A MONOTONE NONLINEAR CELL-CENTERED FINITE ELEMENT METHOD FOR ANISOTROPIC DIFFUSION PROBLEMS

NGUYEN ANH DAO, DUC CAM HAI VO, THANH HAI ONG

Abstract. We present a technique to correct the cell-centered finite element scheme [20] (FECC) for full anisotropic diffusion problems on general meshes, which provides a discrete maximum principle (DMP). The correction scheme, named monotone nonlinear cell centered finite element scheme (MNFECC), is cell-centered in the sense that the solution can be computed from cell unknowns of the general primal mesh. Moreover, its coercivity and convergence are proven in a rigorous theoretical framework. Numerical experiments show that the method is effective and accurate, and it satisfies the discrete maximum principle.

1. Introduction

Heterogeneous anisotropic diffusion problems play an important role in areas of science and engineering such as petroleum engineering [22, 7], image processing [32], plasma physics [13, 25, 29]. Their solutions have been studied in [30, 34]. This work concerns the second order elliptic problem on an open, bounded domain Ω in \( \mathbb{R}^2 \) with Lipschitz boundary \( \partial \Omega \),

\[
- \text{div}(\Lambda(x)\nabla u(x)) = f(x) \quad \text{in } \Omega, \\
u = 0 \quad \text{on } \partial \Omega,
\]

(1.1)

where \( \Lambda : \Omega \rightarrow \mathbb{R}^{2 \times 2} \) is a symmetric, positive definite tensor, and there exist \( \lambda, \Lambda > 0 \) satisfying

\[
\lambda |\xi|^2 \leq \Lambda(x)\xi \cdot \xi \leq \Lambda |\xi|^2,
\]

(1.2)

for a.e. \( x \in \Omega \) and all \( \xi \in \mathbb{R}^2 \). The function \( f \) is the source term and belongs to \( L^2(\Omega) \). For simplicity, the homogeneous Dirichlet boundary conditions are imposed. The analysis given below can be extended to other types of boundary conditions as in [28] Chapter 1, Section 1.4. In addition, the following maximum principle [17, Theorem 1] and [12] can be formulated as follows.

Theorem 1.1. The solution \( u \in C^2(\Omega) \cap C_0(\overline{\Omega}) \) of problem (1.1) attains its maxima on the boundary \( \partial \Omega \) if \( f \) is nonpositive in \( \Omega \) and \( f \in C^\alpha(\Omega) \) for some \( \alpha \in (0, 1) \).

2010 Mathematics Subject Classification. 65N08, 65N30, 65N12, 35J15.
Key words and phrases. Discrete maximum principle; heterogeneous anisotropic diffusion; general grid; finite volume; finite elements; cell-centered scheme.

©2019 Texas State University.
According to [17, Corollary 2], we have the following positivity preservation principle.

**Corollary 1.2.** If \( f \) is nonnegative in \( \Omega \) and \( f \in C^\alpha(\Omega) \) for some \( \alpha \in (0, 1) \), then the solution \( u \in C^2(\Omega) \cap C_0(\Omega) \) of problem (1.1) is also nonnegative.

As is well-known (see [31, Theorem 2.2] and [12]), the positivity preservation and discrete maximum principles are equivalent.

The weak form of problem (1.1) is written as follows: Find \( u \in H^1_0(\Omega) \) such that

\[
\int_{\Omega} (\Lambda(x) \nabla u(x)) \cdot \nabla v(x) \, dx = \int_{\Omega} f(x)v(x) \, dx, \quad \forall v \in H^1_0(\Omega).
\]

(1.3)

It is well-known [5, Chapter 1] that under the assumptions made above, problem (1.3) has a unique solution \( u \in H^1_0(\Omega) \).

There are three main difficulties in solving approximate solutions of these problems: firstly, with a heterogeneous and full anisotropic permeability tensor, it is difficult for numerical methods to obtain an approximate solution which converges to the weak solution of the problem; secondly, it is challenging to design numerical methods on general meshes; and it is hardly possible for numerical methods to achieve the DMP. Violation of the DMP may follow numerical instabilities.

The FECC scheme is introduced with rigorous convergence analysis in [20]. It overcomes the first, second difficulties above since it can be applied to heterogeneous, anisotropic diffusion problems on general (possibly distorted) meshes. Also, based on a technique of dual mesh and multipoint flux approximations, the scheme is cell-centered, and satisfies local continuity of fluxes. In [20], numerical results indicate that on the same primal mesh, the FECC scheme gives more accurate solutions than those by the FEM [2], the mimetic finite difference method (MFD) [21], the mixed finite volume method (MFV) [4], the finite volume method (FVM) [1, 18], the compact-stencil MPFA method [23], the discrete duality finite volume method (DDFV) [14], and the SUSHI method [10]. An extension of the FECC scheme, namely the staggered cell-centered finite element method (SC-FEM), to two- and three-dimensional compressible and nearly-incompressible linear elasticity problems have been studied in [27, 15].

However, the FECC scheme violates the DMP, when it is applied for a strong anisotropic diffusion problem (see Test 2 in Section 5). This drawback also appears in finding approximate solutions of classical finite volume [26] and finite element schemes [8] for strong anisotropic diffusion tensors and/or distorted meshes. In the case of piecewise linear finite element approximations for the Poisson problem on a triangulation and a quadrangulation, it is sufficient to require that all triangles be acute [9] (all angles smaller than or equal to \( \pi/2 \)), and all quadrilaterals be of non-narrow type [11] (aspect ratios smaller than or equal to \( \sqrt{2} \)). However, in [16], the authors state that the higher order finite elements fail to satisfy the DMP in the geometric approach.

To preserve the DMP, a different class of finite volume schemes [24, 19, 33] is corrected by a nonlinear discretization. However, these schemes are required by conditions on the geometry or on the anisotropy ratio to get the coercivity. The algebraic flux correction presented in [17] provides a general framework for constructing monotone discretizations on unstructured meshes. This method uses a combination of algebraic and geometric criteria to enforce the DMP.
Our goal, in the present work, is to construct a nonlinear correction for the FECC scheme (MNFECC), providing a DMP for strong anisotropic diffusion problems, while its important properties included in coercivity, convergence is still retained without conditions on the geometry. In addition, a nonlinear system of the MNFECC is solved by an iterative method (3.25) which can compute with cell unknowns of the primal mesh.

The rest of this article consists of four sections: in section 2, we recall the FECC scheme for the discretization of the problem (1,3). In section 3, we present the MNFECC scheme whose solution is solved by the iterative method (3.25). Its linearized system involves only cell unknowns, the associated matrix is symmetric and positive definite. In section 4, we presented within a rigorous theoretical framework to show the existence of a solution, the coercivity, the convergence properties, and to satisfy the DMP for the MNFECC scheme. In the last section, numerical results show that the proposed scheme is effective in terms of accuracy and satisfies the discrete maximum principle.

2. THE FECC FRAMEWORK

To recall the FECC scheme, we first introduce the following notation and constructions of the primal mesh $T_h$, the dual mesh $T^*_h$ and the dual submesh $T^{**}_h$:

2.1. Meshes. For a polygonal domain $\Omega \subset \mathbb{R}^2$, let us consider a primal mesh $T_h$ of $\Omega$ such that $\Omega = \bigcup_{K \in T_h} K$. From now on we make the assumption: each element $K \in T_h$ is a star-shaped polygon. Its mesh point is a point $K \in \text{int}(K)$.

Next, for constructing the dual mesh $T^*_h$, we have the following geometrical assumption: the line joining two mesh points of any two neighboring elements is inside $\Omega$ and it intersects the common edge of the two elements. The dual mesh $T^*_h$ is constructed from the primal mesh in a way that each dual control volume of $T^*_h$ corresponds to a vertex of $T_h$.

Denote by $N$ the set of all nodes or vertices of $T_h$. For every element $K \in T_h$, we define $N_K = \{P : P$ is a vertex of $K\}$. For each $M \in N$, denote by $T_M := \{K \in T_h : K \text{ shares the vertex } M\}$ the set of primal elements that have $M$ as their vertex. We consider two cases (see Figure 1):

(a) If $M$ is an interior vertex, we obtain the dual control volume $M$ associated with the vertex $M$ by connecting the mesh points of neighboring elements in $T_M$. We choose $M$ the dual mesh point of $M$.

(b) If $M$ is on the boundary $\partial\Omega$ and assume that $K_e$ and $K_{\tilde{e}}$ are two (same or different) elements in $T_M$. We denote by $e \subset \partial K_e$ and $\tilde{e} \subset \partial K_{\tilde{e}}$ the two edges on the boundary that have $M$ as their vertex.

The dual control volume $M$ is defined by joining mesh points of neighboring elements in $T_M$ and the mesh point of $K_e$ (and $K_{\tilde{e}}$) with a chosen interior point (e.g. the midpoint) of $e$ (and $\tilde{e}$ respectively). Note that in this case $M$ has $M$ as its vertex as well. We call this point $M$ a dual mesh point of $M$. 
The collection of all \( M \)'s defines a dual mesh \( \mathcal{T}_h^* \) such that \( \Omega = \bigcup_{M \in \mathcal{N}} M \). Let \( \mathcal{N}^* \) be a set of vertices of elements of \( \mathcal{T}_h^* \). For every dual element \( M \in \mathcal{T}_h^* \), a set \( \mathcal{N}^*_M \) contains all vertices of \( M \).

