

A Nonlinear Finite Volume Element Method Satisfying Maximum Principle for Anisotropic Diffusion Problems on Arbitrary Triangular Meshes

Yanni Gao¹, Shuai Wang¹, Guangwei Yuan^{1,2} and Xudeng Hang^{1,*}

¹ Institute of Applied Physics and Computational Mathematics, Fenghaodong Road,
Haidian district, Beijing 100094, China.

² Laboratory of Computational Physics, P.O.Box 8009, Beijing, 100088, China.

Received 25 December 2017; Accepted (in revised version) 30 July 2018

Abstract. A nonlinear finite volume element scheme for anisotropic diffusion problems on general triangular meshes is proposed. Starting with a standard linear conforming finite volume element approximation, a corrective term with respect to the flux jumps across element boundaries is added to make the scheme satisfy the discrete maximum principle. The new scheme is free of the anisotropic non-obtuse angle condition which is a severe restriction on the grids for problems with anisotropic diffusion. Moreover, this manipulation can nearly keep the same accuracy as the original scheme. We prove the existence of the numerical solution for this nonlinear scheme theoretically. Numerical results and a grid convergence study are presented for both continuous and discontinuous anisotropic diffusion problems.

AMS subject classifications: 65N08, 65N12, 65N15

Key words: Finite volume element method, nonlinear correction, discrete maximum principle, anisotropic diffusion.

1 Introduction

We are concerned with the numerical solution of the diffusion equation:

$$-\nabla \cdot (\Lambda \nabla u) = f \quad \text{in } \Omega, \quad (1.1)$$

$$u = g \quad \text{on } \partial\Omega, \quad (1.2)$$

where

*Corresponding author. *Email addresses:* gaoyan10@163.com (Y. Gao), hang_xudeng@iapcm.ac.cn (X. Hang), yuan_guangwei@iapcm.ac.cn (G. Yuan), wang_shuai@iapcm.ac.cn (S. Wang)

- (a) Ω is an open bounded, convex connected polygonal domain in \mathbb{R}^2 with the boundary $\partial\Omega$;
- (b) f is the source term, belonging to $L^2(\Omega)$;
- (c) g is the Dirichlet boundary data defined on $\partial\Omega$;
- (d) Λ is a symmetric tensor such that Λ is piecewise Lipschitz-continuous on Ω and the set of eigenvalues of Λ is included in $[\lambda_{\min}, \lambda_{\max}]$ with $\lambda_{\min} > 0$.

The boundary value problem (BVP) (1.1)-(1.2) becomes an anisotropic diffusion problem if eigenvalues of Λ are not all equal at least on a portion of Ω . This kind of problem is a model arising in various fields such as plasma physics [19,33], petroleum reservoir simulation [17], and image processing [38]. As typical for diffusion problems, it satisfies the maximum principle

$$\min_{\mathbf{x} \in \Omega \cup \partial\Omega} u(\mathbf{x}) \geq \min_{\mathbf{x} \in \partial\Omega} g(\mathbf{x}) \quad (1.3)$$

provided that $f(\mathbf{x}) \geq 0$ holds for all $\mathbf{x} \in \Omega$. When using a standard numerical method, such as a finite element, a finite difference, or a finite volume method, to solve this problem, spurious oscillations may occur. In order to avoid such spurious oscillations in the numerical solution, a common strategy is to develop numerical schemes guaranteeing the discrete counterpart of (1.3), i.e., the so-called discrete maximum principle (DMP), which are known to produce numerical solutions evading nonphysical local oscillations or preserving positivity.

Development of DMP satisfaction schemes for solving diffusion problems has attracted considerable interest in the past. By virtue of the convex combination of two linear flux approximation and the positivity-preserving interpolation of the auxiliary unknowns, various cell-centered finite volume (FV) schemes circumventing spurious oscillations have been developed. These schemes usually have approximately a second-order accuracy on severely distorted meshes in the highly anisotropic, and/or discontinuous case. However, their extensions to finite element (FE) methods are hard to succeed. We refer readers to [1, 16, 18, 28, 30, 34, 39, 40] and references therein for more details. In the framework of FE methods, the study of DMP-preserving schemes for anisotropic diffusion case is more difficult and relevant results are very limited. In [27], the authors derive an anisotropic non-obtuse angle condition in term of the M-matrix criteria, such that the linear FE scheme guarantees the DMP by employing a suitable mesh. On the other hand, separating the stiffness matrix resulting from the FE discretization into diffusive and anti-diffusive fluxes and limiting the anti-diffusion fluxes by proper limiters lead to the so-called algebraic flux correction scheme. In [23] and [24], the authors propose two types of limiters to make the corrected schemes local extremum diminishing. But they are linearity-preserving only on symmetric meshes, as pointed out in [2].

Finite volume element (FVE) method [12,13,30,31], also called co-volume method [15] or generalized difference method [29], is one of the main numerical methods for solving

partial differential equations. This method has already attracted much attention because of its simple calculation and attractive computational properties, such as local conservation. Recently, many papers have been devoted to developing higher-order FVE schemes and establishing their error estimations, see [10, 11, 14, 18, 21, 41] and references therein. However, their construction and corresponding theoretical analysis are all executed for problems with scalar or even constant diffusion coefficients and mainly focus on the proofs of optimal convergence rates. However, development of FVE schemes satisfying the DMP or positivity preservation and keeping high accuracy simultaneously is far more difficult, especially for more complicated diffusion problems. Numerical results indicate the linear FVE solutions, even obtained on the uniform meshes, dissatisfy the DMP. In this article, we modify the original FVE scheme [29] by adding an extra nonlinear term which is composed of the flux jumps and tangential derivatives along element boundaries. The resultant scheme not only respects the DMP property, but also nearly keeps the numerical accuracy as the original scheme. This nonlinear technique is similar to the one employed in [4, 5]. However, the authors wherein are only concerned with Laplace operator or convective-dominated problems with small constant diffusion coefficients. Therefore, there are some differences in the construction of this nonlinear term. Firstly, the flux jump on each edge is taken as its positive part instead of its absolute value, which can keep numerical accuracy better. Besides, the nonlinear term includes additionally the integral of the tangential derivative, which is useful for the proof of the DMP.

The rest of this paper is organized as follows. In the next section, we introduce some necessary notations, assumptions and definitions. Section 3 is devoted to introducing the conforming linear FVE method on triangular grids and recalling relevant properties. Then, we present the nonlinear FVE method and analyse its solvability and validity of the DMP in Section 4. In Section 5, the performance of this method is illustrated on some numerical tests. Finally, we draw a brief conclusion in Section 6.

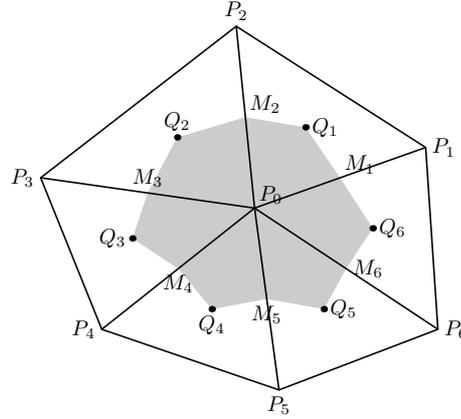
2 Preliminaries

In this paper, we employ standard definitions and notations of the Sobolev spaces. We use C to denote the positive constant independent of the finite element mesh parameter, and the symbol C may take on different values in different places.

