Then (3.2)–(3.3) is equivalent to the problem of finding $(\widetilde{\mathbf{u}}_h, \widetilde{p}_h) \in \mathbf{H}_h \times \mathcal{L}_h$ such that

$$\widetilde{\mathcal{A}}(\widetilde{\mathbf{u}}_h, \widetilde{p}_h; \mathbf{w_h}, q_h) = \phi(\mathbf{w_h}, q_h) \qquad \forall (\mathbf{w_h}, q_h) \in \mathbf{H}_h \times \mathcal{L}_h, \tag{3.5}$$

where

$$\phi(\mathbf{w_h}, q_h) := (f, q_h)$$

is a linear functional on $\mathbf{H}_h \times \mathcal{L}_h$.

This system has the following known convergence result (unscaled version of Eq. 5.3 in [20]):

$$||\mathbf{u} - \widetilde{\mathbf{u}}_h||_0 + \epsilon_1^{1/2}||p - \widetilde{p}_h||_0 \le Ch(||\mathbf{u}||_1 + \epsilon_1^{1/2}h^{-1/2}||p||_1), \tag{3.6}$$

provided that $\mathbf{u} \in \mathbf{H}^1, p \in H^1$. (Note that the proof of this estimate depends on Lemma 2.2, which we have proved even for the mixed boundary condition case not covered by [20].)

On the other hand, our covolume method is equivalent to the problem of finding $(\mathbf{u}_h, p_h) \in \mathbf{H}_h \times \mathcal{L}_h$ such that

$$\mathcal{A}(\mathbf{u}_h, p_h; \mathbf{w_h}, q_h) = \phi(\mathbf{w_h}, q_h) \qquad \forall (\mathbf{w_h}, q_h) \in \mathbf{H}_h \times \mathcal{L}_h, \tag{3.7}$$

where

$$\mathcal{A}(\mathbf{z_h}, s; \mathbf{w_h}, t) := A(\mathbf{z_h}, \mathbf{w_h}) + B(\mathbf{w_h}, s) - B(\mathbf{z_h}, t) + d_h(s, t). \tag{3.8}$$

Using the bilinearity of A and B, (3.7), (3.5) , we have

$$\begin{aligned}\mathcal{A}(\widetilde{\mathbf{u}}_h - \mathbf{u}_h, \widetilde{p}_h - p_h; \mathbf{w_h}, q_h) &= \mathcal{A}(\widetilde{\mathbf{u}}_h, \widetilde{p}_h; \mathbf{w_h}, q_h) - \mathcal{A}(\mathbf{u}_h, p_h; \mathbf{w_h}, q_h) \\ &= \mathcal{A}(\widetilde{\mathbf{u}}_h, \widetilde{p}_h; \mathbf{w_h}, q_h) - \widetilde{\mathcal{A}}(\widetilde{\mathbf{u}}_h, \widetilde{p}_h; \mathbf{w_h}, q_h).\end{aligned}$$

Hence by (3.8) and (3.4) we have

$$\mathcal{A}(\widetilde{\mathbf{u}}_h - \mathbf{u}_h, \widetilde{p}_h - p_h; \mathbf{w_h}, q_h) = a(\widetilde{\mathbf{u}}_h, \gamma_h \mathbf{w}_h) - a(\widetilde{\mathbf{u}}_h, \mathbf{w}_h). \tag{3.9}$$

Since the total error $\mathbf{e}_h = (\mathbf{u} - \widetilde{\mathbf{u}}_h) + (\widetilde{\mathbf{u}}_h - \mathbf{u}_h)$, by the triangle inequality it suffices to estimate $\widetilde{\mathbf{u}}_h - \mathbf{u}_h$. Now set $\mathbf{w}_h = \widetilde{\mathbf{e}}_h := \widetilde{\mathbf{u}}_h - \mathbf{u}_h$ and $q_h = \widetilde{\tau}_h := \widetilde{p}_h - p_h$ in the above equation to get the error equation

$$\mathcal{A}(\widetilde{\mathbf{e}}_h, \widetilde{\tau}_h; \widetilde{\mathbf{e}}_h, \widetilde{\tau}_h) = a(\widetilde{\mathbf{u}}_h, (\gamma_h - I)\widetilde{\mathbf{e}}_h) \tag{3.10}$$