Finally, we construct the dual sub-mesh \( \mathcal{T}_h^{**} \) as a triangular subgrid of the dual mesh \( \mathcal{T}_h^* \) as follows: for an element \( M \in \mathcal{T}_h^* \), we construct elements of \( \mathcal{T}_h^{**} \) by connecting \( M \) to all vertices of \( \mathcal{T}_h^* \) (see Figure 2):

\[ \Omega = \bigcup_{T \in \mathcal{T}_h^{**}} T. \]

Let \( \mathcal{N}^{**} \) be the set of vertices of elements of \( \mathcal{T}_h^{**} \). Let any \( P \in \mathcal{N}^{**} \), we define the set \( \mathcal{T}^*_P = \{ T \in \mathcal{T}_h^{**} : T \text{ has the common vertex } P \} \) whose the number of all elements are denoted by \( \text{card}(\mathcal{T}^*_P) \).

Remark 2.1. By construction, we have:

(a) for all triangular elements \( T \in \mathcal{T}_h^{**} \) (i.e. \( \partial T \cap \partial \Omega = \emptyset \)), there are at most two primal elements \( K \) and \( L \in \mathcal{T}_h \) such that \( T \cap K \neq \emptyset \) and \( T \cap L \neq \emptyset \).

(b) \( \mathcal{N}^{**} \) consists of three sets \( \mathcal{C} \), \( \mathcal{C}^* \) and \( \mathcal{N}_{\partial \Omega}^{**} \) containing mesh points of primal elements, mesh points of dual control volumes and points lying on the boundary respectively:

\[ \mathcal{N}^{**} = \mathcal{C} \cup \mathcal{C}^* \cup \mathcal{N}_{\partial \Omega}^{**}, \]

(2.1)

where \( \mathcal{C} := \{ K : \forall K \in \mathcal{T}_h \} \), \( \mathcal{C}^* := \{ M : \forall M \in \mathcal{T}_h^* \} \) and \( \mathcal{N}_{\partial \Omega}^{**} := \{ P \in \mathcal{N}^{**} \cap \partial \Omega \} \).

(c) There exists an integer constant \( \rho \), independent of size(\( \mathcal{T}_h^{**} \)), satisfying \( \text{card}(\mathcal{T}^*_P) < \rho \) for all \( P \in \mathcal{N}^{**} \).

For each primal element \( K \in \mathcal{T}_h \) and the average of tensor \( \Lambda \) is denoted by \( \Lambda_K = \frac{1}{m_K} \int_K \Lambda(\mathbf{x})d\mathbf{x} \). Moreover, for any \( T \in \mathcal{T}_h^{**} \), we define \( \Lambda_T = \Lambda_K \) on \( T \cap K \neq \emptyset \).
We aim to handle the heterogeneous, anisotropic case where $\Lambda$ is discontinuous across the primal elements, i.e.:

$$\Lambda_K \neq \Lambda_L \quad \text{for any } K, L \in \mathcal{T}_h, \ K \neq L. \quad (2.2)$$

The FECC scheme solves an approximate solution of problem (1.3) by finding its values at all nodes $P \in \mathcal{N}^{**}$. Hence, we define by $V_h$ the set of all vectors $u_h := (u_P)_{P \in \mathcal{N}^{**}}$ where $u_P$ is regarded as the approximate value of the solution $u(P)$ for all $P \in \mathcal{N}^{**}$:

$$\mathcal{H}_h = \{ u_h = (u_P)_{P \in \mathcal{N}^{**}}, \ u_P \in \mathbb{R} \},$$

and its norm is defined by

$$|u_h|_{1,D^{**}}^2 = \sum_{T \in \mathcal{T}_h^{**}, T = (MKL)} \left[ \frac{\tau_{LK}}{d_{LK}}(u_L - u_K)^2 \right.$$

$$\left. + \frac{\tau_{KM}}{d_{KM}}(u_M - u_K)^2 + \frac{\tau_{ML}}{d_{ML}}(u_M - u_L)^2 \right], \quad (2.3)$$

for every $u_h \in \mathcal{H}_h$. By Remark 2.1(b), we have

$$u_h = (u_P)_{P \in \mathcal{N}^{**}} = ((u_K)_{K \in C}, (u_M)_{M \in C^*}, (u_P)_{P \in \mathcal{N}^{**}}). \quad (2.4)$$

To handle Dirichlet boundary conditions, we need to define a subset of $\mathcal{H}_h$ as follows,

$$\mathcal{H}'_h = \{ u_h \in V_h : u_P = 0 \ \forall P \in \mathcal{N}_{\partial\Omega}^{**} \}. \quad (2.5)$$

To obtain the discrete variational formulation associated with problem (1.3), we define a projection operator $\Phi(u_h)$ and the discrete gradient $\nabla_{\Lambda}u_h$ for $u_h \in \mathcal{H}_h$.

2.2. The projection operator and the discrete gradient. The two operators are defined by their restrictions to each element of $\mathcal{T}_h^{**}$. In particular, the projection operator $\Phi(u_h)$ is a function in $L^2(\Omega)$ and it is continuous piecewise linear on each element $T \in \mathcal{T}_h^{**}$; and the discrete gradient is defined in a way to enforce mass
conservation in each element $T \in T^{**}_h$ when the coefficient $\Lambda$ is discontinuous (cf. (2.2) and Remark 2.1(a)).

We consider a triangle $T = (MKL) \in T^{**}_h$ where $K, L$ are two mesh points of two primal elements $K, L \in T_h$, and $M$ is a mesh point of a dual control volume $M \in T^*_h$ (see Figure 3). Denote by $\sigma$ the common edge of $K$ and $L$ and $C_{\sigma}$ the intersecting point between the segment $[KL]$ and $\sigma$. For any $u_h \in H_h$, the restriction of $\Phi(u_h)$ to $T$, denoted by $\Phi_T(u_h)$, is a continuous function and it is linear on each of the two sub-triangles $(MKC_{\sigma})$ and $(MLC_{\sigma})$.

Let $u^M_\sigma$, an auxiliary unknown to be defined later, be an approximation of $u_h$ at $C_{\sigma}$ seeing from $M$.

![Figure 3. Left: an element of the dual sub-grid $T = (MKL)$.
Center and right: outward normal vectors of each sub-triangle.](image)

We denote by $n^K_{[MC_{\sigma}]}$, $n^L_{[MK]}$ and $n^M_{[KC_{\sigma}]}$ the outward normal vectors of the triangle $(MKC_{\sigma})$ such that the lengths of these vectors are equal to the segments $[MC_{\sigma}], [MK]$ and $[KC_{\sigma}]$ respectively (see Figure 3). We also denote by $m_{(MKC_{\sigma})}$ the measure of triangle $(MKC_{\sigma})$. Remark that $n^K_{[MC_{\sigma}]} + n^K_{MC_{\sigma}} = 0$.

For any vector $u_h \in H_h$, the projection operator $\Phi(u_h)$ and the discrete gradient $\nabla \Lambda u_h$ restricted to $T$ are defined as follows:

(i) On the triangle $(MKC_{\sigma})$, we have

$$\Phi_T(u_h)|_{(MKC_{\sigma})}(x) = \begin{cases} u_M & \text{if } x = M, \\
K & \text{if } x = K, \\
M & \text{if } x = C_{\sigma}. \end{cases}$$

Now using multi-point flux approximations, the restriction of $\nabla \Lambda u_h$ on $(MKC_{\sigma})$ is defined as

$$\left(\nabla \Lambda u_h\right)_{(MKC_{\sigma})} = \frac{-u_K n^K_{[MC_{\sigma}]} - u^M_M n^M_{[MK]} - u_M n^M_{[KC_{\sigma}]}}{2m_{(MKC_{\sigma})}}. \quad (2.5)$$

Similarly, the restrictions of $u_h$ and $\nabla \Lambda u_h$ on triangle $(MLC_{\sigma})$ are respectively

$$\Phi_T(u_h)|_{(MLC_{\sigma})}(x) = \begin{cases} u_M & \text{if } x = M, \\
L & \text{if } x = L, \\
M & \text{if } x = C_{\sigma}, \end{cases}$$

and

$$\left(\nabla \Lambda u_h\right)_{(MLC_{\sigma})} = \frac{-u_L n^L_{[MC_{\sigma}]} - u^M_M n^M_{[ML]} - u_M n^M_{[LC_{\sigma}]}}{2m_{(MLC_{\sigma})}}. \quad (2.6)$$
(ii) We choose $u_{\sigma}^M$ to satisfy the continuity of the flux across $[MC_{\sigma}]:$

$$\Lambda_K (\nabla u_h)_{MC_{\sigma}} \cdot n^K_{MC_{\sigma}} + \Lambda_L (\nabla u_h)_{MLC_{\sigma}} \cdot n^L_{MC_{\sigma}} = 0. \tag{2.7}$$

Assume that

$$\Delta := \beta_{1,\sigma}^M + \beta_{2,\sigma}^M = -\frac{(n^K_{MC_{\sigma}})^t \Lambda_K n^K_{MC_{\sigma}}}{2m(MK_{\sigma})} - \frac{(n^L_{MC_{\sigma}})^t \Lambda_L n^L_{MLC_{\sigma}}}{2m(MLC_{\sigma})} \neq 0, \tag{2.8}$$

where $n^t$ is the transpose of vector $n$, then after performing some calculation on (2.7) we deduce that

$$u_{\sigma}^M = \beta_M u^M + \beta_K u^K + \beta_L u^L, \tag{2.9}$$

where

$$\beta_K = \frac{1}{\Delta} \left( \frac{n^K_{MC_{\sigma}}}{2m(C_{\sigma}C_{\sigma}C_K)} \right)^t \Lambda_K n^K_{MC_{\sigma}}, \quad \beta_L = \frac{1}{\Delta} \left( \frac{n^L_{MLC_{\sigma}}}{2m(C_{\sigma}C_{\sigma}C_L)} \right)^t \Lambda_L n^L_{MLC_{\sigma}}, \quad \beta_M = 1 - \beta_K - \beta_L.$$

**Remark 2.2.** For each internal edge $\sigma \equiv [\tilde{M}]$ of the mesh $T_h$, there are two approximate values $u_{\sigma}^M, u_{\sigma}^\tilde{M}$ of $u$ at $C_{\sigma}$. As for $u_{\sigma}^M$, the value $u_{\sigma}^M$ can be expressed as a linear combination of $u^M, u^K, u_L$. Together the difference between measure values of outward normal vectors, areas of two triangles $(M, K, L)$ and $(\tilde{M}, K, L)$, one has two different values of $u$ at $C_{\sigma}$: $u_{\sigma}^M \neq u_{\sigma}^\tilde{M}$. For the homogeneous Dirichlet boundary conditions, $u_{\sigma}^M = 0$ if $C_{\sigma} \in \partial \Omega$.

