Advances in Computational Mathematics **19:** 3–33, 2003.

Development and analysis of higher order finite volume methods over rectangles for elliptic equations

Zhiqiang Cai *, Jim Douglas, Jr., and Moongyu Park

Department of Mathematics, Purdue University, 1395 Mathematical Science Building, West Lafayette, IN 47907-1395, USA

Received 29 October 2001; accepted 11 September 2002
Communicated by C.A. Micchelli

Currently used finite volume methods are essentially low order methods. In this paper, we present a systematic way to derive higher order finite volume schemes from higher order mixed finite element methods. Mostly for convenience but sometimes from necessity, our procedure starts from the hybridization of the mixed method. It then approximates the inner product of vector functions by an appropriate, critical quadrature rule; this allows the elimination of the flux and Lagrange multiplier parameters so as to obtain equations in the scalar variable, which will define the finite volume method. Following this derivation with different mixed finite element spaces leads to a variety of finite volume schemes. In particular, we restrict ourselves to finite volume methods posed over rectangular partitions and begin by studying an efficient second-order finite volume method based on the Brezzi–Douglas–Fortin–Marini space of index two. Then, we present a general global analysis of the difference between the solution of the underlying mixed finite element method and its related finite volume method. Then, we derive finite volume methods of all orders from the Raviart–Thomas two-dimensional rectangular elements; we also find finite volume methods to associate with $BDFM_2$ three-dimensional rectangles. In each case, we obtain optimal error estimates for both the scalar variable and the recovered flux.

Keywords: finite volume method, higher order, rectangular element

AMS subject classification: 65N30

1. Introduction

Numerical methods for partial differential equations are divided into three general categories: finite difference methods (FDMs), finite element methods (FEMs), and finite volume methods (FVMs). Finite difference and finite element methods have received much more attention than finite volume methods and, as a result, have been developed to a higher degree of sophistication than finite volume techniques; thus, there are well-known higher order finite difference and finite element methods. In general, finite dif-

* The research of Professor Cai was sponsored in part by the National Science Foundation.

ference methods are easier to implement than finite element methods, but finite element methods are more easily adapted to general geometries of the underlying domain on which the differential problem is formulated and to the practical treatment of inhomogeneous physical properties of the media. Finite volume procedures are usually easier to implement than finite element procedures and offer most of the advantages of flexibility in geometry of these methods; in a sense, finite volume methods lie in between the other two techniques in concept and implementation.

There has been one fundamental disadvantage for finite volume methods in comparison to finite difference and finite element methods. Essentially all current finite volume methods are low order methods based on approximating the solution of the differential equations by piecewise-constant functions. Consequently, the finite volume approximation can converge to the solution of the differential problem globally only at a rate proportional to the diameter h of the elements in the partition of the domain on which the problem is set. In some cases, it has been possible to show that certain points can be identified at which the approximate solution converges at a more rapid rate than the global rate; this phenomenon is called superconvergence. In the most commonly used finite volume method based on rectangular elements for elliptic boundary value problems, convergence at a rate $\mathrm{O}(h^2)$ takes place at cell centers under some supplementary conditions. Then, second order convergence can be established for a piecewise multilinear interpolation of these cell-center values, again under proper hypotheses.

In [21], Russell and Wheeler established a relationship between a mixed finite element method on rectangular meshes and cell-centered finite differences for diffusion problems with coefficient matrix being diagonal. Specifically, they obtained a positive-definite, cell-centered finite difference method by using the lowest order Raviart–Thomas (RT_0) space (see [20]), applying an appropriate quadrature rule, and eliminating the flux. This work has been extended to triangular meshes and to a full tensor coefficient matrix (see, e.g., [1,2]). For similar work, see also the recent paper by Baranger et al. [6], where cell-centered finite difference procedures were considered to be finite volume methods. On general triangular meshes, the RT_0 space usually leads to quite complicated finite difference stencils (see [1]) since piecewise-constant pressures on two adjacent triangles are not enough to represent the flux across their common edge. These papers are based on the lowest order Raviart–Thomas space and, consequently, lead to essentially low order methods.

The object of this paper is to present examples of a systematic way for deriving higher order finite volume methods. Our derivations are similar to those mentioned above, though we find it convenient to start with the hybridization of the underlying mixed finite element method. Then, we approximate the weighted inner product of vector functions by an appropriate quadrature rule so as to diagonalize the resulting matrix A_h; this allows the immediate elimination of the flux variables. Finally, we employ the consistency relations across interfaces to permit the elimination of the Lagrange multipliers introduced in the hybridization process to obtain a system of equations in the degrees of freedom for the scalar variable in the mixed formulation of the elliptic problem; this will be our finite volume procedure. It will be clear that the resulting method maintains local conservation of whatever is being transported by the flux variable element-by-element, as does the underlying mixed finite element method.

The choice of the quadrature rule is critical in our approach, as it was in [21]. Our guidelines are that the matrix A_h corresponding to the vector inner product is diagonal and that the numerical integration does not decrease the rate of convergence of the original mixed finite element method. Finite volume schemes can be derived in numerous ways (see, e.g., [4,5,11,12,18]), but the point of our derivation is that the same method can be applied to different mixed finite element spaces to obtain a variety of finite volume procedures. In this paper, we briefly review the standard lower order finite volume method based on the rectangular RT_0 mixed finite element space and then turn first to the derivation of a higher order procedure based on the two-dimensional, Brezzi–Douglas–Fortin–Marini space of index two ($BDFM_2$, see [8]) of rectangular elements. It is worth noting that the derivation technique used here and in [21] naturally leads to the employment of the harmonic average of discontinuous diffusion coefficients, which is physically correct and often is not introduced by other derivations of finite volume schemes.

The difference between the solution of the $BDFM_2$ mixed method and its related finite volume method is discussed as a solution of a perturbed mixed method. The analysis is given in a generalized fashion that allows us to treat a variety of mixed methods and their related finite volume methods by analyzing the effect of the quadrature rule used in deriving the finite volume procedure. As a consequence of this analysis, we indicate finite volume schemes related to the Raviart–Thomas–Nedelec rectangular mixed finite elements in two or three dimensions and show that the resulting finite volume methods retain the global order of convergence of the RTN methods. We also treat the three-dimensional $BDFM_2$ case, where optimal order error estimates are obtained, and some higher order $BDFM_k$ elements, where it seems inevitable that the error in the flux variable, at least, cannot obtain the optimal order of accuracy associated with the underlying mixed method. We defer discussing simplicial methods to future work.

This paper is organized as follows. We end this section by introducing some notation. In section 2, we introduce the diffusion equation, an equivalent first-order system, and its mixed variational formulation. In section 3, the mixed finite element approximations based on the RT_0 and $BDFM_2$ are described, along with their hybridizations; sections 4, 5 and 6 are devoted to the derivation of finite volume methods from the two mixed methods. Global error estimates for finite volume methods interpretable as pertubations of mixed finite elements schemes are considered in section 7, along with the application of these estimates to the $BDFM_2$-based method. Two- and three-dimensional RTN-based finite volume methods are indicated and analyzed in section 8. The three-dimensional $BDFM_2$-based method is derived and analyzed in section 9. The $BDFM_3$ case is discussed in section 10; it is an example that so far has not been shown to lead to a satisfactory finite volume scheme derivable by the technique of this paper.

1.1. Notation

We use the standard notation and definition for the Sobolev space $W^{m,p}(B)$ that consists of functions whose partial derivatives are L^p-integrable up to order m; its standard associated norm and seminorm are denoted by $\|\cdot\|_{m,p,B}$ and $|\cdot|_{m,p,B}$, respectively. For $p = 2$, $H^m(B) = W^{m,2}(B)$ is a Hilbert space and its inner product is denoted by $(\cdot,\cdot)_{m,2,B}$. We omit the subscript B from the notation when $B = \Omega$. For $m = 0$, $H^m(B)$ coincides with $L^2(B)$. In this case, the inner product and norm will be denoted by $(\cdot,\cdot)_B$ and $\|\cdot\|_{0,2,B}$ or $(\cdot,\cdot)$ and $\|\cdot\|_{0,2}$ when $B = \Omega$, respectively. Let

$$\mathbf{V} = H(\mathrm{div};\Omega) = \left\{\mathbf{v} \in \left(L^2(\Omega)\right)^2 \colon \nabla\cdot\mathbf{v} \in L^2(\Omega)\right\};$$

$\mathbf{V}$ is a Hilbert space under the norm

$$\|\mathbf{v}\|_{\mathbf{V}}^2 = \|\mathbf{v}\|_{0,2}^2 + \|\nabla\cdot\mathbf{v}\|_{0,2}^2.$$

Set $W = L^2(\Omega)$ and $\mathcal{V} = \mathbf{V} \times W$.

2. The elliptic problem, its mixed formulation, and preliminaries

Consider the homogeneous Dirichlet probem:

$$\begin{cases} -\nabla\cdot(K\nabla p) = f & \text{in } \Omega, \\ p = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.1}$$

where the symbols $\nabla\cdot$ and ∇ stand for the divergence and gradient operators, respectively; $K \in L^\infty(\Omega)$ and $f \in L^2(\Omega)$ are given real-valued functions, and the domain Ω is the unit square $(0,1)\times(0,1)$. Assume that the diffusion coefficient K is bounded below and above by positive constants; i.e., there exist positive constants K_0 and K_1 such that

$$0 < K_0 \leqslant K(x,y) \leqslant K_1$$

for almost all $(x,y) \in \Omega$.

We introduce the flux variable

$$\mathbf{u} = -K\nabla p \quad \text{in } \Omega. \tag{2.2}$$

Set $c(x,y) = K^{-1}(x,y)$; then $0 < K_1^{-1} = c_0 \leqslant c(x,y) \leqslant c_1 = K_0^{-1}$. Then, the mixed form of (2.1) is given by

$$\begin{cases} c\mathbf{u} + \nabla p = 0 & \text{in } \Omega, \\ \nabla\cdot\mathbf{u} = f & \text{in } \Omega, \\ p = 0 & \text{on } \partial\Omega. \end{cases} \tag{2.3}$$

The mixed weak formulation of problem (2.1) is obtained by multiplying the first and second equations of (2.3) by $\mathbf{v} \in \mathbf{V}$ and $q \in W$, respectively, and integrating the two equations over Ω; thus, the weak solution of (2.3) is found by seeking $(\mathbf{u}, p) \in \mathcal{V} \equiv \mathbf{V} \times W$ such that

$$\begin{cases} (c\mathbf{u}, \mathbf{v}) - (\nabla \cdot \mathbf{v}, p) = 0, & \forall \mathbf{v} \in \mathbf{V}, \\ (\nabla \cdot \mathbf{u}, q) = (f, q), & \forall q \in W. \end{cases} \tag{2.4}$$

The analysis below will be carried out assuming the Dirichlet boundary condition imposed above; however, Neumann or Robin boundary conditions can be treated analogously.

3. The mixed finite element method and its hybridization

Let $\mathcal{T}_h$ be the partition of Ω into squares with the side length $h = N^{-1}$:

$$\overline{\Omega} = \bigcup_{i,j=1}^{N} \overline{T}_{ij},$$

where $\overline{T}_{ij} = [(i-1)h, ih] \times [(j-1)h, jh] = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$. (The extension of the methods in this paper to the connected union of rectangular elements is straightforward.) Let $\mathcal{V}_h \equiv \mathbf{V}_h \times W_h$ be an admissible finite-dimensional subspace of $\mathcal{V}$ (see, e.g., [10]). Then, the mixed finite element approximation in $\mathcal{V}_h$ is the solution $(\mathbf{u}_h, p_h) \in \mathcal{V}_h = \mathbf{V}_h \times W_h$ of

$$\begin{cases} (c\mathbf{u}_h, \mathbf{v}) - (\nabla \cdot \mathbf{v}, p_h) = 0, & \forall \mathbf{v} \in \mathbf{V}_h, \\ (\nabla \cdot \mathbf{u}_h, q) = (f, q), & \forall q \in W_h. \end{cases} \tag{3.1}$$

We find it convenient to localize the continuous mixed problem (2.3). By doing so, we can point out exactly the difference between the finite volume method we associate with the underlying mixed finite element space and the reduction to a symmetric, positive-definite algebraic system for the Lagrange multipliers introduced by Fraeijs de Veubeke [17] in his hybridization of (2.3) and analyzed in detail by Arnold and Brezzi [3].

For each element $T \in \mathcal{T}_h$, let T' be an adjacent element with a common edge $e = \partial T \cap \partial T'$. Let $\mathcal{E}_h$ be the set of all internal edges and denote restrictions of $\mathbf{u}$ and p to the element T by $\mathbf{u}_T$ and p_T, respectively. Then, solving the system

$$\begin{cases} c\mathbf{u}_T(\mathbf{x}) + \nabla p_T(\mathbf{x}) = 0, & \mathbf{x} \in T, \ \forall T \in \mathcal{T}_h, \\ \nabla \cdot \mathbf{u}_T(\mathbf{x}) = f, & \mathbf{x} \in T, \ \forall T \in \mathcal{T}_h, \\ p(\mathbf{x}) = 0, & \mathbf{x} \in \partial\Omega, \end{cases} \tag{3.2}$$

is equivalent to solving (2.3) if the consistency conditions

$$\begin{cases} p_T|_e - p_{T'}|_e = 0, & \forall e \in \mathcal{E}_h, \\ \mathbf{u}_T \cdot \mathbf{n}_T|_e + \mathbf{u}_{T'} \cdot \mathbf{n}_{T'}|_e = 0, & \forall e \in \mathcal{E}_h, \end{cases} \tag{3.3}$$

hold, where $\mathbf{n}_T$ and $\mathbf{n}_{T'}$ are the unit outward vectors normal, respectively, to T and T'.

