MATHEMATICS OF COMPUTATION
Volume 68, Number 227, Pages 991–1011
S 0025-5718(99)01090-X
Article electronically published on February 23, 1999

A GENERAL MIXED COVOLUME FRAMEWORK FOR CONSTRUCTING CONSERVATIVE SCHEMES FOR ELLIPTIC PROBLEMS

SO-HSIANG CHOU AND PANAYOT S. VASSILEVSKI

ABSTRACT. We present a general framework for the finite volume or covolume schemes developed for second order elliptic problems in mixed form, i.e., written as first order systems. We connect these schemes to standard mixed finite element methods via a one-to-one transfer operator between trial and test spaces. In the nonsymmetric case (convection-diffusion equation) we show one-half order convergence rate for the flux variable which is approximated either by the lowest order Raviart–Thomas space or by its image in the space of discontinuous piecewise constants. In the symmetric case (diffusion equation) a first order convergence rate is obtained for both the state variable (e.g., concentration) and its flux. Numerical experiments are included.

1. INTRODUCTION

The purpose of this paper is to give a unified approach for analyzing a number of finite volume or covolume schemes developed for second order elliptic problems in mixed form, i.e., written as first order systems. The joint framework we use is based on relating all these schemes to the standard mixed method by utilizing a one-to-one mapping between the lowest order Raviart–Thomas spaces $\widehat{\mathbf{V}}_h$ for the vector unknown (also called velocity) and the corresponding spaces of piecewise (discontinuous) constant vectors that are used in the covolume schemes of the main interest. Covolume schemes are popular ([10, 11, 20, 21, 24]) in practical fluid mechanics computations due to their conservative properties; namely, they represent discrete analogs of the underlying physical conservation laws dictating the behavior of the fluid system. For instance, if the main variables of interest of the underlying fluid system are a state variable (concentration, temperature, pressure, etc.) and a flux variable (gradient of the state variable), the covolume method then uses two partitions of the fluid domain to find approximations of these two variables. A conservation law on the primal volumes is used for the state variable and a constitutive law on the dual volumes or covolumes are used for the flux variable. In the case of porous media flow the conservation law for the primal volumes is the mass conservation law, and the constitutive law for the covolumes is the Darcy law.

Received by the editor June 16, 1997.
1991 *Mathematics Subject Classification.* Primary 65F10, 65N20, 65N30.
Key words and phrases. Conservative schemes, mixed finite elements, covolume methods, finite volume methods, finite volume element, Raviart–Thomas spaces, error estimates, $H(\mathrm{div})$-preconditioning.
The work of the second author was partially supported by the Bulgarian Ministry for Education, Science and Technology under grant I–95 # 504.

©1999 American Mathematical Society

A survey paper ([24]) on the literature of the covolume methods up to 1995 was written by Nicolaides, Porsching and Hall, and the reader can find various fluid mechanics applications therein. Recent theoretic as well as computational works on covolume methods are [12]–[19] and [25].

The organization of this paper is as follows. In §2, we formulate two covolume methods for a nonsymmetric elliptic problem in the state variable p: one for which the approximant of the velocity $\underline{u} = -\mathcal{K}\nabla p, \mathcal{K}$, a matrix, is piecewise constant; and the other for which the approximate velocity field is from a Raviart–Thomas space. In §3 we introduce a transfer operator γ_h to determine the test spaces, and we study its properties in §5. Derived in §4 are a priori estimates useful for proving stability and existence of the saddle-point formulation of the methods. The main convergence results are contained in Theorem 6.1. Numerical experiments are given in §7.

2. Problem formulations

We concentrate on the general second order elliptic problem,

$$-\operatorname{div}(\mathcal{K}\nabla p) + \operatorname{div}(\underline{b}p) + c_0 p = f(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega, \tag{2.1}$$

and to be specific we impose the Dirichlet boundary condition $p = 0$ on $\partial\Omega$. The domain Ω is a polygonal domain in the plane, and we assume that it is either covered by a rectangular or triangular quasi-uniform partition $\mathcal{T}_h$. Generalizations to three-dimensional polytopes is straightforward. The coefficient $\mathcal{K} = (k_{rs}(x,y))_{r,s=1}^2$ is assumed to be symmetric, bounded and positive definite uniformly in $\overline{\Omega}$. For the convection-dominated case, i.e., when

$$\gamma_1 \le \inf_{\underline{\xi}\in\mathbb{R}^2} \frac{\underline{\xi}^T\mathcal{K}\underline{\xi}}{\underline{\xi}^T\underline{\xi}} \le \sup_{\underline{\xi}\in\mathbb{R}^2} \frac{\underline{\xi}^T\mathcal{K}\underline{\xi}}{\underline{\xi}^T\underline{\xi}} \le \gamma_2 \ll |\underline{b}|_\infty, \tag{2.2}$$

in general, as is well known, one has to use local refinement near the boundary layers. This issue will not be pursued in the present paper, and we keep the convection term $\operatorname{div}(\underline{b}p)$ in our considerations for the sake of generality. Moreover, to avoid technical details, we will even assume that the following relation between the $\underline{b}$ and c_0 exists:

$$\frac{1}{2}\operatorname{div}\underline{b} + c_0 \ge \gamma_0 = \text{Const} > 0, \quad \text{in } \Omega. \tag{2.3}$$

This assumption implies coercivity of the elliptic operator $\mathcal{L} \equiv -\operatorname{div}(\mathcal{K}\cdot) + \operatorname{div}(\underline{b}\cdot) + c_0$ and solvability and uniqueness of the boundary value problem.

We next introduce a new (vector) unknown $\underline{u} = -\mathcal{K}\nabla p$ and rewrite (2.1) as the system

$$\begin{aligned} \mathcal{K}^{-1}\underline{u} &= -\nabla p, \\ \operatorname{div}\underline{u} + \operatorname{div}(\underline{b}p) + c_0 p &= f. \end{aligned} \tag{2.4}$$

In the standard mixed finite element method, one would use only $\mathcal{T}_h$ to define the discrete weak formulation. In covolume methods, we will use two partitions: a primal partition $\mathcal{T}_h$ on which the local mass conservation law $(2.4)_2$ holds, and a dual partition $\mathcal{Q}_h$ (a union of covolumes) over which $(2.4)_1$ holds in the average sense. The most well-known example is the MAC (Marker and Cell) scheme ([22]) that uses two staggered rectangular grids. In general, we can classify covolume methods into overlapping and nonoverlapping types, according to whether covolumes overlap or

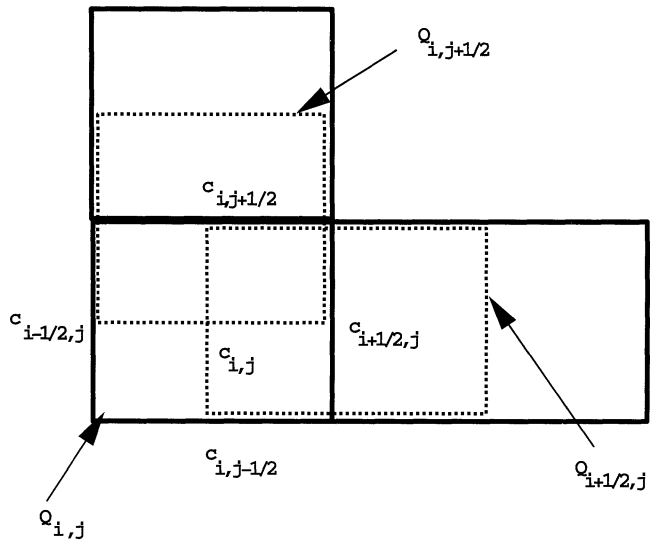

FIGURE 1. Dual domain with overlapping covolumes: the dashed boxes $Q_{i+1/2}$ and $Q_{i,j+1/2}$ are covolumes; $c_{i+1/2,j}$ and $c_{i,j+1/2}$ are the midpoints of the edges of the primal volume $Q_{i,j}$ whose center is $c_{i,j}$.

not. For example, in Figure 1, the dashed covolumes are overlapped. The analysis of such covolume methods was discussed in [8, 14, 17, 19].

On the other hand, we have nonoverlapping covolumes in in Figure 2. For instance, in Figure 1 the primal partition is the union of rectangles. A typical interior covolume in the dual partition is the dashed quadrilateral, the closure of the union of the two triangles $T_E^+ \cup T_E^-$ sharing the common side E. The two vertices in the interiors of the two rectangles are their centers. Note that each edge E of the primal element corresponds to a covolume. On the boundary the covolume is either a T_E^+ or a T_E^-. Such covolume methods were analyzed in [12, 13, 16]. In this paper we develop a general framework for the nonoverlapping case and refer the reader to [19] for the overlapping case.

Let $\mathbf{H}_{\text{loc}}(\text{div})$ be the space of vector functions that are locally in $\mathbf{H}(\text{div}, Q)$ for any $Q \in \mathcal{Q}_h$, and let $H^1_{\text{loc}}(\mathcal{T}_h)$ be the set of all $L^2(\Omega)$ functions that locally are in $H^1(T)$ for all $T \in \mathcal{T}_h$. A natural weak formulation of (2.4) from the above considerations is

(2.5)

$$(\mathcal{K}^{-1}\underline{u}, \underline{v}) - \sum_{Q \in \mathcal{Q}_h} \left(\int_Q \operatorname{div} \underline{v} p - \int_{\partial Q} p \underline{v} \cdot \underline{n} \right) = 0, \quad \text{for all } \underline{v} \in \mathbf{H}_{\text{loc}}(\text{div});$$

$$\sum_{T \in \mathcal{T}_h} \left(-\int_T \nabla q \cdot \underline{u} + \int_{\partial T} \underline{u} \cdot \underline{n} q \right) - \sum_{T \in \mathcal{T}_h} \left(\int_T p \underline{b} \cdot \nabla q + \int_{\partial T} q p \underline{b} \cdot \underline{n} \right)$$
$$+ (c_0 p, q) = (f, q), \qquad \text{for all } q \in H^1_{\text{loc}}(\mathcal{T}_h).$$

The traces of q on ∂T and the traces of $\underline{v} \cdot \underline{n}$ are taken from the interior of T and from the interior of $Q \in \mathcal{Q}_h$, respectively.