Substituting (2.9) into (2.5) and (2.6), we conclude that the discrete gradient $\nabla u_h$ restricted to the triangle $T = (MKL) \in T_h^{**}$ depends linearly on the three nodal values $u^M, u^K$ and $u_L$:

$$\nabla u_h = \frac{-u_K n^K_{MKC_{\sigma}} - u_L n^L_{MLC_{\sigma}} - u_M n^M_{MC_{\sigma}}}{2m(MK_{\sigma})}; \tag{2.10}$$

$$\nabla u_h = \frac{-u_K n^K_{MLC_{\sigma}} - u_L n^L_{MLC_{\sigma}} - u_M n^M_{MLC_{\sigma}}}{2m(MLC_{\sigma})}; \tag{2.11}$$

with

$$\tilde{n}^K_{MC_{\sigma}} = n^K_{MC_{\sigma}} + \beta_K n^K_{MK_{\sigma}}, \quad \tilde{n}^L_{MC_{\sigma}} = n^L_{MC_{\sigma}} + \beta_L n^L_{ML_{\sigma}},$$

$$\tilde{n}^M_{MC_{\sigma}} = n^K_{MC_{\sigma}} + \beta_M n^K_{ML_{\sigma}}, \quad \tilde{n}^L_{MLC_{\sigma}} = n^K_{MLC_{\sigma}} + \beta_L n^K_{ML_{\sigma}}, \quad \tilde{n}^M_{MLC_{\sigma}} = n^K_{MLC_{\sigma}} + \beta_M n^K_{ML_{\sigma}}.$$
for each \( P \in (\mathcal{N}^{**} \setminus \mathcal{N}_{\partial \Omega}^{**}) = \mathcal{C}^* \cup \mathcal{C} \). Then we obtain
\[
\int_{\Omega} (\nabla \Lambda u_h) \cdot \nabla \Lambda v_P^h \, dx = \int_{\Omega} f \Phi(v_P^h) \, dx, \quad \forall P \in (\mathcal{C}^* \cup \mathcal{C}),
\]
(2.14)
in which the discrete gradient depends only on the nodal values \( u_Q, Q \in \mathcal{N}^{**} \) (cf. formulas (2.10) and (2.11)).

Now, we proceed as in [20, p. 12-14] by first choosing \( v_h = v_M^h \) for each \( M \in \mathcal{C}^* \) in (2.14) and obtain the linear system
\[
\mathbf{D} u_h|_{T_h^*} + \mathbf{E} u_h|_{T_h} = \mathbf{F}^*,
\]
(2.15)
where \( u_h|_{T_h^*} := (u_M)_{M \in T_h^*}, u_h|_{T_h} := (u_K)_{K \in T_h}, \mathbf{F}^* = (\int_{T_h} f \Phi(v_P^h) \, dx)_{P \in \mathcal{C}} \) a column matrix depending on \( f \). And the square matrix \( \mathbf{D} \) is diagonal, positive definite, since \( \text{supp}(\nabla \Lambda v_M^h) \subset M \). This implies
\[
u_h|_{T_h^*} = \mathbf{D}^{-1}(\mathbf{F}^* - \mathbf{E} u_h|_{T_h}).
\]
(2.16)
Next, we take \( v_h = v_K^h \) for each \( K \in \mathcal{C} \):
\[
\mathbf{M} u_h|_{T_h^*} + \mathbf{N} u_h|_{T_h} = \mathbf{F},
\]
(2.17)
where \( \mathbf{N} \) is a symmetric, square matrix, \( \mathbf{F} \) a column matrix depending on \( f \) and \( \mathbf{M} \) is the transpose matrix of \( \mathbf{E} \).

Deriving from (2.15) and (2.17), we obtain the following matrix system associated with (2.14)
\[
\begin{pmatrix}
\mathbf{D} & \mathbf{E} \\
\mathbf{M} & \mathbf{N}
\end{pmatrix}
\begin{pmatrix}
u_h|_{T_h^*} \\
\nu_h|_{T_h}
\end{pmatrix} = \begin{pmatrix}\mathbf{F}^* \\
\mathbf{F}\end{pmatrix}.
\]
(2.18)

In addition, substituting (2.16) into (2.17), we obtain another linear system involving only primal cell unknowns \( (u_K)_{K \in T_h} \) as follows
\[
(\mathbf{N} - \mathbf{MD}^{-1}\mathbf{E}) u_h|_{T_h} = \mathbf{F} - \mathbf{MD}^{-1}\mathbf{F}^*.
\]
(2.19)
The matrix \( \mathbf{A} := \mathbf{N} - \mathbf{MD}^{-1}\mathbf{E} \) is a variant of the stiffness matrix. Under assumption (2.8), \( \mathbf{A} \) is symmetric and positive definite on general meshes [20].

System (2.19) verifies that the FECC scheme is a cell-centered one, since this system only have the cell-centered unknowns. We also recall [20, Corollary 5.4] that the FECC scheme is convergent, that is to say, \( \Phi(u_h) \) converges to the exact solution \( u_{exa} \) of problem (1.3) and \( \nabla \Lambda u_h \) tends to \( \nabla u_{exa} \) as \( h \to 0 \).

Now, let us concentrate on the discrete maximum principle for the FECC scheme. We introduce the following definitions based on a straightforward analog of Theorem 1.1 and Corollary 1.2.

**Proposition 2.3.** Problem (2.12) satisfies
(a) the discrete maximum principle if
\[
f \leq 0 \quad \text{a.e. in } \Omega \Rightarrow \max_{\Omega} u_h \leq 0,
\]
(2.20)
(b) the discrete nonnegative preservation if
\[
f \geq 0 \quad \text{a.e. in } \Omega \Rightarrow u_h \geq 0,
\]
(2.21)
where \( u_h \) is the solution of Problem (2.12).

As in [31, Theorem 3.2] and [9, Proposition 1.4], the principles (a) and (b) from Definition 2.3 are equivalent, since the FECC scheme follows the idea of the standard finite element method applying on the dual sub mesh \( T_h^{**} \) (in the isotropic permeability tensor case, these two schemes are equivalent on \( T_h^{**} \)).
**Proposition 2.4.** Assume \( f \geq 0 \) on \( \Omega \). If \( u_h \) is a solution to a scheme satisfying the DMP, then \( u_h \geq 0 \).

Unfortunately, Test 2 in Section 5 states that the FECC violates the DMP. Therefore, in the following section, we present the monotone nonlinear technique to correct the FECC, which establishes a monotone nonlinear cell-centered finite element scheme (MNFECC) satisfying the DMP.

3. The MNFECC frame work

We begin by defining an operator \( A^{D*} : H_h \rightarrow H_h \) for (2.14), such that
\[
-A^{D*}(u_h) \cdot v^o_h = \int_\Omega (\Lambda \nabla u_h) \cdot \nabla v^o_h \, dx, \quad \forall v^o_h \in (C^* \cup C),
\]
where \( A^{D*}(u_h) = (A^{D*}_P(u_h))_{P \in N} \) satisfies \( A^{D*}_P(u_h) = 0 \) for all \( P \in N^*_h \), with the homogeneous Dirichlet boundary condition.

A correction for the FECC scheme defined by \( A^{D*} \) is a family of functions \((\beta_P, Q)_{P \in N^*_h, Q \in V(P)}\) with \( \beta_P, Q : H_h \rightarrow \mathbb{R} \), where the set \( V(P) \) corresponding to the stencil of the FECC scheme is defined as

(a) If \( P \in C \) is a primal mesh point of \( P \in T_h \), then
\[
V(P) = N^*_P \cup \left( \cup_{M \in N^*_P} \{ N^*_M : M \in T^*_h \text{ has the mesh point } M \} \right).
\]

(b) If \( P \in C^* \) is a dual mesh point of \( P \in T^*_h \), then
\[
V(P) = N^*_P \cup \{ P \},
\]

note that \( N^*_P \subset C \cup N^*_h \).

**Remark 3.1.** By construction and Remark 2.1, a set \( V(\cdot) \) is symmetric in the sense that: for \( P, Q \in (C \cup C^*) \), if \( P \in V(Q) \) then \( Q \in V(P) \).