Set

$$\mathbf{V}(T) = H(\mathrm{div}; T), \qquad W(T) = L^2(T),$$
$$\mathbf{V}_h(T) = \{\mathbf{v}|_T\colon\ \mathbf{v} \in \mathbf{V}_h\}, \qquad W_h(T) = \{q|_T\colon\ q \in W_h\}.$$

Let

$$\langle r, t\rangle_{\partial T} = \int_{\partial T} rt\, \mathrm{d}s.$$

Now, test (3.2) against $\mathbf{v} \in \mathbf{V}(T)$ and $q \in W(T)$:

$$\begin{cases} (c\mathbf{u}_T, \mathbf{v}_T)_T - (p_T, \nabla\cdot\mathbf{v}_T)_T + \langle p_T, \mathbf{v}_T\cdot\mathbf{n}_T\rangle_{\partial T} = 0, \\ \qquad\qquad\qquad\qquad (\nabla\cdot\mathbf{u}_T, q_T)_T = (f, q_T)_T. \end{cases} \tag{3.4}$$

Note that, though p was required to belong just to $L^2(\Omega)$, it is actually sufficiently smooth that its trace on ∂T is clearly defined; however, for all commonly used mixed finite element spaces, p_h is discontinuous across internal edges in $\mathcal{E}_h$. But, (3.4) can be used as motivation for localizing (3.1) as follows. Let us replace the value of p on e in the discrete analogue of (3.4) by a single-valued Lagrange multiplier λ_e on e, which can be defined as follows. Let

$$\mathcal{M}_e = \big\{(\mathbf{v}_T\cdot\mathbf{n}_T)|_e\colon\ \mathbf{v}_T \in \mathbf{V}_h(T)\big\},$$

and let

$$\Lambda_h(e) = \{\lambda_e \in \mathcal{M}_e\colon\ e \in \mathcal{E}_h\},$$
$$\Lambda_h = \big\{\lambda\colon\ \forall e \in \mathcal{E}_h,\ \lambda \in \Lambda_h(e)\big\}.$$

Now, let

$$\mathcal{V}_h = \widetilde{\mathbf{V}}_h \times W_h \times \Lambda_h,$$

where $\widetilde{\mathbf{V}}_h = \{\mathbf{v}\colon\ \mathbf{v}|_T \in \mathbf{V}_h(T)\}$. Note that continuity of the normal component of the flux across the interfaces $\partial T \cap \partial T'$ is not imposed on functions in $\widetilde{\mathbf{V}}_h$; but, for any $\mathbf{v} \in \widetilde{\mathbf{V}}_h$, the flux consistency condition can be enforced by requiring that

$$\sum_{T\in\mathcal{T}_h} \langle \mathbf{v}_T\cdot\mathbf{n}_T, \mu\rangle_{\partial T\setminus\partial\Omega} = 0, \quad \forall\mu \in \Lambda_h. \tag{3.5}$$

Thus, (3.1) can be localized (or *hybridized*) by seeking a triple $(\mathbf{u}_h, p_h, \lambda_h) \in \mathcal{V}_h$ such that

$$\begin{cases} (c\mathbf{u}_h, \mathbf{v})_T - (\nabla\cdot\mathbf{v}, p_h)_T + \langle \mathbf{v}\cdot\mathbf{n}_T, \lambda_h\rangle_{\partial T\setminus\partial\Omega} = 0, & \mathbf{v} \in \mathbf{V}_h(T),\ T \in \mathcal{T}_h, \\ (\nabla\cdot\mathbf{u}_h, q)_T = (f, q)_T, & q \in W_h(T),\ T \in \mathcal{T}_h, \\ \sum_{T\in\mathcal{T}_h} \langle \mathbf{u}_h\cdot\mathbf{n}_T, \mu\rangle_{\partial T\setminus\partial\Omega} = 0, & \mu \in \Lambda_h. \end{cases} \tag{3.6}$$

In this paper, we begin by briefly considering the lowest order Raviart–Thomas element (RT_0) (see [20]) and then concentrating on the Brezzi, Douglas, Fortin, and

Marini element (see [8]) of index two ($BDFM_2$). With appropriate quadrature rules applied to the inner product $(c\mathbf{u}_h, \mathbf{v})$, the former leads to the standard low order FVM and the latter to a second-order FVM that is the first object of this paper. Later, we consider several other mixed finite element spaces.

Let $P_k(T)$ be the set of all polynomials on $T \in \mathcal{T}_h$ of total degree not greater than k; let $P_k(z)$ for $z = x$ or y be the set of all polynomials of one variable z of the degree not greater than k. Also, let $P_{k,\ell} = P_k(x) \times P_\ell(y)$. For the RT_0 mixed finite element space,

$$\mathbf{V}_h = \left\{ \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in H(\mathrm{div};\Omega)\colon\ v_1|_T \in P_{1,0},\ v_2|_T \in P_{0,1},\ \forall T \in \mathcal{T}_h \right\},$$
$$W_h = \left\{ q \in L^2(\Omega)\colon\ q|_T \in P_0(T),\ \forall T \in \mathcal{T}_h \right\},$$

while, for the $BDFM_2$ space,

$$\mathbf{V}_h = \left\{ \mathbf{v} \in H(\mathrm{div};\Omega)\colon\ \mathbf{v}|_T \in \left[P_2(T)\backslash\{y^2\}\right] \times \left[P_2(T)\backslash\{x^2\}\right],\ \forall T \in \mathcal{T}_h \right\},$$
$$W_h = \left\{ q \in L^2(\Omega)\colon\ q|_T \in P_1(T),\ \forall T \in \mathcal{T}_h \right\}.$$

It is well known that both the RT_0 and the $BDFM_2$ spaces satisfy the inf–sup condition (see, e.g., [10]) and are admissible mixed finite element spaces. Since

$$\mathbf{v}_T \cdot \mathbf{n}_T|_e \in \begin{cases} P_0(e), & \text{for } RT_0, \\ P_1(e), & \text{for } BDFM_2, \end{cases}$$

the natural choices of the Lagrange multipliers are given by

$$\Lambda_h = \left\{ \lambda\colon\ \forall e \in \mathcal{E}_h,\ \lambda_e \in \begin{cases} P_0(e), & \text{for } RT_0, \\ P_1(e), & \text{for } BDFM_2 \end{cases} \right\}.$$

4. Finite volume methods

We shall derive finite volume schemes based on (3.6) by choosing appropriate quadrature procedures, first globally in matricial form and then in terms of the local equations for the degrees of freedom of the scalar variable p_h.

Denote basis functions by double indices T for elements or e for edges and j for the degrees of freedom on T or e. Let $\{\{\boldsymbol{\xi}_T^j(x,y)\}_{j=1}^{N_1}\}_{T\in\mathcal{T}_h}$, $\{\{\eta_T^j(x,y)\}_{j=1}^{N_2}\}_{T\in\mathcal{T}_h}$, and $\{\{\zeta_e^j(x,y)\}_{j=1}^{N_3}\}_{e\in\mathcal{E}_h}$ be bases for $\widetilde{\mathbf{V}}_h$, W_h, and Λ_h, respectively;

$$N_1 = \begin{cases} 4, & RT_0, \\ 10, & BDFM_2, \end{cases} \qquad N_2 = \begin{cases} 1, & RT_0, \\ 3, & BDFM_2, \end{cases} \qquad \text{and} \quad N_3 = \begin{cases} 1, & RT_0, \\ 2, & BDFM_2. \end{cases}$$

Let

$$\begin{cases} A_T = \left(a_{ij}^T\right)_{N_1\times N_1} = \left(\left(c\boldsymbol{\xi}_T^j, \boldsymbol{\xi}_T^i\right)_T\right)_{N_1\times N_1}, \\ B_T = \left(b_{ij}^T\right)_{N_1\times N_2} = \left(\left(\nabla\cdot\boldsymbol{\xi}_T^i, \eta_T^j\right)_T\right)_{N_1\times N_2}, \\ C_{Te} = \left(c_{ij}^{Te}\right)_{N_1\times N_3} = \left(\left\langle\boldsymbol{\xi}_T^i\cdot\mathbf{n}, \zeta_e^j\right\rangle_{\partial T\setminus\partial\Omega}\right)_{N_1\times N_3}, \end{cases} \tag{4.1}$$

and let

$$\mathbf{A} = \operatorname{diag}(A_T\colon\ T\in\mathcal{T}_h)_{N^2\times N^2},$$
$$\mathbf{B} = \operatorname{diag}(B_T\colon\ T\in\mathcal{T}_h)_{N^2\times N^2},$$
$$\mathbf{C} = (C_{Te}\colon\ T\in\mathcal{T}_h,\ e\in\mathcal{E}_h)_{N^2\times 2N(N-1)}.$$

The matrices **A** and **B** are block diagonal with the blocks being the 4×4 for the RT_0 or 10×10 for the $BDFM_2$ and 4×1 for the RT_0 or 10×3 for the $BDFM_2$, respectively. Let

$$\mathbf{U} = (\mathbf{U}_T)_{N^2\times 1}, \qquad \mathbf{P} = (\mathbf{P}_T)_{N^2\times 1} \quad\text{and}\quad \boldsymbol{\lambda} = (\boldsymbol{\lambda}_e)_{2N(N-1)\times 1},$$

where

$$\mathbf{U}_T = \left(u_T^j\right)_{N_1\times 1}, \qquad \mathbf{P}_T = \left(p_T^j\right)_{N_2\times 1} \quad\text{and}\quad \boldsymbol{\lambda}_e = \left(\lambda_e^j\right)_{N_3\times 1}.$$

Let $\mathbf{F} = (\mathbf{F}_T)_{N^2\times 1}$ be the right-hand side vector with $\mathbf{F}_T = ((f, \eta_T^i))_{N_2\times 1}$. Then, (3.6) can be written in matricial form as

$$\begin{cases} \mathbf{AU} + \mathbf{BP} + \mathbf{C}\boldsymbol{\lambda} = \mathbf{0}, \\ \mathbf{B}^{\mathrm{T}}\mathbf{U} = \mathbf{F}, \\ \mathbf{C}^{\mathrm{T}}\mathbf{U} = \mathbf{0}. \end{cases} \tag{4.2}$$

Inverting **A** in the first equation of (4.2) gives

$$\mathbf{U} = -\mathbf{A}^{-1}\mathbf{BP} - \mathbf{A}^{-1}\mathbf{C}\boldsymbol{\lambda}. \tag{4.3}$$

This is a cell-by-cell calculation and, thus, inexpensive. Substituting (4.3) into the second and third equations in (4.2) yields

$$\begin{cases} \mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{BP} + \mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{C}\boldsymbol{\lambda} = -\mathbf{F}, \\ \mathbf{C}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{BP} + \mathbf{C}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{C}\boldsymbol{\lambda} = \mathbf{0}. \end{cases} \tag{4.4}$$

The next choice is whether to eliminate **P** using the first equation of (4.4) or to eliminate $\boldsymbol{\lambda}$ using the second equation of (4.4). Since each component of $\boldsymbol{\lambda}$ is shared by two elements, there are approximately two (for RT_0) or four (for $BDFM_2$) components of $\boldsymbol{\lambda}$ per element and only one (for RT_0) or three (for $BDFM_2$) components of **P** per element. The well-known Fraeijs de Veubeke [3,17] reduction of the saddle-point problem in (3.6) to a symmetric, positive-definite linear system in edge degrees of freedom is to solve for **P** in terms of $\boldsymbol{\lambda}$ using the first equation of (4.4) and to substitute this relation into the second equation of (4.4) to obtain an equation for $\boldsymbol{\lambda}$. If, instead, the second equation of (4.4) is used to solve for $\boldsymbol{\lambda}$ in terms of $\mathbf{P}$, then we see that

$$\boldsymbol{\lambda} = -\left(\mathbf{C}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{C}\right)^{-1}\left(\mathbf{C}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{B}\right)\mathbf{P}, \tag{4.5}$$

and the following equation (in fewer variables than obtained in the Fraeijs de Veubeke reduction) results:

$$\left\{\mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{B} - \mathbf{B}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{B}\right\}\mathbf{P} = \mathbf{F}. \tag{4.6}$$

With the block diagonal matrix $\mathbf{A}$, the coefficient matrix of the discretization for $\mathbf{P}$ in (4.6) is still complicated. Instead, let us approximate the integrals $(c\boldsymbol{\xi}_j, \boldsymbol{\xi}_i)_T$ by a quadrature,

$$\left(c\boldsymbol{\xi}_T^j, \boldsymbol{\xi}_T^i\right)_T \approx Q_T\left(c\boldsymbol{\xi}_T^j \cdot \boldsymbol{\xi}_T^i\right),$$

and employ exact integration for the remaining integrals. We obtain equations of the form

$$\begin{cases} \widetilde{\mathbf{A}}\widetilde{\mathbf{U}} + \mathbf{B}\widetilde{\mathbf{P}} + \mathbf{C}\tilde{\boldsymbol{\lambda}} = \mathbf{0}, \\ \mathbf{B}^{\mathrm{T}}\widetilde{\mathbf{U}} = \mathbf{F}, \\ \mathbf{C}^{\mathrm{T}}\widetilde{\mathbf{U}} = \mathbf{0}, \end{cases} \tag{4.7}$$

where $\widetilde{\mathbf{A}} = (\widetilde{A}_T)_{N^2\times N^2}$ with $\widetilde{A}_T = (Q_T(c\boldsymbol{\xi}_T^j \cdot \boldsymbol{\xi}_T^i))_{N_1\times N_1}$. Then, elimination as above leads to the perturbed equations

$$\left\{\mathbf{B}^{\mathrm{T}}\widetilde{\mathbf{A}}^{-1}\mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\widetilde{\mathbf{A}}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\widetilde{\mathbf{A}}^{-1}\mathbf{B} - \mathbf{B}^{\mathrm{T}}\widetilde{\mathbf{A}}^{-1}\mathbf{B}\right\}\widetilde{\mathbf{P}} = \mathbf{F}. \tag{4.8}$$

The object is to choose a quadrature rule so that the matrix $\widetilde{\mathbf{A}}$ is diagonal, instead of block diagonal, and such that the numerical integration does not decrease the order of convergence of the original mixed finite element method. Equation (4.8) will be defined as the finite volume method to be considered in this paper. While in some cases it can be derived in other ways, the point of this derivation is that the same procedure can be applied to different finite element spaces to obtain a variety of finite volume procedures. In each case, the choice of the quadrature procedure will be critical.