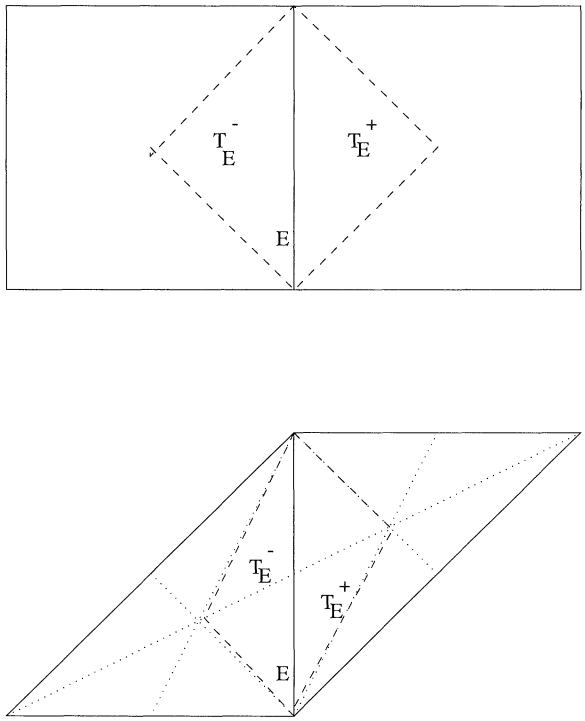

FIGURE 2. Primal and dual domains; the dual element (covolume) $Q_E = T_E^+ \cup E \cup T_E^-$.

To be specific, from now on we assume that each $Q \in \mathcal{Q}_h$ is associated with an edge E of elements from $\mathcal{T}_h$. We will write $Q = Q_E$. Also, the neighboring elements in $\mathcal{T}_h$ that share the same edge E will be denoted by T_E^+ and T_E^-.

Now it is clear that using the divergence theorem one gets the expression

$$\int_{\partial Q} p\underline{v} \cdot \underline{n} = \int_{T_E^+} \nabla p \cdot \underline{v} + \int_{T_E^-} \nabla p \cdot \underline{v} + \int_{T_E^+ \cup T_E^-} \operatorname{div} \underline{v} p - \int_E [p]_E \, \underline{v} \cdot \underline{n}_E. \tag{2.6}$$

Here, $[p]_E = \lim_{t \to +0} (p(x + t\underline{n}_E) - p(x - t\underline{n}_E))$ for $x \in E$, stands for the jump of p across the edge E. The unit vector $\underline{n}_E$ associated with the edge E is considered fixed once chosen, whereas $\underline{n}$ always denotes an outward unit normal vector.

Now we are in a position to derive the finite volume type, or covolume finite element discretizations of the mixed system in the weak formulation (2.5), which can be viewed as nonconforming mixed finite element discretizations of (2.1). Let W_h be the space of piecewise constant functions with respect to the primal triangulation $\mathcal{T}_h$, and let $\mathbf{V}_h$ be the space of piecewise constant vector functions with respect to the dual partition $\mathcal{Q}_h$ that have a continuous normal trace across the interior edges E; more precisely,

$$\begin{aligned} \mathbf{V}_h := \Big\{ \underline{v} : \ & \underline{v}|_K \ \text{ is constant, } K = T_E^+, \, T_E^-, \\ & \underline{v}|_{T_E^+} \cdot \underline{n}_E = \underline{v}|_{T_E^-} \cdot \underline{n}_E, \qquad \text{for all } Q_E = T_E^+ \cup E \cup T_E^- \in \mathcal{Q}_h \Big\}. \end{aligned} \tag{2.7}$$

Then the discrete system reads as follows.

Find $\underline{u}_h \in \mathbf{V}_h$, $p_h \in W_h$ such that

$$(\mathcal{K}^{-1}\underline{u}_h, \underline{v}) + \sum_{E\in\mathcal{E}_h} \int_E \underline{v}\cdot\underline{n}_E\,[p_h]_E = 0, \quad \text{for all } \underline{v}\in\mathbf{V}_h, \tag{2.8}$$

$$\sum_{E\in\mathcal{E}_h} \int_E \underline{u}_h\cdot\underline{n}_E\,[q]_E - \sum_{T\in\mathcal{T}_h} \int_{\partial T} q^i\left(p_h^i(\underline{b}\cdot\underline{n})^+ + p_h^o(\underline{b}\cdot\underline{n})^-\right) - (c_0 p_h, q) = -(f,q), \quad \text{for all } q\in W_h.$$

Here $\mathcal{E}_h$ is the set of all edges (including the boundary ones). For boundary edges E we set $q^o = 0$ for all $q \in W_h$ and in particular $p_h^o = 0$. Note that we have used the weak form (2.5) with the first equation rewritten using expression (2.6) based on the fact that we deal with piecewise constant functions. For the discretization of the convection part we have utilized an upwinding strategy; namely, with the value $(\underline{b}\cdot\underline{n})^- \equiv \min(0, \underline{b}\cdot\underline{n})$ we associate p_h^o, the trace of p_h from the exterior of T, and with $(\underline{b}\cdot\underline{n})^+ \equiv \max(0, \underline{b}\cdot\underline{n})$ we associate the value p_h^i, i.e., the trace of p_h from the interior of T.

One may also consider a Petrov–Galerkin type mixed covolume scheme. That is, we seek the solution $\widehat{\underline{u}}_h \in \widehat{\mathbf{V}}_h$ and $p_h \in W_h$ such that

$$(\mathcal{K}^{-1}\widehat{\underline{u}}_h, \underline{v}) + \sum_{E\in\mathcal{E}_h} \int_E \underline{v}\cdot\underline{n}_E\,[p_h]_E = 0, \quad \text{for all } \underline{v}\in\overline{\mathbf{V}}_h \subset \mathbf{V}_h, \tag{2.9}$$

$$\sum_{E\in\mathcal{E}_h} \int_E \widehat{\underline{u}}_h\cdot\underline{n}_E\,[q]_E - \sum_{T\in\mathcal{T}_h} \int_{\partial T} q^i\left(p_h^i(\underline{b}\cdot\underline{n})^+ + p_h^o(\underline{b}\cdot\underline{n})^-\right) - (c_0 p_h, q) = -(f,q), \quad \text{for all } q\in W_h.$$

Here, $\overline{\mathbf{V}}_h$ is a proper subspace of $\mathbf{V}_h$ of the same dimension as the trial space $\widehat{\mathbf{V}}_h$ (the lowest order Raviart–Thomas space).

3. Determination of a test space by an edge-averaging operator

To analyze the covolume schemes (2.8) we will relate it to the standard mixed finite element method applied to (2.4). Let us take the corresponding mixed finite element spaces $\widehat{\mathbf{V}}_h$ to be the Raviart–Thomas spaces. Note that functions in these spaces have continuous normal components across the edges. For triangular elements the Raviart–Thomas spaces are characterized by the piecewise polynomials of the form $\begin{bmatrix} a_1 + cx \\ b_1 + cy \end{bmatrix}$ that have continuous normal traces across the edges of the triangles $T \in \mathcal{T}_h$. Each function $\underline{v} \in \widehat{\mathbf{V}}_h$ is determined from the interpolation conditions

$$\int_E \underline{v}\cdot\underline{n}_E, \quad \text{for the three edges } E \text{ of a given element } T\in\mathcal{T}_h.$$

Similarly, for rectangular elements the space $\widehat{\mathbf{V}}_h$ is defined as piecewise polynomials of the form $\begin{bmatrix} a_1 + b_1 x \\ a_2 + b_2 y \end{bmatrix}$ on every rectangle $T \in \mathcal{T}_h$. The interpolation conditions here are again specified on the edges of each element T; namely,

$$\int_E \underline{v}\cdot\underline{n}_E, \quad \text{for all four edges } E \text{ of the given element } T\in\mathcal{T}_h.$$

We now define the *transfer* operator $\gamma_h : \widehat{\mathbf{V}}_h \to \mathbf{V}_h$ that relates $\mathbf{V}_h$ with the Raviart–Thomas space $\widehat{\mathbf{V}}_h$ as follows:

$$(\gamma_h \underline{v})|_{Q_E} = \begin{cases} \frac{1}{|E|} \int_E \underline{v}^- \, d\rho, & \text{on } T_E^-, \\ \frac{1}{|E|} \int_E \underline{v}^+ \, d\rho, & \text{on } T_E^+, \end{cases} \tag{3.1}$$

where $|E|$ denotes the measure of E. We make obvious modifications for boundary edges E. Note that $\gamma_h \underline{v}$ is a discontinuous piecewise constant vector function with continuous normal trace and hence belongs to $\mathbf{V}_h$. Therefore, a minimal space with piecewise constant vector functions, locally in $\mathbf{H}(\mathrm{div}, Q)$ and hence locally divergence free, is the space $\overline{\mathbf{V}}_h := \gamma_h \widehat{\mathbf{V}}_h$. It is clear that if $\gamma_h \underline{\widehat{v}} = 0$ for some $\underline{\widehat{v}} \in \widehat{\mathbf{V}}_h$, this will imply $\underline{\widehat{v}} \cdot n_E = 0$ and therefore $\underline{\widehat{v}} = 0$. Hence, γ_h is injective. Actually, the following local estimates for γ_h hold:

$$\|\gamma_h \underline{\widehat{v}}\|_{0,T}^2 \simeq |T| \sum_{E \subset \partial T} (\underline{\widehat{v}}(m_E) \cdot \underline{\widehat{v}}(m_E)) \simeq \|\underline{\widehat{v}}\|_{0,T}^2,$$

where $\underline{\widehat{v}}(m_E)$ is evaluated at the midpoint m_E of the edge E from the interior side of T. The above inequalities in particular imply the global equivalence

$$\|\gamma_h \underline{\widehat{v}}\|_0 \simeq \|\underline{\widehat{v}}\|_0 \cdot \tag{3.2}$$

Therefore, for the Petrov–Galerkin mixed covolume scheme (2.9) we can use the space $\overline{\mathbf{V}}_h = \gamma_h \widehat{\mathbf{V}}_h$.

Let us describe a basis of the space $\mathbf{V}_h$. Since the elements of $\mathbf{V}_h$ are piecewise constant vectors with continuous normal component across each interior edge E for $E \in Q_E$, we have $\underline{v}|_{Q_E} = \alpha \underline{n}_E + \alpha^+ \underline{\tau}_E^+ + \alpha^- \underline{\tau}_E^-$. Here,

$$\underline{\tau}_E^+ = \begin{cases} \underline{\tau}_E, & \text{on } T_E^+, \\ 0, & \text{on } T_E^-, \end{cases}$$

and analogously for

$$\underline{\tau}_E^- = \begin{cases} \underline{\tau}_E, & \text{on } T_E^-, \\ 0, & \text{on } T_E^+. \end{cases}$$

The vector $\underline{\tau}_E$ is the unit vector parallel to E (and orthogonal to $\underline{n}$). We have for any choice of the constants α, α^+ and α^- that $\underline{v} \cdot \underline{n} = \alpha$ is continuous across E. Again, the above argument needs to be modified at a boundary edge. For simplicity, we shall not specifically mention the boundary edge case when the modification is obvious.

Alternatively, one may use the Helmholtz decomposition ([1]) of any piecewise constant vector $\underline{v} = \nabla \psi + \mathbf{curl}\, \phi$, where ϕ is piecewise linear conforming and ψ is a piecewise linear nonconforming Crouzeix–Raviart function (continuous at midpoints). Here the decomposition is considered locally on each dual element $Q = E \cup T_E^+ \cup T_E^-$, and the function ψ vanishes on ∂Q and is linear on each of T_E^+ and T_E^-. This actually implies that $\psi = 0$. Therefore, $\underline{v} = \mathbf{curl}\, \phi$. (This conclusion can also be derived directly without using the Helmholtz decomposition.) Acting in this way one can construct a basis in our spaces $\mathbf{V}_h$.