With the above definition of \( V(P) \), for all \( P \in (C \cup C^*) \), we suppose the discrete linear operator \( A^{D*} \) rewrites in the form
\[
\forall u_h \in H^0_h, \forall P \in (C \cup C^*), \quad A^{D*}_P(u_h) = \sum_{Q \in V(P)} \alpha_{P,Q}(u_Q - u_P).
\]

Now, let us give a parameter \( \eta > 0 \). For all \( u_h \in H^0_h \), all \( P \in C \cup C^* \) and all \( Q \in V(P) \), we define a correction \( \beta \) as follows:

If \( Q \in N^*_h \), then
\[
B_P,Q(u_h) = \eta \left( \frac{|A^{D*}_P(u_h)|}{\sum_{Y \in V(P)} |u_Y - u_P|} \right).
\]

If \( Q \in C \cup C^* \), then
\[
B_P,Q(u_h) = \eta \left( \frac{|A^{D*}_P(u_h)|}{\sum_{Y \in V(P)} |u_Y - u_P|} + \frac{|A^{D*}_Q(u_h)|}{\sum_{Z \in V(Q)} |u_Z - u_Q|} \right).
\]

Note that if one of \( \sum_{Y \in V(P)} |u_Y - u_P| \) or \( \sum_{Z \in V(Q)} |u_Z - u_Q| \) is zero, the corresponding term \( \frac{|A^{D*}_P(u_h)|}{\sum_{Y \in V(P)} |u_Y - u_P|} \) or \( \frac{|A^{D*}_Q(u_h)|}{\sum_{Z \in V(Q)} |u_Z - u_Q|} \) is dropped in (3.5) and (3.6).
Using the above correction $\mathcal{B}$, we define the monotone nonlinear cell-centered finite element scheme (MNFECC) as follows: finding $u_h = (u_P)_{P \in \mathcal{N}^{**}} \in \mathcal{H}_h^0$ such that

$$S_{P}^{D^{**}}(u_h) = \int_{\Omega} f(x) P_1(v_P^h) \, dx, \quad \forall P \in (\mathcal{C}^* \cup \mathcal{C}),$$

with $S_{P}^{D^{**}}(u_h) = (S_{P}^{D^{**}}(u_h))_{P \in \mathcal{N}^{**}}$ and

$$S_{P}^{D^{**}}(u_h) = \begin{cases} -A_{P}^{D^{**}}(u_h) + \sum_{Q \in \mathcal{V}(P)} B_{P,Q}(u_h)(u_P - u_Q), & P \in (\mathcal{C}^* \cup \mathcal{C}), \\ 0, & P \in \mathcal{N}_{\partial \Omega}^{**}, \end{cases}$$

for all $u_h \in \mathcal{H}_h^0$. The Lagrange interpolation function $P_1(v_P^h)$ of degree 1 on the mesh $\mathcal{T}_h^{**}$ is positive since $v_P^h > 0$, which guarantees that if $f \geq 0$, then the right-hand side is also positive.

Next, we present in more detail the operator $A^{D^{**}}$. For this work, we need to introduce and recall the following notation, definitions, lemmas in [20]: let us introduce a triangle $T := (MKL) \in \mathcal{T}_h^{**}$, see Figure 4.

**Figure 4. Trienagle $T$**

- For each pair of points $(P, Q) \in \mathcal{N}^{*}_{T,2} := \{(P, Q) : P, Q \in \mathcal{N}_T\}$, the associated edge $[PQ]$ has a midpoint $C_{[PQ]}$.
- Vectors $\mathbf{n}_{KM}, \mathbf{n}_{ML}$ and $\mathbf{n}_{LK}$ which are orthogonal to the edges $[TC_{MK}], [TC_{ML}]$ and $[TC_{KL}]$, are equal to the lengths of these edges, respectively.
- The vectors $\mathbf{n}_K, \mathbf{n}_L, \mathbf{n}_M, \mathbf{n}_{e1}, \mathbf{n}_{e2}$ and $\mathbf{n}_{M1,2}$ are orthogonal (with the same lengths) to the edges $[ML], [MK], [KL], [MC_e], [KC_e]$ and $[LC_e]$.
- The values $d_{LK}, d_{KM}$ and $d_{LM}$ are three distances from $K$ and $[TC_{KL}]$ ($d_{LK} = d_{KL}$), from $M$ to $[TC_{KM}]$ ($d_{KM} = d_{MK}$), from $L$ to $[TC_{ML}]$ ($d_{LM} = d_{ML}$). Besides, these distances are used to compute the area

$$m_T = \sum_{(P,Q) \in \mathcal{V}^{**}_T} d_{PQ} |\mathbf{n}_{PQ}|,$$

with $\mathcal{V}^{**}_T = \{(K, L), (K, M), (M, L)\}$.
By the property of the permeability tensor $\Lambda$, the dual mesh $\mathcal{T}^{**}_h$ has two subsets:

$$\mathcal{T}^{**}_{\text{const}} = \{ T \in \mathcal{T}^{**}_h : \Lambda \text{ is constant on } T \},$$
$$\mathcal{T}^*_h = \{ T \in \mathcal{T}^{**}_h : \Lambda \text{ is discontinuous on } T \}.$$

According to [20, Lemma 5.1 and 5.2], for any $u_h \in \mathcal{H}$, $T \in (\mathcal{MKL}) \in \mathcal{T}^{**}_h$, the discrete gradient $(\nabla_{\Lambda} u_h)_T$ is written in the following forms

(i) for $T \in \mathcal{T}^{**}_{\text{const}}$

$$m_T(\nabla_{\Lambda} u_h)_T = (u_M - u_K)\eta_{KM} + (u_L - u_M)\eta_{ML} + (u_K - u_L)\eta_{LK},$$

$$= (u_M - u_K)\eta_{KM}^T + (u_L - u_K)\eta_{KL}^T,$$

$$= (u_K - u_M)\eta_{MK}^T + (u_L - u_M)\eta_{ML}^T,$$

with $\eta_{KM}^T = \tau_{KM} - \tau_{ML}$, $\eta_{KL}^T = \tau_{ML} - \tau_{LK}$; $\eta_{MK}^T = \tau_{LK} - \tau_{KM}$ and $\eta_{ML}^T = \tau_{ML} - \tau_{LK}$.

(ii) for $T \in \mathcal{T}^{**}_h \setminus (\mathcal{T}^{**}_{\text{const}} \cup \mathcal{T}^{**}_{\text{const}})$

$$m_T(\nabla_{\Lambda} u_h)_T = (u_M - u_K)(\tau_{KM} + \epsilon_{KM}) + (u_L - u_M)(\tau_{ML} + \epsilon_{ML})$$

$$+ (u_K - u_L)(\tau_{LK} + \epsilon_{LK}),$$

$$= (u_M - u_K)\eta_{KM}^T + (u_L - u_K)\eta_{KL}^T,$$

$$= (u_K - u_M)\eta_{MK}^T + (u_L - u_M)\eta_{ML}^T,$$

with

$$\lim_{h_{\mathcal{T}^{**}} \to 0} \left| \epsilon_{KM} \right| = 0, \quad \lim_{h_{\mathcal{T}^{**}} \to 0} \left| \epsilon_{ML} \right| = 0, \quad \lim_{h_{\mathcal{T}^{**}} \to 0} \left| \epsilon_{LK} \right| = 0.$$

Then we compute the vectors: $\eta_{KM}^T = \tau_{KM} + \epsilon_{KM} - \tau_{ML} - \epsilon_{ML}$, $\eta_{KL}^T = \tau_{ML} - \epsilon_{KL}$; $\eta_{MK}^T = \tau_{LK} - \epsilon_{KM}$ and $\eta_{ML}^T = \tau_{ML} + \epsilon_{ML} - \tau_{LK} - \epsilon_{LK}$.

(iii) for $T \in \mathcal{T}^{**}_A$, $T$ has two subsets $T_K = T \cap K$, $T_L = T \cap L$. On each subset $T_Q = T_K$ and $T_L$, we have

$$m_{T_Q}(\nabla_{\Lambda} u_h)_T = (u_M - u_K)\theta_{\text{KM}}^Q + (u_L - u_M)\theta_{\text{ML}}^Q + (u_K - u_L)\theta_{\text{LK}}^Q,$$

$$= (u_M - u_K)\theta_{\text{KM}}^Q + (u_L - u_K)\theta_{\text{KL}}^Q,$$

$$= (u_K - u_M)\theta_{\text{MK}}^Q + (u_L - u_M)\theta_{\text{ML}}^Q,$$

Furthermore, we have

$$|\theta_{\text{KM}}^Q| \leq C_2|\tau_{KM}|, \quad |\theta_{\text{ML}}^Q| \leq C_2|\tau_{ML}|, \quad |\theta_{\text{LK}}^Q| \leq C_2|\tau_{LK}|,$$

where $C_2$ is a constant. We put

$$\eta_{KM}^T = \begin{cases} \theta_{\text{KM}}^T, & \text{on } T_K, \\ \theta_{\text{MK}}^T, & \text{on } T_L, \end{cases}$$

$$m_T(\nabla_{\Lambda} u_h)_T = \begin{cases} m_{T_K}(\nabla_{\Lambda} u_h)_T, & \text{on } T_K, \\ m_{T_L}(\nabla_{\Lambda} u_h)_T, & \text{on } T_L, \end{cases}$$

with $\theta_{\text{KM}}^Q = \theta_{\text{KM}}^T - \theta_{\text{ML}}^T$, $\theta_{\text{KL}}^T = \theta_{\text{ML}}^T - \theta_{\text{LK}}^T$; $\theta_{\text{MK}}^T = \theta_{\text{LK}}^T - \theta_{\text{KM}}^T$ and $\theta_{\text{ML}}^T = \theta_{\text{ML}}^T - \theta_{\text{LK}}^T$.