5. The standard finite volume method

For $T \in \mathcal{T}_h$, let its four edges (left, right, bottom, and top) be denoted by

$$\alpha \in E_T = \{\ell, r, b, t\},$$

respectively. Let $\mathbf{n}_T$ be the unit outward vector normal to ∂T, so that

$$\mathbf{n}_T|_\ell = \boldsymbol{\ell} = (-1, 0)^{\mathrm{T}}, \qquad \mathbf{n}_T|_r = \mathbf{r} = (1, 0)^{\mathrm{T}},$$
$$\mathbf{n}_T|_b = \mathbf{b} = (0, -1)^{\mathrm{T}}, \qquad \mathbf{n}_T|_t = \mathbf{t} = (0, 1)^{\mathrm{T}}.$$

For the RT_0 space, the natural degrees of freedom are $v_{T,\alpha} = (\mathbf{v}_T \cdot \mathbf{n}_T)|_\alpha$, $\alpha \in E_T$, the (constant) normal components of the flux on the edges α, for $\widetilde{\mathbf{V}}_h$; q_T, the value of q on T for W_h and interpreted as its cell-center value; and $m_{T,\alpha}$, the midpoint values of m on the edges α with the restriction that, for any two adjacent elements T and T', $m_{T',\beta} = m_{T,\alpha}$ if they represent the Lagrange multiplier on the common edge of T and T'. The nodal basis functions $\boldsymbol{\xi}_T^\alpha$ on $T = (x_{i-1}, x_i) \times (y_{j-1}, y_j)$ are given by

$$\boldsymbol{\xi}_T^\ell = \frac{x_i - x}{h}\boldsymbol{\ell}, \qquad \boldsymbol{\xi}_T^r = \frac{x - x_{i-1}}{h}\mathbf{r}, \qquad \boldsymbol{\xi}_T^b = \frac{y_j - y}{h}\mathbf{b}, \quad \text{and} \quad \boldsymbol{\xi}_T^t = \frac{y - y_{j-1}}{h}\mathbf{t}.$$

The basis functions for W_h and Λ_h are (with χ_B the characteristic function for B)

$$\eta_T(x, y) = \chi_T(x, y) \quad \text{and} \quad \zeta_e(x, y) = \chi_e(x, y),$$

respectively.

Let $T \in \mathcal{T}_h$, and let $e = \partial T \cap \partial T'$. The choice $\mu = \zeta_e$ in the third equation of (3.6) implies that

$$u_{h,T,e} + u_{h,T'e} = 0, \tag{5.9}$$

where $u_{h,T,e}$ and $u_{h,T',e}$ are the normal components of $\mathbf{u}_{h,T}$ and $\mathbf{u}_{h,T'}$ on the edge e, respectively.

Recall that we have assumed the coefficient $K(\mathbf{x})$ to the constant K_T on each element $T \in \mathcal{T}_h$, so that $c|_T = c_T$ on T. Now, take $\mathbf{v} = \boldsymbol{\xi}_T^e$ and then $\boldsymbol{\xi}_{T'}^e$ in the first equation of (3.6). Since $\operatorname{div} \boldsymbol{\xi}_T^e = \operatorname{div} \boldsymbol{\xi}_{T'}^e = h^{-1}$ and $\boldsymbol{\xi}_T^e \cdot \mathbf{n}_T = \boldsymbol{\xi}_{T'}^e \cdot \mathbf{n}_T' = 1$,

$$\begin{aligned} c_T\big(\mathbf{u}_{h,T}, \boldsymbol{\xi}_T^e\big)_T - h^2 p_{h,T} \operatorname{div} \boldsymbol{\xi}_T^e + h\lambda_e &= 0, \\ c_{T'}\big(\mathbf{u}_{h,T'}, \boldsymbol{\xi}_{T'}^e\big)_{T'} - h^2 p_{h,T'} \operatorname{div} \boldsymbol{\xi}_{T'}^e + h\lambda_e &= 0, \end{aligned}$$

so that

$$c_T\big(\mathbf{u}_{h,T}, \boldsymbol{\xi}_T^e\big)_T - c_{T'}\big(\mathbf{u}_{h,T'}, \boldsymbol{\xi}_{T'}^e\big)_{T'} = h(p_{h,T} - p_{h,T'}). \tag{5.10}$$

Each integral in (5.10) involves normal components of $\mathbf{u}_{h,T}$ on the edge e and its opposite edge, but with only the two equations (5.9) and (5.10) we cannot solve for the flux on e, thereby motivating the introduction of a quadrature rule for these integrals so that they can be approximated by the normal component of $\mathbf{u}_{h,T}$ on the edge e alone. Given the vector functions in the RT_0 space, it suffices to employ the trapezoidal rule:

$$(c\mathbf{w}_T, \mathbf{v}_T)_T \approx c_T Q_T(\mathbf{w}_T \cdot \mathbf{v}_T) = c_T \frac{h^2}{2} \sum_{\alpha \in E_T} w_{T,\alpha} v_{T,\alpha}. \tag{5.11}$$

With the above numerical integration, (5.10) is replaced by

$$c_T \frac{h^2}{2} \tilde{u}_{h,T,e} - c_{T'} \frac{h^2}{2} \tilde{u}_{h,T',e} = h(\tilde{p}_{h,T} - \tilde{p}_{h,T'}),$$

where we denote approximations after application of the quadrature rule by $\tilde{\mathbf{u}}_h$ and $\tilde{p}_h$. By (5.9), we then have

$$h\tilde{u}_{h,T,e} = \frac{2}{c_T + c_{T'}}(\tilde{p}_{h,T} - \tilde{p}_{h,T'}) = \frac{2K_T K_{T'}}{K_T + K_{T'}}(\tilde{p}_{h,T} - \tilde{p}_{h,T'}). \tag{5.12}$$

Every element $T \in \mathcal{T}_h$ has four adjacent elements, which we denote by T_α, $\alpha \in E_T$; then, denote the restrictions of K and $\tilde{p}_h$ to these elements by

$$K_T = K|_T, \qquad K_\alpha = K|_{T_\alpha}, \qquad \tilde{p}_T = \tilde{p}_h|_T, \quad \text{and} \quad \tilde{p}_\alpha = \tilde{p}_h|_{T_\alpha}.$$

By taking $q = \eta_T$ in the second equation of (3.6) and using the divergence theorem and (5.12), we see that

$$\begin{aligned}(f, \eta_T)_T = (\operatorname{div} \tilde{\mathbf{u}}_{h,T}, \eta_T)_T &= \int_{\partial T} \tilde{\mathbf{u}}_{h,T} \cdot \mathbf{n}_T \, \mathrm{d}s \\ &= \sum_{\alpha \in E_T} h\tilde{u}_{h,T,\alpha} = \sum_{\alpha \in E_T} \frac{2K_T K_\alpha}{K_T + K_\alpha}(\tilde{p}_T - \tilde{p}_\alpha).\end{aligned}$$

Thus, we obtain the desired finite volume scheme:

$$\sum_{\alpha \in E_T} d_{T,\alpha}(\tilde{p}_T - \tilde{p}_\alpha) = (f, \eta_T)_T, \tag{5.13}$$

where $d_{T,\alpha}$ denotes the harmonic average of K_T and K_α:

$$d_{T,\alpha} = \frac{2K_\alpha K_T}{K_\alpha + K_T}.$$

This equation is the classical finite volume method (or cell-centered finite difference equation) with harmonically averaged diffusion coefficients. See [21] for a slightly different derivation of (5.13) and an error analysis. In section 8, we shall derive a finite volume procedure corresponding to the rectangular RTN_k mixed finite element space for all k, thereby finding finite volume methods for all orders of accuracy.

6. A higher order finite volume method based on $BDFM_2$

6.1. Derivation

Let us consider an analogous derivation based on the $BDFM_2$ space. For each element $T = (x_{i-1}, x_i) \times (y_{j-1}, y_j) \in \mathcal{T}_h$, denote its center and vertices by

$$\mathbf{a}_T = (x_{i-1/2}, y_{j-1/2}) = \left(x_i - \frac{h}{2}, y_j - \frac{h}{2}\right),$$

$$\mathbf{a}_{T,1} = (x_{i-1}, y_{j-1}), \qquad \mathbf{a}_{T,2} = (x_{i-1}, y_j), \qquad \mathbf{a}_{T,3} = (x_i, y_{j-1}), \qquad \mathbf{a}_{T,4} = (x_i, y_j),$$

and the end points of the edges in $E_T = \{\ell, r, b, t\}$ by

$$\mathbf{b}_{\ell,1} = \mathbf{b}_{b,1} = \mathbf{a}_{T,1}, \qquad \mathbf{b}_{\ell,2} = \mathbf{b}_{t,1} = \mathbf{a}_{T,2},$$
$$\mathbf{b}_{r,1} = \mathbf{b}_{b,2} = \mathbf{a}_{T,3}, \qquad \mathbf{b}_{r,2} = \mathbf{b}_{t,2} = \mathbf{a}_{T,4}.$$

For the $BDFM_2$ space,

$$\begin{aligned}
\mathbf{V}_h(T) &= \mathrm{Span}\{\boldsymbol{\xi}_T^i,\ i = 1, 2;\ \boldsymbol{\xi}_T^{\alpha,i},\ \alpha = r, \ell, b, t,\ i = 1, 2\},\\
W_h(T) &= P_1(T) = \mathrm{Span}\{1, x - x_{i-1/2}, y - y_{j-1/2}\} = \mathrm{Span}\{\eta_T^0, \eta_T^x, \eta_T^y\},\\
\Lambda_h(\partial T) &= \{m \in L^2(\partial T):\ m|_\alpha \in P_1(\alpha),\ \alpha \in E_T\};
\end{aligned}$$

the degrees of freedom for $\mathbf{V}_h(T)$ are $\mathbf{v}_T(\mathbf{a}_T)$, the value at the center of T and $v_{T,\alpha,i} = (\mathbf{v}_T \cdot \mathbf{n}_T)|_\alpha(\mathbf{b}_{\alpha,i})$, the value of the normal component on the edge $\alpha \in E_T = \{\ell, r, b, t\}$ at its two end points $\mathbf{b}_{\alpha,1}$ and $\mathbf{b}_{\alpha,2}$. Let

$$\theta_i = \frac{2(x - x_{i-1/2})}{h}, \qquad \psi_j = \frac{2(y - y_{j-1/2})}{h}.$$

The two nodal basis functions $\boldsymbol{\xi}_T^j$ $(j = 1, 2)$ associated with the center of T satisfy

$$\boldsymbol{\xi}_T^1(\mathbf{a}_T) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{\xi}_T^2(\mathbf{a}_T) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{and} \quad (\boldsymbol{\xi}_T^j \cdot \mathbf{n}_T)\big|_\alpha = 0 \quad \text{for } \alpha \in E_T; \tag{6.14}$$

thus,

$$\boldsymbol{\xi}_T^1(x, y) = \begin{pmatrix} 1 - \theta_i^2 \\ 0 \end{pmatrix}, \qquad \boldsymbol{\xi}_T^2(x, y) = \begin{pmatrix} 0 \\ 1 - \psi_j^2 \end{pmatrix}, \quad (x, y) \in T. \tag{6.15}$$

The nodal basis functions $\boldsymbol{\xi}_T^{\alpha,i}$ associated with the edge $\alpha \in E_T$ and its end points $\mathbf{b}_{\alpha,i}$ $(i = 1, 2)$ satisfy

$$\begin{aligned}
&\boldsymbol{\xi}_T^{\ell,i}(\mathbf{b}_{\ell,i}) = -\boldsymbol{\xi}_T^{r,i}(\mathbf{b}_{r,i}) = -\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{\xi}_T^{b,i}(\mathbf{b}_{b,i}) = -\boldsymbol{\xi}_T^{t,i}(\mathbf{b}_{t,i}) = -\begin{pmatrix} 0 \\ 1 \end{pmatrix},\\
&\boldsymbol{\xi}_T^{\alpha,i}(\mathbf{a}_T) = \mathbf{0}, \qquad \boldsymbol{\xi}_T^{\alpha,i}(\mathbf{b}_{\beta,j}) = \mathbf{0}, \quad \beta \neq \alpha,\ j \neq i.
\end{aligned} \tag{6.16}$$

For example,

$$\boldsymbol{\xi}_T^{r,1} = \begin{pmatrix} \frac{1}{4}(\theta_i - \psi_j)(1 + \theta_i) \\ 0 \end{pmatrix}, \qquad \boldsymbol{\xi}_T^{r,2} = \begin{pmatrix} \frac{1}{4}(\theta_i + \psi_j)(1 + \theta_i) \\ 0 \end{pmatrix}. \tag{6.17}$$

Then,

$$\begin{aligned}
\mathbf{u}_{h,T} &= \sum_{j=1}^{2} u_{h,T}^j \boldsymbol{\xi}_T^j + \sum_{\alpha \in E_T} \sum_{i=1}^{2} u_{h,T,\alpha}^i \boldsymbol{\xi}_T^{\alpha,i},\\
p_{h,T} &= p_{h,T}^0 \eta_T^0 + p_{h,T}^x \eta_T^x + p_{h,T}^y \eta_T^y,
\end{aligned}$$

with $u^j_{h,T}$ and $u^i_{h,T,\alpha}$ being the values of the jth component at the center and the values of the normal component on the edge α at its end point $\mathbf{b}_{\alpha,i}$, respectively, and $p^0_{h,T}$ and $(p^x_{h,T}, p^y_{h,T})^{\mathrm{T}}$ the value and gradient of $p_{h,T}$ at the center of T, respectively. The nodal basis function ζ^i_α for $\Lambda_h(\partial T)$ associated with the edge $\alpha \in E_T$ and its end point $\mathbf{b}_{\alpha,i}$ is the linear function on the edge such that

$$\zeta^i_\alpha(\mathbf{b}_{\alpha,j}) = \delta_{ij} \quad \text{for } j = 1, 2.$$

Let $T \in \mathcal{T}_h$ and choose the test function in the first equation of (3.6) to be $\boldsymbol{\xi}^k_T$; then, we see that

$$c_T\big(\mathbf{u}_{h,T}, \boldsymbol{\xi}^k_T\big)_T = \big(\nabla \cdot \boldsymbol{\xi}^k_T,\ p_{h,T}\big)_T. \tag{6.18}$$