Note that the basis is local; the support of each basis (vector-) function being the dual element Q_E. Associated with each interior edge E of $\mathcal{E}_h$ are three basis functions and so the number of degrees of freedom of $\mathbf{V}_h$ associated with interior edges are three times the cardinality of the set of all interior edges in $\mathcal{E}_h$. Comparing

this with the dimension of the associated Raviart–Thomas space $\widehat{\mathbf{V}}_h$, we have proved the following main result.

Theorem 3.1. *The space $\overline{\mathbf{V}}_h = \gamma_h \widehat{\mathbf{V}}_h$ is a proper subspace of $\mathbf{V}_h$, and $\overline{\mathbf{V}}_h$ is isomorphic to the lowest order Raviart–Thomas space $\widehat{\mathbf{V}}_h$. Moreover, the operator $\gamma_h : \widehat{\mathbf{V}}_h \to \overline{\mathbf{V}}_h$ is bounded, one-to-one, and L^2-coercive.*

4. A priori estimates

In this section we establish the stability of system (2.8) and its restricted version. The following identity is readily seen ([23]).

Lemma 4.1. *Define the bilinear form $c = c_h$ on $W_h \times W_h$,*

$$c(w,q) = \sum_{T\in\mathcal{T}_h} \int_{\partial T} q^i \left(w^i(\underline{b}\cdot\underline{n})^+ + w^o(\underline{b}\cdot\underline{n})^-\right) + (c_0 w, q). \tag{4.1}$$

Then, for $w = q \in W_h$ one gets the identity,

$$c(q,q) = \frac{1}{2}\sum_{E\in\mathcal{E}_h} \int_E [q]^2 |\underline{b}\cdot\underline{n}|\, d\rho + \left((c_0 + \frac{1}{2}\operatorname{div}\underline{b})q, q\right), \tag{4.2}$$

which due to (2.3) *implies the positive semi-definiteness of the form C_h defined in* (4.5) *below.*

Next we introduce the bilinear forms:

- $A_h : \mathbf{V}_h \mapsto \mathbf{V}_h$, defined by

$$(A_h\underline{v}, \underline{\chi}) = (\mathcal{K}^{-1}\underline{v}, \underline{\chi}), \quad \text{for all } \underline{v}, \underline{\chi} \in \mathbf{V}_h; \tag{4.3}$$

- $B_h : \mathbf{V}_h \mapsto W_h$ defined by

$$(B_h\underline{v}, q) = \sum_{E\in\mathcal{E}_h} \int_E \underline{v}\cdot\underline{n}_E\, [q]_E\,, \quad \text{for all } \underline{v}\in\mathbf{V}_h, \text{ for all } q\in W_h; \tag{4.4}$$

- $C_h : W_h \mapsto W_h$,

$$(C_h q, \psi) = \sum_{T\in\mathcal{T}_h} \int_{\partial T} \psi^i \left(q^i(\underline{b}\cdot\underline{n})^+ + q^o(\underline{b}\cdot\underline{n})^-\right) + (c_0 q, \psi), \quad \text{for all } q, \psi \in W_h. \tag{4.5}$$

We can now formulate (2.8) in the operator saddle-point form

$$\begin{bmatrix} A_h & B_h^T \\ B_h & -C_h \end{bmatrix} \begin{bmatrix} \underline{u}_h \\ p_h \end{bmatrix} = \begin{bmatrix} 0 \\ -f_h \end{bmatrix}, \tag{4.6}$$

where f_h is the L^2 projection of f onto W_h, the space of piecewise constant functions.

Having the explicit representation of a basis of $\mathbf{V}_h$ given by

$$\left\{\alpha_E\underline{n}_E + \alpha_E^+\underline{\tau}_E^+ + \alpha_E^-\underline{\tau}_E^-,\ E\in\mathcal{E}_h\right\}, \tag{4.7}$$

we can prove solvability and uniqueness of the discrete problem (2.8). First observe that

$$\begin{aligned}(A_h\underline{v}, \underline{v}) &= \sum_{E\in\mathcal{E}_h} \int_{T_E^+\cup T_E^-} \mathcal{K}^{-1}\underline{v}\cdot\underline{v} \\ &\le \gamma_1^{-1} \sum_{T\in\mathcal{T}_h}\sum_{E\in\partial T} \left(|T_E^+| + |T_E^-|\right)\left(\alpha_E^2 + (\alpha_E^-)^2 + (\alpha_E^+)^2\right),\end{aligned}$$

where γ_1 is from (2.2), $\alpha_E = \underline{v} \cdot \underline{n}_E$ and $\alpha_E^- = \underline{v} \cdot \tau_E^-$, etc. Then, it is clear that

$$\gamma_2^{-1} \|\gamma_h \widehat{\underline{v}}\|_0^2 \le (A_h \gamma_h \widehat{\underline{v}}, \gamma_h \widehat{\underline{v}}) \le \text{Const } \|\widehat{\underline{v}}\|_0^2, \quad \text{for any } \widehat{\underline{v}} \in \widehat{\mathbf{V}}_h. \tag{4.8}$$

The last estimate also follows from the global equivalence (3.2). Also, note that

$$(B_h \gamma_h \widehat{\underline{v}}, q) = (\operatorname{div} \widehat{\underline{v}}, q), \quad \text{for all } \widehat{\underline{v}} \in \widehat{\mathbf{V}}_h \text{ and } q \in W_h. \tag{4.9}$$

This, together with (4.8) imply the discrete inf-sup condition

$$\beta \|q\|_0 \le \sup_{\underline{v} \in \mathbf{V}_h} \frac{(B_h \underline{v}, q)}{(\|\underline{v}\|_0^2 + \|B_h \underline{v}\|_0^2)^{\frac{1}{2}}} \tag{4.10}$$

for some positive constant β. This is seen from Theorem 3.1, the inf-sup condition on the Raviart–Thomas spaces $(\widehat{\mathbf{V}}_h, W_h)$, the L^2 boundedness (4.8) of γ_h, and $\|B_h \gamma_h \widehat{\underline{v}}\|_0 \le \|\operatorname{div} \widehat{\underline{v}}\|_0$ for any $\widehat{\underline{v}} \in \widehat{\mathbf{V}}_h$ implied by the commutativity property (4.9).

Now the solvability of the discrete problem follows from the a priori estimates in the following lemma.

Lemma 4.2. *Assume that the form C_h is coercive (see Lemma 4.1). Consider the following problem:*

$$\begin{bmatrix} A_h & B_h^T \\ B_h & -C_h \end{bmatrix} \begin{bmatrix} \underline{\chi} \\ w \end{bmatrix} = \begin{bmatrix} g \\ f \end{bmatrix}. \tag{4.11}$$

Here g is a given L^2-bounded linear functional, namely

$$|g(\underline{v})| \le \|g\|_0 \|\underline{v}\|_0, \quad \textit{for all } \underline{v} \in \mathbf{V}_h.$$

Similarly, f is a given L^2-bounded functional,

$$|f(q)| \le \|f\|_0 \|q\|_0, \quad \textit{for all } q \in W_h.$$

Then, the following a priori estimate holds:

$$\begin{aligned} \|\underline{\chi}\|_0 &\le C \left[\frac{\gamma_2}{\gamma_1} \|f\|_0 + \gamma_2 \|g\|_0 \right] \\ \|w\|_0 &\le C \left[\|g\|_0 + \gamma_1^{-1} \|f\|_0 \right] \frac{\gamma_2}{\gamma_1}. \end{aligned} \tag{4.12}$$

For the convection dominated case, i.e., when $\gamma_2 \ll |\underline{b}|_\infty$, the estimate for w deteriorates, but using the strong coercivity of C_h, one gets the alternaiive estimate,

$$\gamma_0 \|w\|_0^2 \le \|f\|_0 \|w\|_0 + \|g\|_0 \|\underline{\chi}\|_0,$$

that is,

$$\|w\|_0 \le C \gamma_0^{-1} \left[\|f\|_0 + \left(\frac{\gamma_2}{\gamma_1} \|f\|_0 + \gamma_2 \|g\|_0 \right) \right]. \tag{4.13}$$

If one considers the problem (4.11) *only in the subspace $\gamma_h \widehat{\mathbf{V}}_h$, then in addition, one gets the estimate,*

$$\|\operatorname{div} \gamma_h^{-1} \underline{\chi}\|_0 \le \|f\|_0 + \left[C_0 h^{-1} + |c_0|_\infty \right] \left[\|g\|_0 + \gamma_1^{-1} \|f\|_0 \right],$$

or the alternative estimate,

$$\|\operatorname{div} \gamma_h^{-1} \underline{\chi}\|_0 \le \|f\|_0 + C \left[C_0 h^{-1} + |c_0|_\infty \right] \gamma_0^{-1} \left[\|f\|_0 + \left(\frac{\gamma_2}{\gamma_1} \|f\|_0 + \gamma_2 \|g\|_0 \right) \right].$$

The constant C_0 can be taken to be zero if the convection term is not present.

Proof. Testing the first equation of (4.11) with $\underline{\chi}$ and the the second equation of (4.11) with $-w$ and adding them together, one arrives at

$$(A_h\underline{\chi},\underline{\chi}) + C_h(w,w) = -f(w) + g(\underline{\chi}) \le \|f\|_0\|w\|_0 + \|g\|_0\|\underline{\chi}\|_0. \tag{4.14}$$

Using the inf-sup condition for $q = w$ in (4.10) yields

$$\begin{aligned}\beta\|w\|_0 &\le \sup_{\underline{v}} \frac{(B_h\underline{v},w)}{(\|\underline{v}\|^2 + \|B_h\underline{v}\|^2)^{\frac{1}{2}}}\\ &\le \sup_{\underline{v}} \frac{g(\underline{v}) - (A_h\underline{v},\underline{\chi})}{(\|\underline{v}\|^2 + \|B_h\underline{v}\|^2)^{\frac{1}{2}}}\\ &\le \|g\|_0 + \gamma_1^{-1}\|\underline{\chi}\|_0,\end{aligned}$$

which together with (4.14) implies,

$$\begin{aligned}(A_h\underline{\chi},\underline{\chi}) + C_h(w,w) &\le \|f\|_0\|w\|_0 + \|g\|_0\|\underline{\chi}\|_0\\ &\le \|f\|\beta^{-1}\left(\|g\|_0 + \gamma_1^{-1}\|\underline{\chi}\|_0\right) + \|g\|_0\|\underline{\chi}\|_0.\end{aligned}$$

Apply the coercivity of C_h and A_h, respectively, and use the generalized arithmetic-geometric inequality to obtain

$$\gamma_2^{-1}(1-\delta)\|\underline{\chi}\|_0^2 \le \beta^{-1}\|f\|_0\|g\|_0 + \frac{\gamma_2}{4}\delta^{-1}\left[\beta^{-1}\gamma_1^{-1}\|f\|_0 + \|g\|_0\right]^2. \tag{4.15}$$