Consider a family $\{\mathcal{T}_h\}_{h>0}$ of shape-regular triangulations of Ω consisting of disjoint triangular element K . Suppose triangulation \mathcal{T}_h is body-fitted if the diffusion tensor Λ is discontinuous across some interfaces. Define $h_K := \text{diam}(K)$, and $h = \max\{h_K : K \in \mathcal{T}_h\}$. We associate with the triangulation \mathcal{T}_h the finite element space

$$U_h = \{u_h \in H^1(\Omega) : u_h|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_h \text{ and } u_h|_{\partial\Omega} = g_h\},$$

where $\mathbb{P}_1(K)$ is the space of polynomials of degree at most 1 on K and g_h is the interpolation approximation of g . The nodes of \mathcal{T}_h are denoted by $\{P_i\}_{i=0}^N$, and the usual associated basis functions of U_h are denoted by $\{\varphi_i\}_{i=0}^N$, where N is a positive integer.

Figure 1: Dual element surrounding node P_0 .

Let \mathcal{E}_h and \mathcal{P}_h be the sets of the interior edges and nodes of \mathcal{T}_h , respectively. For each edge $E \in \mathcal{E}_h$, we define $h_E := |E|$ and $\omega_E := \{K \in \mathcal{T}_h : K \cap E \neq \emptyset\}$, and fix one unit tangent vector, denoted by \mathbf{t}_E . For an interior node $P_i \in \mathcal{P}_h$, we define the associated edges $\mathcal{E}_i := \{E \in \mathcal{E}_h : P_i \in E\}$ and let the subset Ω_i be the union of all elements K sharing the node P_i .

Next we construct the dual partition \mathcal{T}_h^* associated with \mathcal{T}_h . Fig. 1 shows that P_0 is an interior node, P_j ($j=1,2,\dots,6$) are neighboring nodes around P_0 , Q_j ($j=1,2,\dots,6$) are barycenter points of triangular elements, and M_j ($j=1,2,\dots,6$) are edge midpoints of primary elements. Connect M_j and Q_j ($j=1,2,\dots,6$) successively to form a polygonal region $K_{P_0}^*$ surrounding P_0 , called a dual element. If P_0 is located on $\partial\Omega$, the corresponding dual element $K_{P_0}^*$ reduces to a border polygon. Denote the set of all the dual elements by \mathcal{T}_h^* . We associate with the dual partition \mathcal{T}_h^* a piecewise constant space

$$V_h = \{v_h \in L^2(\Omega) : v_h|_{K_{P_i}^*} \text{ is a constant for } i=0,\dots,N\}.$$

$$V_h^0 = \{v_h \in V_h : v_h|_{K_{P_i}^*} = 0, \text{ if } P_i \text{ is a boundary node}\}.$$

Introduce a projection Π_h^* from U_h onto V_h , satisfying

$$\Pi_h^* u_h|_{K_{P_i}^*} = u_h(P_i), \quad \text{for } i=0,\dots,N. \quad (2.1)$$

Let S_Q and $S_{P_i}^*$ denote the areas of the element K_Q and the dual element $K_{P_i}^*$, respectively. Due to the regularity of triangulation \mathcal{T}_h , we have

$$C_1 h^2 \leq S_Q \leq h^2 \quad (2.2)$$

$$C_2 h^2 \leq S_{P_i}^* \leq C_3 h^2, \quad (2.3)$$

where $C_1, C_2,$ and C_3 are positive constants. For element $K_Q \in \mathcal{T}_h$ and $u_h \in U_h$, let P_i, P_j, P_k be the three nodes of K_Q , see Fig. 2, and u_i, u_j, u_k the values of u_h on the three nodes. Define

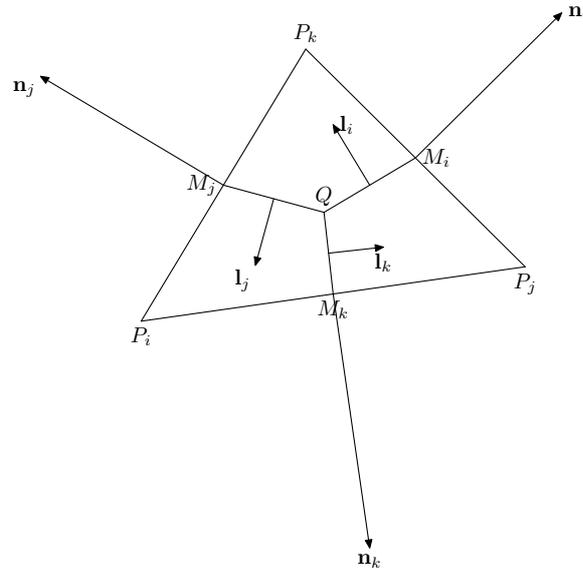


Figure 2: The notations on a triangular element.

the following discrete L^2 -norm and H^1 -seminorm

$$\begin{aligned} \|u_h\|_{0,h} &= \left(\sum_{K \in \mathcal{T}_h} |u_h|_{0,h,K}^2 \right)^{1/2}, \\ |u_h|_{1,h} &= \left(\sum_{K \in \mathcal{T}_h} |u_h|_{1,h,K}^2 \right)^{1/2}, \end{aligned}$$

where

$$\begin{aligned} |u_h|_{0,h,K} &= \left[\frac{1}{3} (u_i^2 + u_j^2 + u_k^2) S_Q \right]^{1/2}, \\ |u_h|_{1,h,K} &= \left\{ \left[\left(\frac{\partial u_h}{\partial x}(Q) \right)^2 + \left(\frac{\partial u_h}{\partial y}(Q) \right)^2 \right] S_Q \right\}^{1/2}. \end{aligned}$$

Obviously the discrete H^1 -seminorm $|\cdot|_{1,h}$ is equal to the continuous H^1 -seminorm $|\cdot|_1$ on the space U_h , because the gradient is constant on each element.

Lemma 2.1. *The discrete L^2 -norm $\|\cdot\|_{0,h}$ is equivalent to the continuous L^2 -norm $\|\cdot\|_0$ on the space U_h , i.e.*

$$C_1 \|u_h\|_{0,h} \leq \|u_h\|_0 \leq C_2 \|u_h\|_{0,h}, \quad \forall u_h \in U_h, \tag{2.4}$$

where C_1, C_2 are two positive constants independent of U_h .

Proof. For $u_h \in U_h$,

$$\|u_h\|_0^2 = \sum_{K \in \mathcal{T}_h} \int_K u_h^2 dx dy. \tag{2.5}$$

On each element $K \in \mathcal{T}_h$, u_h is a linear polynomial. Applying the numerical quadrature with second-order accuracy, we have

$$\begin{aligned} \int_K u_h^2 dx dy &= \frac{1}{3} (u_h(M_i)^2 + u_h(M_j)^2 + u_h(M_k)^2) S_Q \\ &= \frac{1}{6} (u_i^2 + u_j^2 + u_k^2 + u_i u_j + u_i u_k + u_k u_j) S_Q \\ &= \frac{1}{12} [u_i^2 + u_j^2 + u_k^2 + (u_i + u_k + u_j)^2] S_Q, \end{aligned}$$

where M_i, M_j, M_k are the midpoints of edges $P_j P_k, P_k P_i, P_i P_j$, respectively, see Fig. 2.

