Comparing Two Approaches of Analyzing Mixed Finite Volume Methods

So-Hsiang Chou and Shengrong Tang

Abstract

Given the anisotropic Poisson equation \(-\nabla \cdot K \nabla p = f\), one can convert it into a system of two first order PDEs: the Darcy law for the flux \(u = -K \nabla p\) and conservation of mass \(\nabla \cdot u = f\). A very natural mixed finite volume method for this system is to seek the pressure in the nonconforming \(P_1\) space and the Darcy velocity in the lowest order Raviart-Thomas space. The equations for these variables are obtained by integrating the two first order systems over the triangular volumes. In this paper we show that such a method is really a standard finite element method with local recovery of the flux in disguise. As a consequence, we compare two approaches in analyzing finite volume methods (FVM) and shed light on the proper way of analyzing non co–volume type of FVM. Numerical results for Dirichlet and Neumann problems are included.

1 Introduction

Consider the variable-coefficient Poisson equation in a polygonal domain \(\Omega \subset \mathbb{R}^2\)

\[
\begin{aligned}
\left\{ \begin{array}{ll}
-\nabla \cdot K \nabla p &= f & \text{in } \Omega, \\
p &= 0 & \text{on } \partial \Omega,
\end{array} \right.
\end{aligned}
\]

where \(K = K(x)\) is a symmetric positive definite matrix function such that there exist two positive constants \(\alpha_1\) and \(\alpha_2\) with

\[
\alpha_1 \xi^T \xi \leq K(x) \xi^T \xi \leq \alpha_2 \xi^T \xi \quad \forall \xi \in \mathbb{R}^2, x \in \bar{\Omega}. 
\]

Now let us introduce a flux variable \(u := -K \nabla p\) and write the above equation as the system of first order partial differential equations

\[
\begin{aligned}
\left\{ \begin{array}{ll}
\nabla \cdot u - f &= 0 & \text{in } \Omega \\
u + K \nabla p &= 0 & \text{in } \Omega \\
p &= 0 & \text{on } \partial \Omega
\end{array} \right.
\end{aligned}
\]
Problems (1) and (3) are equivalent when the right hand side $f$, the diffusion tensor $K$ and the domain $\Omega$ (e.g. convex) are such that the solution is smooth enough, e.g., $u \in H^1(\Omega)^2, p \in H^1_0(\Omega) \cap H^2(\Omega)$.

This system can be interpreted as modeling an incompressible single phase flow in a reservoir; ignoring gravitational effects. The matrix $K$ is the mobility $\kappa/\mu$, the ratio of permeability tensor to viscosity of the fluid, $u$ is the Darcy velocity and $p$ the pressure. The second equation is the Darcy law and the first represents conservation of mass with $f$ standing for a source or sink term. Since $\kappa$ is in general discontinuous due to different rock formations, separating the Darcy law from the second order equation and discretizing it directly together with the mass conservation may lead to a better numerical treatment on the velocity than just computing it from the pressure via the Darcy law. This approach is well known in the finite element circle [20], but the same approach can be applied in conjunction with the finite volume method (termed mixed finite volume methods) as well [6, 9, 10, 11, 12, 20, 21]. For other similarly related issues, see also [17, 18, 22].

Let $T_h = \{K_j\}_{j=1}^{NT}$ be the usual non–overlapping finite element triangulation of the domain $\Omega = \bigcup_{K \in T_h} K$. Furthermore $T_h$ is assumed to be regular, that is, $\min_{K \in T_h} d(K)/\rho(K) \geq C$ for a constant $C$ independent of $h$. Here $\rho(K)$ is the diameter of triangle $K$; $d(K)$ the diameter of the inscribed circle of $K$, and $h = \max_{K \in T_h} \rho(K)$.

We denote the area of $K$ by $|K|$, by $A = A_i \cup A_b$ the set of all edges of $T_h$ consisting of the interior edge set $A_i$ and boundary edge set $A_b$. We use $N_{A_i}$ and $N_{A_b}$ to denote the number of interior edges and the number of boundary edges, respectively. The total number of edges is $N_A = N_{A_i} + N_{A_b}.$

Define the lowest order Raviart-Thomas space [19]
$$V_h = \{u_h \in H(\text{div}; \Omega) : u_h|_K \in RT_0(K)\}$$
where $RT_0(K) = \{u = (u^1, u^2) : u^1 = a + bx, u^2 = c + by \text{ in } K\}$ and the standard $P1$ nonconforming finite element space
$$Y_h = \{p_h|_K \in P_1(K) : p_h \text{ continuous at the middle point of each } e \in \partial K\}.$$ Consider approximating $u$ by $u_h \in V_h$ and $p$ by $p_h \in Y_h$ via a mixed finite volume approach on (3):

Find $(u_h, p_h) \in V_h \times Y_h$ such that
$$\begin{cases} 
(\nabla \cdot u_h - f, \chi_K) &= 0 \text{ for all } K \in T_h, \\
(u_h + K \nabla p_h, \chi_K) &= 0 \text{ for all } K \in T_h
\end{cases}$$
(4)
where $p_h = 0$ at all midpoints of boundary edges. Here $\chi_K$ is the characteristic function of triangle $K$ and $\chi_K$ is any constant vector multiple of $\chi_K$.

This method was introduced and analyzed for the case of $K = I$, the identity matrix, by Courbet & Croisille [15]. In this paper we study the general full tensor case. The
full tensor case is important in porous media flow applications since it represents the more natural anisotropic case. Mathematically, the full tensor case is by no means a trivial extension of the simple Poisson equation, especially in the finite volume case: the presence of the full tensor often results in a nonsymmetric system and is harder to give physical interpretations [1]. Furthermore, unlike in the mathematical analysis of finite element methods, convergence of finite volume methods for the isotropic and anisotropic cases in general cannot be handled in a unified way. See [16] and the references therein for the difficulties involved. Also see [10, 11] where a uniform treatment is possible, but in these co-volume schemes Chou et al. used two grid systems in discretization. It is therefore surprising that the extension of the Courbet & Croisille method has none of these drawbacks, having only one grid system for both variables.

This paper is the first in a series of two papers on comparing the analysis of the finite volume method from a finite element person’s or a “pure” finite volume person’s viewpoint. It grows out of the informal report [13]. We will present things in an elementary and non–terse way and the point will be how different viewpoints may lead to the same conclusions but in a roundabout way. The reader is referred to follow-up [14] for a more extensive and in-depth mathematical presentation.

First we adopt a pure finite volume person’s point of view. In Sec. 2, we derive the discrete system, and in Sec. 3 we compute explicitly the element and global stiffness matrices whose proper interpretations lead to Thm 4.1, which says that the pressure approximation can be re–interpreted as the solution to the system of the standard \( P_1 \) nonconforming finite elements applied to the second order elliptic problem (1). However, the process leading to that conclusion is long and technical. The advantage in this approach is that conservation laws are obvious from the start and the drawback is that the mathematical analysis is long and very uninspiring.

On the other hand, as shown in [8, 12, 11] it is fruitful to try to relate the analysis of a mixed finite volume method to a close finite element method. Adopting this viewpoint, we can prove Thm. 4.1 in a few short lines. The advantage of this approach is that mathematical analysis is neat, but conservation law is somewhat hidden. To keep the paper short, how to get the best of these two approaches is elucidated in the follow-up paper [14].

The remaining of this paper is organized as follows. In Sec. 5, we show the error estimates in \( p \) and \( u \). Finally, in the last section we provide numerical results for both Dirichlet and Neumann problems.