Next, let T' be adjacent to T and share an edge $e = \partial T \cap \partial T'$ with end points $\mathbf{b}_{e,i}$ $(i = 1, 2)$ with T. Taking first $\mathbf{v} = \boldsymbol{\xi}^{e,k}_T$ and then $\boldsymbol{\xi}^{e,k}_{T'}$ $(k = 1, 2)$ in the first equation of (3.6) gives the relations

$$\begin{aligned} c_T\big(\mathbf{u}_{h,T}, \boldsymbol{\xi}^{e,k}_T\big)_T &= \big(\nabla \cdot \boldsymbol{\xi}^{e,k}_T, p_{h,T}\big)_T - \big\langle \boldsymbol{\xi}^{e,k}_T \cdot \mathbf{n}_T, \lambda_h\big\rangle_e, \\ c_{T'}\big(\mathbf{u}_{h,T'}, \boldsymbol{\xi}^{e,k}_{T'}\big)_{T'} &= \big(\nabla \cdot \boldsymbol{\xi}^{e,k}_{T'}, p_{h,T'}\big)_{T'} - \big\langle \boldsymbol{\xi}^{e,k}_{T'} \cdot \mathbf{n}_{T'}, \lambda_h\big\rangle_e. \end{aligned} \tag{6.19}$$

Since $\boldsymbol{\xi}^{e,k}_T \cdot \mathbf{n}_T = \boldsymbol{\xi}^{e,k}_{T'} \cdot \mathbf{n}_{T'}$ on e, differencing of the two equations in (6.19) gives

$$c_T\big(\mathbf{u}_{h,T}, \boldsymbol{\xi}^{e,k}_T\big)_T - c_{T'}\big(\mathbf{u}_{h,T'}, \boldsymbol{\xi}^{e,k}_{T'}\big)_{T'} = \big(\nabla \cdot \boldsymbol{\xi}^{e,k}_T, p_{h,T}\big)_T - \big(\nabla \cdot \boldsymbol{\xi}^{e,k}_{T'}, p_{h,T'}\big)_{T'}, \quad k = 1, 2. \tag{6.20}$$

The finite volume method will be derived from (6.18) and (6.20). Note that the Lagrange multipliers have been eliminated; in fact, it was not necessary to introduce them [21]. The third equation in (3.6) implies the flux consistency conditions

$$u^1_{h,T,e} + u^1_{h,T',e} = 0 \quad \text{and} \quad u^2_{h,T,e} + u^2_{h,T',e} = 0. \tag{6.21}$$

Next, we shall introduce a quadrature rule to diagonalize the left-hand sides of (6.18) and (6.20). Clearly, it is appropriate to choose a five-point rule using the center and four vertices of each element, and the optimum choice is given by

$$\int_T g(x, y)\,\mathrm{d}x\,\mathrm{d}y \approx Q_T(g) = \frac{h^2}{12}\left(8g(\mathbf{a}_T) + \sum_{i=1}^{4} g(\mathbf{a}_{T,i})\right), \quad T \in \mathcal{T}_h, \tag{6.22}$$

since this is the only rule associated with these nodes that is exact for $P_2(T)$; in fact, it is exact on $P_3(T)$ and will lead to retaining the accuracy of the $BDFM_2$ mixed method. With this numerical integration, (6.18) and (6.20) are replaced by

$$\frac{2h^2}{3} c_T \tilde{u}^k_{h,T} = \big(\nabla \cdot \boldsymbol{\xi}^k_T, \tilde{p}_{h,T}\big)_T \quad \text{and} \tag{6.23}$$

$$\frac{h^2}{12}\big(c_T \tilde{u}^k_{h,T,e} - c_{T'}\tilde{u}^k_{h,T',e}\big) = \big(\nabla \cdot \boldsymbol{\xi}^{e,k}_T, \tilde{p}_{h,T}\big)_T - \big(\nabla \cdot \boldsymbol{\xi}^{e,k}_{T'}, \tilde{p}_{h,T'}\big)_{T'} \tag{6.24}$$

for $k = 1, 2$, where we denote the approximate solution after quadrature by $\tilde{\mathbf{u}}_h$ and $\tilde{p}_h$.

It follows easily from (6.15) and (6.23) that

$$\tilde{u}_{h,T}^{k} = \begin{cases} -K_T \tilde{p}_{h,T}^{x}, & k = 1, \\ -K_T \tilde{p}_{h,T}^{y}, & k = 2. \end{cases} \tag{6.25}$$

Next, it follows from (6.24) and the analogue of the flux consistency condition (6.21) that

$$\tilde{u}_{h,T,e}^{k} = d_{T,T'} \frac{6}{h^2} \big(\big(\nabla \cdot \boldsymbol{\xi}_T^{e,k}, \tilde{p}_{h,T}\big)_T - \big(\nabla \cdot \boldsymbol{\xi}_{T'}^{e,k}, \tilde{p}_{h,T'}\big)_{T'} \big), \tag{6.26}$$

where

$$d_{T,T'} = d_{T,e} = \frac{2K_T K_{T'}}{K_T + K_T'}$$

is the harmonically averaged diffusion coefficient on $e = \partial T \cap \partial T'$.

A straightforward calculation based on (6.16) (or (6.17)) shows that

$$\big(\nabla \cdot \boldsymbol{\xi}_T^{r,1}, \tilde{p}_{h,T}\big)_T = \frac{h}{2} \tilde{p}_{h,T}^{0} + \frac{h^2}{6} \tilde{p}_{h,T}^{x} - \frac{h^2}{12} \tilde{p}_{h,T}^{y}; \tag{6.27}$$

the other integrals of this form can be evaluated similarly. It follows that

$$\tilde{u}_{h,T,r}^{1} = d_{T,r} \left(3 \frac{\tilde{p}_{h,T}^{0} - \tilde{p}_{h,T_r}^{0}}{h} + \big(\tilde{p}_{h,T}^{x} + \tilde{p}_{h,T_r}^{x}\big) - \frac{1}{2} \big(\tilde{p}_{h,T}^{y} - \tilde{p}_{h,T_r}^{y}\big) \right), \tag{6.28a}$$

$$\tilde{u}_{h,T,r}^{2} = d_{T,r} \left(3 \frac{\tilde{p}_{h,T}^{0} - \tilde{p}_{h,T_r}^{0}}{h} + \big(\tilde{p}_{h,T}^{x} + \tilde{p}_{h,T_r}^{x}\big) + \frac{1}{2} \big(\tilde{p}_{h,T}^{y} - \tilde{p}_{h,T_r}^{y}\big) \right), \tag{6.28b}$$

so that

$$\int_{\mathbf{r}} \tilde{\mathbf{u}}_h \cdot \mathbf{n}_T \, \mathrm{d}s = \frac{h}{2} \sum_{j=1}^{2} \tilde{u}_{h,T,r}^{j} = d_{T,r} \big(3\big(\tilde{p}_{h,T}^{0} - \tilde{p}_{h,T_r}^{0}\big) + h\big(\tilde{p}_{h,T}^{x} + \tilde{p}_{h,T_r}^{x}\big) \big). \tag{6.29}$$

See section 6.3 for a complete set of the ten flux coefficients in terms of the scalar coefficients. Equation (6.28a) has the following interpretation. Let $\mathbf{z}$ denote the midpoint of the right edge of T. Then, to higher order in h,

$$3\, \frac{\tilde{p}_{h,T}^{0} - \tilde{p}_{h,T_r}^{0}}{h} + \big(\tilde{p}_{h,T}^{x} + \tilde{p}_{h,T_r}^{x}\big) \approx -\frac{\partial p}{\partial x}(\mathbf{z})$$

and

$$-\frac{1}{2} \big(\tilde{p}_{h,T}^{y} - \tilde{p}_{h,T_r}^{y}\big) \approx \frac{h}{2} \frac{\partial^2 p}{\partial x \partial y}(\mathbf{z}),$$

so that the right-hand side of (6.28a) approximates

$$-\frac{\partial p}{\partial x}(\mathbf{z}) + \frac{h}{2} \frac{\partial^2 p}{\partial x \partial y}(\mathbf{z}) \approx -\frac{\partial p}{\partial x}(\mathbf{a}_{T,3}),$$

and (6.28a) is then a proper approximation of the flux at $\mathbf{a}_{T,3}$.

The second equation of (3.6) implies that

$$\begin{aligned}\left(f,\eta_T^0\right)_T &= (f,1)_T = (\nabla\cdot\tilde{\mathbf{u}}_{h,T},\ 1)_T\\ &= \int_{\partial T}\tilde{\mathbf{u}}_{h,T}\cdot\mathbf{n}_T\,\mathrm{d}s = \frac{h}{2}\sum_{\alpha\in E_T}\sum_{j=1}^{2}\tilde{u}^j_{h,T,\alpha}.\end{aligned}\tag{6.30}$$

Also from the same equation,

$$\left(f,\eta_T^x\right)_T = \left(\operatorname{div}\tilde{\mathbf{u}}_{h,T},\eta_T^x\right)_T = -\left(\tilde{\mathbf{u}}_{h,T},\nabla\eta_T^x\right)_T + \int_{\partial T}(\tilde{\mathbf{u}}_{h,T}\cdot\mathbf{n}_T)(x-x_{i-1/2})\,\mathrm{d}s.\tag{6.31}$$

A trivial, but tedious, calculation leads to the equation

$$\begin{aligned}h^{-2}(f,x-x_{i-1/2})_T &= -\frac{2}{3}\tilde{u}^1_{h,T}+\frac{1}{6}\sum_{i=1}^{2}\left(\tilde{u}^i_{h,T,r}-\tilde{u}^i_{h,T,\ell}\right)\\ &\quad+\frac{1}{12}\left(\tilde{u}^2_{h,T,t}-\tilde{u}^1_{h,T,t}-\tilde{u}^1_{h,T,b}+\tilde{u}^2_{h,T,b}\right).\end{aligned}\tag{6.32}$$

The last term above represents an approximation to the second mixed partial derivative of the y-component of the flux at the center of T. Similarly,

$$\begin{aligned}h^{-2}(f,y-y_{j-1/2})_T &= -\frac{2}{3}\tilde{u}^2_{h,T}+\frac{1}{6}\sum_{i=1}^{2}\left(\tilde{u}^i_{h,T,t}-\tilde{u}^i_{h,T,b}\right)\\ &\quad+\frac{1}{12}\left(\tilde{u}^2_{h,T,r}+\tilde{u}^2_{h,T,\ell}-\tilde{u}^1_{h,T,\ell}-\tilde{u}^1_{h,T,r}\right).\end{aligned}\tag{6.33}$$

6.2. *The BDFM$_2$ finite volume equations*

It follows from (6.30), (6.32), (6.33), and the ten equations given by (6.35) that

$$\begin{aligned}(f,1)_T &= 3\sum_{\alpha} d_{T,\alpha}\left(\tilde{p}^0_{h,T}-\tilde{p}^0_{h,T_\alpha}\right)+h\left[d_{T,r}\left(\tilde{p}^x_{h,T}+\tilde{p}^x_{h,T_r}\right)\right.\\ &\quad\left.-d_{T,\ell}\left(\tilde{p}^x_{h,T}+\tilde{p}^x_{h,T_\ell}\right)\right]+h\left[d_{T,t}\left(\tilde{p}^y_{h,T}+\tilde{p}^y_{h,T_t}\right)\right.\\ &\quad\left.-d_{T,b}\left(\tilde{p}^y_{h,T}+\tilde{p}^y_{T,b}\right)\right],\end{aligned}\tag{6.34a}$$

$$\begin{aligned}h^{-2}(f,x-x_{i-1/2})_T &= \frac{2}{3}K_T\tilde{p}^x_{h,T}+d_{T,r}\frac{\tilde{p}^0_{h,T}-\tilde{p}^0_{h,T_r}}{h}-d_{T,\ell}\frac{\tilde{p}^0_{h,T}-\tilde{p}^0_{h,T_\ell}}{h}\\ &\quad+\frac{1}{3}\sum_{\alpha=\ell,r}d_{T,\alpha}\left(\tilde{p}^x_{h,T}+\tilde{p}^x_{h,T_\alpha}\right)+\frac{1}{12}\sum_{\alpha=b,t}d_{T,\alpha}\left(\tilde{p}^x_{h,T}-\tilde{p}^x_{h,T_\alpha}\right),\end{aligned}\tag{6.34b}$$

$$
\begin{aligned}
h^{-2}(f, y - y_{j-1/2})_T = {} & \frac{2}{3} K_T \tilde{p}^y_{h,T} + d_{T,t} \frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_t}}{h} - d_{T,b} \frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_b}}{h} \\
& + \frac{1}{3} \sum_{\alpha=b,t} d_{T,\alpha} \big(\tilde{p}^y_{h,T} + \tilde{p}^y_{h,T_\alpha}\big) + \frac{1}{12} \sum_{\alpha=\ell,r} d_{T,\alpha} \big(\tilde{p}^y_{h,T} - \tilde{p}^y_{h,T_\alpha}\big).
\end{aligned}
\tag{6.34c}
$$

The system (6.34) is the finite volume method that associate with the $BDFM_2$ mixed finite element method.

These equations can be interpretated as follows. Let the elliptic equation be

$$-\Delta p = f;$$

then dividing (6.34a) by h^2 and letting $h \to 0$ leads back to $-\Delta p = f$. The second equation, (6.34b), has as limiting form the equation

$$-\Delta \frac{\partial p}{\partial x} = \frac{\partial f}{\partial x};$$

analogously, (6.34c) tends to

$$-\Delta \frac{\partial p}{\partial y} = \frac{\partial f}{\partial y}.$$

Note that we can easily obtain the approximate flux coefficients $\tilde{u}^i_{h,T,\alpha}$ on each edge of the blocks $T \in \mathcal{T}_h$ from (6.28) and their corresponding relations on the other edges. The coefficients for the two internal flux components are given by (6.25). These coefficients are collected below.