The above equality can be further estimated as

$$\frac{1}{4} |u_h|_{0,h,K}^2 \leq \int_K u_h^2 dx dy \leq |u_h|_{0,h,K}^2. \quad (2.6)$$

Substituting (2.6) into (2.5) yields the desired conclusion. \square

3 Linear FVE method

In this section, we present a symmetric FVE scheme under the triangular grid \mathcal{T}_h and analyse its relevant properties. For simplicity of exposition, we consider problems with homogenous boundary value condition. The discussion for non-homogenous case can be done analogously.

In the construction of FVE method, U_h and V_h^0 are taken as the trial space and test space, respectively. For any $v_h \in V_h^0$, multiply Eq. (1.1) by v_h and integrate the resultant equation on Ω . By Green's formula, we get

$$a(u, v_h) = (f, v_h), \quad (3.1)$$

where

$$a(u, v_h) = \sum_{P_i \in \mathcal{P}_h} -v_h(P_i) \int_{\partial K_{P_i}^*} (\Lambda \nabla u) \cdot \mathbf{n} ds. \quad (3.2)$$

Here and below, \mathbf{n} denotes the outward unit normal vector, with respect to its relevant edge. Generally speaking, the surface integrals in (3.2) cannot be carried out analytically and numerical quadrature is often necessary, especially when the diffusion tensor is heterogeneous. By employing appropriate numerical quadrature, we can get a symmetric and conservative FVE scheme. Let Λ_h be the piecewise constant interpolant of diffusion tensor Λ , satisfying

$$\Lambda_h|_{K_Q} = \Lambda(Q), \quad \forall K_Q \in \mathcal{T}_h, \quad (3.3)$$

where Q is the barycenter of element K_Q . Then, the desired FVE method is: Find $u_h \in U_h$ such that,

$$a_h(u_h, v_h) = (f, v_h), \quad \forall v_h \in V_h^0, \tag{3.4}$$

where

$$a_h(u_h, v_h) = \sum_{P_i \in \mathcal{P}_h} -v_h(P_i) \int_{\partial K_{P_i}^*} (\Lambda_h \nabla u_h) \cdot \mathbf{n} ds. \tag{3.5}$$

Due to $\Pi_h^* U_h = V_h^0$, the FVE method (3.4) is equivalent to: Find $u_h \in U_h$ such that,

$$a_h(u_h, \Pi_h^* v_h) = (f, \Pi_h^* v_h), \quad \forall v_h \in U_h. \tag{3.6}$$

Here and below, we view $a_h(\cdot, \Pi_h^* \cdot)$ as a bilinear form defined on $U_h \times U_h$ and use this equivalent problem to carry out analysis.

For a given interior node P_i , by taking $v_h = \varphi_i$ in (3.6), the equation associated with the dual element $K_{P_i}^*$ reduce to

$$a_h(u_h, \Pi_h^* \varphi_i) = (f, \Pi_h^* \varphi_i), \tag{3.7}$$

with

$$a_h(u_h, \Pi_h^* \varphi_i) = - \int_{\partial K_{P_i}^*} (\Lambda_h \nabla u_h) \cdot \mathbf{n} ds. \tag{3.8}$$

Take the interior point P_0 shown in Fig. 1 as an example to illustrate the calculation of the corresponding integral in detail. As shown in Fig. 1, the surface integral along $\partial K_{P_0}^*$ can be written as a summation of the integrals along the broken segments $\partial K_{P_0}^* \cap K_{Q_j}$ ($j=1, \dots, 6$), i.e.,

$$a_h(u_h, \Pi_h^* \varphi_0) = \sum_{j=1}^6 - \int_{\partial K_{P_0}^* \cap K_{Q_j}} (\Lambda_h \nabla u_h) \cdot \mathbf{n} ds. \tag{3.9}$$

For a given triangular element $\triangle P_0 P_1 P_2$, we denote the normal vectors to its edges by $\mathbf{n}_i, \mathbf{l}_i$ ($i=0,1,2$), as shown in Fig. 3. We assume that these vectors have the same length with their corresponding edges. Then, the following identities hold:

$$\mathbf{n}_0 + \mathbf{n}_1 + \mathbf{n}_2 = 0, \tag{3.10}$$

$$\mathbf{l}_0 + \mathbf{l}_1 + \mathbf{l}_2 = 0. \tag{3.11}$$

On element $\triangle P_0 P_1 P_2$, the gradient of u_h is formulated as

$$\nabla u_h = -\frac{u_h(P_0)}{2S_{Q_1}} \mathbf{n}_0 - \frac{u_h(P_1)}{2S_{Q_1}} \mathbf{n}_1 - \frac{u_h(P_2)}{2S_{Q_1}} \mathbf{n}_2,$$

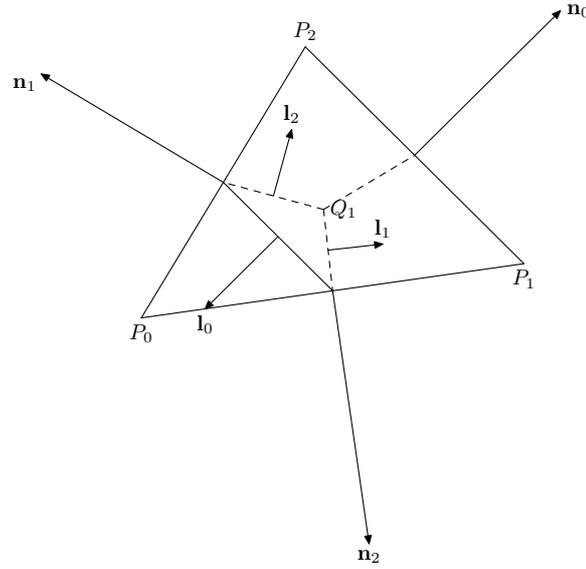


Figure 3: The notations for the dual cells of the triangle.

where S_{Q_1} is the area of $\triangle P_0P_1P_2$. By virtue of (3.10), we have

$$\nabla u_h = \frac{u_h(P_0) - u_h(P_1)}{2S_{Q_1}} \mathbf{n}_1 + \frac{u_h(P_0) - u_h(P_2)}{2S_{Q_1}} \mathbf{n}_2. \quad (3.12)$$

Through simple calculation, we get

$$\begin{aligned} - \int_{\partial K_{P_0}^* \cap K_{Q_1}} (\Lambda_h \nabla u_h) \cdot \mathbf{n} ds &= - \int_{M_1 Q_1} (\Lambda_h \nabla u_h) \cdot \mathbf{n} ds - \int_{Q_1 M_2} (\Lambda_h \nabla u_h) \cdot \mathbf{n} ds \\ &= - (\Lambda_h \nabla u_h) \cdot \mathbf{l}_1 - (\Lambda_h \nabla u_h) \cdot \mathbf{l}_2 \\ &= (\Lambda_h \nabla u_h) \cdot \mathbf{l}_0 \\ &= \frac{-1}{4S_{Q_1}} (\Lambda(Q_1) \mathbf{n}_1) \cdot \mathbf{n}_0 (u_h(P_0) - u_h(P_1)) \\ &\quad + \frac{-1}{4S_{Q_1}} (\Lambda(Q_1) \mathbf{n}_2) \cdot \mathbf{n}_0 (u_h(P_0) - u_h(P_2)), \end{aligned} \quad (3.13)$$

where (3.11) and the fact $\mathbf{l}_0 = -\frac{\mathbf{n}_0}{2}$ have been used in the third and last equalities, respectively. The surface integrals on the other elements are manipulated in a similar way.