Therefore, for w one gets the estimate

$$\beta\|w\|_0 \le \|g\|_0 + \gamma_1^{-1}\|\underline{\chi}\|_0 \le C(\gamma_1^{-1}\|f\|_0 + \|g\|_0)\frac{\gamma_2}{\gamma_1}. \tag{4.16}$$

In the case of seeking a solution in the subspace $\boldsymbol{\gamma}_h\widehat{\mathbf{V}}_h$, we test the second equation of (4.11) with $q = \operatorname{div}\boldsymbol{\gamma}_h^{-1}\underline{\chi}$ and use the commutativity relation (4.9) to obtain

$$\begin{aligned}\|\operatorname{div}\boldsymbol{\gamma}_h^{-1}\underline{\chi}\|_0^2 &= (B_h\underline{\chi}, \operatorname{div}\boldsymbol{\gamma}_h^{-1}\underline{\chi})\\ &= (f, \operatorname{div}\boldsymbol{\gamma}_h^{-1}\underline{u}_h) + (C_hw, \operatorname{div}\boldsymbol{\gamma}_h^{-1}\underline{\chi})\\ &\le \|f\|_0\|\operatorname{div}\boldsymbol{\gamma}_h^{-1}\underline{u}_h\|_0 + \left[C_0h^{-1} + |c_0|_\infty\right]\|w\|_0\|\operatorname{div}\boldsymbol{\gamma}_h^{-1}\underline{u}_h\|_0.\end{aligned}$$

Here we have used an inverse inequality for C_hw, which follows from Lemma 4.1. Hence

$$\|\operatorname{div}\boldsymbol{\gamma}_h^{-1}\underline{\chi}\|_0 \le \|f\|_0 + \left[C_0h^{-1} + |c_0|_\infty\right]\|w\|_0.$$

To bound $\|w\|_0$ one can either use (4.13) or (the second estimate in) (4.12). □

Corollary 4.1. *Assume that $\underline{b} = 0$ and $c_0 = 0$ and that we solve the problem* (4.11) *in the subspace $\boldsymbol{\gamma}_h\widehat{\mathbf{V}}_h$. Then the following a priori estimate holds:*

$$\left(\|\widehat{\underline{\chi}}\|_{\mathbf{H}(\mathrm{div})}^2 + \|w\|_0^2\right)^{\frac{1}{2}} \le C\left(\|A_0^{-\frac{1}{2}}\underline{g}\|_0 + \|f\|_0\right).$$

Here, A_0 is defined as

$$(A_0\,\boldsymbol{\gamma}_h\widehat{\underline{v}}, \boldsymbol{\gamma}_h\widehat{\underline{\theta}}) = (\operatorname{div}\widehat{\underline{v}}, \operatorname{div}\widehat{\underline{\theta}}) + (\mathcal{K}^{-1}\widehat{\underline{v}}, \widehat{\underline{\theta}}), \quad \textit{for all } \widehat{\underline{\theta}},\ \widehat{\underline{v}} \in \widehat{\mathbf{V}}_h.$$

5. Properties of γ_h

In this section we prove some technical properties of the main operator γ_h that relates $\widehat{\mathbf{V}}_h$ and its image $\overline{\mathbf{V}}_h = \gamma_h \widehat{\mathbf{V}}_h \subset \mathbf{V}_h$. We will actually show a little more; namely, in view of (3.1), one can extend the domain of γ_h to the discontinuous Raviart–Thomas space $\widehat{\mathbf{V}}_h^d$ (the superscript stands for "discontinuous"). In other words, we do not require $\widehat{\mathbf{V}}_h^d$ to be a subspace of $\mathbf{H}(\mathrm{div}, \Omega)$. In particular, we show symmetry and coercivity of γ_h in the discontinuous Raviart–Thomas space $\widehat{\mathbf{V}}_h^d$.

We first show that γ_h is a self-adjoint operator with respect to the L^2 inner product on $\widehat{\mathbf{V}}_h^d$.

Lemma 5.1. *Let $\widehat{\mathbf{V}}_h^d$ be the discontinuous Raviart–Thomas space on rectangular or triangular elements; i.e., we do not require that the space be a subspace of $\mathbf{H}(\mathrm{div}, \Omega)$. The following relation holds:*

$$(\gamma_h \widehat{\mathbf{v}}, \widehat{\mathbf{w}}) = (\widehat{\mathbf{v}}, \gamma_h \widehat{\mathbf{w}}) \quad \textit{for all } \widehat{\mathbf{v}},\ \widehat{\mathbf{w}} \in \widehat{\mathbf{V}}_h^d \cdot \tag{5.1}$$

Proof. It suffices to show that the x-component result holds over a reference element T, since the y-component result can be handled similarly. In other words, we show that $\int_T \gamma_h \widehat{v}\, \widehat{w}\, dxdy = \int_T \widehat{v}\, \gamma_h \widehat{w}\, dxdy$, where $\widehat{v}$ and $\widehat{w}$ are the x-components of $\widehat{\mathbf{v}}$ and $\widehat{\mathbf{w}}$, respectively, and where $\gamma_h \widehat{v}$ is the x-component of $\gamma_h \widehat{\mathbf{v}}$.

We first demonstrate the result for $T = (0,1) \times (0,1)$. Then $\widehat{\mathbf{V}}_h^d$ is the Raviart–Thomas space on rectangles. Let $\lambda_1(x,y) = 1 - x$ and $\lambda_2(x,y) = x$ be the two nodal basis functions and let $T_i, 1 \le i \le 4$, be the four triangles formed by drawing the two diagonals. It is straightforward to show that for $M_{ij} := \int (\gamma_h \lambda_i)\, \lambda_j dxdy = \Sigma_{k=1}^4 \int_{T_k} (\gamma_h \lambda_i)\, \lambda_j dxdy, 1 \le i, j \le 2$,

$$M_{ij} = \begin{cases} \dfrac{|T|}{3} & \text{if } i = j, \\[2ex] \dfrac{|T|}{6} & \text{if } i \neq j. \end{cases}$$

Hence M is a symmetric matrix.

As for the triangular elements, let T be the reference triangle with the vertices $\mathbf{a}^1 = (0,0)$, $\mathbf{a}^2 = (1,1)$ and $\mathbf{a}^3 = (0,1)$ and the barycenter $\mathbf{c} = (1/3, 2/3)$. Let T_1, T_2, T_3 be the subtriangles $\Delta \mathbf{c}\mathbf{a}^2\mathbf{a}^3, \Delta \mathbf{c}\mathbf{a}^3\mathbf{a}^1, \Delta \mathbf{a}^1\mathbf{c}\mathbf{a}^2$, respectively.

Obviously, any x-component of a vector field $\widehat{\mathbf{v}}$ in $\widehat{\mathbf{V}}_h^d$, when restricted to T, has the form $a + bx + cy$. Thus let λ_i be the Lagrange nodal basis function associated with $\mathbf{a}^i$ such that $\lambda_i(\mathbf{a}^j) = \delta_{ij}, 1 \le i, j \le 3$. Then with the above ordering it is easy to show that for $M_{ij} := \int (\gamma_h \lambda_i)\, \lambda_j dxdy = \Sigma_{k=1}^3 \int_{T_k} (\gamma_h \lambda_i)\, \lambda_j dxdy, 1 \le i, j \le 3$,

$$M_{ij} = \begin{cases} \dfrac{5|T|}{54} & \text{if } i \neq j, \\[2ex] \dfrac{4|T|}{27} & \text{if } i = j. \end{cases} \tag{5.2}$$

Hence M is a symmetric matrix. This completes the proof. □

Lemma 5.2. *Let $\widehat{\mathbf{V}}_h^d$ be the discontinuous Raviart–Thomas space on rectangular or triangular elements. Then the following coercivity estimate holds:*

$$(\gamma_h \widehat{\mathbf{v}}, \widehat{\mathbf{v}}) \ge \delta \|\widehat{\mathbf{v}}\|_0^2, \quad \textit{for all } \widehat{\mathbf{v}} \in \widehat{\mathbf{V}}_h^d \cdot$$

Proof. It is sufficient to show that the result holds on a reference element T. We will only show the triangular case, since the easier rectangular case can be done in a similar fashion.

For triangular elements, let $\widehat{\mathbf{v}} = (a + bx + cy, p + qx + ry)^t$ in $\widehat{\mathbf{V}}_h^d$. Then from (5.2) we obtain after simple calculations that

$$\int_T \boldsymbol{\gamma}_h \widehat{\mathbf{v}} \cdot \widehat{\mathbf{v}}\, dxdy = \frac{|T|}{3}\left[3a^2 + \frac{4}{9}b^2 + 2ab + 4ac + \frac{13}{9}bc + \frac{13}{9}c^2 \right.$$
$$\left. + 3p^2 + \frac{4}{9}q^2 + 2pq + 4pr + \frac{13}{9}rq + \frac{13}{9}r^2\right] \cdot$$

The positive-definiteness is then seen by looking at the coefficient matrix of the above quadratic forms,

$$\begin{bmatrix} 3 & 1 & 2 \\ 1 & \frac{4}{9} & \frac{13}{18} \\ 2 & \frac{13}{18} & \frac{13}{9} \end{bmatrix} \cdot$$

By direct computation one sees that the eigenvalues are 0.1136, 0.0518 and 4.7235. Hence

$$\int_T \boldsymbol{\gamma}_h \widehat{\mathbf{v}} \cdot \widehat{\mathbf{v}}\, dxdy \geq 0.0518 \cdot \frac{|T|}{3}\left[a^2 + b^2 + c^2 + p^2 + q^2 + r^2\right] \geq \delta \|\widehat{\mathbf{v}}\|_{0,\,T}^2 \cdot$$

□

We finally show the coercivity of $\boldsymbol{\gamma}_h$ in the weighted inner product $(\mathcal{K}^{-1}\cdot, \cdot)$.