$$= a(\widetilde{\mathbf{u}}_h - \mathbf{u}, (\gamma_h - I)\widetilde{\mathbf{e}}_h) + a(\mathbf{u}, (\gamma_h - I)\widetilde{\mathbf{e}}_h) \tag{3.11}$$

$$\leq C\epsilon_1^{-1} ||\widetilde{\mathbf{u}}_h - \mathbf{u}||_0 ||\widetilde{\mathbf{e}}_h||_0 + C\epsilon_1^{-1} h ||\mathbf{u}||_1 ||\widetilde{\mathbf{e}}_h||_0 \tag{3.12}$$

$$\leq Ch\epsilon_1^{-1} \left[||\mathbf{u}||_1 + \epsilon_1^{1/2} h^{-1/2} ||p||_1 \right] ||\widetilde{\mathbf{e}}_h||_0, \tag{3.13}$$

where we have used the boundedness of $\gamma_h - I$, (2.12) in deriving (3.12), and (3.6) in deriving (3.13). Applying (2.14) to the left side of (3.10), we get from (3.13) that

$$a\ (\widetilde{\mathbf{e}}_h, \gamma_h \widetilde{\mathbf{e}}_h) + \left(\left(\alpha + \frac{1}{2} \nabla \cdot \mathbf{b} \right) \widetilde{\tau}_h, \widetilde{\tau}_h \right) + \frac{1}{2} |||\widetilde{\tau}_h|||^2 \tag{3.14}$$

$$\leq C\epsilon_1^{-1} h \left[||\mathbf{u}||_1 + h^{-1/2} \epsilon_1^{1/2} ||p||_1 \right] ||\widetilde{\mathbf{e}}_h||_0. \tag{3.15}$$

Invoking (1.4) and Lemma 2.5 completes the estimate for $\widetilde{\mathbf{e}}_h$. As for the estimate for $p - p_h$, it suffices to estimate $\widetilde{\tau}_h = \tilde{p}_h - p_h$ due to the triangle inequality and (3.6). From (3.14), we have

$$||\widetilde{\tau}_h||_0^2 \leq C\epsilon_1^{-1} h \left[||\mathbf{u}||_1 + h^{-1/2} \epsilon_1^{1/2} ||p||_1 \right] ||\widetilde{\mathbf{e}}_h||_0.$$

Now applying the just-obtained upper bound of $||\tilde{\mathbf{e}}_h||_0$ completes the proof. ∎

4. Numerical results. In this section we use quadratures to obtain the final discretization equations and show some numerical results to demonstrate the error behavior of the studied mixed covolume method. We shall deviate from the notation of the previous sections slightly for the convenience of reporting numerical results. We consider in this section the error behavior of the following problem:

$$\nabla \cdot (-\epsilon \mathcal{K} \nabla p + \mathbf{b}p) + c_0 p = f(x, y), \quad (x, y) \in \Omega = (0, 1)^2. \tag{4.1}$$

The exact solution chosen is $p = x(1 - x)y(1 - y)$ and Dirichlet boundary conditions are imposed. The coefficients of the operator are $\mathcal{K} = diag(k_1, k_2)$, $k_1 = 1 + 10x^2 + y^2$, $k_2 = 1 + x^2 + 10y^2$, $c_0 = 1$, and $\mathbf{b} = (b_1, b_2)$, where

$$\begin{aligned} b_1 &= -\cos\alpha(1 - x\cos\alpha), \\ b_2 &= -\sin\alpha(1 - y\sin\alpha), \end{aligned} \tag{4.2}$$

for angles $\alpha = -\frac{\pi}{2}, -\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}$. Note that $\nabla \cdot \mathbf{b} = 1$ so that condition (1.4) is satisfied and that $\epsilon = 1, 10^{-2}, 10^{-4}, 10^{-6}$.

For the flux variable $\mathbf{u} = (u_1, u_2)$ we used the lowest order Raviart–Thomas piecewise polynomial spaces on isosceles rectangles (squares) of size $h \times h$, for $h = 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}$. The concentration variable p corresponded to piecewise constants on the same rectangular elements.