If generating equation on the dual element $K_{P_1}^*$, from the symmetry of displacement we know that the coefficient associated to $u_h(P_0)$ is $\frac{1}{4S_{Q_1}} (\Lambda(Q_1) \mathbf{n}_0) \cdot \mathbf{n}_1$. Due to the symmetry of $\Lambda(Q_1)$, $\frac{1}{4S_{Q_1}} (\Lambda(Q_1) \mathbf{n}_0) \cdot \mathbf{n}_1$ is equal to $\frac{1}{4S_{Q_1}} (\Lambda(Q_1) \mathbf{n}_1) \cdot \mathbf{n}_0$ which is the coefficient associated to $u_h(P_1)$ in (3.13). Therefore, we conclude that the FVE scheme (3.6) is symmetric.

Let U be the vector consisting of unknown variables. Denote the linear system resulting from FVE scheme (3.6) by

$$AU = F, \quad (3.14)$$

where the global stiffness matrix A and the right-hand side F are defined as

$$A = \begin{pmatrix} A_\Omega & A_\Gamma \\ 0 & I \end{pmatrix}, \quad F = \begin{pmatrix} F_\Omega \\ g \end{pmatrix}. \quad (3.15)$$

The subscripts Ω and Γ refer to row/column numbers corresponding to the interior and boundary nodes, respectively. According to the discussion provided in [23], the discrete extremum principle holds if the matrix A satisfies

- diagonal coefficients are strictly positive,
- off-diagonal coefficients are non-positive,
- the row sums associated to interior nodes are equal to zero.

In (3.13), we know that the row sum corresponding to interior node P_0 is equal to zero and the positive definiteness of diffusion tensor Λ guarantees the positiveness of the diagonal coefficient. However, the off-diagonal coefficients are likely to go against the second criteria. Analysing the coefficients presented in (3.13), we could obtain the following conclusion:

1. For a special case $\Lambda = I$, the requirement for the linear FVE scheme (3.6) to satisfy DMP is that the triangulation \mathcal{T}_h satisfies the non-obtuse angle condition, i.e. the dihedral angles of mesh elements are non-obtuse.
2. For the anisotropic case, the restriction on the meshes become so-called anisotropic non-obtuse angle condition which requires the dihedral angles of mesh elements to be non-obtuse, when measured in a metric depending on Λ [27].

The above restrictions on the meshes are usually too severe to satisfy in practice, especially for the anisotropic case. In order to circumvent these angle conditions, we propose a nonlinear FVE scheme which guarantees the discrete maximum principle on both distorted and uniform meshes for anisotropic diffusion problems.

4 Nonlinear corrected FVE method

In this section, we modify the FVE method (3.6) by adding a nonlinear term. This nonlinear technique makes the new scheme guarantee the DMP on general meshes.

For $u_h \in U_h$ and an interior edge $E = K_{Q^+} \cap K_{Q^-}$, where K_{Q^+} and K_{Q^-} are two distinct elements of \mathcal{T}_h with respective unit outwards normals \mathbf{n}^+ and \mathbf{n}^- , we introduce

$$\delta_E(u_h) = \max\{0, h_E [(\Lambda(Q^+) \nabla u_h|_{K_{Q^+}}) \cdot \mathbf{n}^+ + (\Lambda(Q^-) \nabla u_h|_{K_{Q^-}}) \cdot \mathbf{n}^-]\}. \quad (4.1)$$

Define the following sign function:

$$\text{sign}(x) = \frac{x}{|x|} = \begin{cases} 1, & x > 0; \\ -1, & x \leq 0. \end{cases} \quad (4.2)$$

For $u_h, v_h \in U_h$, and an interior edge $E \in \mathcal{E}_h$, the nonlinear corrective term involving tangential derivatives along element edges and the jumps of the diffusion flux across interfaces is defined as follows:

$$j(u_h, v_h) = C' \sum_{E \in \mathcal{E}_h} \delta_E(u_h) \psi_E(u_h, v_h) + C'' \sum_{E \in \mathcal{E}_h} h_E^2 (\nabla u_h \cdot \mathbf{t}_E, \nabla v_h \cdot \mathbf{t}_E)_E, \quad (4.3)$$

where

$$\psi_E(u_h, v_h) = h_E \text{sign}(h_E \nabla u_h \cdot \mathbf{t}_E) \nabla v_h \cdot \mathbf{t}_E, \quad (4.4)$$

$(\cdot, \cdot)_E$ and C', C'' denote the scalar product on edge E and positive constants, respectively.

The corrected FVE approximation to (1.1)-(1.2) is: Find $u_h \in U_h$, such that

$$a_h(u_h, \Pi_h^* v_h) + j(u_h, v_h) = (f, \Pi_h^* v_h), \quad \forall v_h \in U_h. \quad (4.5)$$

The specific values of the constants C' and C'' would be determined in the subsequent analysis.

It is known that the standard FVE methods are local conservative on the edges of dual elements. In fact, the corrected term $j(u_h, v_h)$ in (4.5) does not break the conservation respected by the original FVE scheme. Take the nodes P_0, P_1 and $E = P_0 P_1$ shown in Figure 1 as an example to illustrate it. In (4.5), applying $v_h = \varphi_0$ and $v_h = \varphi_1$ as the test functions respectively, we obtain the corresponding corrected terms

$$j(u_h, \varphi_0) = C' \sum_{E \in \mathcal{E}_0} \delta_E(u_h) \psi_E(u_h, \varphi_0) + C'' \sum_{E \in \mathcal{E}_0} h_E^2 (\nabla u_h \cdot \mathbf{t}_E, \nabla \varphi_0 \cdot \mathbf{t}_E)_E, \quad (4.6)$$

$$j(u_h, \varphi_1) = C' \sum_{E \in \mathcal{E}_1} \delta_E(u_h) \psi_E(u_h, \varphi_1) + C'' \sum_{E \in \mathcal{E}_1} h_E^2 (\nabla u_h \cdot \mathbf{t}_E, \nabla \varphi_1 \cdot \mathbf{t}_E)_E. \quad (4.7)$$

On the edge $E = P_0 P_1$ which belongs to $\mathcal{E}_0 \cap \mathcal{E}_1$, the unit tangent vector \mathbf{t}_E is fixed. It is easily to observe that

$$\nabla \varphi_0 \cdot \mathbf{t}_E = -\nabla \varphi_1 \cdot \mathbf{t}_E.$$

Accordingly, from the definitions of $\psi_E(u_h, \varphi_0)$ and $\psi_E(u_h, \varphi_1)$, we have

$$\psi_E(u_h, \varphi_0) = -\psi_E(u_h, \varphi_1).$$

Hence, on the common edge $E = P_0 P_1$, $C' \delta_E(u_h) \psi_E(u_h, \varphi_0) + C'' h_E^2 (\nabla u_h \cdot \mathbf{t}_E, \nabla \varphi_0 \cdot \mathbf{t}_E)_E = -C' \delta_E(u_h) \psi_E(u_h, \varphi_1) - C'' h_E^2 (\nabla u_h \cdot \mathbf{t}_E, \nabla \varphi_1 \cdot \mathbf{t}_E)_E$, which implies that the corrected FVE scheme is still local conservative on the edges of dual elements.