Lemma 5.3. *Assume that $\mathcal{K} \in W^{1,\infty}$. Then for sufficiently small h*

$$(\mathcal{K}^{-1}\widehat{\mathbf{v}}, \boldsymbol{\gamma}_h \widehat{\mathbf{v}}) \geq \delta(\mathcal{K}^{-1}\widehat{\mathbf{v}}, \widehat{\mathbf{v}}), \quad \textit{for all } \widehat{\mathbf{v}} \in \widehat{\mathbf{V}}_h^d.$$

Proof. Assume for the time being that $\mathcal{K}$ is a piecewise constant matrix with respect to the elements $T \in \mathcal{T}_h$. Then, it is clear that $\mathcal{K}^{-\frac{1}{2}}$ is also a piecewise constant matrix and that

$$\mathcal{K}^{-\frac{1}{2}} \boldsymbol{\gamma}_h \widehat{\mathbf{v}} = \boldsymbol{\gamma}_h \mathcal{K}^{-\frac{1}{2}} \widehat{\mathbf{v}}.$$

Therefore, by Lemma 5.2 one gets

$$(\mathcal{K}^{-1}\boldsymbol{\gamma}_h \widehat{\mathbf{v}}, \widehat{\mathbf{v}}) = (\boldsymbol{\gamma}_h \mathcal{K}^{-\frac{1}{2}}\widehat{\mathbf{v}},\ \mathcal{K}^{-\frac{1}{2}}\widehat{\mathbf{v}}) \geq \delta(\mathcal{K}^{-\frac{1}{2}}\widehat{\mathbf{v}},\ \mathcal{K}^{-\frac{1}{2}}\widehat{\mathbf{v}}) = \delta(\mathcal{K}^{-1}\widehat{\mathbf{v}}, \widehat{\mathbf{v}}) \cdot$$

For the variable coefficient case, consider the piecewise constant interpolant $\mathcal{K}_0 = \frac{1}{|T|}\int_T \mathcal{K}\, dxdy$. Then, for sufficiently small h,

$$(\mathcal{K}^{-1}\boldsymbol{\gamma}_h \widehat{\mathbf{v}}, \widehat{\mathbf{v}}) = ((\mathcal{K}^{-1} - \mathcal{K}_0^{-1})\boldsymbol{\gamma}_h \widehat{\mathbf{v}}, \widehat{\mathbf{v}}) + (\mathcal{K}_0^{-1}\boldsymbol{\gamma}_h \widehat{\mathbf{v}}, \widehat{\mathbf{v}})$$
$$\geq (-Ch + \delta)(\mathcal{K}_0^{-1}\widehat{\mathbf{v}}, \widehat{\mathbf{v}}) \geq \delta_1(\mathcal{K}^{-1}\widehat{\mathbf{v}}, \widehat{\mathbf{v}}) \cdot$$

□

Lemma 5.4. *Let $\widehat{\mathbf{V}}_h^d$ be the discontinuous Raviart–Thomas space on rectangular or triangular elements. If $\mathcal{K} \in W^{1,\infty}$, then there exists a constant C independent of h such that*

$$a(\mathbf{u}, (I - \boldsymbol{\gamma}_h)\widehat{\mathbf{w}}_h) \leq Ch||\mathbf{u}||_1||\widehat{\mathbf{w}}_h||_0, \qquad \textit{for all } \widehat{\mathbf{w}}_h \in \widehat{\mathbf{V}}_h^d, \tag{5.3}$$

and for all $\mathbf{u}$ such that $\mathbf{u} \in \mathbf{H}^1$.

Proof. Let $\mathbf{Q}_h : (L^2(\Omega))^2 \mapsto \mathbf{V}_h^d$ be the $(L^2(\Omega))^2$ projection onto the space of piecewise constant vectors and consider the coefficient $\mathcal{K}_0$, the piecewise constant average of $\mathcal{K}$. One has,

$$\begin{aligned} a(\mathbf{u}, (I-\boldsymbol{\gamma}_h)\widehat{\mathbf{w}}_h) &= ((\mathcal{K}^{-1} - \mathcal{K}_0^{-1})\mathbf{u}, (I-\boldsymbol{\gamma}_h)\widehat{\mathbf{w}}_h) + (\mathcal{K}_0^{-1}\mathbf{u}, (I-\boldsymbol{\gamma}_h)\widehat{\mathbf{w}}_h) \\ &\le Ch\|\mathbf{u}\|_0\|\widehat{\mathbf{w}}_h\|_0 + (\mathbf{u}, \mathcal{K}_0^{-1}(I-\boldsymbol{\gamma}_h)\widehat{\mathbf{w}}_h) \\ &= Ch\|\mathbf{u}\|_0\|\widehat{\mathbf{w}}_h\|_0 + (\mathbf{u}, (I-\boldsymbol{\gamma}_h)\mathcal{K}_0^{-1}\widehat{\mathbf{w}}_h) \\ &= Ch\|\mathbf{u}\|_0\|\widehat{\mathbf{w}}_h\|_0 + (\mathbf{u}-\mathbf{Q}_h\mathbf{u}, (I-\boldsymbol{\gamma}_h)\mathcal{K}_0^{-1}\widehat{\mathbf{w}}_h) \\ &\le Ch\|\mathbf{u}\|_0\|\widehat{\mathbf{w}}_h\|_0 + Ch|\mathbf{u}|_1\|\widehat{\mathbf{w}}_h\|_0 \\ &\le Ch\|\mathbf{u}\|_1\|\widehat{\mathbf{w}}_h\|_0. \end{aligned}$$

Here we used the symmetry and boundedness of $\boldsymbol{\gamma}_h$ and that $\boldsymbol{\gamma}_h\mathbf{Q}_h\mathbf{u} = \mathbf{Q}_h\mathbf{u}$. Also, we used the fact that $\mathcal{K}_0^{-1}\boldsymbol{\gamma}_h\widehat{\mathbf{w}}_h = \boldsymbol{\gamma}_h\mathcal{K}_0^{-1}\widehat{\mathbf{w}}_h$, and that the coefficient $\mathcal{K}^{-1} \in W^{1,\infty}$. □

6. Some particular examples of covolume schemes

We are now in a position to formulate two main covolume schemes: one symmetric and the other nonsymmetric. Consider the space $\mathbf{V}_h'$ that is a certain collection of piecewise polynomials (constants or linear functions) with continuous normal components across the edges E of the primal elements $T \in \mathcal{T}_h$. In our application, $\mathbf{V}_h'$ will be either $\overline{\mathbf{V}}_h = \boldsymbol{\gamma}_h\widehat{\mathbf{V}}_h$ or $\widehat{\mathbf{V}}_h$, the lowest order Raviart–Thomas space.

Define the bilinear form $\mathcal{A}(\mathbf{v}, p;\ \mathbf{w}, q)$ on $(\mathbf{V}_h', W_h) \times (\overline{\mathbf{V}}_h, W_h)$,

$$\mathcal{A}(\mathbf{v}, p;\ \mathbf{w}, q) = a(\mathbf{v}, \mathbf{w}) + b(\mathbf{w}, p) - b(\mathbf{v}, q) + c(p, q). \tag{6.1}$$

Recall that,

$$\begin{aligned} a(\mathbf{v}, \mathbf{w}) &= (\mathcal{K}^{-1}\mathbf{v}, \mathbf{w}), \\ b(\mathbf{w}, q) &= \sum_{E\in\mathcal{E}_h} \int_E \mathbf{w}\cdot\underline{n}_E[q]\, d\rho, \\ c(p, q) &= (C_h p, q) = \sum_{T\in\mathcal{T}_h} \int_{\partial T} q^i \left(p^i(\underline{b}\cdot\underline{n})^+ + p^o(\underline{b}\cdot\underline{n})^-\right)\, d\rho + (c_0 p, q). \end{aligned}$$

Let $\widehat{\mathbf{V}}_h$ be the lowest order Raviart–Thomas space on triangular or rectangular elements and W_h be the space of piecewise constants associated with the primal partition $\mathcal{T}_h$ of the given polygonal domain Ω.

- "A nonsymmetric covolume scheme": Here we let $\mathbf{V}_h' = \widehat{\mathbf{V}}_h$ and $\overline{V}_h = \boldsymbol{\gamma}_h\widehat{\mathbf{V}}_h$. Find $\widehat{\mathbf{u}}_h \in \widehat{\mathbf{V}}_h$ and $p_h \in W_h$ such that

$$\mathcal{A}(\widehat{\mathbf{u}}_h, p_h;\ \boldsymbol{\gamma}_h\widehat{\mathbf{w}}, q) = (f, q), \qquad \text{for all } \widehat{\mathbf{w}} \in \widehat{\mathbf{V}}_h,\ q \in W_h\,. \tag{6.2}$$

  (Note that the above system is equivalent to (2.9).)
- "a symmetric covolume scheme": Here we let $\mathbf{V}_h' = \overline{\mathbf{V}}_h = \boldsymbol{\gamma}_h\widehat{\mathbf{V}}_h$. Find $\widehat{\mathbf{u}}_h \in \widehat{\mathbf{V}}_h$ and $p_h \in W_h$ such that

$$\mathcal{A}(\boldsymbol{\gamma}_h\widehat{\mathbf{u}}_h, p_h;\ \boldsymbol{\gamma}_h\widehat{\mathbf{w}}, q) = (f, q), \qquad \text{for all } \widehat{\mathbf{w}} \in \widehat{\mathbf{V}}_h,\ q \in W_h\,. \tag{6.3}$$

  (Note that this system is obtained from (2.8) by restricting its trial and test spaces to $\boldsymbol{\gamma}_h\widehat{\mathbf{V}}_h$.)

The uniqueness and hence the existence of a solution of the system (6.3) was demonstrated in Lemma 4.2. For the nonsymmetric system (6.2), we have the following lemma.

Lemma 6.1. *For h sufficiently small, there is a unique $(\mathbf{u}_h, p_h) \in \widehat{\mathbf{V}}_h \times W_h$ for the system*

$$a(\mathbf{u}_h, \boldsymbol{\gamma}_h \mathbf{w}_h) + b(\gamma_h \mathbf{w}_h, p_h) = 0, \quad \forall \mathbf{w}_h \text{ in } \widehat{\mathbf{V}}_h,$$
$$b(\boldsymbol{\gamma}_h \mathbf{u}_h, q_h) - c(p_h, q_h) = -(f, q_h), \quad \forall q_h \text{ in } W_h.$$

Remark 6.1. Since

$$b(\boldsymbol{\gamma}_h \widehat{\mathbf{v}}, q) = b(\widehat{\mathbf{v}}, q) \quad \forall \widehat{\mathbf{v}} \in \widehat{\mathbf{V}}_h, \tag{6.4}$$

the above system is the same as system (6.2).