Using the above formulas, (2.14) and (3.1), we define the operator $A^{**}$ as follows: for every $M \in C^*$ and $K \in C^*$,
In (3.1), choosing $\mathbf{P} \equiv \mathbf{M} \in \mathcal{C}^*$, its left hand side is rewritten as
\[
\int_{\Omega} \Lambda \nabla u_h \cdot \nabla v^M \, dx
= \sum_{T \in \mathcal{T}_h^*, T := (MKL)} \left( m_{TK} [\Lambda_K (\nabla_A u_h)_T] \cdot (\nabla_A v^M)_{T_K} + m_{TL} [\Lambda_L (\nabla_A u_h)_T] \cdot (\nabla_A v^M)_{T_L} \right) + m_T [\Lambda_T (\nabla_A u_h)_T] \cdot (\nabla_A v^M)_T.
\] (3.15)

Substituting the formulas (3.14)-(3.15) into (3.15), we obtain
\[
A_{K,M}^{**}(u_h) = \sum_{T \in \mathcal{T}_h^*} A_{K,M,T}^{**}(u_h),
\]
where the coefficients are defined as follows: on $T := (MKL) \in \mathcal{T}_h^* \setminus \mathcal{T}_h^*$,
\[
\alpha_{KM}^T = \frac{\Lambda_T \eta_{MK} \cdot (\eta_{MK} + \eta_{ML}^T)}{4m_T}, \quad \alpha_{LM}^T = \frac{\Lambda_T \eta_{ML} \cdot (\eta_{MK} + \eta_{ML}^T)}{4m_T},
\]
and on $T := (MKL) \in \mathcal{T}_h^*$,
\[
\alpha_{KM}^T = \sum_{Q=K,L} \frac{\Lambda_Q \eta_{MQ} \cdot (\eta_{MQ} + \eta_{ML}^T)}{4m_{TQ}}, \quad \alpha_{KL}^T = \sum_{Q=K,L} \frac{\Lambda_Q \eta_{QM} \cdot (\eta_{MQ} + \eta_{ML}^T)}{4m_{TQ}}.
\]

As for obtaining (3.16), we choose $\mathbf{P} \equiv \mathbf{K} \in \mathcal{C}$ in (3.1) and perform similar calculations to have the equation
\[
A_{K}^{**}(u_h) = \sum_{T \in \mathcal{T}_h^*} A_{K,T}^{**}(u_h),
\] (3.17)
whose coefficients are defined as follows: on $T := (MKL) \in \mathcal{T}_h^* \setminus \mathcal{T}_h^*$,
\[
\alpha_{KM}^T = \frac{\Lambda_T \eta_{KM} \cdot (\eta_{KM} + \eta_{KL}^T)}{4m_T}, \quad \alpha_{KL}^T = \frac{\Lambda_T \eta_{KL} \cdot (\eta_{KM} + \eta_{KL}^T)}{4m_T},
\]
and on $T := (MKL) \in \mathcal{T}_h^*$,
\[
\alpha_{KM}^T = \sum_{Q=K,L} \frac{\Lambda_Q \eta_{MQ} \cdot (\eta_{MQ} + \eta_{ML}^T)}{m_{TQ}}, \quad \alpha_{KL}^T = \sum_{Q=K,L} \frac{\Lambda_Q \eta_{QM} \cdot (\eta_{MQ} + \eta_{ML}^T)}{m_{TQ}}.
\]

For the non-linear equation (3.7), we apply an iterative algorithm to solve it. For each iteration step $(i)$, we denote $u_h^{(i)}$ its solution. We fix $u_h = u_h^{(i)}$ in $B_{P,Q}(u_h)$ in (3.8) and the iterative scheme for (3.7) can be written as
\[
-A_{P}^{**}(u_h^{(i+1)}) + \sum_{Q \in \mathcal{V}(P)} B_{P,Q}(u_h^{(i)}) (u_h^{(i+1)} - u_h^{(i)}) = \int_{\Omega} f_{P}(u_h^{(i)}) \, dx,
\] (3.18)
for all $\mathbf{P} \in \mathcal{C} \cup \mathcal{C}^*$. To construct the linear system associated with (3.15), we process the following two steps:
Step 1: Taking $v_h^M \in H_0^1$ with $M \in \mathcal{C}^*$, we rewrite (3.18) as
\[-A_M^{D^{i+1}}(u_h^{(i+1)}) + \sum_{Y \in V(M)} B_{M,Y}(u_h^{(i)})(u_{M}^{(i+1)} - u_Y^{(i+1)}) = \int_{\Omega} f P_1(v_h^M) \, dx,
\]
and obtain
\[
\overrightarrow{D} u_h^{(i+1)}|_{\mathcal{T}_h} + \overrightarrow{E} u_h^{(i+1)}|_{\mathcal{T}_h} = \overrightarrow{F}^*,
\]
with
\[
\overrightarrow{E} = E - (B_{M,Y}(u_h^{(i)}))_{M \in \mathcal{C}^*, Y \in \mathcal{C}},
\]
\[
\overrightarrow{D} = D + \left( \sum_{Y \in V(M)} B_{M,Y}(u_h^{(i)}) \right)_{M \in \mathcal{C}^*, M \in \mathcal{C}^*},
\]
in which the matrix $(B_{M,Y}(u_h^{(i)}))_{M \in \mathcal{C}^*, Y \in \mathcal{C}}$ belongs to $\mathbb{R}^{\text{card}(\mathcal{C}) \times \text{card}(\mathcal{C}^*)}$, and the square matrix $(\sum_{Y \in V(M)} B_{M,Y}(u_h^{(i)}))_{M \in \mathcal{C}^*, M \in \mathcal{C}^*}$ belongs to $\mathbb{R}^{\text{card}(\mathcal{C}^*) \times \text{card}(\mathcal{C}^*)}$, and is diagonal and positive definite, because of property (a) in Remark 3.1 and (2.16), and all coefficients $B_{M,Y}0$ are positive. It follows that the square matrix $\overrightarrow{D}$ is also diagonal and positive definite.

Step 2: We proceed as in Step 1 by choosing $v_h^K \in H_0^1$ with $K \in \mathcal{C}$, then (3.18) is rewritten as
\[-A_K^{D^{i+1}}(u_h^{(i+1)}) + \sum_{Z \in V(K)} B_{K,Z}(u_h^{(i)})(u_{K}^{(i+1)} - u_Z^{(i+1)}) = \int_{\Omega} f P_1(v_h^K) \, dx,
\]
From the above equation, we get the linearized system
\[
\overrightarrow{M} u_h^{(i+1)}|_{\mathcal{T}_h} + \overrightarrow{N} u_h^{(i+1)}|_{\mathcal{T}_h} = \overrightarrow{F},
\]
with
\[
\overrightarrow{M} = M - (B_{K,Z})_{K \in \mathcal{C}, Z \in \mathcal{C}^*}, \quad \overrightarrow{N} = N + \text{big}(\sum_{Z \in V(K)} B_{K,Z})_{K \in \mathcal{C}, K \in \mathcal{C}^*},
\]
in which the matrix $(B_{K,Z})_{K \in \mathcal{C}, Z \in \mathcal{C}^*}$ belongs to $\mathbb{R}^{\text{card}(\mathcal{C}) \times \text{card}(\mathcal{C}^*)}$ and the matrix $(\sum_{Z \in V(K)} B_{K,Z})_{K \in \mathcal{C}, K \in \mathcal{C}}$ is a diagonal, and positive definite in $\mathbb{R}^{\text{card}(\mathcal{C}^*) \times \text{card}(\mathcal{C})}$. Therefore, the matrix system associated with (3.18) and (3.22) is
\[
\begin{pmatrix}
\overrightarrow{D} & \overrightarrow{E} \\
\overrightarrow{M} & \overrightarrow{N}
\end{pmatrix}
\begin{pmatrix}
\begin{bmatrix} u_h^{(i+1)}|_{\mathcal{T}_h} \\ u_h^{(i+1)}|_{\mathcal{T}_h}
\end{bmatrix}
\end{pmatrix}
= 
\begin{pmatrix}
\begin{bmatrix} \overrightarrow{F}^* \\ \overrightarrow{F}
\end{bmatrix}
\end{pmatrix},
\]
Moreover, since $\overrightarrow{D}$ is diagonal and positive definite, we can compute $u_h^{(i+1)}|_{\mathcal{T}_h}$ from (3.20) as
\[
u_h^{(i+1)}|_{\mathcal{T}_h} = \overrightarrow{D}^{-1}(\overrightarrow{F}^* - \overrightarrow{E}u_h^{(i+1)}|_{\mathcal{T}_h}).
\]
Substituting this into (3.22), we get the following linearized system involving only primal cell unknowns $u_h^{(i+1)}|_{\mathcal{T}_h}$,
\[
(\overrightarrow{N} - \overrightarrow{M} \overrightarrow{D}^{-1}\overrightarrow{E}) u_h^{(i+1)}|_{\mathcal{T}_h} = \overrightarrow{F} - \overrightarrow{M} \overrightarrow{D}^{-1}\overrightarrow{F}^*.
\]
We stop the above algorithm when the criterion $\frac{|u_h^{(i+1)}|_{\mathcal{T}_h} - u_h^{(i)}|_{\mathcal{T}_h}|}{|u_h^{(i)}|_{\mathcal{T}_h}} \leq 10^{-4}$ is satisfied.
4. Mathematical properties of the MNFECC

From now on, we make all assumptions for the FECC scheme stated in [20, Section 5]. With these assumptions, we begin by estimating the solution of the MNFECC scheme in the \( \| \cdot \|_{1, \mathcal{D}^{*}} \) norm, it suffices for this scheme to fulfill the following coercivity property.