2 Problem formulation

Let us now represent the system (4) using proper basis functions. Notice that by divergence theorem, Eq. (4)_1 can be written as

\[
\int_{\partial K} \mathbf{u}_h \cdot \mathbf{n} \, d\mathbf{x} - |K| f_K = 0
\]
where \( f_K = \frac{1}{|K|} \int_K f \, dx \) is the average of \( f(x) \) over triangle \( K \). Since \( \nabla p_h \) is constant over \( K \), Eq. (4) can be written as

\[
\int_K u_h \, dx + |K| A_K \nabla p_h = 0 \tag{6}
\]

where the matrix \( A_K = \frac{1}{|K|} \int_K K \, dx \) is the average of matrix \( K(x) \) over triangle \( K \).

---

Figure 1: Local elements based on an interior edge \( e_1 \).

With reference to Fig. 1, let \( \lambda_S \) be the usual nodal linear basis function associated with the vertex \( S \) of \( K = K_L \). Recall that \( \lambda_S \) is one at \( S \) and zero at other two vertices and is also called barycentric or area coordinate function. For any \( p_h(x) \in Y_h \), we have the local representation on \( K \)

\[
p_h(x)|_K = \sum_{e \in \partial K} p_e \varphi_e(x) \tag{7}
\]

where \( \varphi_e(x) = 1 - 2\lambda_S(x) \) is the local basis function of space \( Y_h \) on edge \( e \) with \( \lambda_S(x) \) being barycentric coordinate of \( x \) with respect to vertex \( S \) opposite to \( e \) in triangle \( K \). (Note that \( e = e_1 \) in Fig. 1.) It is easy to see \( \nabla \varphi_e(x) = \frac{|e|}{|K|} \mathbf{n}_e = const \), where \( |e| \) is the length of edge \( e \).

Given any triangular element \( K \in T_h \), we always orient \( K \) counterclockwise as shown in Fig. 1 (e.g. \( K = K_L \) there). Then the three local basis functions associated
with the three edges are as follows. For example, for the edge $e = e_1$ of $K = K_L$ in Fig. 1, we define $P_{K,e}$

$$P_{K,e}(x) = \frac{1}{2|K|} \left[ \begin{array}{c} x - x_S \\ y - y_S \end{array} \right] \quad \forall \ (x, y) \in K. \quad (8)$$

Note that

$$P_{K,e}(x) \cdot n = \begin{cases} 1/|e| & \forall \ x \in e = S'S'' \\ 0 & \forall \ x \in SS' \\ 0 & \forall \ x \in SS''. \end{cases} \quad (9)$$

The other two basis functions $P_{K,e_i}, i = 2, 3$ are defined similarly.

For any $u_h(x) \in V_h$, we have the local representation on $K$

$$u_h(x)|_K = \sum_{e \in \partial K} u_e P_{K,e}(x) \quad (10)$$

where $u_e = \int_e u \cdot n ds$ is the flux across edge $e$. For any edge $a \in A_i$ we define the global canonical basis of $V_h$ associated with edge $a$ as follows. If the edge $a$ corresponds to edge $S'S''$ in the local ordering (cf. Fig. 1), then

$$P_a(x) = P_{K_L,e}(x)\chi_{K_L}(x) - P_{K_R,e}(x)\chi_{K_R}(x) \quad (11)$$

where $a$ is oriented from $K_L$ towards $K_R$. The global basis functions based on boundary edges are defined similarly.

Finally, by Taylor’s expansion at the barycenter $B$ of $K$ we have on $K$

$$u_h(x) = u_K + (\nabla \cdot u_h)_K P_K(x) \quad (12)$$

where $u_K = \frac{1}{|K|} \int_K u_h dx = -A_K \nabla p_h$, the average of $u_h(x)$ on $K$, and

$$P_K(x) = \frac{1}{2} \left[ \begin{array}{c} x - x_B \\ y - y_B \end{array} \right] = \frac{|K|}{3} \sum_{e \in \partial K} P_{K,e}(x) \quad \forall \ (x, y) \in K \quad (13)$$

with $(x_B, y_B)$ being the coordinates of $B$. Alternatively, one can also write

$$u_h(x) = -A_K \nabla p_h + |K| f_K P_K(x). \quad (14)$$

Note that $|K| f_K P_K(x)$ has zero mean on triangle $K$.

Let $u_h \in V_h$ and $p_h \in Y_h$ have the local representations (10) and (7), respectively. Then for each $K \in T_h$, Eq. (5) implies

$$\sum_{e \in \partial K} u_e - |K| f_K = 0 \quad (NT \text{ equations}) \quad (15)$$
and Eq. (6) becomes

\[ \sum_{e \in \partial K} u_e \int_K P_{K,e}(x)dx + p_e |K| A_K \nabla \varphi_e(x) = 0. \]

Define \( Q_e = \int_K P_{K,e}(x)dx \), \( N_e = |e| A_K n_e \), and recall that \( \nabla \varphi_e(x) = \left| e \right| n_e = \text{const} \) to get

\[ \sum_{e \in \partial K} u_e Q_e + p_e N_e = 0 \quad (2NT \text{ equations}). \quad (16) \]

Referring to Fig. 1, we see that by the one-point quadrature using the barycenter \( B \)

\[ Q_{e_1} = P_{K,e_1}(B)|K| = 1/6(\overrightarrow{SS'} + \overrightarrow{SS''}) := 1/6(e_3 - e_2) \]

where \( SS' = e_3 \) and \( S'S'' = e_2 \). Hence

\[ \sum_{e \in \partial K} Q_e = 0. \quad (17) \]

Also note

\[ \sum_{e \in \partial K} N_e = A_K (\sum_i |e_i| n_{e_i}) = 0. \quad (18) \]

On the boundary

\[ p_a = 0 \quad (N A_b \text{ equations}). \quad (19) \]

Clearly

\[ 3NT + NA_b = 2NA. \quad (20) \]

and we have as many equations as unknowns: the number of unknowns \((u_a, p_a)_{a \in A}\) being \(2NA\) and the total number of equations in (15), (16) and (19) being \(3NT + N A_b = 2NA\).

Combining (15), (16) and (19), we see that system (4) becomes: Find \( u_h = \sum_{a \in A} u_a P_a(x) \), \( p_h(x) = \sum_{a \in A} p_a \varphi_a(x) \) such that

\[ \begin{cases} 
\sum_{e \in \partial K} u_e = |K| f_K & \forall K \in T_h, \\
\sum_{e \in \partial K} (u_e Q_e + p_e N_e) = 0 & \forall K \in T_h, \\
p_a = 0 & \forall a \in A_b. 
\end{cases} \quad (21) \]

A remark about notation is in order here. We emphasize that the notation \( u_a, a \in A \) is reserved for the component with respect to the global basis whereas the notation \( u_{K,e}, e \in \partial K \) is the component with respect to the local basis. When there is no danger of confusion we simply use \( u_e \) instead of \( u_{K,e} \). Later in sec. 5, we shall show the existence and uniqueness of the above system.
3 Element and global stiffness matrices

In this section we shall work out the details of implementation of the resulting discrete systems.