Proof. Define the bilinear form on $\widehat{\mathbf{V}}_h \times W_h$

$$\mathcal{H}(\boldsymbol{z}_h, s; \mathbf{w}_h, t) := a(\boldsymbol{z}_h, \boldsymbol{\gamma}_h \mathbf{w}_h) + b(\boldsymbol{\gamma}_h \mathbf{w}_h, s) - b(\boldsymbol{\gamma}_h \boldsymbol{z}_h, t) + c(s, t).$$

Obviously, the above system is equivalent to

$$\mathcal{H}(\mathbf{u}_h, p_h; \mathbf{w}_h, q_h) = \phi(\mathbf{w}_h, q_h) \qquad \forall (\mathbf{w}_h, q_h) \in \widehat{\mathbf{V}}_h \times W_h,$$

where

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

is a linear functional on $\widehat{\mathbf{V}}_h \times W_h$. By Lemma 4.1

$$\mathcal{H}(\mathbf{w}_h, q_h; \mathbf{w}_h, q_h) = a(\mathbf{w}_h, \gamma_h \mathbf{w}_h) + ((\alpha + \frac{1}{2}\nabla \cdot \underline{b}) q_h, q_h) + \frac{1}{2} \|\!| q_h \|\!|^2. \tag{6.5}$$

Here,

$$\|\!| w \|\!|^2 \equiv \sum_{E \in \mathcal{E}_h} \int_E [w]^2 |\underline{b} \cdot \underline{n}| \, d\rho \cdot$$

It suffices to show that $\mathcal{H}(\mathbf{w}_h, q_h; \mathbf{w}_h, q_h) = 0$ admits only a zero solution, which can be inferred by the coercivity of $\mathcal{H}(\mathbf{w}_h, q_h; \mathbf{w}_h, q_h)$ implied by Lemma 5.3 and (2.3). □

We now provide error estimates for the systems (6.2) and (6.3)

Theorem 6.1. *Assume the coefficient $K \in W^{1,\infty}$. Let $\mathbf{u}$ be the solution of the weak form* (2.5)*, and let $\mathbf{u}_h$ be the solution of either the symmetric method* (6.2) *or nonsymmetric covolume method* (6.3)*. Then there exist constants $C_1 > 0$ and $C_2 \geq 0$ independent of h*

$$\|\mathbf{u} - \mathbf{u}_h\|_0 + \|p - p_h\|_0 \leq C_1 h(\|\mathbf{u}\|_1 + \|p\|_1) + C_2 h^{1/2} ||p||_1 \tag{6.6}$$

provided that $\mathbf{u} \in \mathbf{H}^1$ and $p \in H^1$. Furthermore, the constant C_2 can be taken as zero in the case of pure diffusion problems; i.e., $\underline{b} = 0$ in (2.1).

Proof. The proof is simple but long. The basic idea is to first prove the error estimate for the nonsymmetric problem by comparing it with a standard mixed finite element method, and then prove the error estimate for the symmetric problem by comparing it with the nonsymmetric problem.

Introduce the auxiliary mixed formulation to (2.4): find $(\tilde{\mathbf{u}}_h, \tilde{p}_h) \in \widehat{V}_h \times W_h$ such that

$$a(\tilde{\mathbf{u}}_h, \mathbf{w}_h) + b(\gamma_h \mathbf{w}_h, \tilde{p}_h) = 0, \quad \forall \mathbf{w}_h \text{ in } \widehat{V}_h \tag{6.7}$$

$$b(\gamma_h \tilde{\mathbf{u}}_h, q_h) - c(\tilde{p}_h, q_h) = -(f, q_h), \quad \forall q_h \text{ in } W_h. \tag{6.8}$$

Let

$$\tilde{\mathcal{H}}(\mathbf{z}_h, s; \mathbf{w}_h, t) := a(\mathbf{z}_h, \mathbf{w}_h) + b(\gamma_h \mathbf{w}_h, s) - b(\gamma_h \mathbf{z}_h, t) + c(s, t) \tag{6.9}$$

be a bilinear form on $\widehat{V}_h \times W_h$. Then (6.7)–(6.8) is equivalent to the problem of finding $(\tilde{\mathbf{u}}_h, \tilde{p}_h) \in \widehat{V}_h \times W_h$ such that

$$\tilde{\mathcal{H}}(\tilde{\mathbf{u}}_h, \tilde{p}_h; \mathbf{w}_h, q_h) = \phi(\mathbf{w}_h, q_h) \qquad \forall (\mathbf{w}_h, q_h) \in \widehat{V}_h \times W_h, \tag{6.10}$$

where

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

is a linear functional on $\widehat{V}_h \times W_h$.

Once again if we observe (6.4), then this system has the following convergence result (See equation (5.5) in [23])

$$||\mathbf{u} - \tilde{\mathbf{u}}_h||_0 + ||p - \tilde{p}_h||_0 \leq C_1 h(||\mathbf{u}||_1 + \|p\|_1) + C_2 h^{1/2} ||p||_1 \tag{6.11}$$

provided that $\mathbf{u} \in \mathbf{H}^1, p \in H^1$. Here C_2 can be taken to be zero for the pure diffusion problem (i.e., when $\underline{b} = 0$).

On the other hand, consider the nonsymmetric problem (6.2) of finding $(\bar{\mathbf{u}}_h, \bar{p}_h) \in \widehat{V}_h \times W_h$ such that

$$\bar{\mathcal{H}}(\bar{\mathbf{u}}_h, \bar{p}_h; \mathbf{w}_h, q_h) = \phi(\mathbf{w}_h, q_h) \qquad \forall (\mathbf{w}_h, q_h) \in \widehat{V}_h \times W_h, \tag{6.12}$$

where

$$\bar{\mathcal{H}}(\boldsymbol{z}_h, s; \mathbf{w}_h, t) := a(\boldsymbol{z}_h, \gamma_h \mathbf{w}_h) + b(\gamma_h \mathbf{w}_h, s) - b(\gamma_h \mathbf{z}_h, t) + c(s, t). \tag{6.13}$$

Note that the right-hand side is exactly $\mathcal{A}(\bar{\mathbf{u}}_h, \bar{p}_h; \gamma_h \mathbf{w}_h, q_h)$, where $\mathcal{A}$ is as in (6.2). Using the bilinearity, (6.12), (6.10), we have

$$\begin{aligned} \bar{\mathcal{H}}(\tilde{\mathbf{u}}_h - \bar{\mathbf{u}}_h, \tilde{p}_h - \bar{p}_h; \mathbf{w}_h, q_h) &= \bar{\mathcal{H}}(\tilde{\mathbf{u}}_h, \tilde{p}_h; \mathbf{w}_h, q_h) - \bar{\mathcal{H}}(\bar{\mathbf{u}}_h, \bar{p}_h; \mathbf{w}_h, q_h) \\ &= \bar{\mathcal{H}}(\tilde{\mathbf{u}}_h, \tilde{p}_h; \mathbf{w}_h, q_h) - \tilde{\mathcal{H}}(\tilde{\mathbf{u}}_h, \tilde{p}_h; \mathbf{w}_h, q_h). \end{aligned}$$

Hence by (6.13) and (6.9) we have

$$\bar{\mathcal{H}}(\tilde{\mathbf{u}}_h - \bar{\mathbf{u}}_h, \tilde{p}_h - \bar{p}_h; \mathbf{w}_h, q_h) = a(\tilde{\mathbf{u}}_h, \gamma_h \mathbf{w}_h) - a(\tilde{\mathbf{u}}_h, \mathbf{w}_h). \tag{6.14}$$

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

$$\bar{\mathcal{H}}(\tilde{\mathbf{e}}_h, \tilde{\tau}_h; \tilde{\mathbf{e}}_h, \tilde{\tau}_h) = a(\tilde{\mathbf{u}}_h, (\gamma_h - I)\tilde{\mathbf{e}}_h) \tag{6.15}$$

$$= a(\tilde{\mathbf{u}}_h - \mathbf{u}, (\gamma_h - I)\tilde{\mathbf{e}}_h) + a(\mathbf{u}, (\gamma_h - I)\tilde{\mathbf{e}}_h) \tag{6.16}$$

$$\leq C||\tilde{\mathbf{u}}_h - \mathbf{u}||_0 ||\tilde{\mathbf{e}}_h||_0 + Ch||\mathbf{u}||_1 ||\tilde{\mathbf{e}}_h||_0 \tag{6.17}$$

$$\leq \left[Ch(||\mathbf{u}||_1 + \|p\|_1) + C_2 h^{1/2} ||p||_1\right] ||\tilde{\mathbf{e}}_h||_0, \tag{6.18}$$

where we have used (5.3) in deriving (6.17), and (6.11) in deriving (6.18). Applying (6.5) to the left side of (6.15), we get from (6.18) that

$$a(\tilde{\mathbf{e}}_h, \gamma_h \tilde{\mathbf{e}}_h) + ((\alpha + \frac{1}{2}\nabla \cdot \underline{b})\tilde{\tau}_h, \tilde{\tau}_h) + \frac{1}{2}|||\tilde{\tau}_h|||^2 \tag{6.19}$$

$$\leq \left[Ch\left(||\mathbf{u}||_1 + \|p\|_1\right) + C_2 h^{1/2}||p||_1\right] ||\tilde{\mathbf{e}}_h||_0. \tag{6.20}$$

Invoking (2.3) completes the estimate on $\bar{\mathbf{u}}_h$ and

$$||\mathbf{u} - \bar{\mathbf{u}}_h||_0 + ||p - \bar{p}_h||_0 \leq (Ch(||\mathbf{u}||_1 + \|p\|_1) + C_2 h^{1/2}||p||_1) \tag{6.21}$$

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

Now the covolume method (6.3) is equivalent to the problem of

$$\mathcal{H}(\mathbf{u}_h, p_h; \mathbf{w}_h, q_h) = \phi(\mathbf{w}_h, q_h) \qquad \forall (\mathbf{w}_h, q_h) \in \widehat{V}_h \times W_h, \tag{6.22}$$

where

$$\mathcal{H}(\boldsymbol{z}_h, s; \mathbf{w}_h, t) := a(\gamma_h \boldsymbol{z}_h, \gamma_h \mathbf{w}_h) + b(\gamma_h \mathbf{w}_h, s) - b(\gamma_h \boldsymbol{z}_h, t) + c(s, t). \tag{6.23}$$

Using the bilinearity, (6.22), (6.23), we have

$$\begin{aligned} \mathcal{H}(\mathbf{u}_h - \bar{\mathbf{u}}_h, p_h - \bar{p}_h; \mathbf{w}_h, q_h) &= \mathcal{H}(\mathbf{u}_h, p_h; \mathbf{w}_h, q_h) - \mathcal{H}(\bar{\mathbf{u}}_h, \bar{p}_h; \mathbf{w}_h, q_h) \\ &= \bar{\mathcal{H}}(\bar{\mathbf{u}}_h, \bar{p}_h; \mathbf{w}_h, q_h) - \mathcal{H}(\bar{\mathbf{u}}_h, \bar{p}_h; \mathbf{w}_h, q_h) \\ &= a(\bar{\mathbf{u}}_h, \gamma_h \mathbf{w}_h) - a(\gamma_h \bar{\mathbf{u}}_h, \gamma_h \mathbf{w}_h). \end{aligned}$$

Now set $\mathbf{w}_h = \bar{\mathbf{e}}_h := \bar{\mathbf{u}}_h - \mathbf{u}_h$ and $q_h = \bar{\tau}_h := \bar{p}_h - p_h$ in the above equation to get the error equation

$$\begin{aligned} \mathcal{H}(\bar{\mathbf{e}}_h, \bar{\tau}_h; \bar{\mathbf{e}}_h, \bar{\tau}_h) &= a\left((I - \gamma_h)(\bar{\mathbf{u}}_h - \mathbf{u}), \gamma_h \bar{\mathbf{e}}_h\right) \\ &\quad + a\left((I - \gamma_h)\mathbf{u}, \gamma_h \bar{\mathbf{e}}_h\right) && (6.24) \\ &\leq C||\bar{\mathbf{u}}_h - \mathbf{u}||_0 ||\bar{\mathbf{e}}_h||_0 + Ch||\mathbf{u}||_1 ||\bar{\mathbf{e}}_h||_0 && (6.25) \\ &\leq \left[Ch(||\mathbf{u}||_1 + \|p\|_1) + C_2 h^{1/2}||p||_1\right] ||\bar{\mathbf{e}}_h||_0. && (6.26) \end{aligned}$$