**Proposition 4.1.** Let \( \mathcal{G} = (\mathcal{T}_h, \mathcal{H}_h, \nabla_{\Lambda}, \Phi) \) be a discretization on \( \Omega \), then the MNFECC scheme is coercive; it means that there exists a positive constant \( \rho_1 \) such that

\[
\forall u_h \in \mathcal{H}_h, \quad \sum_{P \in (C \cup C^*)} S_{p}^{D^{*}}(u_h)u_P \geq \rho_1 \| u_h \|_{1, \mathcal{D}^{*}}^2
\]  

(4.1)

**Proof.** Multiplying (3.8) by \( u_P \), for \( P \in (C \cup C^*) \), summing over all, and using the property (1.2) with (3.1), we obtain

\[
\sum_{P \in (C \cup C^*)} S_{p}^{D^{*}}(u_h)u_P = - \sum_{P \in (C \cup C^*)} A_{p}^{D^{*}}(u_h)u_P + \sum_{P \in (C \cup C^*)} u_P \sum_{Q \in \mathcal{V}(P)} B_{P,Q}(u_h)(u_P - u_Q)
\]  

(4.2)

\[
\geq \lambda \| \nabla_{\Lambda} u_h \|_{(L^2(\Omega))}^2.
\]

From the symmetric property of the set \( \mathcal{V}(P) \) (see Remark 3.1), we have

\[
\sum_{P \in (C \cup C^*)} u_P \sum_{Q \in \mathcal{V}(P)} B_{P,Q}(u_h)(u_P - u_Q) = \sum_{P \in (C \cup C^*)} \sum_{Q \in \mathcal{V}(P)} B_{P,Q}(u_h)(u_P - u_Q)^2,
\]

By [20, inequality (21)] there exists \( C_4 \) such that \( \| u_h \|_{1, \mathcal{D}^{*}}^2 \leq C_4 \| \nabla_{\Lambda} u_h \|_{(L^2(\Omega))}^2 \). We combine the above equation with (4.2) to get (4.1) with \( \rho_1 = \lambda/C_4 \).

This coercivity property allows us to estimate a solution of the MNFECC scheme using the following result.

**Proposition 4.2.** Let \( \mathcal{G} = (\mathcal{T}_h, \mathcal{H}_h, \nabla_{\Lambda}, \Phi) \) be a discretization on \( \Omega \), and \( u_h \) be a solution to (3.7). Then there exists a positive constant \( \rho_2 \) depending only on \( \Omega \) such that

\[
\| u_h \|_{1, \mathcal{D}^{*}} \leq \rho_2 \| f \|_{L^2(\Omega)}.
\]  

(4.3)

**Proof.** Multiplying (3.7) by \( u_P \), summing over \( P \in (C^* \cup C) \), and using (4.2) and [20, inequality (26)], we obtain \( C_{22} \) such that

\[
\| P_1(u_h) \|_{L^2(\Omega)} \leq C_{22} \| \nabla_{\Lambda} u_h \|_{(L^2(\Omega))}^2
\]

and

\[
\lambda \| \nabla_{\Lambda} u_h \|_{(L^2(\Omega))}^2 \leq \sum_{P \in (C^* \cup C)} S_{p}^{D^{*}}(u_h)u_P
\]  

(4.4)

\[
= \int_{\Omega} f(x) P_1(u_h) dx \leq \| f \|_{L^2(\Omega)} \| P_1(u_h) \|_{L^2(\Omega)} \leq C_{22} \| f \|_{L^2(\Omega)} \| \nabla_{\Lambda} u_h \|_{(L^2(\Omega))}^2.
\]

From inequalities (4.4) and (20), we obtain

\[
\lambda \| u_h \|_{1, \mathcal{D}^{*}} \leq \| \nabla_{\Lambda} u_h \|_{(L^2(\Omega))}^2 \leq C_{22} \| f \|_{L^2(\Omega)}.
\]

We can choose \( \rho_2 = \frac{C_{22} \sqrt{\lambda}}{\lambda} \) to obtain (4.3).
In addition, the coercivity property guarantees that there exists one solution to the MNFECC scheme.

**Proposition 4.3.** Let $\mathcal{G} = (\mathcal{T}_h, \mathcal{H}_h, \nabla, \Phi)$ be a discretization on $\Omega$. Then there exists one solution $u_h$ to (3.7).

**Proof.** For $u_h \in \mathcal{H}^0_h$, we first define the map $H$ joining $A^{D^{**}}$ and $S^{D^{**}}$ by $H(t, u_h) = -(1 - t)A^{D^{**}}(u_h) + tS^{D^{**}}(u_h)$ for all $t \in [0, 1]$. Obviously, we have

$$H(0, u_h) = A^{D^{**}}(u_h) \quad \text{and} \quad H(1, u_h) = S^{D^{**}}(u_h). \quad (4.5)$$

Also we need to estimate the solution of the equation

$$H(t, u_h) = \mathcal{F}_h. \quad (4.6)$$

with $\mathcal{F}_h = \left( \int_\Omega f(x) P_1(v^h) dx \right)_{P \in (C^* \cup \mathcal{C})}$. Based on the coercivity of $-A^{D^{**}}$ and $S^{D^{**}}$, for all $t \in [0, 1]$ and any solution to (4.6), we have the coercivity

$$H(t, u_h) \cdot u_h \geq (1 - t)\Delta|\nabla u_h|^2_{L^2(\Omega)} + t\Delta|\nabla u_h|^2_{L^2(\Omega)} \geq \rho_1 |u_h|^2_{1, D^{**}}.$$ 

It follows that any solution to (4.6) is also bounded by the value $\rho_2 \|f\|_{L^2(\Omega)}$ in the norm $\| \cdot \|_{1, D^{**}}$. This result and (4.5) indicate that the map $H$ is a homotopy joining $A^{D^{**}}$ and $S^{D^{**}}$. Therefore, the Brower’s topological degree of $H(0, \cdot) = S^{D^{**}}$ is the same the degree of $H(1, \cdot) = -A^{D^{**}}$ which is non zero (since $-A^{D^{**}}(u_h) = \mathcal{F}_h$ always has a unique solution). Consequently, there exists one solution to the MNFECC scheme (3.7).

□

**Proposition 4.4.** Let $\mathcal{G} = (\mathcal{T}_h, \mathcal{H}_h, \nabla, \Phi)$ be a discretization on $\Omega$, and let the parameter $\eta > 1$. The the MNFECC scheme satisfies the DMP (see Definition 2.4); this means that if $f \geq 0$ on $\Omega$, then a solution $u_h = (u_P)_{P \in N^{**}}$ to (3.7) satisfies $\min_{P \in N^{**}} u_P \geq 0$.

**Proof.** For a solution $u_h = (u_P)_{P \in N^{**}} \in \mathcal{H}_h^0$ to (3.7), we put

$$u_{P_0} = \min_{P \in N^{**}} u_P.$$ 

We consider the first assumption: $P_0 \in (C^* \cup \mathcal{C})$ is a mesh point of a grid element $P_0 \subseteq (\mathcal{T}_h \cup \mathcal{T}_h^0)$ such that $\partial P_0 \cap \partial \Omega = \emptyset$. At the point $P_0$, note that $V(P_0) \subseteq (C^* \cup \mathcal{C})$, we have

$$S^{D^{**}}_{P_0}(u_h) = A^{D^{**}}_{P_0}(u_h) + \sum_{P_0 \in V(P_0)} B_{P_0, Y}(u_{P_0} - u_Y)$$

$$= \sum_{Y \in V(P_0)} A^{D^{**}}_{P_0}(u_h) \frac{(|u_{P_0} - u_Y|)}{\sum_{Q \in V(P_0)}(|u_{P_0} - u_Q|)}$$

$$+ \eta \sum_{Y \in V(P_0)} \left[ \frac{|A^{D^{**}}_{P_0}(u_h)|}{\sum_{Q \in V(P_0)}(|u_Q - u_{P_0}|)} \right] (u_{P_0} - u_Y)$$

$$+ \sum_{Y \in V(Y)} \theta_{P_0, Y}(u_{P_0} - u_Y) \quad (4.7)$$
with
\[
\theta_{p_0, Y} = \frac{A_{p_0}^{D^*}(u_h) \sgn(u_{p_0} - u_Y)}{\sum_{Q \in V(p_0)}(|u_{p_0} - u_Q|)} + \frac{\eta|A_{p_0}^{D^*}(u_h)|}{\sum_{Q \in V(p_0)}(|u_Q - u_{p_0}|)} + \frac{\eta|A_Y^{D^*}(u_h)|}{\sum_{Q \in V(Y)}(|u_Q - u_Y|)} > 0.
\]

Substituting (4.7) into the left-hand side of (3.7), we obtain
\[
\sum_{Y \in V(p_0)} \theta_{p_0, Y}(u_{p_0} - u_Y) = \int_{\Omega} f(x) P_1(u^P_0) \, dx.
\]

Obviously, we have
\[
\sum_{Y \in V(p_0) \setminus \Lambda_{\Omega}^*} \theta_{p_0, Y}(u_{p_0} - u_Y) < 0, \quad \text{and} \quad S_{p_0}^{D^*}(u_h) = \int_{\Omega} f(x) P_1(u^P_0) \, dx > 0,
\]

thus the value \(u_{p_0}\) must be greater than 0.

Next we show the convergence of the MFNECC scheme. For any function \(\varphi \in C^\infty_c(\Omega)\), we put \(\varphi_h = (\varphi_P)_{P \in \mathcal{N}^{D^*}} \in H^1_0(\Omega)\), with \(\varphi_P = \varphi(P)\). We multiply (4.7) at \(P\) by \(\varphi_P\) and sum over \(P \in (C^* \cup C)\) to obtain
\[
- \sum_{P \in (C^* \cup C)} A_P^{D^*}(u_h) \varphi_P + \sum_{P \in (C^* \cup C)} \varphi_P \sum_{Q \in V(P)} B_{P,Q}(u_h)(u_P - u_Q) = \int_{\Omega} f(x) P_1(\varphi_h) \, dx.
\]

Obviously, the right-hand side tends to \(\int_{\Omega} f(x) \varphi_h \, dx\), as \(\text{size}(\mathcal{T}_{\Omega}^{D^*}) \to 0\). The convergence of the FECC and the FECCB schemes are shown in [20 Proposition 5.3 and Corollary 5.4]; These results and (3.1) ensure that
\[
- \sum_{P \in (C^* \cup C)} A_P^{D^*}(u_h) \varphi_P = \int_{\Omega} \Lambda \nabla \varphi_h \cdot \nabla \varphi_h \, dx \to \int_{\Omega} \Lambda \nabla u \cdot \nabla \varphi_h \, dx,
\]
as \( \text{size}(\mathcal{T}_h^{**}) \to 0 \), where \( u \) is the unique solution to problem \((1.3)\).