Denote by \( U = (u_a)_{a \in A} \) the global vector of \( u_h(x) \) onto basis \( \{P_a(x)\} \) and \( P = (p_a)_{a \in A} \) the global vector of \( p_h(x) \) onto basis \( \{\varphi_a(x)\} \). Also define \( U_K \) and \( P_K \) to be the local vectors of \( u_h(x) \) and \( p_h(x) \) onto local basis functions in the triangle \( K \)

\[
\begin{align*}
U_K &= \begin{bmatrix} u_{e_1}, u_{e_2}, u_{e_3} \end{bmatrix}^T, \\
P_K &= \begin{bmatrix} p_{e_1}, p_{e_2}, p_{e_3} \end{bmatrix}^T
\end{align*}
\]

where \( e_1, e_2 \) and \( e_3 \) are three edges of \( K \).

Then system (21) can be written as:

\[
\bar{L}_K U_K + \bar{M}_K P_K = \bar{F}_K \quad \forall K \in T_h
\]

where \( \bar{L}_K, \bar{M}_K \in \mathbb{R}^{3 \times 3}, \bar{F}_K \in \mathbb{R}^{3} \) with

\[
\begin{align*}
\bar{L}_K &= \begin{bmatrix} 1 & 1 & 1 \\
Q_{e_1} & Q_{e_2} & Q_{e_3} \\
\end{bmatrix} = \begin{bmatrix} \frac{1}{6}(e_3 - e_2) & \frac{1}{6}(e_1 - e_3) & \frac{1}{6}(e_2 - e_1) \\
\end{bmatrix}, \\
\bar{M}_K &= \begin{bmatrix} 0 & 0 & 0 \\
N_{e_1} & N_{e_2} & N_{e_3} \\
\end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\
|e_1| A_K n_{e_1} & |e_2| A_K n_{e_2} & |e_3| A_K n_{e_3} \\
\end{bmatrix}, \\
\bar{F}_K &= \begin{bmatrix} |K| f_K \\
0 \\
0 \\
\end{bmatrix}.
\end{align*}
\]

We know that the matrix \( \bar{L}_K \) is nonsingular by (17) and hence (24) can be rewritten as

\[
U_K = F_K - M_K P_K \quad \forall K \in T_h
\]

where \( M_K = \bar{L}_K^{-1} \bar{M}_K \) and \( F_K = \bar{L}_K^{-1} \bar{F}_K \). It is easy to check

\[
F_K = \frac{|K|}{3} f_K \begin{bmatrix} 1 \\
1 \\
1 \\
\end{bmatrix}.
\]

We can eliminate the unknowns \((u_a)_{a \in A}\) to obtain a system in the pressures alone. If \( a \) is an interior edge with orientation from \( K_L(a) \) towards \( K_R(a) \), \( a = e_L \) in \( K_L(a) \), \( a = e_R \) in \( K_R(a) \), then the continuity of the flux across \( a \) gives the identity \( U_{K_L,e_L} = -U_{K_R,e_R} \) holds. (Here \( U_{K_L,e_L} \) means the component of \( U_{K_L} \) in (22) corresponding to the edge \( e_L \).) Thus we have the (scalar) identity

\[
[M_{K_L} P_{K_L}]_{e_L} + [M_{K_R} P_{K_R}]_{e_R} = F_{K_L,e_L} + F_{K_R,e_R} \quad \forall a \in A_i.
\]
This assembly along with \( p_a = 0 \) for all \( a \in A \) then gives rise to an \( NA \times NA \) linear system in the unknown \( P = (p_a)_{a \in A} \)

\[
\mathcal{M} P = \mathcal{F}.
\]  

(28)

We denote \( P_i \in \mathbb{R}^{N_{A_i}} \), \( F_i \in \mathbb{R}^{N_{A_i}} \), and \( M_i \in \mathbb{R}^{N_{A_i} \times N_{A_i}} \) the sub-vectors or submatrix of \( P, F \) and \( M \) corresponding to interior edge set \( A_i \). The resulting system is \( M_i P_i = F_i \). Next, we will give a simple form of local stencil \( M_K \) which is a symmetric positive matrix. Then we will introduce a way to assemble global stiffness matrix \( M_i \) from local stencil \( M_K \). Finally we show that the matrix \( M_i \) is symmetric positive definite.

**Lemma 3.1** The local stencil \( M_K \) is symmetric positive semi-definite.

**Proof** Let us compute \( \tilde{L}_K^{-T} \) first. Referring to the notation in Fig. 1, we have

\[
\tilde{L}_K^{-T} = \begin{bmatrix}
1 & 1/6(e_3 - e_2)^T \\
1 & 1/6(e_1 - e_3)^T \\
1 & 1/6(e_2 - e_1)^T
\end{bmatrix}.
\]

Using the fact \( e_1 + e_2 + e_3 = 0 \) and elementary geometry, we can easily see that

\[
\tilde{L}_K^{-T} = \begin{bmatrix}
\frac{1}{3} e_1^T \mathbf{n}_{e_1} & \frac{1}{3} e_2^T \mathbf{n}_{e_2} & \frac{1}{3} e_3^T \mathbf{n}_{e_3}
\end{bmatrix}
\]

where \( \mathbf{n}_{e_i} \) are the unit outward normal to \( e_i \). Thus \( M_K = \tilde{L}_K^{-1} \tilde{M}_K \) has entries

\[
(M_K)_{ij} = \frac{|e_i||e_j|}{|K|} \mathbf{A}_K \mathbf{n}_{e_j}^T \mathbf{n}_{e_i} \mathbf{A}_K \mathbf{n}_{e_i} = (M_K)_{ji}
\]

and hence \( M_K \) is symmetric.

Now let \( R(\theta) \) be the rotation matrix through an angle of \( \theta \), then \( n_{e_i} = R(-\frac{\pi}{2}) e_i / |e_i| \) and so

\[
(M_K)_{ij} = \frac{1}{|K|} e_i^T \tilde{A}_K e_j
\]

(29)

where

\[
\tilde{A}_K = \begin{bmatrix}
c & -b \\
-b & a
\end{bmatrix} \quad if \quad A_K = \begin{bmatrix}
a & b \\
b & c
\end{bmatrix}.
\]

Next we show \( M_K \) is positive semi-definite. Denote \( E = [e_1, e_2, e_3] \), then

\[
M_K = \frac{1}{|K|} E^T \tilde{A}_K E.
\]

(31)
and

\[
\xi^T M_K \xi = \frac{1}{|K|^2} \xi^T E^T \bar{A}_K E \xi = \frac{1}{|K|^2} \xi^T E^T R(-\pi/2)^T A_K R(-\pi/2) E \xi \geq \alpha_1 \frac{1}{|K|} |R(-\pi/2) E \xi|^2 = \alpha_1 \frac{1}{|K|} |E \xi|^2.
\]

Therefore \( M_K \) is positive semi-definite.