4.1 Existence of a discrete solution

Theorem 4.1. *The nonlinear problem (4.5) admits at least one solution.*

Proof. Let ε be a given positive constant. For any interior edge $E \in \mathcal{E}_h$, define the functional

$$\phi_{E,\varepsilon}: U \in U_h \mapsto h_E \frac{\delta_E(U)}{|h_E \nabla U \cdot \mathbf{t}_E| + \varepsilon} \in \mathbb{R}.$$

For any $\tilde{U} \in U_h$, consider the following regularized problem: Find $u_h \in U_h$, such that

$$A_h(\tilde{U}; u_h, \Pi_h^* v_h) = (f, \Pi_h^* v_h), \quad \forall v_h \in U_h, \tag{4.8}$$

where

$$A_h(\tilde{U}; u_h, \Pi_h^* v_h) = a_h(u_h, \Pi_h^* v_h) + \sum_{E \in \mathcal{E}_h} (C' \phi_{E,\varepsilon}(\tilde{U}) + C'' h_E^2) (\nabla u_h \cdot \mathbf{t}_E, \nabla v_h \cdot \mathbf{t}_E)_E. \tag{4.9}$$

Define an operator $T_\varepsilon: \tilde{U} \in U_h \mapsto u_h \in U_h$, where u_h is the solution of problem (4.8).

The bilinear form $a_h(\cdot, \Pi_h^* \cdot)$ is coercive and bounded on $U_h \times U_h$, i.e.

$$a_h(u_h, \Pi_h^* u_h) \geq C \|u_h\|_1^2, \quad \forall u_h \in U_h, \tag{4.10}$$

$$a_h(u_h, \Pi_h^* v_h) \leq C \|u_h\|_1 \|v_h\|_1, \quad \forall u_h, v_h \in U_h. \tag{4.11}$$

In [29], the authors have given a detailed proof of (4.10) and (4.11). Readers are referred to the appendix where a more concise proof is provided. Then, it is obviously that

$$A_h(u_h, \Pi_h^* u_h) \geq C \|u_h\|_1^2, \quad \forall u_h \in U_h. \tag{4.12}$$

Since the regularity (2.2), we get that

$$\begin{aligned} \sum_{E \in \mathcal{E}_h} h_E (\nabla u_h \cdot \mathbf{t}_E, \nabla v_h \cdot \mathbf{t}_E)_E &\leq C \|\nabla u_h\|_0 \|\nabla v_h\|_0 \\ &\leq C \|u_h\|_1 \|v_h\|_1. \end{aligned} \tag{4.13}$$

Then, combining (4.11) and (4.13) yields

$$A_h(u_h, \Pi_h^* u_h) \leq C \|u_h\|_1 \|v_h\|_1.$$

Therefore, the regularised problem (4.8) has a unique solution owing to the Lax-Milgram Lemma. Therefore, the operator T_ε is well defined.

Taking $v_h = u_h$ in (4.8) and using (4.12) lead to

$$\|u_h\|_1^2 \leq C (f, \Pi_h^* u_h). \tag{4.14}$$

By simple calculation and Lemma 2.1, we have

$$\begin{aligned} \|\Pi_h^* u_h\|_0 &= \|u_h\|_{0,h} \\ &\leq C \|u_h\|_0 \\ &\leq C \|u_h\|_1. \end{aligned} \quad (4.15)$$

By virtue of Cauchy-Schwarz inequality and (4.15), we further have

$$\|u_h\|_1 \leq C \|f\|_0. \quad (4.16)$$

Therefore, if \tilde{U} is such that $\|\tilde{U}\|_1 \leq C \|f\|_0$, $u_h = T_\varepsilon \tilde{U}$ will also be in this ball. Next we prove that the operator T_ε is continuous on this ball

For any given $\tilde{U}_1, \tilde{U}_2 \in U_h$, let $u_{h,1} = T_\varepsilon \tilde{U}_1, u_{h,2} = T_\varepsilon \tilde{U}_2$. Subtracting (4.8) for $u_{h,1}$ from (4.8) for $u_{h,2}$ and testing with $v_h = u_{h,1} - u_{h,2}$ yield

$$\begin{aligned} a_h(u_{h,1} - u_{h,2}, \Pi_h^*(u_{h,1} - u_{h,2})) + \sum_{E \in \mathcal{E}_h} \{C' [\phi_{E,\varepsilon}(\tilde{U}_1)(\nabla u_{h,1} \cdot \mathbf{t}_E, \nabla(u_{h,1} - u_{h,2}) \cdot \mathbf{t}_E)_E \\ - \phi_{E,\varepsilon}(\tilde{U}_2)(\nabla u_{h,2} \cdot \mathbf{t}_E, \nabla(u_{h,1} - u_{h,2}) \cdot \mathbf{t}_E)_E] + C'' h_E^2 \|\nabla(u_{h,1} - u_{h,2})\|_{L^2(E)}^2\} = 0. \end{aligned}$$

Above equality can be further formulated as

$$\begin{aligned} a_h(u_{h,1} - u_{h,2}, \Pi_h^*(u_{h,1} - u_{h,2})) + \sum_{E \in \mathcal{E}_h} (C' \phi_{E,\varepsilon}(\tilde{U}_1) + C'' h_E^2) \|\nabla(u_{h,1} - u_{h,2})\|_{L^2(E)}^2 \\ = C' \sum_{E \in \mathcal{E}_h} (\phi_{E,\varepsilon}(\tilde{U}_2) - \phi_{E,\varepsilon}(\tilde{U}_1)) (\nabla u_{h,2} \cdot \mathbf{t}_E, \nabla(u_{h,1} - u_{h,2}) \cdot \mathbf{t}_E)_E. \end{aligned} \quad (4.17)$$

From the coercivity of $a_h(\cdot, \Pi_h^* \cdot)$, it holds that the left-hand side of this equation has a lower bound $C \|u_{h,1} - u_{h,2}\|_1^2$. Furthermore, for the term $\phi_{E,\varepsilon}(\tilde{U}_2) - \phi_{E,\varepsilon}(\tilde{U}_1)$, we have following estimation

$$\begin{aligned} \phi_{E,\varepsilon}(\tilde{U}_2) - \phi_{E,\varepsilon}(\tilde{U}_1) &= h_E \frac{\delta_E(\tilde{U}_2) - \delta_E(\tilde{U}_1)}{|h_E \nabla \tilde{U}_2 \cdot \mathbf{t}_E| + \varepsilon} \\ &\quad + h_E \delta_E(\tilde{U}_1) \frac{|h_E \nabla \tilde{U}_1 \cdot \mathbf{t}_E| - |h_E \nabla \tilde{U}_2 \cdot \mathbf{t}_E|}{(|h_E \nabla \tilde{U}_2 \cdot \mathbf{t}_E| + \varepsilon)(|h_E \nabla \tilde{U}_1 \cdot \mathbf{t}_E| + \varepsilon)} \\ &\leq h_E \frac{\delta_E(\tilde{U}_2 - \tilde{U}_1)}{\varepsilon} + h_E \delta_E(\tilde{U}_1) \frac{|h_E \nabla(\tilde{U}_1 - \tilde{U}_2) \cdot \mathbf{t}_E|}{\varepsilon^2} \\ &\leq Ch_E (1 + \|\tilde{U}_1\|_{H^1(\omega_E)}) \|\nabla(\tilde{U}_1 - \tilde{U}_2)\|_{L^2(\omega_E)}, \end{aligned}$$

where the regularity of \mathcal{T}_h has been used in the last inequality. Again from the Cauchy-Schwarz inequality and regularity of \mathcal{T}_h , it holds that

$$h_E (\nabla u_{h,2} \cdot \mathbf{t}_E, \nabla(u_{h,1} - u_{h,2}) \cdot \mathbf{t}_E)_E \leq C \|u_{h,2}\|_{H^1(\omega_E)} \|\nabla(u_{h,1} - u_{h,2})\|_{L^2(\omega_E)}.$$