As before,

$$a(\gamma_h \bar{\mathbf{e}}_h, \gamma_h \bar{\mathbf{e}}_h) + ((\alpha + \frac{1}{2}\nabla \cdot \underline{b})\bar{\tau}_h, \bar{\tau}_h) + \frac{1}{2}|||\bar{\tau}_h|||^2 \tag{6.27}$$

$$\leq \left[Ch(||\mathbf{u}||_1 + \|p\|_1) + C_2 h^{1/2}||p||_1\right] ||\bar{\mathbf{e}}_h||_0. \tag{6.28}$$

Invoking (2.3) and coercivity completes the estimate on $\mathbf{u}_h$. □

Remark 6.2. We note that

$$\|\mathbf{u} - \gamma_h \mathbf{u}_h\|_0 + \|p - p_h\|_0 \leq C_1 h(\|\mathbf{u}\|_1 + \|p\|_1) + C_2 h^{1/2}||p||_1, \tag{6.29}$$

which is obtained by simply observing that

$$||\mathbf{u}_h - \gamma_h \mathbf{u}_h||_0 \leq ||\mathbf{u}_h - \mathbf{u}||_0 + ||\mathbf{u} - \gamma_h \mathbf{u}||_0 + ||\gamma_h(\mathbf{u} - \mathbf{u}_h)||_0.$$

7. Numerical experiments

In this section we present numerical results that illustrate the error behavior of the studied mixed covolume method for two cases: the pure diffusion problem ($\underline{b} = 0$ and $c_0 = 0$ in (2.1)) and the convection dominated problem. Extensive tests on the nonsymmetric (i.e., Petrov–Galerkin) scheme using rectangular elements for convection-diffusion problems on axi-parallel domains have been presented in Chou, Kwak and Vassilevski ([17]).

Numerical tests for diffusion problems. Here we used the proper subspace $\gamma_h \widehat{\mathbf{V}}_h$ as our discretization space. The problem was

$$\nabla \cdot (-\mathcal{K}\nabla p) = f(x,y), \quad (x,y) \in \Omega = (0,1)^2. \tag{7.1}$$

The exact solution was chosen $p = x(1-x)y(1-y)$ and Dirichlet boundary conditions were imposed. The coefficients of the operator were

$$\mathcal{K} = \begin{bmatrix} 1+10x^2+y^2 & \frac{1}{2}+x^2+y^2 \\ \frac{1}{2}+x^2+y^2 & 1+x^2+10y^2 \end{bmatrix}.$$

For the velocity variable $\underline{u} = (u_1, u_2)$ we used the special piecewise constant vectors that corresponded to $\gamma_h \underline{v}$ for $\underline{v}$ in the lowest order Raviart–Thomas piecewise polynomial space $\widehat{\mathbf{V}}_h$ on isosceles right-angled triangles of size h, for $h = 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}$.

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

$$\mathcal{A} \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \mathbf{P} \end{bmatrix} = \underline{\mathbf{f}} = \begin{bmatrix} \boldsymbol{rhs}_{\mathbf{U}_1} \\ \boldsymbol{rhs}_{\mathbf{U}_2} \\ \boldsymbol{rhs}_P \end{bmatrix}, \tag{7.2}$$

with the saddle-point like stiffness matrix

$$\mathcal{A} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}. \tag{7.3}$$

We used the fact that $\mathcal{A}$ satisfies the inf-sup condition,

$$\sup_{\gamma_h \underline{v}, q} \frac{\mathcal{A}(\gamma_h \underline{u}, p;\ \gamma_h \underline{v}, q)}{\left[\|\underline{v}\|^2_{H(\mathrm{div})} + \|q\|^2_0\right]^{\frac{1}{2}}} \geq \beta \left[\|\underline{u}\|^2_{H(\mathrm{div})} + \|p\|^2_0\right]^{\frac{1}{2}}, \quad \text{for all } (\underline{u},\ p) \in \mathbf{V}_h \times W_h, \tag{7.4}$$

which in matrix form reduces to the spectral equivalence relations

$$(\mathcal{A}^T \mathcal{A}_0^{-1} \mathcal{A} \underline{\boldsymbol{x}}, \underline{\boldsymbol{x}}) \geq \beta(\mathcal{A}_0 \underline{\boldsymbol{x}}, \underline{\boldsymbol{x}}), \quad \text{for all } \underline{\boldsymbol{x}} = (\mathbf{U}_1, \mathbf{U}_2, \mathbf{P}). \tag{7.5}$$

Here, $\mathcal{A}_0 = \begin{bmatrix} A_0 & 0 \\ 0 & I \end{bmatrix}$, where A_0 corresponds to the stiffness matrix arising from the $H(\mathrm{div})$-bilinear form $\int \mathcal{K}^{-1} \underline{u} \cdot \underline{v} + \int \mathrm{div}\, \underline{u}\ \mathrm{div}\, \underline{v}$, restricted to the Raviart–Thomas space for the velocity variable.

Now any preconditioner M of optimal order for A_0 will define an optimal order preconditioner $\mathcal{M} = \begin{bmatrix} M & 0 \\ 0 & I \end{bmatrix}$ for $\mathcal{A}$. Since $\mathcal{A}$ is symmetric but indefinite, one can either use $\mathcal{M}$ as a preconditioner in the MINRES method for $\mathcal{A}$, or one can use $\mathcal{M}$ as a preconditioner to $\mathcal{A}^T \mathcal{M}^{-1} \mathcal{A}$ in the standard CG method. Alternatively, one may use the transformation proposed in [5] to have the saddle-point problem coercive in a certain inner product. In our experiments we have chosen the preconditioned MINRES method. The search vectors in the MINRES method were constructed orthogonally in the $(\mathcal{M}^{-1}\mathcal{A}\cdot, \mathcal{A}\cdot)$-inner product.

Choices of M, a preconditioner for the $H(\mathrm{div})$-bilinear form are found in [2, 7, 30]. In the experiments reported in Table 1, we used an algebraically stabilized version of the hierarchical method from [7]. Details on the algebraic stabilization of the HB methods are found in [31].

TABLE 1. Error behavior and iteration counts for the covolume scheme

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$\approx$ order
δ_p	2.30e-4	5.78e-5	1.44e-5	3.62e-6	2
δ_{u_1}	4.48e-3	1.12e-3	2.87e-4	7.35e-5	2
δ_{u_2}	4.48e-3	1.12e-3	2.87e-4	7.35e-5	2
$\delta_{u_{\text{int}}}$	1.96e-3	5.19e-4	1.35e-4	3.52e-5	2
# unknowns	1312	5184	20 608	82 176	
# iterations	29	30	31	31	
ϱ	0.47	0.48	0.49	0.50	
κ	2.00	2.09	2.20	2.25	

The stopping criterion in the MINRES method was

$$\|\mathcal{M}^{-\frac{1}{2}}\mathcal{A}\mathbf{r}\| \leq 10^{-9}\|\mathcal{M}^{-\frac{1}{2}}\mathcal{A}\mathbf{r}_0\|,$$

where $\|\mathbf{v}\|^2 = \mathbf{v}^T\mathbf{v}$, and $\mathbf{r}_0$ stands for the initial residual, $\mathbf{r}$ is the current one. The initial iterate was chosen as $\mathbf{x}_0 = \mathcal{M}^{-1}\mathbf{f}$, where $\mathbf{f}$ was the right-hand side of the discrete problem $\mathcal{A}\mathbf{x} = \mathbf{f}$.

We show in Table 1, in addition to the error behavior of the covolume discretization method, also ϱ, κ and the number of iterations, where

$$\varrho = \left(\frac{\|\mathcal{M}^{-\frac{1}{2}}\mathcal{A}\boldsymbol{r}\|}{\|\mathcal{M}^{-\frac{1}{2}}\mathcal{A}\boldsymbol{r}_0\|}\right)^{\frac{1}{\#\text{ iterations}}} \tag{7.6}$$

was an average reduction factor, and κ was the condition number of $\mathcal{M}^{-1}\mathcal{A}_0$. Recall that $\mathcal{A}_0 = \begin{bmatrix} A_0 & 0 \\ 0 & I \end{bmatrix}$, where A_0 stands for the matrix corresponding to the $H(\text{div})$-bilinear form $(\mathcal{K}^{-1}\underline{u}, \underline{v}) + (\text{div}\,\underline{u}, \text{div}\,\underline{v})$ computed from the triangular Raviart–Thomas velocity space.

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 Table 1, we show:

(i)

$$\begin{aligned}\delta_p &= \|I_h p - p_h\|_h \\ &\equiv \left[\sum_{i=1}^{n_x}\sum_{j=1}^{n_y} h_x h_y (p(x_i - \frac{1}{2}h_x, y_j - \frac{1}{2}h_y) - p_h(x_i - \frac{1}{2}h_x, y_j - \frac{1}{2}h_y))^2\right]^{\frac{1}{2}},\end{aligned}$$

i.e., a discrete L^2-norm of the error $p - p_h$;

(ii)

$$\begin{aligned}\delta_{u_1} &= \|I_h u_1 - u_{h,\,1}\|_h \\ &\equiv \left[\sum_{i=0}^{n_x}\sum_{j=1}^{n_y} h_x h_y (u_1(x_i, y_j - \frac{1}{2}h_y) - u_{h,\,1}(x_i, y_j - \frac{1}{2}h_y))^2\right]^{\frac{1}{2}},\end{aligned}$$

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)

$$\delta_{u_{\text{int}}} = \|I_h(\mathbf{u} - \mathbf{u}_h)\|_h$$
$$\equiv \left[\sum_{i=1}^{n_x}\sum_{j=0}^{n_y} h_x h_y \left((\mathbf{u}\cdot\mathbf{n})(x_i - \frac{1}{2}h_x, y_j - \frac{1}{2}h_y) - (\mathbf{u}_h\cdot\mathbf{n})(x_i - \frac{1}{2}h_x, y_j - \frac{1}{2}h_y)\right)^2\right]^{\frac{1}{2}},$$

i.e., a discrete L^2-norm of the error $\mathbf{u}\cdot\mathbf{n} - \mathbf{u}_h\cdot\mathbf{n}$, where $\mathbf{n}$ is the unit normal vector to the edge (x_{i-1}, y_{j-1}), (x_i, y_j);

(v) the number of iterations of the preconditioned MINRES method;

(vi) the average reduction factors ϱ, (7.6);

(vii) the condition number κ of $\mathcal{M}^{-1}\mathcal{A}_0$;

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

It turns out, that our experiments suggest second-order approximation for all variables, i.e., a superconvergence behavior of the covolume scheme. Notice also the constant number of iterations (and corresponding average reduction factors ϱ) in the preconditioned MINRES method.