**Remark:** Note if we take \( A_K = I \), from (29) a simple computation shows that

\[
M_K = 2 \begin{bmatrix} d_2 + d_3 & -d_3 & -d_2 \\
-d_3 & d_1 + d_3 & -d_1 \\
-d_2 & -d_1 & d_1 + d_2 \end{bmatrix}
\]

where \( d_i = \cot \theta_i \), where \( \theta_i \) is the angle opposite to the edge \( e_i \). It is 4 times the local stencil in the standard FVM for Poisson problem. (see [3]) and same as that of the mixed covolume scheme. (see [14])

Now let \( I := \{1, 2, \ldots, N \mathcal{A}_i\} \) be a global ordering of the interior edges in \( \mathcal{A}_i \), and for \( l = 1, \ldots, NT \), let \( \{1, 2, 3\} \) be a local ordering of the edges of triangular element \( K^{(l)} \). We use \( g^{(l)}_j \) to denote the global edge number of the edge in element \( K^{(l)} \) that has local edge number \( j \). Also we use the notation \( M^{(l)} \) for the local stencil \( M_K \) when \( K = K^{(l)} \). Suppose that \( a \in \mathcal{A}_i \) is the intersection of two elements \( K^{(l)} \) and \( K^{(m)} \) with orientation from \( K^{(l)} \) to \( K^{(m)} \). Let \( a = e^{(m)}_s \) on \( K^{(m)} \) and \( a = e^{(l)}_t \) where \( s \) and \( t \) are the local edge numbers. Now let \( K_L = K^{(l)} \) and \( K_R = K^{(m)} \) in (27), then its left hand side can be expressed as

\[
LHS = \sum_{j=1}^{3} m^{(l)}_{ij} p_{g^{(l)}_j} + \sum_{j=1}^{3} m^{(m)}_{sj} p_{g^{(m)}_j}
\]

in terms of pressures \( p_r, r \in I \) (globally indexed) and the entries \( m^{(l)}_{ij} \) of the three by three matrix \( M^{(l)} \) and so on. The above suggests we define a global matrix for each element as follows. For each \( l, 1 \leq l \leq NT \) define the matrix \( \hat{M}^{(l)} = \hat{M}_K \in \mathcal{R}^{N\mathcal{A}_i \times N\mathcal{A}_i} \) associated with \( K = K^{(l)} \) so that \( \hat{m}^{(l)}_{ij} = 0 \) if the (global) edges \( i \) and \( j \) are not in \( K^{(l)} \). Otherwise we set \( \hat{m}^{(l)}_{ij} = m^{(l)}_{st} \) where \( i = g^{(l)}_s \) and \( j = g^{(l)}_t \). Obviously, \( \hat{M}_K \) is a symmetric matrix since \( M_K \) is.

We can now write (32) in terms of this new matrix and global indices. It is then not hard to conclude that

\[
\mathcal{M}_i = \sum_{K \in T_h} \hat{M}_K,
\]

\[
\mathcal{F}_i = \sum_{K \in T_h} \hat{F}_K.
\]
In fact, let $\alpha \in I$ and the edge $e_\alpha = a = K^{(l)} \cap K^{(m)}$. then

\[
\left( \sum_{K \in T_h} \hat{M}_K P_i \right)_\alpha = \sum_{\beta = 1}^{N A_i} \sum_{k = 1}^{N T} \hat{m}_{\alpha\beta}^{(k)}(p_i)_\beta
\]

\[
= \sum_{\beta = 1}^{N A_i} \left\{ \hat{m}_{\alpha\beta}^{(l)}(p_i)_\beta + \hat{m}_{\alpha\beta}^{(m)}(p_i)_\beta \right\}
\]

\[
= \sum_{j = 1}^{3} m_{lj}^{(l)}(p_i)_{g_j} + m_{sj}^{(m)}(p_i)_{g_j}
\]

\[
= (M_i P_i)_\alpha. \quad (35)
\]

Hence $M_i$ is symmetric. Moreover

\[
P_i^T M_i P_i = \sum_{K \in T_h} P_i^T \hat{M}_K P_i = \sum_{K \in T_h} P_K^T M_K P_K. \quad (36)
\]

So from Lemma (3.1), we have $P_i^T M_i P_i \geq 0$. Later in Theorem 5.2 we will show the uniqueness of solution for the system (21). Therefore we obtain the following theorem:

**Theorem 3.1** The global stiffness matrix $M_i$ corresponding to the interior edge set $A_i$ is symmetric positive definite.

**Algorithm:**

\begin{enumerate}
\item do for each $K \in T_h$
\item evaluate $\hat{A}_K$, $f_K$
\item evaluate $M_K$ by (29), $F_K$ by (26)
\item assemble $M_K$ to $M_i$, $F_K$ to $F_i$
\end{enumerate}

enddo

\item solve $M_i P_i = F_i$
\item evaluate $U$ from (25), with boundary condition.

**4 a Mixed box method is a FEM plus local flux recovery**

In this section we first prove in Thm. 4.1 equivalence between the pressure approximation in our box method and a $P1$ nonconforming finite element method applied to the elliptic problem (1) with a modified right hand side. Then we show this leads to a better understanding of the box method. It turns out that the mixed box method is nothing but a standard FEM with accurate local recovery of the flux. This is the content of Thm. 4.2 with zero absorption.
Throughout the rest of this paper, we use the standard notation $W^{1,p}$ for the usual Sobolev spaces and $| \cdot |_{m,K}$, $\| \cdot \|_{m,K}$ for the semi and full $H^m$–norm, $m = 0, 1, 2$. We omit the subscript $K$ when $K = \Omega$ and sometimes use $\| \cdot \|$ when writing an $L^2$ norm. Also $| \cdot |_{H(\text{div};\Omega)}$ is the $H(\text{div};\Omega)$ semi-norm.

We now show the equivalence theorem.

**Theorem 4.1** The linear system (28) is the same as the discrete system resulting from the standard $P_1$ nonconforming FEM: Find $p_h \in Y_{h,0}$ such that

$$a_h(p_h, q_h) = (f_h, q_h) \quad \forall q_h \in Y_{h,0}$$

(37)

where $f_h = P_h f$ is the $L^2$ projection to the piecewise constant space $L_h$.

**Proof 2** It suffices to show the two methods have the same element stiffness matrix and the same right hand side. The element stiffness matrix associated with element $K$ from (41) can be obtained as follows. Let $\varphi_i = 1 - 2\lambda_i$, $i = 1, 2, 3$ with $\lambda_i$ being the barycentric coordinates. Noting that $\nabla \varphi_i = \frac{|e_i|}{|K|} n_i$, we have

$$\int_K (K\nabla \varphi_i) \cdot \nabla \varphi_j dx = \int_K (K|e_i|/|K|) \cdot \frac{|e_j|}{|K|} n_j dx$$

$$= \frac{|e_i|}{|K|} n_i^T (\int_K K dx) \frac{|e_j|}{|K|} n_j$$

$$= \frac{1}{|K|} e_i^T A_K |e_j| n_j$$

$$= \frac{1}{|K|} e_i^T \tilde{A}_K e_j$$

which is exactly (29).

Since $f_h = P_h f$, the $L^2$ projection, $f_h|_K = \frac{1}{|K|} \int_K f$. Thus

$$(f_h, \varphi_j) = \int_K f_h \varphi_j dx$$

$$= \frac{f_h|_K |K|}{3}$$

$$= \frac{|K| f_K}{3}$$

which is exactly (26). This completes the proof.

**Remark.** We must emphasize this theorem is proved because of the hard work in the previous sections. Our line of thinking follows that of [15]. In other words, we try to generalize their method to the anisotropic case, but our crucial new observation is the above theorem. It says the mixed finite volume method is related to a finite element
method. The next theorem says the hard work in the previous sections can be avoided altogether if we have a change of viewpoint.