Then, the right-hand side of (4.17) satisfies

$$\begin{aligned} & C' \sum_{E \in \mathcal{E}_h} (\phi_{E,\varepsilon}(\tilde{U}_2) - \phi_{E,\varepsilon}(\tilde{U}_1)) (\nabla u_{h,2} \cdot \mathbf{t}_E, \nabla(u_{h,1} - u_{h,2}) \cdot \mathbf{t}_E)_E \\ & \leq C \sum_{E \in \mathcal{E}_h} \|u_{h,2}\|_{H^1(\omega_E)} (1 + \|\tilde{U}_1\|_{H^1(\omega_E)}) \|\nabla(\tilde{U}_1 - \tilde{U}_2)\|_{L^2(\omega_E)} \|\nabla(u_{h,1} - u_{h,2})\|_{L^2(\omega_E)}. \end{aligned}$$

Since

$$\|u_{h,2}\|_{H^1(\omega_E)} \leq C \|f\|_0, \quad \|\tilde{U}_1\|_{H^1(\omega_E)} \leq C \|f\|_0,$$

we further have

$$\begin{aligned} & C' \sum_{E \in \mathcal{E}_h} (\phi_{E,\varepsilon}(\tilde{U}_2) - \phi_{E,\varepsilon}(\tilde{U}_1)) (\nabla u_{h,2} \cdot \mathbf{t}_E, \nabla(u_{h,1} - u_{h,2}) \cdot \mathbf{t}_E)_E \\ & \leq C(1 + \|f\|_0) \|f\|_0 \sum_{E \in \mathcal{E}_h} \|\nabla(\tilde{U}_1 - \tilde{U}_2)\|_{L^2(\omega_E)} \|\nabla(u_{h,1} - u_{h,2})\|_{L^2(\omega_E)} \\ & \leq C(1 + \|f\|_0) \|f\|_0 \|\nabla(\tilde{U}_1 - \tilde{U}_2)\|_0 \|\nabla(u_{h,1} - u_{h,2})\|_0 \\ & \leq C(1 + \|f\|_0) \|f\|_0 \|\tilde{U}_1 - \tilde{U}_2\|_1 \|u_{h,1} - u_{h,2}\|_1, \end{aligned}$$

where the discrete Cauchy inequality has been used in the second inequality. Combining above analysis yields

$$\begin{aligned} \|T_\varepsilon \tilde{U}_1 - T_\varepsilon \tilde{U}_2\|_1 &= \|u_{h,1} - u_{h,2}\|_1 \\ &\leq C(1 + \|f\|_0) \|f\|_0 \|\tilde{U}_1 - \tilde{U}_2\|_1. \end{aligned}$$

Therefore, the operator T_ε is continuous.

Due to Brouwer's Theorem, the operator T_ε admits a fixed point, denoted by $u_{h,\varepsilon}$ in the ball of radius $C \|f\|_0$ in U_h . Since the sequence $\{u_{h,\varepsilon}\}_\varepsilon$ is in a finite-dimensional ball, there is a subsequence, still denoted by $\{u_{h,\varepsilon}\}_\varepsilon$, such that $u_{h,\varepsilon} \rightarrow u_h$ as $\varepsilon \rightarrow 0$. In (4.8) taking the limit $\varepsilon \rightarrow 0$, we could conclude that u_h solves (4.5). \square

4.2 Discrete maximum principle

Theorem 4.2. *If $C' \geq \frac{1}{2}$, $C'' > 0$ and $f \geq 0$, then the solution u_h of the nonlinear finite volume element scheme (4.5) reaches its minimum on the boundary $\partial\Omega$.*

Proof. Assume $f \geq 0$ and the corresponding solution u_h reaches its minimum at the interior node P_0 , as shown in Fig. 1.

Take $v_h = \varphi_0$ in (4.5), we have

$$a_h(u_h, \Pi_h^* \varphi_0) + j(u_h, \varphi_0) = (f, \Pi_h^* \varphi_0). \tag{4.18}$$

Here, $a_h(u_h, \Pi_h^* \varphi_0)$ and $j(u_h, \varphi_0)$ reduce to

$$\begin{aligned} a_h(u_h, \Pi_h^* \varphi_0) &= \sum_{i=1}^6 - \int_{M_i M_{i+1}} (\Lambda(Q_i) \nabla u_h) \cdot \mathbf{n} ds, \\ j(u_h, \varphi_0) &= C' \sum_{E \in \mathcal{E}_0} \delta_E(u_h) \psi_E(u_h, \varphi_0) + C'' \sum_{E \in \mathcal{E}_0} h_E^2 (\nabla u_h \cdot \mathbf{t}_E, \nabla \varphi_0 \cdot \mathbf{t}_E)_E, \end{aligned}$$

where $M_7 = M_1$. On element K_{Q_i} ($i = 1, \dots, 6$), $\Lambda(Q_i) \nabla u_h$ is a constant vector. Then applying Green's formula on each element leads to

$$\begin{aligned} a_h(u_h, \Pi_h^* \varphi_0) &= \sum_{i=1}^6 \int_{P_0 M_i} (\Lambda(Q_i) \nabla u_h) \cdot \mathbf{n} ds + \int_{P_0 M_{i+1}} (\Lambda(Q_i) \nabla u_h) \cdot \mathbf{n} ds \\ &= \sum_{i=1}^6 \frac{|P_0 P_i|}{2} [(\Lambda(Q_i) \nabla u_h) \cdot \mathbf{n}_i - (\Lambda(Q_{i-1}) \nabla u_h) \cdot \mathbf{n}_i], \end{aligned}$$

where $Q_0 = Q_6$, and \mathbf{n}_i is a fixed unit normal vector with respect to edge $P_0 P_i$. From the definition of $\delta_E(u_h)$, it holds that

$$a_h(u_h, \Pi_h^* \varphi_0) \leq \sum_{E \in \mathcal{E}_0} \frac{\delta_E(u_h)}{2}. \quad (4.19)$$

Moreover, for $E = P_0 P_i$, owing to $u_h(P_0) \leq u_h(P_i)$ ($i = 1, \dots, 6$), we have

$$\begin{aligned} \psi_E(u_h, \varphi_0) &= h_E \text{sign}(h_E \nabla u_h \cdot \mathbf{t}_E) \nabla \varphi_0 \cdot \mathbf{t}_E \\ &= \text{sign}(u_h(P_0) - u_h(P_i)) \\ &= -1. \end{aligned}$$

Then, the corrective term $j(u_h, \varphi_0)$ can be formulated as

$$j(u_h, \varphi_0) = -C' \sum_{E \in \mathcal{E}_0} \delta_E(u_h) + C'' \sum_{i=1}^6 |P_0 P_i| (u_h(P_0) - u_h(P_i)). \quad (4.20)$$

Substituting (4.19) and (4.20) into (4.18), we get

$$(f, \Pi_h^* \varphi_0) \leq \left(\frac{1}{2} - C'\right) \sum_{E \in \mathcal{E}_0} \delta_E(u_h) + C'' \sum_{i=1}^6 |P_0 P_i| (u_h(P_0) - u_h(P_i)). \quad (4.21)$$