Numerical tests for convection-diffusion problems. In this section we consider the error behavior of the problem

$$\nabla\cdot(-\epsilon\mathcal{K}\nabla p + \underline{b}p) + c_0 p = f(x,y), \quad (x,y)\in\Omega = (0,1)^2. \tag{7.7}$$

The exact solution chosen is $p = x(1-x)y(1-y)$, and Dirichlet boundary conditions are imposed. The coefficients of the operator are $\mathcal{K} = (k_{ij})$, $k_{11} = 1 + 10x^2 + y^2$, $k_{12} = k_{21} = \frac{1}{2} + x^2 + y^2$, and $k_{22} = 1 + x^2 + 10y^2$, $c_0 = 1$ and $\underline{b} = (b_1, b_2)^T$, where

$$\begin{aligned} b_1 &= -\cos\alpha(1 - x\cos\alpha),\\ b_2 &= -\sin\alpha(1 - y\sin\alpha), \end{aligned} \tag{7.8}$$

for angle $\alpha = -\frac{3}{4}\pi$.

Note that $\nabla\cdot\underline{b} = 1$ so that the condition (2.3) is satisfied, i.e., $c_0 + \frac{1}{2}\nabla\cdot\underline{b} \geq \gamma_0 > 0$. We have chosen in the experiments $\epsilon = 1, 10^{-2}$, and $\epsilon = 10^{-3}$.

For the discretization we use the nonsymmetric (Petrov–Galerkin) scheme, i.e., the test spaces are $\gamma_h\widehat{\mathbf{V}}_h$ and W_h, whereas the trial space is $(\widehat{\mathbf{V}}_h,\ W_h)$—the lowest order Raviart–Thomas spaces on squares of size $h = 2^{-\ell}$, $\ell = 4, 5, \ldots, 8$. The saddle-point system of equations is solved via inner-outer iteration method as described in [4], where the major block A is solved within certain high accuracy (the relative residual reduction tolerance was $\epsilon_1 = 10^{-6}$) and the approximately obtained Schur complement system is solved by inner iterations using a steepest descent type iteration method (as described in [4]) using a block-ILU preconditioner

TABLE 2. Error behavior and iteration counts for the covolume scheme; $\epsilon = 0.001$, $\alpha = -\frac{3}{4}\pi$

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$h = 1/256$	$\approx$ order
δ_p	3.47e-3	1.78e-3	9.06e-4	4.61e-4	2.37e-4	1
δ_{u_1}	1.54e-4	8.10e-5	4.25e-5	2.21e-5	1.14e-5	1
δ_{u_2}	6.54e-4	3.82e-4	2.03e-4	1.02e-4	5.08e-5	1
# unknowns	800	3136	12 416	49 408	197120	
# outer it.	6	5	6	6	6	
# inner it.	2	3	4	5	8	
ϱ	7.96e-2	5.27e-2	8.43e-2	8.11e-2	8.74e-2	

TABLE 3. Error behavior and iteration counts for the covolume scheme; $\epsilon = 0.01$, $\alpha = -\frac{3}{4}\pi$

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$h = 1/256$	$\approx$ order
δ_p	2.91e-3	1.57e-3	8.39e-4	4.42e-4	2.30e-4	1
δ_{u_1}	8.14e-4	4.18e-4	2.16e-4	1.11e-4	5.68e-5	1
δ_{u_2}	2.22e-3	1.11e-3	5.52e-4	2.74e-4	1.36e-4	1
# unknowns	800	3 136	12 416	49 408	197 120	
# outer it.	6	6	7	7	7	
# inner it.	4	4	6	14	36	
ϱ	0.08	0.06	0.11	0.12	0.13	

TABLE 4. Error behavior and iteration counts for the covolume scheme; $\epsilon = 1$, $\alpha = -\frac{3}{4}\pi$

	$h = 1/16$	$h = 1/32$	$h = 1/64$	$h = 1/128$	$h = 1/256$	$\approx$ order
δ_p	1.30e-4	8.36e-5	4.72e-5	2.49e-5	1.28e-5	1
δ_{u_1}	5.32e-3	2.10e-3	9.20e-4	4.30e-4	2.08e-4	1
δ_{u_2}	6.06e-3	2.42e-3	1.06e-3	4.98e-4	2.41e-4	1
# unknowns	800	3 136	12 416	49 408	197 120	
# outer it.	6	7	7	7	8	
# inner it.	6	8	24	88	324	
ϱ	0.09	0.09	0.12	0.13	0.135	

for the five-point cell-centered finite difference scheme, obtained by diagonalizing the weighted mass-matrix based on the diagonal part of the coefficient $\mathcal{K}$. The relative residual reduction tolerance here was $\epsilon_2 = 10^{-2}$. The outer iterations were of defect correction type until the relative residual reduction of 10^{-6} was reached. The particular tolerances and respective number of iterations (denoted in Tables 2–4 by "# outer it." and "# inner it.") are not of major interest here since we are mainly interested in the error behavior of the discrete solutions. Our experiments for the tests presented suggest the approximation is of first order.

ACKNOWLEDGMENTS

The authors are thankful to Professor Junping Wang for providing reference [23].

References

[1] D. N. Arnold and R. S. Falk, *A uniformly accurate finite element method for the Reissner–Mindlin plate*, SIAM J. Numer. Anal. **26** (1989), 1276–1290. MR **91c:**65068

[2] D. N. Arnold, R. S. Falk and R. Winther, *Preconditioning in H(div) and applications*, Math. Comp. **66** (1997), 957-984. MR **97i:**65177

[3] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994. MR **95f:**65005

[4] O. Axelsson and P. S. Vassilevski, "*Construction of variable-step preconditioners for inner-outer iteration methods*", Proceedings of IMACS Conference on Iterative methods, April 1991, Brussels, Belgium, *Iterative Methods in Linear Algebra*, (R. Beauwens and P. de Groen, eds.), North Holland, 1992, 1-14. MR **92m:**65053

[5] J. H. Bramble and J. E. Pasciak, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp. **50** (1988), 1–17. MR **89m:**65097a

[6] J. H. Bramble, J. E. Pasciak, and A. T. Vassilev, *Analysis of inexact Uzawa algorithm for saddle point problems*, SIAM J. Numer. Anal. **34** (1997), 1072–1092. MR **98c:**65182

[7] Z. Cai, C. I. Goldstein and J. E. Pasciak, *Multilevel iteration for mixed finite element systems with penalty*, SIAM J. Sci. Comput. **14** (1993), 1072–1088. MR **94h:**65116

[8] Z. Cai, J. E. Jones, S. F. McCormick and T. F. Russell, *Control-Volume mixed finite element methods*, Computational Geosciences **1** (1997), 289–315.

[9] G. F. Carey, A. I. Pehlivanov, and P. S. Vassilevski, *Least–squares mixed finite elements for nonselfadjoint elliptic problems: II, Performance of block–ILU factorization methods*, SIAM J. Sci. Comput. **16** (1995), 1126–1136. MR **97f:**65069

[10] J. C. Cavendish, C. A. Hall and T. A. Porsching, *A complementary volume approach for modeling three-dimensional Navier-Stokes equations using dual Delaunay/Voronoi tessellations*, Internat. J. Numer. Methods Heat Fluid Flow, **4** (1994) 329–345. MR **95d:**76088

[11] Q. Du, R. Nicolaides, and X. Wu, *Analysis and convergence of a covolume approximation of the Ginzburg-Landau model of superconductivity*, SIAM J. Num. Anal. **35** (1998), 1049-1072. CMP 98:11

[12] S. H. Chou, *Analysis and convergence of a covolume method for the generalized Stokes problem*, Math. Comp. **66** (1997), 85-104. MR **97e:**65109

[13] S. H. Chou and D. Y. Kwak, *Analysis and convergence of a MAC-like scheme for the generalized Stokes problem*, Numer. Meth. Partial Diff. Eqns. **13** (1997), 147–162. MR **98a:**65154

[14] S. H. Chou and D. Y. Kwak, *Mixed covolume methods on rectangular grids for elliptic problems*, SIAM J. Num. Anal. (1998), to appear.

[15] 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), submitted.

[16] 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), 497-507. CMP 98:11

[17] S. H. Chou, D. Y. Kwak and P. Vassilevski, *Mixed covolume methods on rectangular grids for convection dominated problems*, SIAM J. Sci. Computing, (1998), to appear.

[18] S. H. Chou, D. Y. Kwak and P. Vassilevski, *Mixed covolume methods for elliptic problems on triangular grids*, SIAM J. Numer. Anal. **35** (1998), 1850–1861. CMP 98:17

[19] S. H. Chou and P. Vassilevski, *An upwinding cell–centered method with piecewise constant velocity over covolumes*, Numer. Meth. Partial Diff. Eqns. (1997), to appear.

[20] C. A. Hall and T. A. Porsching, *A characteristic-like method for thermally expandable flow on unstructured triangular grids*, Internat. J. Numer. Methods Fluids **22** (1996), 731–754. MR **97d:**76029

[21] 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).

[22] 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), 2181.

[23] M. Liu, J. Wang and N. Yan, *New error estimates for approximate solutions of convection-diffusion problems by mixed and discontinuous Galerkin methods*, (1997) preprint.

[24] 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 ed., John Wiley and Sons, (1995), 279–299.

[25] R. Nicolaide and X. Wu, *Covolume solutions of three-dimensional div-curl equations*, SIAM. J. Numer. Anal. **34** (1997), 2195–2203. MR **98f:**65096

[26] T. Rusten and R. Winther, *A preconditioned iterative method for saddlepoint problems*, SIAM J. Matrix Anal. Appl. **13** (1992), 887–904. MR **93a:**65043

[27] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PSW Kent, 1995.

[28] D. Silvester and A. Wathen, *Fast iterative solution of stabilised Stokes systems, II. Using general block preconditioners.* SIAM J. Numer. Anal. **31** (1994), 1352–1367. MR **95g:**65132

[29] P. S. Vassilevski and R. D. Lazarov, *Preconditioning mixed finite element saddle-point elliptic problems*, Numer. Linear Alg. Appl. **3** (1996), 1–20. MR **96m:**65111

[30] P. S. Vassilevski and J. Wang, *Multilevel iterative methods for mixed finite element discretizations of elliptic problems*, Numer. Math. **63** (1992), 503–520. MR **93j:**65187

[31] P. S. Vassilevski, *On two ways of stabilizing the hierarchical basis methods*, SIAM Rev. **39** (1997), 18–53. MR **98a:**65178

Department of Mathematics, Bowling Green State University, Bowling Green, Ohio 43403, U.S.A.

E-mail address: chou@zeus.bgsu.edu

Center of Informatics and Computing Technology, Bulgarian Academy of Sciences, "Acad. G. Bontchev" street, Block 25 A, 1113 Sofia, Bulgaria

E-mail address: panayot@iscbg.acad.bg