6.3. *The flux coefficients for $BDFM_2$*

The full set of equations relating the flux variables to the scalar variables are as follows:

$$\tilde{u}^1_{h,T} = -K_T \tilde{p}^x_{h,T}, \tag{6.35a}$$

$$\tilde{u}^2_{h,T} = -K_T \tilde{p}^y_{h,T}, \tag{6.35b}$$

$$\tilde{u}^1_{h,T,r} = d_{T,r} \left(3 \frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_r}}{h} + \big(\tilde{p}^x_{h,T} + \tilde{p}^x_{h,T_r}\big) - \frac{1}{2} \big(\tilde{p}^y_{h,T} - \tilde{p}^y_{h,T_r}\big) \right), \tag{6.35c}$$

$$\tilde{u}^2_{h,T,r} = d_{T,r} \left(3 \frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_r}}{h} + \big(\tilde{p}^x_{h,T} + \tilde{p}^x_{h,T_r}\big) + \frac{1}{2} \big(\tilde{p}^y_{h,T} - \tilde{p}^y_{h,T_r}\big) \right), \tag{6.35d}$$

$$\tilde{u}^1_{h,T,\ell} = d_{T,\ell} \left(3 \frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_\ell}}{h} - \big(\tilde{p}^x_{h,T} + \tilde{p}^x_{h,T_\ell}\big) - \frac{1}{2} \big(\tilde{p}^y_{h,T} - \tilde{p}^y_{h,T_\ell}\big) \right), \tag{6.35e}$$

$$\tilde{u}^2_{h,T,\ell} = d_{T,\ell} \left(3 \frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_\ell}}{h} - \big(\tilde{p}^x_{h,T} + \tilde{p}^x_{h,T_\ell}\big) + \frac{1}{2} \big(\tilde{p}^y_{h,T} - \tilde{p}^y_{h,T_\ell}\big) \right), \tag{6.35f}$$

$$\tilde{u}^1_{h,T,t} = d_{T,t}\left(3\frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_t}}{h} + \left(\tilde{p}^y_{h,T} + \tilde{p}^y_{h,T_t}\right) - \frac{1}{2}\left(\tilde{p}^x_{h,T} - \tilde{p}^x_{h,T_t}\right)\right), \tag{6.35g}$$

$$\tilde{u}^2_{h,T,t} = d_{T,t}\left(3\frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_t}}{h} + \left(\tilde{p}^y_{h,T} + \tilde{p}^y_{h,T_t}\right) + \frac{1}{2}\left(\tilde{p}^x_{h,T} - \tilde{p}^x_{h,T_t}\right)\right), \tag{6.35h}$$

$$\tilde{u}^1_{h,T,b} = d_{T,b}\left(3\frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_b}}{h} - \left(\tilde{p}^y_{h,T} + \tilde{p}^y_{h,T_b}\right) - \frac{1}{2}\left(\tilde{p}^x_{h,T} - \tilde{p}^x_{h,T_b}\right)\right), \tag{6.35i}$$

$$\tilde{u}^2_{h,T,b} = d_{T,b}\left(3\frac{\tilde{p}^0_{h,T} - \tilde{p}^0_{h,T_b}}{h} - \left(\tilde{p}^y_{h,T} + \tilde{p}^y_{h,T_b}\right) + \frac{1}{2}\left(\tilde{p}^x_{h,T} - \tilde{p}^x_{h,T_b}\right)\right). \tag{6.35j}$$

7. Error estimates for finite volume methods related to mixed finite element methods

7.1. A general convergence analysis

We shall deduce error estimates from classical results for mixed finite element methods for finite volume methods derived from mixed finite element methods by applying an appropriate quadrature rule to the $(c\mathbf{u}, \mathbf{v})$-term. Then, we shall employ these estimates to analyze the $BDFM_2$-based FVM. (The RT_0 case has been satisfactorily treated in [21]; the analysis below, however, does apply to it.)

Now, let $\mathcal{V}_h = \mathbf{V}_h \times W_h$ be an admissible mixed finite element space, and let $(\mathbf{u}_h, p_h) \in \mathcal{V}_h$ be the solution of the discrete mixed problem (3.1). Let $(\mathbf{u}, p) \in \mathcal{V}$ be the solution of (2.4). For $BDFM_2$, we know that [8]

$$\|\mathbf{u} - \mathbf{u}_h\|_{0,2,\Omega} \leqslant Ch^k\left(\sum_{T\in\mathcal{T}_h} \|p\|^2_{k+1,2,T}\right)^{1/2}, \quad k = 1, 2, \tag{7.1a}$$

$$\|p - p_h\|_{0,2,\Omega} \leqslant Ch^2\left(\sum_{T\in\mathcal{T}_h} \|p\|^2_{2,2,T}\right)^{1/2}; \tag{7.1b}$$

analogous error estimates are known for all mixed finite element methods considered below.

Since the fluxes evaluated by (6.35a) from (6.34) satisfy the flux consistency relations, we can drop the Lagrange multipliers in the analysis of the convergence of (6.34). It will be the case that the flux consistency conditions will be applied explicitly or implicitly in all other examples considered below. Consequently, we can consider (6.34) to be a special case of the perturbed mixed finite element method of finding $(\tilde{\mathbf{u}}_h, \tilde{p}_h) \in \mathcal{V}_h$ such that

$$\mathcal{Q}_h(c\tilde{\mathbf{u}}_h \cdot \mathbf{v}) - (\nabla \cdot \mathbf{v}, \tilde{p}_h) = 0, \quad \mathbf{v} \in \mathbf{V}_h, \tag{7.2a}$$

$$(\nabla \cdot \tilde{\mathbf{u}}_h, q) = (f, q), \quad q \in W_h, \tag{7.2b}$$

where

$$\mathcal{Q}_h(g) = \sum_{T\in\mathcal{T}_h} Q_T(g) \tag{7.3}$$

is the quadrature rule associated with derivation of the FVM related to $\mathcal{V}_h$. We shall assume that the quadrature points for Q_T coincide with nodal points for a basis for $\mathbf{V}_h(T)$ and that Q_T reduces the matrix A_T (see (4.1)) to diagonal form; this constraint seems not to be possible for a number of efficient mixed finite element spaces if the additional constraint that it cause no reduction in the optimal order of convergence of the resulting FVM, but we shall indicate several interesting examples where such a Q_T is easily constructed.

Let $(\mathbf{u}_h, p_h)$ denote the solution of (3.1) and set

$$\mathbf{e}^h_{\tilde{\mathbf{u}}} = \mathbf{u}_h - \tilde{\mathbf{u}}_h, \qquad e^h_{\tilde{p}} = p_h - \tilde{p}_h.$$

By differencing (7.2) and (3.1), we see that

$$(c\mathbf{u}_h, \mathbf{v}) - \mathcal{Q}_h(c\tilde{\mathbf{u}}_h \cdot \mathbf{v}) - \big(\nabla \cdot \mathbf{v}, e^h_{\tilde{p}}\big) = 0, \quad \mathbf{v} \in \mathbf{V}_h, \tag{7.4a}$$

$$\big(\nabla \cdot \mathbf{e}^h_{\tilde{\mathbf{u}}}, q\big) = 0, \quad q \in W_h. \tag{7.4b}$$

The convergence argument below makes serious use of the commuting diagram constructed in [8,14] (and implicitly in [16]) and for many other families of mixed finite element spaces; thus, we assume that there exists a map $\Pi_h : H^1(\Omega)^2 \to \mathbf{V}_h$, along with the $L^2(\Omega)$-projection $P_h : L^2(\Omega) \to W_h$, such that

$$\operatorname{div} \Pi_h = P_h \operatorname{div} : H^1(\Omega)^2 \longrightarrow W_h \tag{7.5}$$

and such that

$$\|\mathbf{v} - \Pi_h \mathbf{v}\|_0 \leqslant M\|\mathbf{v}\|_s h^s, \quad 0 \leqslant s \leqslant r + 1, \tag{7.6a}$$

$$\|q - P_h q\| \leqslant M\|q\|_s h^s, \quad 0 \leqslant s \leqslant r^*, \tag{7.6b}$$

where $r^* = r$ in some cases [7,9,15] and $r^* = r + 1$ in others [8,19,20]. We also require the "inverse" property

$$|\mathbf{v}|_{s,T} \leqslant Ch^{-(s-s')}\|\mathbf{v}\|_{s',T}, \quad 0 \leqslant s' \leqslant s \leqslant t,\ \mathbf{v} \in \mathbf{V}_h, \tag{7.7}$$

where C is independent of h and $\mathbf{V}_h(T) \subset P_t(T)^d$, with d being the dimension of Ω and t being independent of T; (7.7) holds for quasi-regular partitions, so that it holds for the uniform partition $\mathcal{T}_h$ in particular. It should be noted that Π_h is usually defined element-by-element in terms of boundary and interior moments, so that it is defined when $\mathbf{u} \in H^1(\Omega)^d$, or, equivalently, $p \in H^2(\Omega)$. Consequently, the norms on the right-hand sides in (7.6) can be broken norms over the partition. The two inequalities in (7.6) imply that

$$P_r(T)^d \subset \mathbf{V}_h(T), \qquad P_{r^*-1}(T) \subset W_h(T), \quad \forall T \in \mathcal{T}_h.$$

Begin the analysis by taking $q = \nabla \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^{h,\mathrm{T}} \in W_h$ in (7.4b) to see that

$$\left\| \nabla \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^{h} \right\|^2 = 0,$$

so that $\nabla \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^{h} \equiv 0$.

Next, let $\mathbf{v} \in \widetilde{\mathbf{V}}_h = \{\mathbf{v}\colon\ \mathbf{v} \in \mathbf{V}_h(T),\ T \in \mathcal{T}_h\}$, and assume that

$$\mathbf{v}_T = \sum_{j=1}^{m} v_{h,T}^{j} \boldsymbol{\xi}_T^{j},$$

where $\{\boldsymbol{\xi}_T^j,\ j = 1, \dots, m\}$ is a nodal basis for $\mathbf{V}_h(T)$ over the nodes $\mathbf{a}_{j,T},\ j = 1, \dots, m$, with the nodes being similarly placed in T as in the reference element $[-1, 1]^d$; i.e.,

$$\boldsymbol{\xi}_T^j(\mathbf{a}_{k,T}) = \delta_{j,k}, \quad j, k = 1, \dots, m.$$

Also, assume that

$$Q_T(g) = h^d \sum_{j=1}^{m} \omega_j g(\mathbf{a}_{j,T}).$$

Then, there exist positive constants γ_1 and γ_2, independent of h, such that

$$\gamma_1^2 \|\mathbf{v}_T\|_{0,T}^2 \leqslant Q_T\big(|\mathbf{v}_T|^2\big) = h^d \sum_{j=1}^{m} \omega_j \big(v_{h,T}^j\big)^2 \leqslant \gamma_2^2 \|\mathbf{v}_T\|_{0,T}^2,$$

since $\mathbf{V}_h(T)$ is a finite-dimensional subset of $P_t(T)^d$. Thus,

$$\gamma_1 \|\mathbf{v}\| \leqslant \big(\mathcal{Q}_h\big(|\mathbf{v}|^2\big)\big)^{1/2} \leqslant \gamma_2 \|\mathbf{v}\|, \quad \mathbf{v} \in \widetilde{\mathbf{V}}_h, \tag{7.8}$$

so that the perturbed $L^2(\Omega)$-like norm based on $\sqrt{\mathcal{Q}_h}$ is equivalent to the ordinary $L^2(\Omega)$ norm. Also, assume that Q_T is exact on $P_n(T)$:

$$Q_T(q) = \int_T q(\mathbf{x})\, \mathrm{d}\mathbf{x}, \quad q \in P_n(T);$$

actually, we need a slight generalization of this relation. If q corresponds to $c\mathbf{u} \cdot \mathbf{v}$, let $q_i = c\mathbf{u}_i \mathbf{v}_i,\ i = 1, \dots, d$. Then, let

$$\int_T q\, \mathrm{d}\mathbf{x} = \sum_{i=1}^{d} \int_T q_i\, \mathrm{d}\mathbf{x} \approx \sum_{i=1}^{d} Q_{i,T}(q_i), \tag{7.9}$$

and assume that

$$Q_{i,T}(q_i) = \int_T q\, \mathrm{d}\mathbf{x}, \quad q_i \in P_n(T),\ i = 1, \dots, d. \tag{7.10}$$

It is not necessary that c be constant on each T in the analysis below, but it will be assumed that $c(\mathbf{x})$ is sufficiently smooth on each T, uniformly in T, and that it is bounded away from zero and infinity on Ω.

Rewrite the error equation (7.4a) in the form

$$\mathcal{Q}_h\big(c\mathbf{e}_{\tilde{\mathbf{u}}}^h \cdot \mathbf{v}\big) - \big(\nabla \cdot \mathbf{v}, e_{\tilde{p}}^h\big) = \mathcal{Q}_h(c\mathbf{u}_h \cdot \mathbf{v}) - (c\mathbf{u}_h, \mathbf{v}), \quad \mathbf{v} \in \mathbf{V}_h. \tag{7.11}$$

Let $\mathbf{v} = \mathbf{e}_{\tilde{\mathbf{u}}}^h$ and recall that $\nabla \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h = 0$, so that

$$c_{\min}\gamma_1^2\big\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big\|^2 \leqslant \mathcal{Q}_h\big(c\big|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big|^2\big) = \sum_{T\in\mathcal{T}_h}\big(Q_T\big(c\mathbf{u}_h \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h\big) - \big(c\mathbf{u}_h, \mathbf{e}_{\tilde{\mathbf{u}}}^h\big)_T\big). \tag{7.12}$$

Let

$$|g|_{\tau,T}^2 = \sum_{k+\ell=\tau}\int_T\left(\frac{\partial^\tau g}{\partial x^k \partial y^\ell}\right)^2 \mathrm{d}x\,\mathrm{d}y.$$

Then, the Bramble–Hilbert lemma [13] implies that there exists a constant M, independent of h, such that

$$\big|Q_T\big(c\mathbf{u}_h \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h\big) - \big(c\mathbf{u}_h, \mathbf{e}_{\tilde{\mathbf{u}}}^h\big)_T\big| \leqslant M\big|c\mathbf{u}_h \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h\big|_{n+1,T}h^{n+1}, \tag{7.13}$$

since the quadrature rule is exact on $P_n(T)$. (The quadrature rule is exact on $P_m(T)$ for $m < n$, as well; the argument below applies with n replaced by $m < n$ and is of considerable interest for $m = n - 1$ in the analysis of e_p^h.)