In general, to evaluate the integrals

$$\int_{Q_{i+1/2,j}} k_1^{-1}(x,y)u_1(x,y)dxdy = \int_{x_i}^{x_i+\frac{1}{2}h_{x_{i+1}}} \int_{y_{j-1/2}}^{y_{j+1/2}} k_1^{-1}(x,y)u_1(x,y)dxdy + \int_{x_i+\frac{1}{2}h_{x_{i+1}}}^{x_{i+1}} k_1^{-1}(x,y)u_1(x,y)dxdy,$$

where $Q_{i+1/2,j} = [x_i, x_{i+1}] \times [y_{j-1/2}, y_{j+1/2}]$, $h_{x_r} = x_r - x_{r-1}$, we used the midpoint quadrature formula in y-direction and the trapezoidal rule in x-direction. This led, for example, for the first integral above, to the following expression:

$$\begin{aligned} &\approx \frac{h_x h_y}{4}\left[k_1^{-1}(c_{i,j})u_1(c_{i,j}) + k_1^{-1}(c_{i+1/2,j})u_1(c_{i+1/2,j})\right] \\ &= \frac{h_x h_y}{4}\left[k_1^{-1}(c_{i,j})\frac{u_1(c_{i-1/2,j}) + u_1(c_{i+1/2,j})}{2} + k_1^{-1}(c_{i+1/2,j})u_1(c_{i+1/2,j})\right] \end{aligned}$$

Note that the definition of $Q_{i+1/2,j}$ at the boundary needs to be modified and we take either the left half or the right half, depending on the location of the boundary. One of the pressures in (1.12) is taken from the boundary value.

We remark that the above quadrature formula actually symmetrizes the bilinear form $a(\mathbf{u}_h, \mathbf{v}_h)$, for $\mathbf{u}_h \in \mathbf{H}_h$, $\mathbf{v}_h \in \mathbf{Y}_h$; namely, the resulting matrix A is symmetric. Also, we note that A is block-diagonal (diagonal blocks A_i, $i = 1, 2$, corresponding to the variables u_1 and u_2, respectively) and each block A_i is block-diagonal itself, with tridiagonal (band) matrices if the unknowns corresponding to u_1 are ordered linewise along the x-direction (each horizontal line giving rise to a block) and the unknowns corresponding to u_2 are ordered linewise along y-directions (with blocks being the unknowns grouped on a given vertical line).

The edge integrals involving $\mathbf{b} \cdot \mathbf{n}$ are computed exactly, which is simple for our particular choice of the field $\mathbf{b}$.

After the discretization one ends up with the following linear system of equations to be solved,

$$\mathcal{A}\begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \mathbf{P} \end{bmatrix} = \begin{bmatrix} \mathbf{rhs}_{\mathbf{U}_1} \\ \mathbf{rhs}_{\mathbf{U}_2} \\ \mathbf{rhs}_P \end{bmatrix}, \tag{4.3}$$

with the saddle-point–like stiffness matrix:

$$\mathcal{A} = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix}. \tag{4.4}$$

We use the fact that A has simple structure, block-diagonal with blocks being tridiagonal matrices (assuming the linewise ordering explained as above), to effectively compute the inverse actions of A. That is, one can eliminate the unknown $\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix}$ and solve the reduced system,

$$(C + BA^{-1}B^T)\mathbf{P} = -\mathbf{rhs}_P + BA^{-1}\begin{bmatrix} \mathbf{rhs}_{\mathbf{U}_1} \\ \mathbf{rhs}_{\mathbf{U}_2} \end{bmatrix}. \tag{4.5}$$

TABLE 4.1
Error behavior for $\alpha = 0$, $\epsilon = 1$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	5.33e-5	3.62e-5	2.07e-5	1.10e-5	1
δ_{u_1}	2.72e-3	9.09e-4	3.49e-4	1.49e-4	> 1
δ_{u_2}	3.04e-3	1.06e-3	4.29e-4	1.90e-4	> 1
# unknowns	800	3 136	12 416	49 408	
# iterations	17	23	45	112	
ϱ	0.28	0.40	0.62	0.83	

We do not compute the reduced Schur matrix $S \equiv C + BA^{-1}B^T$ explicitly, but rather take advantage of the fact that the actions of S on vectors are inexpensively available through solutions of tridiagonal problems to get the inverse action of A. To solve the reduced problem, we used a generalized conjugate gradient least squares method (GCG-LS) as derived in Axelsson [1] (a mathematically-equivalent-to-the-GMRES method, see Saad [24]) with block-ILU preconditioning. The preconditioned matrix M was obtained from $C + A_0$, where A_0 was a simple five-point finite difference approximation of the term $-\nabla \cdot \epsilon \mathcal{K} \nabla(.)$ on the given cell-centered grid.