Due to $f \geq 0$ and $\Pi_h^* \varphi_0 = 1$ on dual element $K_{P_0}^*$, the left-hand side of (4.21) is non-negative. If the constant $C' \geq \frac{1}{2}$, $C'' \geq 0$, the right-hand side of (4.21) is non-positive. Therefore,

$$\left(\frac{1}{2} - C'\right) \sum_{E \in \mathcal{E}_0} \delta_E(u_h) + C'' \sum_{i=1}^6 |P_0 P_i| (u_h(P_0) - u_h(P_i)) = 0.$$

Since $(\frac{1}{2} - C') \sum_{E \in \mathcal{E}_0} \delta_E(u_h)$ and $\sum_{i=1}^6 |P_0 P_i| (u_h(P_0) - u_h(P_i))$ are both non-positive, we get

$$\sum_{i=1}^6 |P_0 P_i| (u_h(P_0) - u_h(P_i)) = 0.$$

The above equality indicates that $u_h(P_0) = u_h(P_i)$ ($i = 1, \dots, 6$). We can infer that u_h is a constant on the whole domain Ω , if the minimum is reached on an interior node. It could be eventually deduced that the minimum is reached on the boundary. \square

Remark 4.1. In this theorem, the requirement $C' \geq \frac{1}{2}$ is a sufficient condition for the non-linear scheme (4.5) to respect DMP. In the numerical solution, the value of C' could be smaller than $\frac{1}{2}$. Besides, it could be observed that the proof is valid, only if C'' is positive. Hence, we take $C'' = h^\alpha$ ($\alpha \geq 1$), which would not affect the accuracy.

Remark 4.2. Still take the node P_0 as an example to estimate the magnitude of the corrected term $j(u_h, \varphi_0)$. First, the set \mathcal{E}_0 is finite. If $C'' = h^\alpha$, we easily have that $C'' \sum_{E \in \mathcal{E}_0} h_E^2 (\nabla u_h \cdot \mathbf{t}_E, \nabla \varphi_0 \cdot \mathbf{t}_E)_E$ is $\mathcal{O}(h^{\alpha+2})$. On the other hand, on each $E \in \mathcal{E}_0$, the flux jump $\delta_E(u_h) \psi_E(u_h, \varphi_0)$ is equivalent to $h_E \int_E [\Lambda \frac{\partial u_h}{\partial \mathbf{n}}] \frac{\partial \varphi_0}{\partial \mathbf{t}_E} ds$, which has the same magnitude with $h_E \int_E [\Lambda \frac{\partial u_h}{\partial \mathbf{n}}] [\frac{\partial \varphi_0}{\partial \mathbf{n}}] ds$. Here, $[\cdot]$ denotes the usual jump on the edge E , as that in Discontinuous Galerkin methods. For conforming finite element spaces, $h_E \int_E [\Lambda \frac{\partial u_h}{\partial \mathbf{n}}] [\frac{\partial \varphi_0}{\partial \mathbf{n}}] ds$ can keep the second-order accuracy. Therefore, the corrected term $j(u_h, \varphi_0)$ is $\mathcal{O}(h^2)$.

5 Numerical examples

In this section, we examine the numerical performance of the corrected FVE method proposed in this article. We report some numerical results yielded by both the standard FVE method and the corrected FVE method.

The nonlinear system resulted from discrete equations is solved by the fixed-point or Picard iteration combined with Anderson Acceleration method [37]. Let U be the vector of discrete unknowns, M denote a matrix corresponding to the first term in the nonlinear scheme (4.5), and $J(U)$ and F be the vectors corresponding to the corrected term and source term. Then, the nonlinear system resulted from (4.5) is

$$MU + J(U) = F. \tag{5.1}$$

Choose a small positive value ε_{non} and initial vector U^0 . The Anderson Acceleration of the Picard iteration employed in this paper is as follows:

1. Solve the system (5.1) by Picard iteration to get U^1, U^2 and set $\tilde{U}^k = U^k, \Delta U^k = \tilde{U}^k - U^{k-1}, k = 1, 2;$
2. For $k = 2, \dots$ do

Determine weights α_1, α_2 by solving the minimization problem

$$\min \|\alpha_1 \Delta U^{k-1} + \alpha_2 \Delta U^k\|$$

subjected to the constraint

$$\alpha_1 + \alpha_2 = 1;$$

3. Set new iterate

$$U^{k+1} = \alpha_1 \tilde{U}^{k-1} + \alpha_2 \tilde{U}^k;$$

4. Stop if $\|MU^{k+1} + J(U^{k+1}) - F\| \leq \varepsilon_{non} \|MU^0 + J(U^0) - F\|$, otherwise solve the system (5.1) by U^{k+1} to obtain \tilde{U}^{k+1} and set $\Delta U^{k+1} = \tilde{U}^{k+1} - U^{k+1}$.

In this paper, the prescribed tolerance is 10^{-10} . Numerical tests indicate that on uniform meshes the Picard iteration converges as fast as that with Anderson Acceleration, but on distorted meshes the number of nonlinear iteration can be reduced significantly by Anderson Acceleration (nearly by half).

We investigate the following discrete L^∞ and L^2 errors for the solutions u_h and u :

$$E_\infty(u - u_h) = \max_{P_i \in \mathcal{P}_h} |u(P_i) - u_h(P_i)|,$$

$$E_2(u - u_h) = \left(\sum_{K_Q \in \mathcal{T}_h} S_Q |u(Q) - u_h(Q)|^2 \right)^{1/2},$$

where Q is the barycenter of element K_Q . The rate of convergence is obtained by the following formula

$$Rate = \frac{\log[E(h_2)/E(h_1)]}{\log(h_2/h_1)},$$

where h_1, h_2 denote the mesh sizes of the two successive meshes, and $E(h_1), E(h_2)$ the corresponding discrete errors.

Example 5.1. We consider the following strong heterogeneous and anisotropic problem:

$$-\nabla \cdot (\Lambda(\mathbf{x})) \nabla u = f(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad (5.2)$$

$$u = g(\mathbf{x}), \quad \mathbf{x} \in \partial\Omega, \quad (5.3)$$

where the domain $\Omega = [0,1] \times [0,1]$, and diffusion tensor $\Lambda(\mathbf{x})$, source term $f(\mathbf{x})$ and boundary value condition $g(\mathbf{x})$ are respectively defined as:

$$\Lambda(\mathbf{x}) = \begin{bmatrix} \alpha x^2 + y^2 & (\alpha - 1)xy \\ (\alpha - 1)xy & x^2 + \alpha y^2 \end{bmatrix}, \quad \alpha = 100,$$

$$f(\mathbf{x}) = 0,$$

$$g(x, y) = \begin{cases} 2, & \text{on } x=0 \text{ or } x=1, \\ 0, & \text{else.} \end{cases}$$

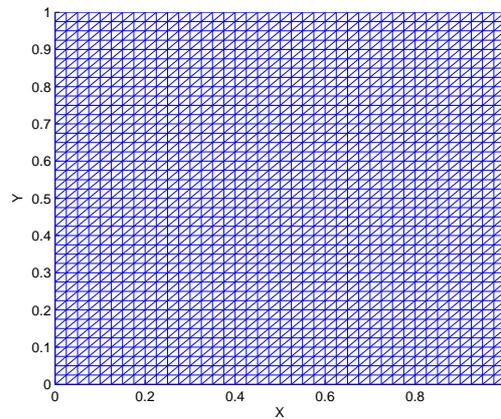


Figure 4: Uniform grid.