We must break the proof into two cases. If

$$2t \leqslant n + 1,$$

then since each of the factors $\mathbf{u}_h$ and $\mathbf{e}_{\tilde{\mathbf{u}}}^h$ is a polynomial in $P_t(T)$ and c has sufficiently many uniformly bounded derivatives on T,

$$\big|Q_T\big(c\mathbf{u}_h \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h\big) - \big(c\mathbf{u}_h, \mathbf{e}_{\tilde{\mathbf{u}}}^h\big)_T\big| \leqslant M\|\mathbf{u}_h\|_{t,T}\big\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big\|_{t,T}h^{n+1};$$

hence, it follows that

$$\big|\mathcal{Q}_h\big(c\mathbf{u}_h \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h\big) - \big(c\mathbf{u}_h, e_{\tilde{u}}^h\big)\big| \leqslant M\left(\sum_{T\in\mathcal{T}_h}\|\mathbf{u}_h\|_{t,T}^2\right)^{1/2}\left(\sum_{T\in\mathcal{T}_h}\big\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big\|_{t,T}^2\right)^{1/2}h^{n+1}. \tag{7.14}$$

Here, we need error estimates analogous to (7.1) for the underlying mixed finite element method; i.e., assume that

$$\|\mathbf{u} - \mathbf{u}_h\|_{0,\Omega} \leqslant M\|\mathbf{u}\|_{s,\Omega}\,h^s, \quad 0 \leqslant s \leqslant r + 1, \tag{7.15a}$$

$$\|p - p_h\|_{0,\Omega} \leqslant M\|p\|_{\tau,\Omega}\,h^\tau, \quad 0 \leqslant \tau \leqslant r^*. \tag{7.15b}$$

Moreover, in all practical cases of mixed finite element spaces,

$$t = r \quad \text{or} \quad t = r + 1.$$

By (7.15a), the inverse property (7.7) and a simple approximation property, we shall see that the first term on the right-hand side of (7.14) is bounded. Given the partition $\mathcal{T}_h$, it is a standard polynomial approximation property that there exists a piecewise-polynomial χ with $\chi|_T = \chi_T \in P_t(T)^d$ such that

$$\|\mathbf{u} - \chi_T\|_{s,T} \leqslant M\|\mathbf{u}\|_{t,T} h^{t-s}, \quad 0 \leqslant s \leqslant t,$$

with M independent of T, since $\mathcal{T}_h$ is (quasi-)regular. Then,

$$\begin{aligned}
\Big(\sum_{T\in\mathcal{T}_h} \|\mathbf{u}_h\|_{t,T}^2\Big)^{1/2} &\leqslant \Big(\sum_{T\in\mathcal{T}_h} \|\mathbf{u}_h - \chi_T\|_{t,T}^2\Big)^{1/2} + \Big(\sum_{T\in\mathcal{T}_h} \|\chi_T - \mathbf{u}\|_{t,T}^2\Big)^{1/2} \\
&\quad + \Big(\sum_{T\in\mathcal{T}_h} \|\mathbf{u}\|_{t,T}^2\Big)^{1/2} \\
&\leqslant M\big(\|\mathbf{u}_h - \chi\|_{0,\Omega}\, h^{-t} + \|\mathbf{u}\|_{t,\Omega}\big) \\
&\leqslant M\big(\big(\|\mathbf{u} - \mathbf{u}_h\|_{0,\Omega} + \|\mathbf{u} - \chi\|_{0,\Omega}\big)h^{-t} + \|\mathbf{u}\|_{t,\Omega}\big) \\
&\leqslant M\|\mathbf{u}\|_{t,\Omega},
\end{aligned} \tag{7.16}$$

as claimed. The inequality above is valid with t replaced by s, $0 \leqslant s < t$ and will be used for such s below.

By the inverse property (7.7), the second term satisfies the inequality

$$\Big(\sum_{T\in\mathcal{T}_h} \big\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big\|_{t,T}^2\Big)^{1/2} \leqslant M'\big\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big\|_{0,\Omega}\, h^{-t},$$

where M' is also independent of h. Thus, it follows that

$$\big\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big\|_{0,\Omega} \leqslant C\|\mathbf{u}\|_{t,\Omega}\, h^{n+1-t}, \tag{7.17}$$

which implies that (assuming $t \leqslant r + 1$)

$$\|\mathbf{u} - \tilde{\mathbf{u}}_h\|_{0,\Omega} \leqslant C\|\mathbf{u}\|_{r+1,\Omega}\, h^{\min(r+1,n+1-t)}. \tag{7.18}$$

Next, assume that

$$t \leqslant n + 1 < 2t.$$

Then, only $n + 1$ derivatives, rather than $2t$, can be applied to $\mathbf{u}_h \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h$. Consequently,

$$\big|c\mathbf{u}_h \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h\big|_{n+1,T} \leqslant \sum_{j=n+1-t}^{t} \|\mathbf{u}_h\|_{j,T}\,\big\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big\|_{n+1-j,T}, \tag{7.19}$$

and applying the inequality (7.16) with j in place of t and the inverse property (7.7) leads to the bound

$$\big\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\big\|_{0,\Omega} \leqslant M \sum_{j=n+1-t}^{t} \|\mathbf{u}\|_{j,\Omega}\, h^{j}. \tag{7.20}$$

We have proved the following theorem.

Theorem 7.1. Let $\mathbf{e}^h_{\tilde{\mathbf{u}}} = \mathbf{u}_h - \tilde{\mathbf{u}}_h$ denote the perturbation in the flux caused by the introduction of the quadrature rule Q_T. Then,

$$\nabla \cdot \mathbf{e}^h_{\tilde{\mathbf{u}}} = 0, \tag{7.21a}$$

$$\left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|_{0,\Omega} \leqslant \begin{cases} C\|\mathbf{u}\|_{t,\Omega}\, h^{n+1-t}, & \text{if } n+1 \geqslant 2t, \\ C \displaystyle\sum_{j=n+1-t}^{t} \|\mathbf{u}\|_{j,\Omega}\, h^j, & \text{if } t \leqslant n+1 < 2t, \end{cases} \tag{7.21b}$$

$$\|\mathbf{u} - \tilde{\mathbf{u}}_h\|_{0,\Omega} \leqslant C\|\mathbf{u}\|_{r+1,\Omega}\, h^{r+1}, \quad \text{if } (t,n) = \begin{cases} (r+1, 2r+1), \\ (r, 2r). \end{cases} \tag{7.21c}$$

If (t,n) equals either $(r+1, 2r+1)$ or $(r, 2r)$, we can obtain a collection of other estimates for $\mathbf{e}^h_{\tilde{\mathbf{u}}}$.

Corollary 7.1. If $(t,n) = (r, 2r)$ or $(r+1, 2r+1)$, then

$$\left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|_{0,\Omega} \leqslant C\|\mathbf{u}\|_{s,\Omega}\, h^{s+1}, \quad 0 \leqslant s \leqslant r. \tag{7.22}$$

Proof. We shall treat the $(r, 2r)$-case first. From (7.12), (7.14), the inverse property (7.7), and (7.16) for t replaced by $s \in [0, t]$,

$$\begin{aligned} \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|^2_{0,\Omega} &\leqslant M\Bigl(\sum_T \|\mathbf{u}_h\|^2_{r,T}\Bigr)^{1/2}\Bigl(\sum_T \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|^2_{r,T}\Bigr)^{1/2} h^{2r+1} \\ &\leqslant M\Bigl(\sum_T \|\mathbf{u}_h\|^2_{s,T}\Bigr)^{1/2} h^{s-r} \cdot \Bigl(\sum_T \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|^2_{0,T}\Bigr)^{1/2} h^{-r} h^{2r+1} \\ &\leqslant M\|\mathbf{u}\|_{s,\Omega} \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|_{0,\Omega} h^{s+1}, \end{aligned} \tag{7.23}$$

which completes the proof for this case.

When $(t,n) = (r+1, 2r+1)$,

$$\begin{aligned} \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|^2_{0,\Omega} &\leqslant M\biggl[\Bigl(\sum_T \|\mathbf{u}_h\|^2_{r,T}\Bigr)^{1/2}\Bigl(\sum_T \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|^2_{r+1,T}\Bigr)^{1/2} \\ &\qquad + \Bigl(\sum_T \|\mathbf{u}_h\|^2_{r+1,T}\Bigr)^{1/2}\Bigl(\sum_T \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|^2_{r,T}\Bigr)^{1/2}\biggr] h^{2r+2} \\ &\leqslant M\Bigl(\sum_T \|\mathbf{u}_h\|^2_{s,T}\Bigr)^{1/2}\Bigl(\sum_T \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|^2_{0,T}\Bigr)^{1/2} h^{-(2r-s+1)} h^{2r+2} \\ &\leqslant M\|\mathbf{u}\|_{s,\Omega} \left\|\mathbf{e}^h_{\tilde{\mathbf{u}}}\right\|_{0,\Omega} h^{s+1}, \end{aligned} \tag{7.24}$$

and the corollary has been established. □

It is worth noting that, if $(t, n) = (r, 2r)$, (7.23) indicates that the the two approximate solutions, $\mathbf{u}_h$ and $\tilde{\mathbf{u}}_h$, are closer together under a slightly less restrictive regularity than either can be globally to $\mathbf{u}$ for that regularity.

Global error estimates for $p - \tilde{p}_h$ can be derived using a modification of the duality argument given in [14]; here, the heavy use of the projections Π_h and P_h and the bounds given by (7.6) is involved. Since all of our examples are related to $(t, n) = (r, 2r)$ or $(t, n) = (r + 1, 2r + 1)$, we shall restrict our attention to these two cases.

Write the error equations in the form

$$\big(c\mathbf{e}^h_{\tilde{\mathbf{u}}}, \mathbf{v}\big) - \big(\nabla \cdot \mathbf{v}, e^h_{\tilde{p}}\big) = f(\mathbf{v}), \quad \mathbf{v} \in \mathbf{V}_h, \tag{7.25a}$$

$$\big(\nabla \cdot \mathbf{e}^h_{\tilde{\mathbf{u}}}, w\big) = 0, \quad w \in W_h, \tag{7.25b}$$

where

$$f(\mathbf{v}) = \mathcal{Q}_h(c\tilde{\mathbf{u}}_h \cdot \mathbf{v}) - (c\tilde{\mathbf{u}}_h, \mathbf{v}), \tag{7.26}$$

and recall that $\nabla \cdot \mathbf{e}^h_{\tilde{\mathbf{u}}} = 0$. Let $\varphi \in H^2(\Omega) \cap H^1_0(\Omega)$ satisfy the (self-adjoint) equation

$$-\nabla \cdot (a\nabla\varphi) = e^h_{\tilde{p}}. \tag{7.27}$$

Then, by (7.25a) and (7.5),

$$\begin{aligned}
\|e^h_{\tilde{p}}\|^2_{0,\Omega} &= \big(e^h_{\tilde{p}}, -\nabla \cdot (a\nabla\varphi)\big) \\
&= \big(e^h_{\tilde{p}}, -\nabla \cdot \big(\Pi_h(a\nabla\varphi)\big)\big) + \big(e^h_{\tilde{p}}, \nabla \cdot \big(\Pi_h(a\nabla\varphi) - a\nabla\varphi\big)\big) \\
&= f\big(\Pi_h(a\nabla\varphi)\big) - \big(c\mathbf{e}^h_{\tilde{\mathbf{u}}}, \Pi_h(a\nabla\varphi)\big) + \big(e^h_{\tilde{p}}, P_h e^h_{\tilde{p}} - e^h_{\tilde{p}}\big) \\
&= f\big(\Pi_h(a\nabla\varphi)\big) - \big(c\mathbf{e}^h_{\tilde{\mathbf{u}}}, \Pi_h(a\nabla\varphi)\big) \\
&= f\big(\Pi_h(a\nabla\varphi)\big) + \big(c\mathbf{e}^h_{\tilde{\mathbf{u}}}, a\nabla\varphi - \Pi_h(a\nabla\varphi)\big) - \big(c\mathbf{e}^h_{\tilde{\mathbf{u}}}, a\nabla\varphi\big) \\
&= f\big(\Pi_h(a\nabla\varphi)\big) + \big(c\mathbf{e}^h_{\tilde{\mathbf{u}}}, a\nabla\varphi - \Pi_h(a\nabla\varphi)\big) + \big(\nabla \cdot \mathbf{e}^h_{\tilde{\mathbf{u}}}, \varphi\big) \\
&= f\big(\Pi_h(a\nabla\varphi)\big) + \big(c\mathbf{e}^h_{\tilde{\mathbf{u}}}, a\nabla\varphi - \Pi_h(a\nabla\varphi)\big).
\end{aligned}$$

Now,

$$\begin{aligned}
f\big(\Pi_h(a\nabla\varphi)\big) &= \mathcal{Q}_h\big(c\tilde{\mathbf{u}}_h \cdot \Pi_h(a\nabla\varphi)\big) - \big(c\tilde{\mathbf{u}}_h, \Pi_h(a\nabla\varphi)\big) \\
&= \sum_{T \in \mathcal{T}_h} \Big(Q_T\big(c\tilde{\mathbf{u}}_h \cdot \Pi_h(\nabla\varphi)\big) - \big(c\tilde{\mathbf{u}}_h, \Pi_h(\nabla\varphi)\big)_T\Big),
\end{aligned}$$

so that, if $(t, n) = (r, 2r)$ and $0 \leqslant s \leqslant r$,

$$\begin{aligned}
\big|f\big(\Pi_h(a\nabla\varphi)\big)\big| &\leqslant M_1 \sum_T \big|\tilde{\mathbf{u}}_h \cdot \Pi_h(\nabla\varphi)\big|_{2r+1,T}\, h^{2r+1} \\
&\leqslant M_2 \Big(\sum_t \|\tilde{\mathbf{u}}_h\|^2_{r,T}\Big)^{1/2} \Big(\sum_T \big\|\Pi_h(\nabla\varphi)\big\|^2_{r,T}\Big)^{1/2} h^{2r+1}
\end{aligned}$$

$$
\begin{aligned}
&\leqslant M_3\Bigl(\sum_t \|\tilde{\mathbf{u}}_h\|_{s,T}^2\Bigr)^{1/2}\Bigl(\sum_T \bigl\|\Pi_h(\nabla\varphi)\bigr\|_{1,T}^2\Bigr)^{1/2} h^{s+2}\\
&\leqslant M_4\|\mathbf{u}\|_{s,\Omega}\,\bigl\|e_{\tilde{p}}^h\bigr\|_{0,\Omega}\,h^{s+2},
\end{aligned}
$$

where the same justifications have been used here as in the analyses of the error in the flux, plus the fact that

$$
\Bigl(\sum_T \bigl\|\Pi_h(\nabla\varphi)\bigr\|_{1,T}^2\Bigr)^{1/2} \leqslant C_1\|\varphi\|_2 \leqslant C_2\bigl\|e_{\tilde{p}}^h\bigr\|_0.
$$

The same result holds for the case that $(t,n) = (r+1, 2r+1)$; see the proof of corollary 7.1.