We use the block partition of $A_0 + C$, using as blocks the unknowns within each strip of width h_x along the vertical direction. Let $A_0 + C = C_D - L - U$, where C_D, $-L$, and $-U$ stand for the respective block-diagonal, lower triangular, and upper triangular parts of $A_0 + C$. Then, since both C and the thus-obtained finite difference matrix A_0 were M-matrices, the construction of $M = (D - L)D^{-1}(D - U)$ with blocks of $D = \mathrm{diag}[D_1, D_2, \ldots, D_n]$ being banded matrices was well-defined and turned out to be very effective preconditioners. For the construction of such block-factorization preconditioners we refer to Concus, Golub, and Meurant [14] when D_i are tridiagonal matrices, and for general banded blocks D_i, to Axelsson and Polman [3]; see also, Vassilevski [25] for some particular constructions of approximate band inverses to band matrices that are required in the block-ILU methods. Other block-ILU type preconditioners (not necessarily only for M-matrices) are found in Chan and Vassilevski [6].

The stopping criterion in the GCG-LS method was

$$\|M^{-1}\mathbf{r}\| \leq 10^{-9}\|M^{-1}\mathbf{r}_0\|,$$

where $\|\mathbf{v}\|^2 = \mathbf{v}^T\mathbf{v}$; $\mathbf{r}_0$ stands for the initial residual and $\mathbf{r}$ stands for the current one. The initial iterate was chosen as $\mathbf{x}_0 = M^{-1}\mathbf{f}$, where $\mathbf{f} \equiv -\mathbf{rhs}_P + BA^{-1}\begin{bmatrix} \mathbf{rhs}_{\mathbf{U}_1} \\ \mathbf{rhs}_{\mathbf{U}_2} \end{bmatrix}$ was the right-hand side of the reduced problem (4.5).

We show in Tables 4.1–4.14, in addition to the error behavior of the covolume discretization method, also ϱ and the number of iterations, where

$$\varrho = \left(\frac{\|M^{-1}\mathbf{r}\|}{\|M^{-1}\mathbf{r}_0\|}\right)^{\frac{1}{\#\text{ iterations}}} \tag{4.6}$$

was an average reduction factor.

More specifically, denote $x_i = ih_x$, $y_j = jh_y$, $i = 0, 1, 2, \ldots, n_x$, $j = 0, 1, 2, \ldots, n_y$, $h_x = h_y = h$, $n_x = n_y = n = 1/h$, for a given $h = 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}$. In Tables 4.1–4.14, for a given ϵ and angle α, we show the following items:

TABLE 4.2
Error behavior for $\alpha = \frac{\pi}{4}$, $\epsilon = 1$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	9.86e-5	6.07e-5	3.33e-5	1.74e-5	1
δ_{u_1}	3.56e-3	1.33e-3	5.64e-4	2.57e-4	> 1
δ_{u_2}	3.56e-3	1.33e-3	5.64e-4	2.57e-4	> 1
# unknowns	800	3 136	12 416	49 408	
# iterations	17	23	45	111	
ϱ	0.28	0.40	0.62	0.82	

TABLE 4.3
Error behavior for $\alpha = \frac{\pi}{2}$, $\epsilon = 1$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	5.33e-5	3.62e-5	2.07e-5	1.10e-5	1
δ_{u_1}	3.04e-3	1.06e-3	4.29e-4	1.90e-4	> 1
δ_{u_2}	2.72e-3	9.09e-4	3.49e-4	1.49e-4	> 1
# unknowns	800	3 136	12 416	49 408	
# iterations	17	22	45	113	
ϱ	0.28	0.39	0.62	0.83	

TABLE 4.4
Error behavior for $\alpha = 0$, $\epsilon = 0.01$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	1.77e-3	9.09e-4	4.62e-4	2.33e-4	1
δ_{u_1}	4.09e-4	2.09e-4	1.03e-4	5.00e-5	1
δ_{u_2}	3.46e-4	1.67e-4	8.25e-5	4.11e-5	1
# unknowns	800	3 136	12 416	49 408	
# iterations	15	17	24	55	
ϱ	0.24	0.29	0.40	0.67	