To bring out an important difference between the mixed box method (finite volume) method and the finite element method, we will do more and actually show the equivalence theorem for the original problem (1) with an added absorption term i.e.,

\[
\begin{aligned}
\begin{cases}
-\nabla \cdot K\nabla p + \alpha_0 p &= f & \text{in } \Omega, \\
p &= 0 & \text{on } \partial \Omega,
\end{cases}
\end{aligned}
\]

(38)

where \(\alpha_0\) is a nonnegative piecewise constant function with respect to \(T_h\). The associated weak formulation is to find \(p \in H^1_0\) such that

\[
a(p, q) = (f, q) \quad \forall q \in H^1_0
\]

(39)

where

\[
a(p, q) := \int_\Omega (K\nabla p) \cdot \nabla q + \alpha_0 pq \, dx.
\]

(40)

Let

\[
Y_{h,0} := \{ q \in L^2(\Omega) : q|_K \in P_1(K), \forall K \in T_h; q \text{ is continuous at the midpoints of interior edges and vanishes at the midpoints of boundary edges} \}.
\]

The standard \(P1\) nonconforming FEM discretization is: Find \(\tilde{p}_h \in Y_{h,0}\) such that

\[
a_h(\tilde{p}_h, q_h) = (f, q_h) \quad \forall q_h \in Y_{h,0}
\]

(41)

where

\[
a_h(\tilde{p}_h, q_h) := \sum_{j=1}^{NT} (K\nabla \tilde{p}_h, \nabla q_h)_K + (\alpha_0 p_h, q_h)_K,
\]

(42)

where \((\cdot, \cdot)_K\) is the \(L^2\) inner product on \(K\). Define the semi-norm

\[
|q|_h := \left( \sum_{K \in T_h} |q|_{1,K}^2 \right)^{1/2} \quad \forall q \in H^1_0 \oplus Y_{h,0}.
\]

(43)

It is clear \(| \cdot |_h\) is a full norm on space \(Y_{h,0}\) in Dirichlet Case. It is well known \([7, 4, 5]\) that the solution \(p_h\) of system (41) converges to solution \(p\) of system (39): there exists a constant \(C\) independent of \(h\) such that

\[
\|p - \tilde{p}_h\|_0 + h \|p - \tilde{p}_h\|_h \leq Ch^2 \|p\|_2
\]

(44)

provided that the problem data is smooth enough so that the elliptic regularity estimate \(\|p\|_2 \leq C\|f\|_0\) holds. For example, if \(f \in L^2\) and \(K \in C^1(\Omega)\) on a convex domain \(\Omega\), then \(p \in H^2\) is guaranteed. (See p. 4 of \([4]\) and the references therein.) We now show the equivalence theorem.
Theorem 4.2 Consider the problem of finding \((u_h, p_h) \in V_h \times Y_{h,0}\) such that
\[
(K\nabla p_h, \nabla q_h) + \sum_K \alpha_K (p_K, q_h)_K = (f_h, q_h) \quad \forall q_h \in Y_{h,0}
\] (45)
and over each \(K\) in \(T_h\)
\[
u_h = -A_K \nabla p_h + (f_K - \alpha_K p_K) P_K
\] (46)
where \(f_h = P_h f\) is the \(L_2\) projection to the piecewise constant space \(L_h\) and \(p_K = \frac{1}{|K|} \int_K p_h dx, f_K = \frac{1}{|K|} \int_K f dx = f_h|_K\) along with the mixed box method of finding \((u_h, p_h) \in V_h \times Y_{h,0}\) such that
\[
(\nabla \cdot \nu_h + \alpha_0 p_h - f, \chi_K) = 0,
\] (47)
\[
(\nu_h + K \nabla p_h, \chi_K) = 0.
\] (48)

Then the two above problems are equivalent.

Proof 3 We first show that (45)–(46) implies (47)–(48).

Take divergence on (46), recall (13), and integrate against the characteristic function \(\chi_K\) to see (46) implies (47). Now integrate (46) against \(\chi_K\) and use the fact \((P_K, \chi_K) = 0\) (one point quadrature rule using the barycenter \(B\)) to get (48).

Secondly, we prove that (47)–(48) implies (45)–(46).

From (47) we see that on \(K\)
\[
\nabla \cdot \nu_h = f_K - \alpha_K p_K.
\]
Since \(\nu_h \in V_h\), by Taylor’s expansion
\[
u_h = \nu_K + \frac{1}{2} \nabla \cdot \nu_h (x - x_B),
\]
where \(\nu_K = \frac{1}{|K|} \int_K \nu_h dx = -A_K \nabla p_h\) by (48). So
\[
u_h = -A_K \nabla p_h + (f_K - \alpha_K p_K) P_K
\]
which is (46). On the other hand by (48), integration by parts and (47)
\[
\sum_K (K \nabla p_h, \nabla q_h)_K = \sum_K (-\nu_h, \nabla q_h)_K
\]
\[
= \sum_K (\nabla \cdot \nu_h, q_h)_K - (\nu_h \cdot n, q_h)_{\partial K}
\]
\[
= \sum_K (f_K - \alpha_K p_K, q_h)
\]
which is (45). Notice that the \((\cdot, \cdot)_{\partial K}\) terms cancelled upon summation since \(\nu_h \cdot n\) and \(q_h\) are continuous at midpoints of edges.
Remark. One notes that the mixed finite volume method is equivalent to the standard nonconforming method (with a modified right hand side) when there is no absorption term ($\alpha_0 = 0$). In the presence of the absorption term, the mixed method is equivalent to a nonconforming method with the absorption term evaluated by a one-point quadrature at the element barycenter. The approximate $p_h$ so produced by (45) generates automatically a flux with continuous normal components across edges via (46). In other words, the usual nonconforming method does not produce continuous flux unless $\alpha_0 = 0$.

5 Error estimates

In this section we will prove the existence and uniqueness of solution for the system (21) and some error estimates in an energy norm. Through this section the letter $C$ denotes a generic positive constant, independent of $h$ and not necessarily the same in each occurrence. Let us define two energy norms which by (2) are equivalent:

$$|q|^2_h = \sum_K |q|^2_{h,K} := \sum_K \int_K |\nabla q|^2 \, dx \quad \forall q \in H^1(\Omega) \oplus Y_h.$$  

$$|q|^2_E = \sum_K |q|^2_{E,K} := \sum_K \int_K \nabla q^T K(x) \nabla q \, dx \quad \forall q \in H^1(\Omega) \oplus Y_h.$$  

The next lemma shows they are actually full norms on the space $Y_{h,0} = \{ q_h \in Y_h, q_h = 0 \text{ at all midpoints of boundary edges} \}$.  

Lemma 5.1 The discrete energy semi-norm $|q_h|$ is a norm on the space $Y_{h,0}$.  

Proof 4 Let $q_h \in Y_{h,0}$ such that $|q_h| = 0$. The gradient of $q_h$ is zero in each cell $K \in T_h$. Hence $q_h$ is constant in each cell $K$. Since $q_h$ is continuous at the middle point of each edge $e$ of $T_h$ and $q_h = 0$ on $\partial \Omega$, we have $q_h = 0$ in $\Omega$.

The next theorem shows the existence and uniqueness of solution for system (21).

Theorem 5.1 The discrete system (21) has a unique solution $(u_h, p_h) \in V_h \times Y_{h,0}$.

Proof 5 Since the number of unknowns equals the number of equations in (21) we only need to show that $u_h = 0$ and $p_h = 0$ when $f = 0$.

By Thm. 4.1 or 4.2 and the preceding lemma $p_h = 0$ when $f = 0$. On the other hand, $u_h = u_K = -A_K \nabla p_h = 0$.