By (7.6a),

$$
\bigl|\bigl(c\mathbf{e}_{\tilde{\mathbf{u}}}^h, a\nabla\varphi - \Pi_h(a\nabla\varphi)\bigr)\bigr| \leqslant M_5\bigl\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\bigr\|_0\,|\varphi|_2\,h \leqslant M_6\bigl\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\bigr\|_0\,\bigl\|e_{\tilde{p}}^h\bigr\|_0\,h.
$$

Thus, we have shown that

$$
\begin{aligned}
\|p_h - \tilde{p}_h\|_{0,\Omega} = \bigl\|e_{\tilde{p}}^h\bigr\|_{0,\Omega} &\leqslant M\bigl(\|\mathbf{u}\|_{s,\Omega}\,h^{s+2} + \bigl\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\bigr\|_{0,\Omega}\,h\bigr), \quad 0 \leqslant s \leqslant r,\\
&\leqslant M\|\mathbf{u}\|_{s,\Omega}\,h^{s+2}\\
&\leqslant M\|p\|_{k,\Omega}\,h^{k+1}, \qquad\qquad 1 \leqslant k \leqslant r+1.
\end{aligned}
\tag{7.28}
$$

Again, we have shown that the approximations p_h and $\tilde{p}_h$ are closer together than they are to p; here, the estimate of their difference is better with respect both to regularity and to the maximum exponent on h.

We can summarize the error estimates for $\tilde{p}_h$ in the following theorem.

Theorem 7.2. Assume that $(t,n) = (r, 2r)$ or $(r+1, 2r+1)$. Then,

$$
\|p_h - \tilde{p}_h\|_{0,\Omega} = \bigl\|e_{\tilde{p}}^h\bigr\|_{0,\Omega} \leqslant M\|p\|_{k,\Omega}\,h^{k+1}, \quad 1 \leqslant k \leqslant r+1, \tag{7.29a}
$$

$$
\|p - \tilde{p}_h\|_{0,\Omega} \leqslant M\|p\|_{k,\Omega}\,h^{k}, \qquad 0 \leqslant k \leqslant r^*. \tag{7.29b}
$$

The bound (7.29b) is the optimal global rate of convergence for $\tilde{p}_h$ and requires the minimal regularity for this rate.

A slightly more careful argument would have allowed $\|p\|_{k,\Omega}$ to be replaced by the broken norm

$$
\Bigl(\sum_{T\in\mathcal{T}_h} \|p\|_{k,T}^2\Bigr)^{1/2};
$$

the same remark applies to the bounds for $\mathbf{u} - \tilde{\mathbf{u}}_h$.

7.2. *Error estimates for the $BDFM_2$-based FVM*

Error bounds for $(\tilde{\mathbf{u}}_h, \tilde{p}_h)$, where $\tilde{p}_h$ is the solution of the $BDFM_2$-based FVM equations (6.34) can be obtained directly from corollary 7.1 and theorem 7.2. For this method,

$$r = 1, \quad t = 2, \quad n = 3, \quad (t, n) = (r + 1, 2r + 1), \quad r^* = 2. \tag{7.30}$$

Thus,

$$\|p - \tilde{p}_h\|_{0,\Omega} \leqslant C\|p\|_{s,\Omega} h^s, \quad 1 \leqslant s \leqslant 2, \tag{7.31a}$$

$$\|\mathbf{u} - \tilde{\mathbf{u}}_h\|_{0,\Omega} \leqslant C\|\mathbf{u}\|_{s,\Omega} h^s, \quad 0 \leqslant s \leqslant 2. \tag{7.31b}$$

These are optimal estimates for the errors in both variables.

8. Finite volume methods based on *RTN*-rectangular elements

The Raviart–Thomas–Nedelec rectangular mixed finite elements of index k have as local bases the tensor product spaces $RTN_k = \mathbf{V}_k \times W_k$, where

$$\mathbf{V}_k(T) = \begin{cases} P_{k+1,k}(T) \times P_{k,k+1}(T), & \dim(\Omega) = 2, \\ P_{k+1,k,k}(T) \times P_{k,k+1,k}(T) \times P_{k,k,k+1}(T), & \dim(\Omega) = 3, \end{cases} \tag{8.1a}$$

$$W_k(T) = \begin{cases} P_{k,k}(T), & \dim(\Omega) = 2, \\ P_{k,k,k}(T), & \dim(\Omega) = 3. \end{cases} \tag{8.1b}$$

Note that, in hybridizing the mixed finite element equations, the space Λ_k would consist of a copy of P_k for $d = 2$ and $P_{k,k}$ for $d = 3$ on each interface between elements; thus, the Fraeijs de Veubeke reduction to a positive-definite system in the Lagrange multipliers would lead to a block system with the unknowns on an interface seeing those on the other six faces of the two elements generating the interface if $d = 2$. If $d = 3$, then each block of interface unknowns would see those for the other ten faces of the adjacent elements. Also, note that there are two $(k + 1)$-blocks for $d = 2$ or three $(k + 1)^2$-blocks for $d = 3$ associated with each element for the Fraeijs de Veubeke reduced equations, while there would be a single $(k+1)^2$- or $(k+1)^3$-block per element in a FVM reduction, if feasible. The FVM-block equations would have the same 5-point or 7-point structures as are associated with the simplest finite difference equations for the elliptic equation (2.1) in two or three space variables. We shall show that it is easy to construct a convenient nodal basis for RTN_k and an associated quadrature relation that maintains the global accuracy of the underlying mixed method, so that we can achieve its FVM.

Is it desirable to have the FVM corresponding to RTN_k? Look first at the three-dimensional problem. For $k = 0$, the finite volume method associates one parameter for the scalar function to each element, while there are three Lagrange multipliers associated with an element; moreover, the graph structure is simpler for the finite volume method. Clearly, the finite volume route is superior for $k = 0$. For $k = 1$, there are 12 Lagrange multipliers per element versus eight scalar parameters, plus a simpler graph structure, so that the finite volume approach is again superior. For $k = 2$, there are 27 parameters per element either way; here there is a slight advantage to the finite volume technique. For $k > 2$, the number of scalar parameters exceeds that for Lagrange multipliers, and it is doubtful that the finite volume approach is very helpful.

In the two-dimensional case, there are $2(k + 1)$ Lagrange parameters per element versus $(k+1)^2$ scalar parameters. It appears that the finite volume reduction is better than the Lagrange multiplier procedure for RT_0 and RT_1, but not for $k > 1$. For $k = 1$, the resulting FVM has four scalar parameters per element (i.e., a bilinear function), while the competing $BDFM_2$ finite volume method has three (a linear function). Since both lead to the same error estimates, it would appear that the FVM based on $BDFM_2$ is somewhat more efficient than the RT_1-based one.

As a consequence of these remarks, we shall treat the three-dimensional case; the two-dimensional case is an obvious specialization of it. Though the higher order finite volume methods based on RTN_k may be of lesser practical interest, the analysis is effectively independent of k and will be given for arbitrary k. Consider the x-component $v(\mathbf{x})$ of a vector $\mathbf{v} \in \mathbf{V}_h$. Then, $v \in P_{k+1,k,k}$. Thus, there are $(k + 2)(k + 1)(k + 1)$ degrees of freedom for v, with $(k+1)^2$ necessarily associated with each face of the form $x = const$. Consider first the reference element $T_{\text{ref}} = [-1, 1]^3$. Denote the (one-dimensional) Lobatto quadrature rule with $k + 2$ nodes by $\mathcal{L}_{k+2}$, where

$$\mathcal{L}_{k+2}(g) = \sum_{i=0}^{k+1} g(\eta_i)\omega_i, \quad -1 = \eta_0 < \eta_1 < \cdots < \eta_{k+1} = 1,$$

and the $k + 1$ node Gauss rule by $\mathcal{G}_{k+1}$, where

$$\mathcal{G}_{k+1}(g) = \sum_{j=1}^{k+1} g(\zeta_j)w_j, \quad -1 < \zeta_1 < \cdots < \zeta_{k+1} < 1.$$

It is well known that

$$\mathcal{L}_{k+2}(q) = \mathcal{G}_{k+1}(q) = \int_{-1}^{1} q(x)\,\mathrm{d}x, \quad q \in P_{2k+1}\big([-1, 1]\big);$$

thus,

$$\begin{aligned} Q_{x,T_{\text{ref}}}(g) = \mathcal{L}_{k+2} \times \mathcal{G}_{k+1} \times \mathcal{G}_{k+1}(g) &= \sum_{i=0}^{k+1}\sum_{j=1}^{k+1}\sum_{\ell=1}^{k+1} g(\eta_i, \zeta_j, \zeta_\ell)\omega_i w_j w_\ell \\ &= \int_{T_{\text{ref}}} g(\mathbf{x})\,\mathrm{d}\mathbf{x}, \quad g \in P_{k+1,k,k}. \end{aligned} \tag{8.2}$$

Now, let $Q_{x,T}$ be defined in the obvious way by mapping the nodes $(\eta_i, \zeta_j, \zeta_\ell)$ to T affinely and multiplying its weight by h^3, and define $Q_{y,T}$ and $Q_{z,T}$ analogously. Then, if $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$, let

$$Q_T(c\mathbf{u}\cdot\mathbf{v}) = Q_{x,T}(cu_1v_1) + Q_{y,T}(cu_2v_2) + Q_{z,T}(cu_3v_3). \tag{8.3}$$

Take $\{\eta_i\}\times\{\zeta_j\}\times\{\zeta_\ell\}$ as nodes for the x-component $v_1 \in P_{k+1,k,k}$ of $\mathbf{v}$ on the reference element, and define the (i,j,ℓ)-basis element for this component by requiring that

$$\boldsymbol{\xi}^x_{i,j,\ell}(\eta_{i'},\zeta_{j'},\zeta_{\ell'}) = \begin{cases} 1, & (i',j',\ell') = (i,j,\ell),\\ 0, & (i',j',\ell') \neq (i,j,\ell). \end{cases} \tag{8.4}$$

Similarly, take $\{\zeta_i\}\times\{\eta_j\}\times\{\zeta_\ell\}$ and $\{\zeta_i\}\times\{\zeta_j\}\times\{\eta_\ell\}$ as nodes for the y- and z-components, with corresponding basis elements.

Clearly, the quadrature rule

$$\mathcal{Q}_h(c\mathbf{u}_h\cdot\mathbf{v}_h) = \sum_{T\in\mathcal{T}_h} Q_T(c\mathbf{u}_h\cdot\mathbf{v}_h)$$

diagonalizes each matrix $A_{h,T}$, thereby generating a FVM in our sense to be associated with $RTN_k(\mathcal{T}_h)$.

Let us consider error estimates for the resulting FVM. Employ the same notation as in section 7. First, Q_T is exact on $P_{2k+1,2k+1,2k+1}(T)$. Thus,

$$r = k,\quad t = k+1,\quad n = 2k+1,\quad (t,n) = (r+1, 2r+1),\quad r^* = k+1.$$

So, corollary 7.1 and theorem 7.2 imply the following optimal error bounds:

$$\|p-\tilde{p}_h\|_{0,\Omega} \leqslant C\|p\|_{s,\Omega}\,h^s,\qquad 1\leqslant s\leqslant k+1, \tag{8.5a}$$

$$\|\mathbf{u}-\tilde{\mathbf{u}}_h\|_{0,\Omega} \leqslant C\|\mathbf{u}\|_{s,\Omega}\,h^{s+1},\quad 0\leqslant s\leqslant k+1. \tag{8.5b}$$

The estimates (8.5) hold in the two-dimensional case, as well.

9. The three-dimensional $BDFM_2$-based finite volume method

Here, we derive a higher order finite volume scheme based on the three-dimensional $BDFM_2$ space using an appropriate quadrature formula. The error analysis below is applicable for $c = c(x,y,z)$ being variable, but the quadrature rule (9.1), which differs in type from those applied to the other mixed finite element spaces discussed in this paper, can be gauranteed to diagonalize the $(c\mathbf{u},\mathbf{v})$-matrix only if c is constant on each element.

Let $\mathcal{T}_h$ be the partition of Ω into cubes with the length $h = N^{-1}$ of each edge:

$$\overline{\Omega} = \bigcup_{i,j,k=1}^{N} \overline{T}_{ijk},$$

where

$$\begin{aligned}\overline{T} = \overline{T}_{ijk} &= \big[(i-1)h, ih\big]\times\big[(j-1)h, jh\big]\times\big[(k-1)h, kh\big]\\ &= [x_{i-1},x_i]\times[y_{j-1},y_j]\times[z_{k-1},z_k].\end{aligned}$$

Let $P_k(\theta)$ for $\theta = x$, y or z be the set of all polynomials in the single variable θ of the degree not greater than k. Set

$$\theta_i' = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_3)$$

and denote by $P_j(\mathrm{hom}, T, i)$ the homogeneous polynomials of degree j in the variables θ_i'.

For each element $T \in \mathcal{T}_h$, $BDFM_2$ is given by

$$\begin{aligned}
&\mathbf{V}_h(T) = \big[P_2(T)\backslash P_2(\mathrm{hom}, T, 1)\big] \times \big[P_2(T)\backslash P_2(\mathrm{hom}, T, 2)\big] \times \big[P_2(T)\backslash P_2(\mathrm{hom}, T, 3)\big],\\
&W_h(T) = P_1(T),\\
&\Lambda_h(e) = \big\{m \in L^2(e)\colon\ m \in P_1(e),\ e \in E_T\big\},
\end{aligned}$$

where $E_T = \{rr, f, \ell, r, b, t\}$ for rear, front, left, right, bottom, and top faces.