TABLE 4.5
Error behavior for $\alpha = \frac{\pi}{4}$, $\epsilon = 0.01$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	2.88e-3	1.51e-3	7.78e-4	3.95e-4	1
δ_{u_1}	5.01e-4	2.49e-4	1.21e-4	5.91e-5	1
δ_{u_2}	5.01e-4	2.49e-4	1.21e-4	5.91e-5	1
# unknowns	800	3 136	12 416	49 408	
# iterations	15	17	25	49	
ϱ	0.24	0.28	0.42	0.65	

TABLE 4.6
Error behavior for $\alpha = -\frac{\pi}{4}$, $\epsilon = 0.01$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	3.10e-3	1.68e-3	8.90e-4	4.64e-4	1
δ_{u_1}	5.77e-4	2.95e-4	1.49e-4	7.50e-5	1
δ_{u_2}	1.70e-3	8.43e-4	4.15e-4	2.04e-4	1
# unknowns	800	3 136	12 416	49 408	
# iterations	15	17	24	46	
ϱ	0.24	0.27	0.41	0.63	

TABLE 4.7
Error behavior for $\alpha = 0$, $\epsilon = 10^{-4}$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	2.29e-3	1.14e-3	5.74e-4	2.87e-4	1
δ_{u_1}	2.40e-5	1.10e-5	5.22e-6	2.89e-6	1
δ_{u_2}	5.36e-6	2.50e-6	1.20e-6	5.95e-7	1
# unknowns	800	3 136	12 416	49 408	
# iterations	10	13	15	15	
ϱ	0.10	0.18	0.22	0.25	

TABLE 4.8
Error behavior for $\alpha = \frac{\pi}{4}$, $\epsilon = 10^{-4}$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	3.57e-3	1.84e-3	9.33e-4	4.68e-4	1
δ_{u_1}	4.20e-5	2.91e-5	1.66e-5	8.26e-6	1
δ_{u_2}	4.20e-5	2.91e-5	1.66e-5	8.26e-6	1
# unknowns	800	3 136	12 416	49 408	
# iterations	6	8	10	13	
ϱ	0.03	0.07	0.11	0.19	

TABLE 4.9
Error behavior for $\alpha = \frac{\pi}{2}$, $\epsilon = 10^{-4}$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	2.29e-3	1.14e-3	5.74e-4	2.87e-4	1
δ_{u_1}	5.36e-6	2.50e-6	1.20e-6	5.95e-7	1
δ_{u_2}	2.40e-5	1.10e-5	5.22e-6	2.89e-6	1
# unknowns	800	3 136	12 416	49 408	
# iterations	10	13	15	21	
ϱ	0.10	0.19	0.23	0.36	

TABLE 4.10
Error behavior for $\alpha = -\frac{\pi}{4}$, $\epsilon = 10^{-4}$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	3.93e-3	2.04e-3	1.03e-3	5.22e-4	1
δ_{u_1}	3.99e-5	2.88e-5	1.67e-5	8.33e-6	1
δ_{u_2}	7.13e-5	5.06e-5	3.36e-5	2.08e-5	< 1
# unknowns	800	3 136	12 416	49 408	
# iterations	6	9	11	15	
ϱ	0.03	0.07	0.13	0.23	

TABLE 4.11
Error behavior for $\alpha = 0$, $\epsilon = 10^{-6}$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	2.36e-3	1.18e-3	5.95e-4	2.97e-4	1
δ_{u_1}	3.57e-7	2.54e-7	1.71e-7	1.00e-7	< 1
δ_{u_2}	5.61e-8	2.63e-8	1.27e-8	6.23e-9	< 1
# unknowns	800	3 136	12 416	49 408	
# iterations	4	5	6	9	
ϱ	0.001	0.006	0.02	0.07	

TABLE 4.12
Error behavior for $\alpha = \frac{\pi}{4}$, $\epsilon = 10^{-6}$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	3.64e-3	1.91e-3	9.86e-4	4.99e-4	1
δ_{u_1}	5.06e-7	4.31e-7	3.33e-7	2.44e-7	$\frac{1}{2}$
δ_{u_2}	5.06e-7	4.31e-7	3.33e-7	2.44e-7	$\frac{1}{2}$
# unknowns	800	3 136	12 416	49 408	
# iterations	3	3	4	4	
ϱ	2.8e-4	7.3e-4	0.001	0.003	

TABLE 4.13
Error behavior for $\alpha = \frac{\pi}{2}$, $\epsilon = 10^{-6}$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	2.36e-3	1.18e-3	5.95e-4	2.97e-4	1
δ_{u_1}	5.61e-8	2.63e-8	1.27e-8	6.23e-9	1
δ_{u_2}	3.57e-7	2.54e-7	1.71e-7	1.00e-7	< 1
# unknowns	800	3 136	12 416	49 408	
# iterations	4	5	6	8	
ϱ	0.001	0.006	0.02	0.07	

TABLE 4.14
Error behavior for $\alpha = -\frac{\pi}{4}$, $\epsilon = 10^{-6}$.