For the corrected term, we use its symmetric property and positive property
\( (B_{P,Q}(u_h) = B_{Q,P}(u_h), B_{P,Q}(u_h) > 0) \) to compute

\[
\sum_{P \in (C^* \cup C)} \sum_{Q \in V(P)} B_{P,Q}(u_h)(u_P - u_Q)
= \sum_{P \in (C^* \cup C)} \sum_{Q \in (V(P) \cap N_{Q0}^{**})} B_{P,Q}(u_h)u_P \varphi_P
+ \sum_{P \in (C^* \cup C)} \sum_{Q \in (V(P) \cap N_{Q0}^{**})} B_{P,Q}(u_h)(u_P - u_Q)(\varphi_P - \varphi_Q)
\leq \sum_{P \in (C^* \cup C)} \sum_{Q \in V(P)} B_{P,Q}(u_h)|u_P - u_Q|\|\varphi_P - \varphi_Q|.
\]

Lemma 4.5 implies
\[
\sum_{P \in (C^* \cup C)} \sum_{Q \in V(P)} B_{P,Q}(u_h)|u_P - u_Q|\|\varphi_P - \varphi_Q| \to 0, \quad \text{as \( \text{size}(\mathcal{T}_h^{**}) \to 0 \)}
\]

Therefore, for any \( \varphi \in C_c^\infty(\Omega) \), as \( \text{size}(\mathcal{T}_h^{**}) \to 0 \), equation \((4.8)\) converges to

\[
\int_\Omega \Lambda \nabla u \cdot \nabla \varphi \, dx = \int_\Omega f \varphi \, dx,
\]

in which \( u \) must be equal to the unique solution of \((1.3)\) on almost every \( \Omega \).

**Lemma 4.5.** Let \( \mathcal{G} = (\mathcal{T}_h, \mathcal{H}_h, \nabla_\Lambda, \Phi) \) be a discretization on \( \Omega \), and \( u_h \) be a solution to the MNFECC scheme \((3.7)\). For any \( \varphi \in C_c^\infty(\Omega) \) setting \( \varphi_h = (\varphi_P)_{P \in N_{Q0}^{**}} \in \mathcal{H}_h^0 \) with \( \varphi_P = \varphi(P) \), we have

\[
\sum_{P \in (C^* \cup C)} \sum_{Q \in V(P)} B_{P,Q}(u_h)|u_Q - u_P|\|\varphi_P - \varphi_Q| \to 0,
\]

as \( \text{size}(\mathcal{T}_h^{**}) \to 0 \), where any \( P \in (C^* \cup C) \) is a mesh point of \( P \in (\mathcal{T}_h^{*} \cup \mathcal{T}_h^{**}) \). Note that if \( Q \in (V(P) \cap N_{Q0}^{**}) \) then \( u_Q = \varphi_Q = 0 \), since \( u_h \in \mathcal{H}_h^0 \).

**Proof.** With definitions \((3.5)\) and \((3.6)\), we have a positive constant \( \rho_3 \), only depended on \( \eta \), such that

\[
\sum_{P \in (C^* \cup C)} \sum_{Q \in V(P)} B_{P,Q}(u_h)|u_P - u_Q|\|\varphi_P - \varphi_Q| \leq \rho_3 \sum_{P \in (C^* \cup C)} \text{diam}(P) \sum_{Q \in V(P)} |A_P^{**}(u_h)|,
\]

Since we have two positive constants \( \rho_4 \) and \( \rho_5 \), independent of the sizes of three meshes, such that

\[
|\varphi_P - \varphi_Q| = |\varphi_P| \leq \rho_4 \text{diam}(P), \quad \forall Q \in (V(P) \cap N_{Q0}^{**}) \text{ and } \varphi_Q = 0,
|\varphi_P - \varphi_Q| \leq \rho_5 \text{diam}(P), \quad \varphi_P - \varphi_Q \leq \rho_5 \text{diam}(Q), \quad \forall Q \in (V(P) \setminus N_{Q0}^{**}),
\]

we obtain

\[
\frac{|u_P - u_Q|}{\sum_{Z \in V(Q)} |u_Z - u_Q|} < 1, \quad \text{with } Q \in V(P), \ P \in V(Q).
\]
Considering on a primal mesh point \( P \equiv K \in \mathcal{C} \) of \( K \in \mathcal{T}_h \), we use (3.1) to write

\[
| \sum_{K \in \mathcal{C}} \text{diam}(K)A_{K}^{P \rightarrow \star}(u_h)| \\
= | \sum_{K \in \mathcal{C}} \text{diam}(K) \sum_{T \in T_{K}^*} \int_{T} \Lambda \nabla_{A} u_h \cdot \nabla v_{h}^{K} \, dx | \\
\leq \sum_{K \in \mathcal{C}} \sum_{T \in T_{K}^*} \text{diam}(K) \int_{T} |\Lambda \nabla_{A} u_h \cdot \nabla v_{h}^{K}| \, dx \\
\leq \lambda \sqrt{\text{size}(\mathcal{T}_h)} \sum_{K \in \mathcal{C}} \sum_{T \in T_{K}^*} \sqrt{m_{T}|(\nabla_{A} u_{h})_{T}|} \sqrt{\text{diam}(K)} \sqrt{m_{T}|(\nabla v_{h}^{K})_{T}|} \\
\leq \lambda \sqrt{\text{size}(\mathcal{T}_h)} \left( \sum_{K \in \mathcal{C}} \sum_{T \in T_{K}^*} m_{T}|(\nabla_{A} u_{h})_{T}|^{2} \right)^{1/2} \\
\times \left( \sum_{K \in \mathcal{C}} \sum_{T \in T_{K}^*} \text{diam}(K) m_{T}|(\nabla v_{h}^{K})_{T}|^{2} \right)^{1/2},
\]

with size(\( \mathcal{T}_h \)) = \( \max_{K \in \mathcal{T}_h} \text{diam}(K) \).

By Remark 2.1 (a) and (4.4), we have

\[
\left( \sum_{K \in \mathcal{C}} \sum_{T \in T_{K}^*} m_{T}|(\nabla_{A} u_{h})_{T}|^{2} \right)^{1/2} \leq 2\|\nabla_{A} u_{h}\|_{(L^{2}(\Omega))^{2}} \leq \frac{2C_{22}}{\lambda} \|f\|_{L^{2}(\Omega)},
\]

and

\[
\left( \sum_{K \in \mathcal{C}} \sum_{T \in T_{K}^*} \text{diam}(K) m_{T}|(\nabla v_{h}^{K})_{T}|^{2} \right)^{1/2} \leq \rho_{6}\rho_{T},
\]

where \( \|\nabla v_{h}^{K}\|_{(L^{2}(\Omega))^{2}} = \sum_{T \in T_{K}^*} m_{T}|(\nabla v_{h}^{K})_{T}|^{2} \). Also there exist two positive constants \( \rho_{6} \) and \( \rho_{T} \), independent of \( h \) such that

\[
\|\nabla v_{h}^{K}\|_{(L^{2}(\Omega))^{2}} \leq \rho_{6} \quad \forall K \in \mathcal{C}, \quad \left( \sum_{K \in \mathcal{T}_h} \text{diam}(K) \right)^{1/2} \leq \rho_{T}.
\]

By (4.10)–(4.12), we obtain

\[
\left| \sum_{K \in \mathcal{C}} \text{diam}(K)A_{K}^{P \rightarrow \star}(u_h) \right| \leq \sqrt{\text{size}(\mathcal{T}_h)} \left( \frac{2\lambda C_{22} \|f\|_{L^{2}(\Omega)} \rho_{6}}{\lambda} \right) \rightarrow 0,
\]

as size(\( \mathcal{T}_h \)) \( \rightarrow 0 \). Considering \( P \equiv M \in \mathcal{C}^{*} \) a dual element \( M \in \mathcal{T}_h^{*} \), we proceed as above to obtain

\[
\left| \sum_{M \in \mathcal{C}^{*}} \text{diam}(M)A_{M}^{P \rightarrow \star}(u_h) \right| \leq \sqrt{\text{size}(\mathcal{T}_h^{*})} \left( \frac{\lambda C_{22} \|f\|_{L^{2}(\Omega)} \rho_{6}}{\lambda} \right) \rightarrow 0,
\]

as size(\( \mathcal{T}_h^{*} \)) \( \rightarrow 0 \). By (4.4),

\[
\left( \sum_{M \in \mathcal{C}^{*}} \sum_{T \in T_{M}^{*}} m_{T}|(\nabla_{A} u_{h})_{T}|^{2} \right)^{1/2} \leq \|\nabla_{A} u_{h}\|_{(L^{2}(\Omega))^{2}} \leq \frac{C_{22}}{\lambda} \|f\|_{L^{2}(\Omega)},
\]
\[ \left\| \nabla \Lambda v_h^M \right\|^2_{L^2(\Omega)} = \sum_{T \in \mathcal{T}_h^*} m_T \left| (\nabla \Lambda v_h^M)_T \right|^2, \]

there exist two positive constants \( \rho_8 \) and \( \rho_9 \), independent with \( h \) such that
\[ \left\| \nabla \Lambda v_h^M \right\|^2_{L^2(\Omega)} \leq \rho_8 \forall M \in C^*, \quad \left( \sum_{M \in \mathcal{T}_h^*} \text{diam}(M) \right)^{1/2} \leq \rho_9. \]

From (4.11), (4.15) and (4.16), it follows that (4.10) is satisfied. \( \square \)

5. Numerical tests

In this section, we carry out two numerical tests \([3]\) to verify the results for convergence and the DMP of the MNFECC scheme. To present this work, we need to introduce the following notation:

The algorithm (3.25) is stopped at iteration number \( n_i \), while the numerical solution \( u_h = (u_P)_{P \in \mathcal{N}^{**}} \) is \( u_h(n_i) \). The quantities \( u_{\text{min}} \) and \( u_{\text{max}} \) are defined as the minimum and maximum values of the approximate solution \( u_h \).