Assume the coefficient c to be constant on each T; for convenience, take $c = c_T = 1$ on T. It suffices to discuss one component in the inner product, say, the x-component. So, consider the reference element $\widehat{T} = [-1, 1]^3$ and let

$$\mathbf{V} = \mathbf{V}_x = \mathbf{V}_x\big(\widehat{T}\big) = \mathrm{Span}\big\{1, x, y, z, x^2, xy, xz\big\} = \mathrm{Span}\{\boldsymbol{\xi}_j,\ j = 1, \dots, 7\},$$

where

$$\begin{aligned}
&\boldsymbol{\xi}_1 = 1 - x^2, && \boldsymbol{\xi}_2 = -\frac{1}{2}x(1-x), && \boldsymbol{\xi}_3 = \frac{1}{2}x(1+x),\\
&\boldsymbol{\xi}_4 = \frac{1}{2}y(1-x), && \boldsymbol{\xi}_5 = \frac{1}{2}z(1-x), &&\\
&\boldsymbol{\xi}_6 = \frac{1}{2}y(1+x), && \boldsymbol{\xi}_7 = \frac{1}{2}z(1+x). &&
\end{aligned}$$

These basis functions correspond to the following degrees of freedom: the values at $(0, 0, 0)$ and $(\pm 1, 0, 0)$ and y and z components of the gradient at $(\pm 1, 0, 0)$. Basis functions for $W_h(\widehat{T})$ can be specified simply as

$$\begin{aligned}
&\eta^0_{\widehat{T}}(x, y, z) = 1, && \eta^x_{\widehat{T}}(x, y, z) = x - x_{i-1/2},\\
&\eta^y_{\widehat{T}}(x, y, z) = y - y_{j-1/2}, && \eta^z_{\widehat{T}}(x, y, z) = z - z_{k-1/2},
\end{aligned}$$

so that

$$p_{h,\widehat{T}}(x, y) = p^0_{h,\widehat{T}}\eta^0_{\widehat{T}} + p^x_{h,\widehat{T}}\eta^x_{\widehat{T}} + p^y_{h,\widehat{T}}\eta^y_{\widehat{T}} + p^z_{h,\widehat{T}}\eta^z_{\widehat{T}}.$$

Now, what is needed is a rule assigning approximations for integrals over $\mathbf{V} \otimes \mathbf{V}$ (i.e., for products $\boldsymbol{\xi}_j\boldsymbol{\xi}_k$) which diagonalizes the matrix $a_{i,j} = (\xi_j, \xi_i)$. For the other mixed finite element spaces, the quadrature rule was based on the degrees of freedom for the space; unfortunately, there does not exist a quadrature rule based on the seven degrees of freedom for $\mathbf{V}_x$ that is exact on $P_3(\widehat{T})$. Thus, we are led to a rule involving second derivatives of normal components at facial midpoints, given by

$$Q_{\widehat{T}}(\mathbf{u}\cdot\mathbf{v}) = Q^x_{\widehat{T}}(g_1) + Q^y_{\widehat{T}}(g_2) + Q^z_{\widehat{T}}(g_3), \quad \text{if } \mathbf{u}\cdot\mathbf{v} = \sum_{k=1}^{3} u_k v_k = \sum_{k=1}^{3} g_k,$$

where

$$Q^x_{\widehat{T}}(g) = \frac{2}{3}\left(8g(0,0,0) + 2\sum_{\mathbf{x}\in\{(\pm1,0,0)\}} g(\mathbf{x}) + \sum_{\mathbf{x}\in\{(\pm1,0,0)\}}\left(\frac{\partial^2 g}{\partial y^2}(\mathbf{x}) + \frac{\partial^2 g}{\partial z^2}(\mathbf{x})\right)\right), \quad (9.1)$$

and $Q^y_{\widehat{T}}$ and $Q^z_{\widehat{T}}$ are defined analogously; $Q^x_{\widehat{T}}$ is exact on $P_3(\widehat{T})$ and vanishes for $g = \xi_j\xi_i$ for $i \neq j$, as desired. When scaled for a cube of side length h, $Q^x_{\widehat{T}}$ is replaced by

$$Q^x_T(g) = \frac{h^3}{12}\left(8g(\mathbf{a}_T) + 2\sum_{\mathbf{x}=\mathbf{a}_f,\mathbf{a}_{rr}} g(\mathbf{x}) + \frac{h^2}{4}\sum_{\mathbf{x}=\mathbf{a}_f,\mathbf{a}_{rr}}\left(\frac{\partial^2 g}{\partial y^2}(\mathbf{x}) + \frac{\partial^2 g}{\partial z^2}(\mathbf{x})\right)\right).$$

The authors are unaware of any previous appearance of the quadrature formula (9.1).

Now, it is clear that there exist positive α_1 and α_2 such that

$$\alpha_1(\mathbf{u},\mathbf{u})_T \leqslant Q_T\left(|\mathbf{u}|^2\right) \leqslant \alpha_2(\mathbf{u},\mathbf{u})_T, \quad \mathbf{u}\in\mathbf{V}(T),$$

so that Q_T generates a norm on $\mathbf{V}\otimes\mathbf{V}$ that is equivalent to the ordinary L^2-norm. In order to make use of the general convergence argument, it is necessary to show that

$$\left|Q_T(\mathbf{uv}) - (\mathbf{u},\mathbf{v})_T\right| \leqslant M\|\mathbf{u}\|_{2,T}\,\|\mathbf{v}\|_{2,T}h^4; \quad (9.2)$$

however, this follows trivially from the exactness of the quadrature rule on $P_3(T)$. Thus, the same error estimates hold for the $BDFM_2$-based three-dimensional procedure as for the two-dimensional one; that is, the errors in the scalar and flux approximations are $O(h^2)$.

10. A finite volume method based on $BDFM_3$

The $BDFM_k = \mathbf{V}^k_h \times W^k_h$ mixed finite element space over rectangles is defined locally by

$$\mathbf{V}^k_h(T) = \left[P_k(T)\setminus\left\{y^k\right\}\right]\times\left[P_k(T)\setminus\left\{x^k\right\}\right], \qquad W^k_h(T) = P_{k-1}(T).$$

Thus,

$$r = k-1, \qquad t = k, \qquad r^* = k;$$

in order that the approximation $\tilde{\mathbf{u}}^k_h$ to $\mathbf{u}$ be of optimal order, it is necessary that the quadrature rule Q^k_T be exact on polynomials of degree

$$n = 2k-1,$$

since $t = r + 1$, so that $2r + 1 = 2k - 1$. We shall consider the cases $k = 3$ and $k = 4$, for which we need quadrature rules exact on $P_5(T)$ and $P_7(T)$, respectively.

For $k = 3$, recall [8] that the projection Π_h for $BDFM_3$ is determined by the degrees of freedom

$$\begin{aligned} \big\langle(\varphi - \Pi_h\varphi)\cdot\nu, q\big\rangle_e &= 0, \quad q \in P_2(e),\ \forall e \in \partial T, \\ (\varphi - \Pi_h\varphi, \chi)_T &= 0, \quad \chi \in P_1(T)^2. \end{aligned}$$

So, for the x-component v_1 of $\mathbf{v} \in \mathbf{V}_h^3(T)$, it would be natural to take the following degrees of freedom (with $\boldsymbol{\ell}$ and $\mathbf{r}$ denoting the left and right edges of T, respectively)

$$v_1(\mathbf{x}_i), \quad i = 0, \ldots, 6, \tag{10.1a}$$

$$\frac{\partial v_1}{\partial x}(\mathbf{x}_0), \qquad \frac{\partial v_1}{\partial y}(\mathbf{x}_0), \tag{10.1b}$$

where $\mathbf{x}_0$ is the center of T, $\mathbf{x}_i$, $i = 1, \ldots, 4$, is a vertex of T, and $\mathbf{x}_5$ and $\mathbf{x}_6$ are the midpoints of $\boldsymbol{\ell}$ and $\mathbf{r}$, respectively. It is easy to show that this set of degrees of freedom determine the first component of $\mathbf{v} \in \mathbf{V}_h^3(T)$.

To these degrees of freedom, we should like to associate a quadrature rule of the form (on the reference element $[-1, 1]^2$)

$$Q_{\widehat{T}}(g) = \sum_{i=0}^{6} g(\mathbf{x}_i)w_i + \omega_{x,x}g_x(\mathbf{x}_0) + \omega_{x,y}g_y(\mathbf{x}_0). \tag{10.2}$$

Unfortunately, the best quadrature rule for these degrees of freedom is the same rule as we applied for $BDFM_2$. It is exact on $P_3(T)$, but it obviously fails for $q(\mathbf{x}) = x^2(1-x^2)$. Clearly, this quadrature rule kills four of the basis functions for each component of $\mathbf{v} \in \mathbf{V}_h^k(T)$. Thus, it is a bit surprising that any convergence result can occur with this rule, but let us show a suboptimal result.

With Q_T defined by (6.22), $n = 3$ and, following the general argument of section 7, we see that

$$\left|c\mathbf{u}_h \cdot \mathbf{e}_{\tilde{\mathbf{u}}}^h\right|_{4,T} \leqslant M \sum_{j=0}^{3} \|\mathbf{u}_h\|_{j,T} \left\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\right\|_{3-j,T} h^4 \leqslant M\|\mathbf{u}_h\|_{0,T} \left\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\right\|_{0,T} h, \tag{10.3}$$

from which it follows that

$$\left\|\mathbf{e}_{\tilde{\mathbf{u}}}^h\right\|_{0,\Omega} \leqslant M\|\mathbf{u}\|_{0,\Omega}\, h. \tag{10.4}$$

All that follows from (10.4) is that

$$\|\mathbf{u} - \tilde{\mathbf{u}}_h\|_{0,\Omega} \leqslant M\|\mathbf{u}\|_{1,\Omega}\, h, \tag{10.5}$$

a distinctly suboptimal result.

The authors have so far not found a set of degrees of freedom for $\mathbf{V}_h^3(T)$ and a corresponding quadrature rule to improve on this $\mathrm{O}(h)$ error estimate for $\tilde{\mathbf{u}}_h$. Currently, $BDFM_3$ serves as an example of a mixed finite element space for which our procedure for deriving a satisfactory finite volume procedure fails. In a sense, the difficulty lies in the basic efficiency of the $BDFM_3$ mixed method; the dimension of $\mathbf{V}_h^3(T)$ is too small to support a quadrature rule that is exact for $P_5(T)$, whereas the less efficient RTN_2 method does. On the other hand, this finite volume method could be used to find an initial guess for an Uzawa (or other) iteration for the solution of the original mixed method.

References

[1] T. Arbogast, C.N. Dawson, P.T. Keenan, M.F. Wheeler and I. Yotov, Enhanced cell-centered finite differences for elliptic equations on general geometry, SIAM J. Numer. Anal. 19 (1998) 404–425.

[2] T. Arbogast, M.F. Wheeler and I. Yotov, Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences, SIAM J. Numer. Anal. 34 (1997) 828–852.

[3] D.N. Arnold and F. Brezzi, Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates, RAIRO Modél. Math. Anal. Numér. 19 (1985) 7–32.

[4] B.R. Baliga and S.V. Patankar, A new finite-element formulation for convection–diffusion problems, Numer. Heat Transfer 3 (1980) 393–409.

[5] R.E. Bank and D.J. Rose, Some error estimates for the box method, SIAM J. Numer. Anal. 24 (1987) 777–787.

[6] J. Baranger, J.F. Maitre and F. Oudin, Connection between finite volume and mixed finite element methods, Modél. Math. Anal. Numér. 30(4) (1996) 445–465.

[7] F. Brezzi, J. Douglas, Jr., R. Durán and M. Fortin, Mixed finite elements for second order elliptic problems in three variables, Numer. Math. 51 (1987) 237–250.

[8] F. Brezzi, J. Douglas, Jr., M. Fortin and L.D. Marini, Efficient rectangular mixed finite elements in two and three space variables, RAIRO Anal. Numér. 21 (1987) 581–604.

[9] F. Brezzi, J. Douglas, Jr. and L.D. Marini, Two families of mixed finite elements for second order elliptic problems, Numer. Math. 47 (1985) 217–235.

[10] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods* (Springer, New York, 1991).

[11] Z. Cai, On the finite volume element method, Numer. Math. 58 (1991) 713–735.

[12] Z. Cai, J. Mandel and S. McCormick, he finite volume element method for diffusion equations on general triangulations, SIAM J. Numer. Anal. 28(2) (1991) 392–402.

[13] P.G. Ciarlet, *The Finite Element Method for Elliptic Problems* (North-Holland, Amsterdam, 1978).

[14] J. Douglas, Jr. and J.E. Roberts, Global estimates for mixed methods for second order elliptic equations, Math. Comp. 44 (1985) 39–52.

[15] J. Douglas, Jr. and J. Wang, A new family of mixed finite element spaces over rectangles, Mat. Apl. Comput. 12 (1993) 183–197.

[16] R.S. Falk and J.E. Osborn, Error estimates for mixed methods, RAIRO Anal. Numér. 14 (1980) 309–324.

[17] B.X. Fraeijs de Veubeke, Stress function approach, in: *Internat. Congress on the Finite Element Method in Structural Mechanics*, Bournemouth, 1975.

[18] B. Heinrich, *Finite Difference Methods on Irregular Networks* (Birkhäuser, Basel, 1987).

[19] J.C. Nedelec, Mixed finite elements in $\mathbb{R}^3$, Numer. Math. 35 (1980) 315–341.

[20] P.A. Raviart and J.M. Thomas, A mixed finite element method for 2nd order elliptic problems, in: *Mathematical Aspects of Finite Element Methods*, Lecture Notes in Mathematics, Vol. 606, eds. I. Galligani and E. Magenes (Springer, New York, 1977) pp. 292–315.

[21] T.F. Russell and M.F. Wheeler, Finite element and finite difference methods for continuous flows in porous media, in: *The Mathematics of Reservoir Simulation*, ed. R.E. Ewing (SIAM, Philadelphia, PA, 1983).