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	3.97e-3	2.07e-3	1.06e-3	5.39e-4	< 1
δ_{u_1}	4.75e-7	4.22e-7	3.33e-7	2.45e-7	$\frac{1}{2}$
δ_{u_2}	7.49e-7	5.58e-7	4.06e-7	2.90e-7	$\frac{1}{2}$
# unknowns	800	3 136	12 416	49 408	
# iterations	3	3	4	4	
ϱ	3.67e-4	9.02e-4	0.001	0.003	

(i) $\delta_p = \|I_h p - p_h\|_h \equiv \left[\sum_{i=1}^{n_x} \sum_{j=1}^{n_y} h_x h_y (p(x_i - \frac{1}{2}h_x, y_j - \frac{1}{2}h_y) - p_h(x_i - \frac{1}{2}h_x, y_j - \frac{1}{2}h_y))^2 \right]^{\frac{1}{2}}$, i.e., a discrete L^2-norm of the error $p - p_h$;

(ii) $\delta_{u_1} = \|I_h u_1 - u_{h,1}\|_h \equiv \left[\sum_{i=0}^{n_x} \sum_{j=1}^{n_y} h_x h_y (u_1(x_i, y_j - \frac{1}{2}h_y) - u_{h,1}(x_i, y_j - \frac{1}{2}h_y))^2 \right]^{\frac{1}{2}}$, i.e., a discrete L^2-norm of the error $u_1 - u_{h,1}$;

(iii) $\delta_{u_2} = \|I_h u_2 - u_{h,2}\|_h \equiv \left[\sum_{i=1}^{n_x} \sum_{j=0}^{n_y} h_x h_y (u_2(x_i - \frac{1}{2}h_x, y_j) - u_{h,2}(x_i - \frac{1}{2}h_x, y_j))^2 \right]^{\frac{1}{2}}$, i.e., a discrete L^2–norm of the error $u_2 - u_{h,2}$;

(iv) the number of iterations of the preconditioned GCG–LS method;

(v) the average reduction factors ϱ, (4.6);

(vi) the total number of unknowns (for both $\mathbf{U}$ and $\mathbf{P}$).

It is noticeable that the order $h^{\frac{1}{2}}$ of the error estimate proven in the present paper is attained for small $\epsilon = 10^{-6}$; see, for example, Tables 4.12 and 4.14. It turns out that our experiments suggest first order approximation of the concentration variable for all values of ϵ we have tested. The fast convergence of the preconditioned GCG-

LS method for small ϵ is explained due to the fact that, in that case, the reduced matrix $S \simeq C$ is almost triangular and because our block-ILU factorization matrix turned out to be very accurate in that case. This also agrees with the experience of the performance of block-ILU methods in the case of streamline diffusion finite element discretization of convection-diffusion problems reported in Axelsson, Polman, Eijkhout, and Vassilevski [2]. See also, Lazarov, Mishev, and Vassilevski [19] and Iliev, Makarov, and Vassilevski [18] for finite difference or finite volume discretizations, as well as in the least-squares mixed finite element discretization of convection–diffusion problems; cf., Carey, Pehlivanov, and Vassilevski [5]. We finally remark that the preconditioners used in the present paper are not the main issue; rather, they are tools to solve the resulting systems.

Acknowledgment. The authors are grateful to one of the referees for her/his constructive comments on the original manuscript.