Also we have the following stability condition:
Lemma 5.2 If \((u_h, p_h) \in V_h \times Y_{h,0}\) is the solution of system (21), then there exist positive constants \(C_1\) and \(C_2\), independent of \(h\), such that

\[
C_1 |p_h|_h \leq \|u_h\|_{0, \Omega} \leq C_2 (|p_h|_h + h\|f\|_{0, \Omega}) \tag{49}
\]

\[
|u_h|_h := \left( \sum_K \|\nabla u_h\|_{0,K}^2 \right)^{1/2} \leq \frac{1}{\sqrt{2}} \|f\|_{0, \Omega} \tag{50}
\]

Proof 6 We first show (50). From (14) \(\nabla u_h = |K| f_K \nabla P_K(x)\) and from (13) \(\|\nabla P_K\|_{0,K}^2 = \frac{1}{2|K|}\), so

\[
|u_h|_h^2 = \sum_K \|\nabla u_h\|_{0,K}^2 = \sum_K |K|^2 |f_K|^2 \|\nabla P_K\|_{0,K}^2
\]

\[
\leq 1/2 \sum_K |K| |f_K|^2 \leq 1/2 \|f\|_{0, \Omega}^2
\]

and we obtain (50).

Next we show (49). By the linearity of \(p_h\), bounds (2) on \(A_K\), and the Cauchy-Schwarz inequality, we get

\[
|p_h|_h^2 = \sum_K \int_K |\nabla p_h|^2 dx = \sum_K |K| \|\nabla p_h\|^2
\]

\[
= \sum_K |K| |(A_K^{-1} u_K)^T (A_K^{-1} u_K)| \leq C \sum_K |K| |u_K^T u_K|
\]

\[
\leq C \sum_K \int_K |u_h(x)|^2 dx = C \|u_h\|_{0, \Omega}^2,
\]

whence

\[
C_1 |p_h|_h \leq \|u_h\|_{0, \Omega}.
\]

Again by (14), bound (2) on \(A_K\),

\[
\|u_h\|_{0,K} \leq \|A_K \nabla p_h\|_{0,K} + |K| |f_K| \|P_K\|_{0,K}
\]

\[
= \left( \int_K \nabla p_h^T A_K^2 \nabla p_h dx \right)^{1/2} + |K| |f_K| \|P_K\|_{0,K}
\]

\[
\leq C \|\nabla p_h\|_{0,K} + |K| |f_K| \|P_K\|_{0,K}
\]

\[
= C |p_h|_h + |K| |f_K| \|P_K\|_{0,K}
\]

\(\)From (13) we have

\[
\|P_K\|_{0,K}^2 = \frac{1}{4|K|^2} \int_K (x - x_B)^2 + (y - y_B)^2 dxdy
\]

\[
\leq \frac{h_K^2}{4|K|}.\]
By the regularity assumption on the triangulations we know $h_K / |K|^{1/2} \leq C$ for all $K$. Also notice $|f_K| \leq \|f\|_{0,K} / |K|^{1/2}$, we obtain

$$\|u_h\|_{0,K} \leq C(|p_h|_{h,K} + |K|^{1/2}\|f\|_{0,K}) \leq C(|p_h|_{h,K} + h\|f\|_{0,K}).$$

Summing over $K$, we get

$$\|u_h\|_{0,\Omega} \leq C(|p_h|_h + h\|f\|_{0,\Omega}).$$

This completes the proof.

Our main result in this section is the following error estimate theorem.

**Theorem 5.2** Let the problem data of (1) be smooth enough so that the pressure solution $p \in H^2 \cap H^1_0$ and $u(x) = -K(x) \nabla p(x) \in H^1(\Omega)^2$. Then there exists a constant $C$ independent of $h$ such that

$$\|p - p_h\|_0 \leq Ch^2(|f|_h + \|f\|_0), \text{ if } f_K \in H^1(K) \quad \forall K \in T_h, \tag{51}$$

$$|p - p_h|_h \leq Ch\|f\|_0 \tag{52}$$

$$\|u - u_h\|_0 \leq Ch(\|f\|_0 + \|u\|_0) \tag{53}$$

$$|u - u_h|_{H(\text{div};\Omega)} \leq Ch|f|_h \quad \text{if } f_K \in H^1(K) \quad \forall K \in T_h. \tag{54}$$

Also we assume that $K \in W^{1,\infty}$.

**Remark.** Note that the condition of the function $f \in H^1(K)$ for all $K$ can be satisfied if we impose it on the coarsest mesh in the context of local refinement.

**Proof 7** i) proof of (51) and (52). First we show that

$$\|\bar{p} - p\|_0 \leq Ch^2|f|_h$$

where $\bar{p} \in H^1_0$ is the solution of $a(\bar{p}, q) = (f_h, q) \quad \forall q \in H^1_0$ and $p$ is the solution of $a(p, q) = (f, q) \quad \forall q \in H^1_0$. Subtracting these two bilinear forms, we get

$$a(p - \bar{p}, q) = (f - f_h, q) \quad \forall q \in H^1_0.$$  

Noticing that $\int_K f - f_h \, dx = 0$, we have $(f - f_h, q)_K = (f - f_h, q - q_h)_K$ where $q_h|_K = 1/|K| \int_K q \, dx$ is constant on each $K$. Then by the Cauchy–Schwarz inequality and an interpolation theorem, we have

$$|(f - f_h, q - q_h)| = \left| \sum_K (f - f_h, q - q_h)_K \right| \tag{55}$$

$$\leq Ch^2 \sum_K |f|_{1,K} |q|_{1,K}$$

$$\leq Ch^2 |f|_h |q|_1$$
Thus
\[ |a(p - \tilde{p}, q) - C h^2 |f_h|q| \leq Ch^2 \sum K \|f\|_0, K \|q\|_1, K. \]

Taking \( q = p - \tilde{p} \) and using the coercivity of \( a(\cdot, \cdot) \), we have
\[ |p - \tilde{p}|_1^2 \leq Ch^2 |f|_h |p - \tilde{p}|_1 \]
and hence
\[ |p - \tilde{p}|_1 \leq Ch^2 |f|_h. \]

By the Poincaré inequality, we get
\[ \|p - \tilde{p}\|_0 \leq Ch^2 |f|_h. \]

Now note that \( p_h \in Y_{h,0} \) is the solution of the nonconforming method
\[ a_h(p_h, q_h) = (f_h, q_h) \quad \forall q_h \in Y_{h,0} \]
associated with the variational problem of finding \( \tilde{p} \in H^1 \) such that
\[ a(\tilde{p}, q) = (f_h, q) \quad \forall q \in H^1. \]
(The right hand side contains \( f_h \) instead of \( f \!) Hence
\[ \|p_h - \tilde{p}\|_0 + h|\tilde{p} - p_h|_h \leq Ch^2 \|\tilde{p}\|_2, \quad (58) \]
where we have applied (44) with the right side \( f_h \) and all the remarks concerning regularity of the solution made there then also apply here.

Now by the triangle inequality, (57), (58), the stability condition and the fact that \( f_h \) is the \( L_2 \) projection of \( f \), we have
\[
\begin{align*}
\|p_h - p\|_0 & \leq \|p_h - \tilde{p}\|_0 + \|\tilde{p} - p\|_0 \\
& \leq Ch^2 \|\tilde{p}\|_2 + Ch^2 |f_h|_h \\
& \leq Ch^2 |f_h|_0 + Ch^2 |f|_h \\
& \leq Ch^2 |f|_0 + Ch^2 |f|_h.
\end{align*}
\]
This completes the proof of (51).

As for the proof of (52), first note that from (55) we can use the fact that \( \|f - f_h\|_{0, K} \leq \|f\|_{0, K} \) and an interpolation theorem to obtain
\[
\begin{align*}
|(f - f_h, q - q_h)| & \leq C h \sum K \|f\|_{0, K} |q|_{1, K} \\
& \leq C h \|f\|_0 |q|_1
\end{align*}
\]
and consequently instead of (56) we have $|\tilde{p} - p_1| \leq Ch\|f\|_0$; arguing as before. Now

$$
|p_h - p|_h \leq |p_h - \tilde{p}|_h + |\tilde{p} - p|_h \\
\leq Ch\|	ilde{p}\|_2 + Ch\|f\|_0.
$$

ii) proof of (53).