The relative error on the subdual mesh \( \mathcal{T}_h^{**} \) in \( L^2 \) of the MNFECC scheme which is
\[ \text{erl}^2_{\mathcal{T}_h^{**}} = \left( \frac{\sum_{T \in \mathcal{T}_h^{**}} \int_T |u_h - u|^2 \, dx}{\sum_{T \in \mathcal{T}_h^*} \int_T |u|^2 \, dx} \right)^{1/2}, \]

where \( u \) is the analytic solution. Its rate of convergence is expressed for each number of mesh \( i \geq 2 \), as
\[ \text{ratio}l^2_{\mathcal{T}_h^{**}} = -2 \frac{\log \left( \frac{\text{erl}^2_{\mathcal{T}_h^{**}}(i)}{\text{erl}^2_{\mathcal{T}_h^{**}}(i-1)} \right)}{\log \left( \frac{n_{u_{\mathcal{T}_h^{**}}}(i)}{n_{u_{\mathcal{T}_h^{**}}}(i-1)} \right)}, \]

where \( n_{u_{\mathcal{T}_h^{**}}} \) is number of unknowns in the linear system (3.23). Moreover, \( n_{u_{\mathcal{T}_h}} \) is number of unknowns in the linear system (3.25).

In the following two tests, the domain \( \Omega \) is partitioned by a uniform rectangular mesh.

5.1. Stationary analytical function. We begin by considering the following problem in order to estimate the convergence of the FECC and the MNFECC schemes,
\[ -\text{div}(\Lambda \nabla u) = f, \quad \text{in} \ \Omega = (0, 0.5) \times (0, 0.5), \]
\[ u(x, y) = \sin(\pi x) \sin(\pi y), \quad \text{on} \ \partial \Omega, \quad (5.1) \]

where
\[ \Lambda = \frac{1}{x^2 + y^2} \begin{pmatrix} x^2 + \alpha x^2 & -(1 - \alpha) xy \\ -(1 - \alpha) xy & x^2 + \alpha y^2 \end{pmatrix}, \quad \text{for} \,(x, y) \in \Omega, \]
\[ u(x, y) = \sin(\pi x) \sin(\pi y), \quad \text{for} \,(x, y) \in \Omega. \quad (5.2) \]

We see that the anisotropy ratio of \( \Lambda \) is \( 10^6 \). Moreover, the source term \( f \) which is computed from the exact solution \( u \) and the first equation of (5.1), is positive.

Table 1 shows the numerical results of the FECC and the MNFECC schemes (with \( \eta = 0.5, 1, 1.25 \)). We see that: (i) these schemes satisfy the DMP; (ii) for the FECC scheme, its rate of convergence is near to 2 and for the MNFECC scheme, its rate is close to 1 (this result is similar as ones of the nonlinear correction schemes
Table 1. Numerical results for (5.1).

<table>
<thead>
<tr>
<th>$\text{card}(T_h)$</th>
<th>16</th>
<th>64</th>
<th>256</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{card}(T_h^{**})$</td>
<td>80</td>
<td>288</td>
<td>1088</td>
<td>4224</td>
</tr>
<tr>
<td>$\nu_{T_h}$</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>1024</td>
</tr>
<tr>
<td>$\nu_{T_h^{**}}$</td>
<td>41</td>
<td>145</td>
<td>545</td>
<td>2113</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FECC scheme</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{erl}_2{T_h^{**}}$</td>
<td>8.999e-03</td>
<td>2.119e-03</td>
<td>5.491e-04</td>
<td>1.431e-04</td>
</tr>
<tr>
<td>$\text{ratio}_{l2}{T_h^{**}}$</td>
<td>2.29</td>
<td>2.04</td>
<td>1.98</td>
<td></td>
</tr>
<tr>
<td>$\text{umin}$</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MNFECC scheme with $\eta = 0.5$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{erl}_2{T_h^{**}}$</td>
<td>6.697e-02</td>
<td>3.265e-02</td>
<td>1.615e-02</td>
<td>8.200e-03</td>
</tr>
<tr>
<td>$\text{ratio}_{l2}{T_h^{**}}$</td>
<td>1.137</td>
<td>1.063</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>$\text{umin}$</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>$\text{nit}$</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MNFECC scheme with $\eta = 1$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{erl}_2{T_h^{**}}$</td>
<td>1.191e-01</td>
<td>6.323e-02</td>
<td>3.149e-02</td>
<td>1.604e-02</td>
</tr>
<tr>
<td>$\text{ratio}_{l2}{T_h^{**}}$</td>
<td>1.003</td>
<td>1.053</td>
<td>0.996</td>
<td></td>
</tr>
<tr>
<td>$\text{umin}$</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>$\text{nit}$</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MNFECC scheme with $\eta = 1.25$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{erl}_2{T_h^{**}}$</td>
<td>1.424e-01</td>
<td>7.813e-02</td>
<td>3.904e-02</td>
<td>1.985e-02</td>
</tr>
<tr>
<td>$\text{ratio}_{l2}{T_h^{**}}$</td>
<td>0.950</td>
<td>1.048</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>$\text{umin}$</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>$\text{nit}$</td>
<td>8</td>
<td>12</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

(iii) the FECC and the MNFECC schemes are more accurate than the original and the corrected schemes proposed by [3] in the same test and the same sizes of the primal mesh (see [3, Table 3]); (iv) for the MNFECC schemes, if the coefficient $\eta$ is smaller, the associated numerical results are more precise.

5.2. Stationary non analytical solution. To check the discrete maximum principle, the second test is proposed as

\[-\text{div}(\Lambda \nabla u) = f, \quad \text{in } \Omega = (0, 0.5) \times (0, 0.5),\]

\[u(x, y) = 0, \quad \text{on } \partial \Omega,\]

where the tensor $\Lambda$ is similar as (5.2), and the source term function is

\[f(x, y) = \begin{cases} 
10 & (x, y) \in (0.25, 0.5) \times (0.25, 0.5), \\
0 & \text{otherwise}.
\end{cases}\]

However, its solution is not determined.
Table 2. Numerical results for (5.3).

<table>
<thead>
<tr>
<th>$\text{card}(\mathcal{T}_h)$</th>
<th>16</th>
<th>64</th>
<th>256</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{card}(\mathcal{T}_h^{**})$</td>
<td>80</td>
<td>288</td>
<td>1088</td>
<td>4224</td>
</tr>
<tr>
<td>$\nu T_h$</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>1024</td>
</tr>
<tr>
<td>$\nu T_h^{**}$</td>
<td>41</td>
<td>145</td>
<td>545</td>
<td>2113</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scheme</th>
<th>$\text{umin}$</th>
<th>nit</th>
</tr>
</thead>
<tbody>
<tr>
<td>FECC scheme</td>
<td>-0.005138</td>
<td>9</td>
</tr>
<tr>
<td>MNFECC scheme with $\eta = 0.5$</td>
<td>-0.006901</td>
<td>9</td>
</tr>
<tr>
<td>MNFECC scheme with $\eta = 1$</td>
<td>-0.003449</td>
<td>9</td>
</tr>
<tr>
<td>MNFECC scheme with $\eta = 1.25$</td>
<td>-0.002069</td>
<td>10</td>
</tr>
<tr>
<td>MNFECC scheme with $\eta = 1.5$</td>
<td>-0.000802</td>
<td>9</td>
</tr>
<tr>
<td>MNFECC scheme with $\eta = 1$</td>
<td>-0.001681</td>
<td>9</td>
</tr>
<tr>
<td>MNFECC scheme with $\eta = 1.25$</td>
<td>-0.002147</td>
<td>9</td>
</tr>
<tr>
<td>MNFECC scheme with $\eta = 1.25$</td>
<td>-0.001705</td>
<td>10</td>
</tr>
</tbody>
</table>

For Test 5.3, its numerical results expressed in Table 5.2 verifies that the FECC and the MNFECC schemes (with $\eta = 0.5$, 1) violate the DMP. And the MNFECC scheme with $\eta = 1.25$ satisfies this principle, however its number of iteration steps is much larger than the other schemes.

Acknowledgments. This work is supported by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 101.02-2017.06. Part of this article was completed during a stay of the authors at the Vietnam Institute for Advanced Study in Mathematics (VIASM), whose hospitality is gratefully appreciated.

References


Nguyen Anh Dao
Applied Analysis Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam
Email address: daonguyenanh@tdtu.edu.vn

Duc Cam Hai Vo
Faculty of Mathematics and Computer Science, University of Science, VNU HCMC, Ho Chi Minh City, Vietnam
Email address: vdchai@hcmus.edu.vn

Thanh Hai Ong
Faculty of Mathematics and Computer Science, University of Science, VNU HCMC, Ho Chi Minh City, Vietnam
Email address: othai@hcmus.edu.vn