REFERENCES

[1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994.
[2] O. Axelsson, V. Eijkhout, B. Polman, and P. S. Vassilevski, *Incomplete block–matrix factorization iterative methods for convection–diffusion problems*, BIT, 29 (1989), pp. 867–889.
[3] O. Axelsson and B. Polman, *On approximate factorization methods for block matrices suitable for vector and parallel processors*, Linear Algebra Appl., 77 (1986), pp. 3–26.
[4] Z. Cai, J. E. Jones, S. F. McCormick, and T. F. Russell, *Control-volume mixed finite element methods*, Computational Geosciences, 1 (1997), pp. 289–315.
[5] G. F. Carey, A. I. Pehlivanov, and P. S. Vassilevski, *Least-squares mixed finite elements for non-selfadjoint elliptic problems.* II. *Performance of block-ILU factorization methods*, SIAM J. Sci. Comput., 16 (1995), pp. 1126–1136.
[6] T. F. Chan and P. S. Vassilevski, *A Framework for block-ILU factorizations using block-size reduction*, Math. Comp., 64 (1995), pp. 129–156.
[7] S. H. Chou, *Analysis and convergence of a covolume method for the generalized Stokes problem*, Math. Comp., 66 (1997), pp. 85–104.
[8] S. H. Chou and D. Y. Kwak, *Analysis and convergence of a MAC-like scheme for the generalized Stokes problem*, Numer. Methods Partial Differential Equations., 13 (1997), pp. 141–167.
[9] S. H. Chou and D. Y. Kwak, *A Covolume method based on rotated bilinears for the generalized Stokes problem*, SIAM J. Numer. Anal., 35 (1998), pp. 494–507.
[10] S. H. Chou and D. Y. Kwak, *Mixed covolume methods on rectangular grids for elliptic problems*, SIAM J. Numer. Anal, to appear.
[11] S. H. Chou, D. Y. Kwak, and P. S. Vassilevski, *Mixed covolume methods for elliptic problems on triangular grids*, SIAM J. Numer. Anal., 35 (1998), pp. 1850–1861.
[12] S. H. Chou and Q. Li, *Error Estimates in L^2, H^1 and L^∞ in Covolume methods for elliptic and parabolic problems: A Unified Approach*, Math. Comp., 1996, to appear.
[13] S. H. Chou and P. S. Vassilevski, *An upwinding cell–centered method with piecewise constant velocity over covolumes*, Numer. Methods Partial Differential Equations., 1997, 15 (1999), pp. 49–62.
[14] P. Concus, G. H. Golub, and G. Meurant, *Block preconditioning for the conjugate gradient method*, SIAM J. Sci. Stat. Comput., 6 (1985), pp. 220–252.
[15] V. Girault and H. Lopez, *Finite-element error estimates for the MAC scheme,* IMA J. Numer. Anal., 16 (1996), pp. 247–379.
[16] C. A. Hall, T. A. Porsching, and P. Hu, *Covolume-dual variable method for thermally expandable flow on unstructured triangular grids*, Comp. Fluid Dyn, 2 (1994), pp. 111–139.
[17] F. H. Harlow and F. E. Welch, *Numerical calculations of time dependent viscous incompressible flow of fluid with a free surface*, Phys. Fluids, 8 (1965), p. 2181.
[18] O. P. Iliev, M. M. Makarov, and P. S. Vassilevski, *Performance of certain iterative methods in solving implicit difference schemes for 2–D Navier–Stokes equations*, Internat. J. Numer. Methods Engrg., 33 (1992), pp. 1465–1480.

[19] R. D. Lazarov, I. D. Mishev, and P. S. Vassilevski, *Finite volume methods for convection–diffusion problems*, SIAM J. Numer. Anal., 32 (1995), pp. 235–248.
[20] M. Liu, J. Wang, and N. Yan, *New error estimates for approximate solutions of convection-diffusion problems by mixed and discontinuous Galerkin methods*, preprint, Department of Mathematics, University of Wyoming, Laramie, Wyoming, 1997.
[21] R. A. Nicolaides, T. A. Porsching, and C. A. Hall, *Covolume methods in Computational Fluid Dynamics*, in Computational Fluid Dynamics Review, M. Hafez and K. Oshma, eds., John Wiley and Sons, New York, 1995, pp. 279–299.
[22] J. Roberts and J. Thomas, *Mixed and hybrid methods*, in Handbook of Numerical Analysis Vol II, Ch. 4, P. G. Ciarlet and J. L. Lions, eds., North-Holland, Amsterdam, 1991.
[23] T. F. Russell, *Rigorous Block-Centered Discretizations on Irregular Grids: Improved Simulation of Complex Reservoir Systems*, Technical Report 3, Reservoir Simulation Research Corporation, 1995.
[24] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PSW, Kent, UK, 1995.
[25] P. S. Vassilevski, *On some ways of approximating inverses of banded matrices in connection with deriving preconditioners based on incomplete block factorizations*, Computing, 43 (1990), pp. 277–296.