From (14) we have

$$
\textbf{u}_h(x) - \textbf{u}(x)|_K = -A_K\nabla p_h + \mathcal{K}(x)\nabla p + |f_K|P_K(x).
$$

So

$$
\|\textbf{u}_h - \textbf{u}\|_{0,K} \leq \|A_K\nabla p_h - \mathcal{K}(x)\nabla p\|_{0,K} + |f_K|\|P_K\|_{0,K},
$$

and since $|K| |f_K| \|P_K\|_{0,K} \leq Ch\|f\|_{0,K}$ as shown in Lemma 5.2 we have

$$
\|\textbf{u}_h - \textbf{u}\|_{0,K} \leq \|A_K\nabla p_h - \mathcal{K}(x)\nabla p\|_{0,K} + Ch\|f\|_{0,K},
$$

whereas by the triangle inequality and the interpolation theorem

$$
\|A_K\nabla p_h - \mathcal{K}\nabla p\|_{0,K} \leq \|A_K\nabla p_h - \mathcal{K}\nabla p\|_{0,K} + \|\mathcal{K}\nabla p_h - \mathcal{K}\nabla p\|_{0,K} - \|\mathcal{K}\nabla p_h - \mathcal{K}\nabla p\|_{0,K}
$$

$$
= (\nabla p_h^T \int_K (A_K - \mathcal{K})^2 d\mathbf{x} \nabla p_h)^{1/2} + \left( \int_K |\mathcal{K}(\nabla p_h - \nabla p)|^2 d\mathbf{x} \right)^{1/2}
$$

$$
\leq Ch|p_h|_{h,K}|\mathcal{K}|_{1,\infty,K} + C|\nabla p_h - p|_{E,K}
$$

$$
\leq Ch\|\textbf{u}_h\|_{0,K} + C|p_h - p|_{E,K}
$$

$$
\leq C\{h\|\textbf{u} - \textbf{u}_h\|_{0,K} + h\|\textbf{u}\|_{0,K} + C|p_h - p|_{E,K}\}
$$

where we have used the stability (49) restricted on $K$ as shown in its proof. Now taking $h$ small enough to move the first term on the right side we have

$$
\|\textbf{u}_h - \textbf{u}\|_{0,K} \leq C\{|p_h - p|_{E,K} + h\|\textbf{u}\|_{0,K} + h\|f\|_{0,K}\}.
$$

Summing over $K$ and using (52) we have

$$
\|\textbf{u}_h - \textbf{u}\|_0 \leq Ch(\|f\|_0 + \|\textbf{u}\|_0).
$$

iii) proof of (54).

By (4), $\nabla \cdot \textbf{u}_h(x) = f_K$ and hence

$$
\|\nabla \cdot \textbf{u}_h - \nabla \cdot \textbf{u}\|_{0,K} = \|f - f_K\|_{0,K} \leq Ch|f|_{1,K}
$$

(59)

where an interpolation theorem is used. Summing over $K$, we get

$$
|\textbf{u} - \textbf{u}_h|_{H(\text{div};\Omega)} \leq \|\nabla \cdot \textbf{u}_h - \nabla \cdot \textbf{u}\|_{0,\Omega} \leq Ch|f|_h.
$$
6 Numerical Examples

Notice that the error estimate theorem and the algorithm are valid for both Dirichlet and Neumann problems. We present numerical results for both cases. We partition the unit square $[0, 1] \times [0, 1]$ into squares evenly in both directions with the diagonals running from the upper-left corner of each triangle to its lower-right corner. We use the incomplete LU preconditioned conjugate gradient method to solve all the problems. The integral of $f$ over element $K$ is computed by the midpoint rule using the three edges of the triangle. Our experiments suggest second order approximation in all cases. The discrete norms in which the errors are estimated are as follows.

6.1 Choice of discrete norms

In Thm. 5.2 we predicted first order convergence in the $H(\text{div}; \Omega)$ norm for flux $u$ and second order convergence in the $L_2$ norm for the pressure $p$. We need to choose proper discrete norms to measure the error between true solution and computed solution.

Let $(x_i, y_j)$ be the center of square $(i, j)$ with $x_i = (i - 1/2)h$, $y_j = (j - 1/2)h$, $h = 1/n, i, j = 1, 2, \ldots , n$. Let $p_{ij}$ be the computed pressure at $(x_i, y_j)$. We define

$$p_{\text{Err}}_{\text{abs}} = \|p - p_h\| := \left[ \sum_{i,j=1}^{n} h^2(p(x_i, y_j) - p_{ij})^2 \right]^{1/2},$$

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

Due to (59), the $H(\text{div}; \Omega)$–seminorm of the error in the flux is directly related to $f - f_K$, and so for the flux we will use only an equivalent $L_2$ discrete norm:

$$u_{\text{Err}}_{\text{abs}} = \|u - u_h\| := \left[ \sum_{K}(\int_{\partial K}[(u - u_h) \cdot n] ds)^2 \right]^{1/2},$$

where the edge integrals are evaluated by the midpoint rule [11]. (Note that the $x$ components are picked out by the vertical edges, the $y$ components by the horizontal edges, etc..)

We also compute the relative error

$$p_{\text{Err}}_{\text{rel}} := p_{\text{Err}}_{\text{abs}}/\|p\|$$

and

$$u_{\text{Err}}_{\text{rel}} := u_{\text{Err}}_{\text{abs}}/\|u\|.$$
6.2 Dirichlet problems

We consider the following Dirichlet problem

\[
\begin{align*}
- \nabla \cdot K \nabla p &= f \quad \text{in } \Omega, \\
p &= 0 \quad \text{on } \partial \Omega.
\end{align*}
\]

(60)

In examples 1–4, the true pressure is \( p = (x^2 - x)(y^2 - y) \) on the unit square.

**Example 1.** The mobility tensor \( K = \text{diag}(1 + 10x^2 + y^2, 1 + x^2 + 10y^2) \).

<table>
<thead>
<tr>
<th>Example 1.</th>
<th>( h = 1/16 )</th>
<th>( h = 1/32 )</th>
<th>( h = 1/64 )</th>
<th>( h = 1/128 )</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>pErr_abs</td>
<td>8.7748e-5</td>
<td>2.2318e-5</td>
<td>5.6041e-6</td>
<td>1.4026e-6</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>pErr_rel</td>
<td>0.0026</td>
<td>6.6952e-4</td>
<td>1.6812e-4</td>
<td>4.2077e-5</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>uErr_abs</td>
<td>0.0113</td>
<td>0.0029</td>
<td>7.1931e-4</td>
<td>1.8029e-4</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>uErr_rel</td>
<td>0.0086</td>
<td>0.0021</td>
<td>5.1929e-4</td>
<td>1.2905e-4</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>Length of ( P )</td>
<td>736</td>
<td>3,008</td>
<td>12,160</td>
<td>48,896</td>
<td></td>
</tr>
<tr>
<td>Length of ( U )</td>
<td>736</td>
<td>3,008</td>
<td>12,160</td>
<td>48,896</td>
<td></td>
</tr>
</tbody>
</table>

**Example 2.** The mobility tensor \( K = \text{diag}(10^4, 1) \).

<table>
<thead>
<tr>
<th>Example 2.</th>
<th>( h = 1/16 )</th>
<th>( h = 1/32 )</th>
<th>( h = 1/64 )</th>
<th>( h = 1/128 )</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>pErr_abs</td>
<td>9.1815e-5</td>
<td>2.3286e-5</td>
<td>5.8558e-6</td>
<td>1.4657e-6</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>pErr_rel</td>
<td>0.0026</td>
<td>6.9859e-4</td>
<td>1.7567e-4</td>
<td>4.3971e-5</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>uErr_abs</td>
<td>8.0759</td>
<td>2.0653</td>
<td>0.5218</td>
<td>0.1311</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>uErr_rel</td>
<td>0.0057</td>
<td>0.0014</td>
<td>3.5417e-4</td>
<td>8.8436e-5</td>
<td>( \approx 2 )</td>
</tr>
<tr>
<td>Length of ( P )</td>
<td>736</td>
<td>3,008</td>
<td>12,160</td>
<td>48,896</td>
<td></td>
</tr>
<tr>
<td>Length of ( U )</td>
<td>736</td>
<td>3,008</td>
<td>12,160</td>
<td>48,896</td>
<td></td>
</tr>
</tbody>
</table>

**Example 3.** In this example, we study the effect of discontinuous mobility matrix \( K \). Let

\[
K = \begin{bmatrix} 10000 & 0 \\ 0 & 1 \end{bmatrix}
\]
on the left half unit square, and

\[
K = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}
\]
on the right half unit square. We find the approximation is as good as continuous cases.

**Example 4.** Let

\[
K = \begin{bmatrix} 1 + 10x^2 + y^2 & 1/2 + x^2 + y^2 \\ 1/2 + x^2 + y^2 & 1 + x^2 + 10y^2 \end{bmatrix}.
\]
Table 3: Error behavior for Dirichlet problem

<table>
<thead>
<tr>
<th>Example 3.</th>
<th>$h = 1/16$</th>
<th>$h = 1/32$</th>
<th>$h = 1/64$</th>
<th>$h = 1/128$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{Err}}_{\text{abs}}$</td>
<td>1.6425e-4</td>
<td>4.1581e-5</td>
<td>1.0439e-5</td>
<td>2.6128e-6</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$p_{\text{Err}}_{\text{rel}}$</td>
<td>0.0049</td>
<td>0.0012</td>
<td>3.1317e-4</td>
<td>7.8382e-5</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$u_{\text{Err}}_{\text{abs}}$</td>
<td>6.2337</td>
<td>1.5882</td>
<td>0.4005</td>
<td>0.1005</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$u_{\text{Err}}_{\text{rel}}$</td>
<td>0.0062</td>
<td>0.0015</td>
<td>3.8441e-4</td>
<td>9.5877e-5</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>Length of $P$</td>
<td>736</td>
<td>3,008</td>
<td>12,160</td>
<td>48,896</td>
<td></td>
</tr>
<tr>
<td>Length of $U$</td>
<td>736</td>
<td>3,008</td>
<td>12,160</td>
<td>48,896</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Error behavior for Dirichlet problem

<table>
<thead>
<tr>
<th>Example 4.</th>
<th>$h = 1/16$</th>
<th>$h = 1/32$</th>
<th>$h = 1/64$</th>
<th>$h = 1/128$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{Err}}_{\text{abs}}$</td>
<td>1.4595e-4</td>
<td>3.7458e-5</td>
<td>9.4305e-6</td>
<td>2.3620e-6</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$p_{\text{Err}}_{\text{rel}}$</td>
<td>0.0044</td>
<td>0.0011</td>
<td>2.8292e-4</td>
<td>7.0861e-5</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$u_{\text{Err}}_{\text{abs}}$</td>
<td>0.0168</td>
<td>0.0043</td>
<td>0.0011</td>
<td>2.7533e-4</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$u_{\text{Err}}_{\text{rel}}$</td>
<td>0.0114</td>
<td>0.0028</td>
<td>7.0927e-4</td>
<td>1.7784e-4</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>Length of $P$</td>
<td>736</td>
<td>3,008</td>
<td>12,160</td>
<td>48,896</td>
<td></td>
</tr>
<tr>
<td>Length of $U$</td>
<td>736</td>
<td>3,008</td>
<td>12,160</td>
<td>48,896</td>
<td></td>
</tr>
</tbody>
</table>

6.3 Neumann problems

\[
\begin{align*}
-\nabla \cdot K \nabla p &= f \text{ in } \Omega, \\
K \nabla p \cdot n &= 0 \text{ on } \partial \Omega.
\end{align*}
\]  

(61)

Example 5. The true pressure is $p = \cos(2\pi x) \cos(2\pi y)$ and $K = \text{diag}(\cos(2\pi y) + 2, \cos(2\pi x) + 2)$.

Table 5: Error behavior for Neumann problem

<table>
<thead>
<tr>
<th>Example 5.</th>
<th>$h = 1/16$</th>
<th>$h = 1/32$</th>
<th>$h = 1/64$</th>
<th>$h = 1/128$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{\text{Err}}_{\text{abs}}$</td>
<td>0.0110</td>
<td>0.0028</td>
<td>6.9570e-4</td>
<td>1.7399e-4</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$p_{\text{Err}}_{\text{rel}}$</td>
<td>0.0221</td>
<td>0.0056</td>
<td>0.0014</td>
<td>3.4798e-4</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$u_{\text{Err}}_{\text{abs}}$</td>
<td>0.1383</td>
<td>0.0353</td>
<td>0.0089</td>
<td>0.0022</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>$u_{\text{Err}}_{\text{rel}}$</td>
<td>0.0101</td>
<td>0.0026</td>
<td>6.4814e-4</td>
<td>1.6226e-4</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>Length of $P$</td>
<td>800</td>
<td>3,136</td>
<td>12,416</td>
<td>49,408</td>
<td></td>
</tr>
<tr>
<td>Length of $U$</td>
<td>800</td>
<td>3,136</td>
<td>12,416</td>
<td>49,408</td>
<td></td>
</tr>
</tbody>
</table>

Example 6. The true (oscillatory) pressure is $p = \cos(2\pi x) \cos(10\pi y)$ and $K = I$. 
Table 6: Error behavior for Neumann problem

<table>
<thead>
<tr>
<th>Example 6.</th>
<th>$h = 1/16$</th>
<th>$h = 1/32$</th>
<th>$h = 1/64$</th>
<th>$h = 1/128$</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>pErr_abs</td>
<td>0.0441</td>
<td>0.0130</td>
<td>0.0034</td>
<td>8.4912e-4</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>pErr_rel</td>
<td>0.0882</td>
<td>0.0260</td>
<td>0.0067</td>
<td>0.0017</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>uErr_abs</td>
<td>4.1789</td>
<td>1.1274</td>
<td>0.2882</td>
<td>0.0725</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>uErr_rel</td>
<td>0.1845</td>
<td>0.0498</td>
<td>0.0127</td>
<td>0.0032</td>
<td>$\approx 2$</td>
</tr>
<tr>
<td>Length of $P$</td>
<td>800</td>
<td>3,136</td>
<td>12,416</td>
<td>49,408</td>
<td></td>
</tr>
<tr>
<td>Length of $U$</td>
<td>800</td>
<td>3,136</td>
<td>12,416</td>
<td>49,408</td>
<td></td>
</tr>
</tbody>
</table>

References


Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, Ohio, U.S.A
email: chou@zeus.bgsu.edu
http://www-math.bgsu.edu/~chou/.

Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, Ohio, U.S.A
email: tang@bgnet.bgsu.edu