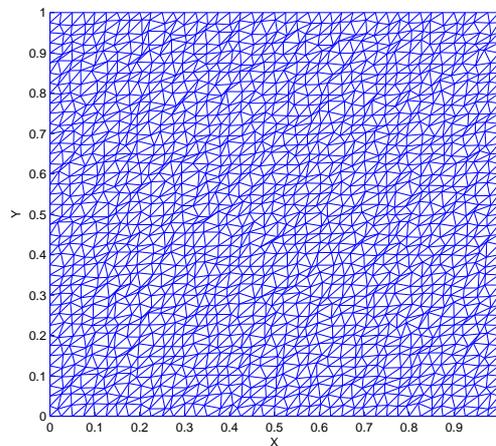


Figure 5: Random distorted grid.

In Example 5.1, the two constants C' and C'' in scheme (4.5) are taken as $\frac{1}{80}$ and h , respectively. We employ the uniform mesh and its random distorted counterpart as the computational grids, shown in Fig. 4 and Fig. 5, respectively. These meshes no longer satisfy the anisotropic non-obtuse angle condition. Numerical results for the standard FVE method on these two kinds of grids are presented in Fig. 6 and Fig. 7, respectively. The counterparts for the corrected FVE method are presented in Fig. 8 and Fig. 9. We note that the results produced by these two FVE methods look alike but the standard FVE solutions exhibit undershoots and overshoots on both meshes. Whereas the corrected solutions are within the range $[0,2]$ of admissible values, which indicates that this non-linear FVE scheme guarantees the DMP even if a mesh dissatisfying the angle condition is employed.

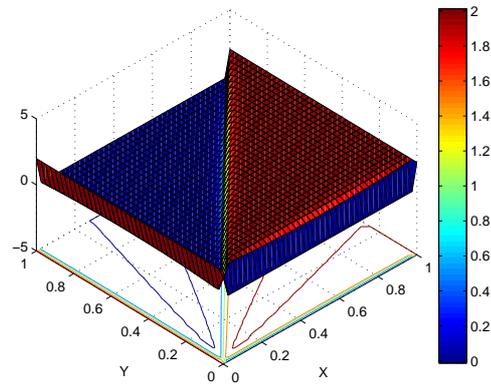


Figure 6: Solution profile for the standard FVE method on the uniform grid, maximum=2.0146, minimum= $-1.5491e-3$.

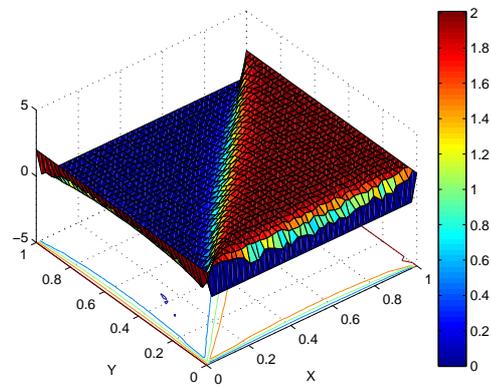


Figure 7: Solution profile for the standard FVE method on the random distorted grid, maximum=2.0072, minimum= $-7.5300e-3$.

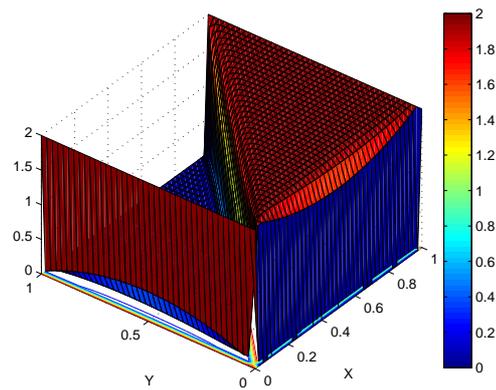


Figure 8: Solution profile for the corrected FVE method on the uniform grid, maximum=1.9999, minimum= $4.3808e-6$.

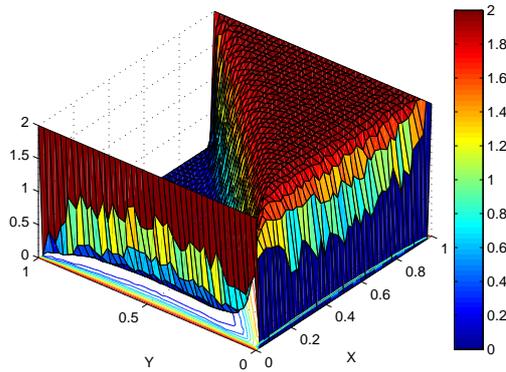


Figure 9: Solution profile for the corrected FVE method on the random distorted grid, maximum=1.9996, minimum= $3.2532e-4$.

Example 5.2. This example is used to examine the convergence of this new scheme. The computational domain Ω , diffusion tensor $\Lambda(\mathbf{x})$ are the same as those employed in Example 5.1. The exact solution is chosen as:

$$u(x,y) = \sin(\pi x)\sin(\pi y).$$

In Example 5.2, the two constants C' and C'' in scheme (4.5) are taken as $\frac{1}{100}$ and h , respectively. We still take uniform triangular meshes and their random distorted counterparts to carry out the solution. Numerical errors and convergence rates are presented in Tables 1-4. We can observe that these two methods have the same convergence and the numerical errors in both norms are nearly comparable. From Table 3 and Table 7 in next example we note that the convergence rate in the discrete L^∞ norm descends noticeably, which seems to be less compatible with the results in [30,31]. That is because the generation of the refined distorted meshes is different. In fact, the refined distorted meshes employed in [30,31] are obtained by symmetric refinement of quadrilaterals via bisection on edges. Therefore the resultant meshes get better gradually, such that the corresponding numerical solutions could keep nearly optimal convergence.

Table 1: Standard FVE method on uniform mesh for Example 5.2.

$1/h$	E_2	Rate	E_∞	Rate
5	2.9245E-2	-	9.4906E-2	-
10	8.4787E-3	1.7863	2.5937E-2	1.8715
20	2.5332E-3	1.7429	6.6571E-3	1.9620
40	7.4304E-4	1.7694	1.8639E-3	1.8366
80	2.0983E-4	1.8242	5.6116E-4	1.7318
160	5.7260E-5	1.8736	1.7440E-4	1.6860

Table 2: Corrected FVE method on uniform mesh for Example 5.2.

$1/h$	E_2	Rate	E_∞	Rate
5	3.0260E-2	-	6.9536E-2	-
10	8.8141E-3	1.7795	2.1267E-2	1.7091
20	2.6296E-3	1.7450	5.7289E-3	1.8923
40	7.7214E-4	1.7679	1.7832E-3	1.6838
80	2.1825E-4	1.8229	5.9890E-4	1.5741
160	5.9593E-5	1.8728	1.8434E-4	1.6999

Table 3: Standard FVE method on random distorted mesh for Example 5.2.

$1/h$	E_2	Rate	E_∞	Rate
5	3.4903E-2	-	1.3669E-1	-
10	1.0233E-2	1.7701	2.9359E-2	2.2190
20	3.0009E-3	1.7698	7.6865E-3	1.9334
40	8.6505E-4	1.7945	2.6944E-3	1.5124
80	2.1177E-4	2.0303	9.4951E-4	1.5047
160	5.4310E-5	1.9632	2.2687E-4	2.0653

Table 4: Corrected FVE method on random distorted mesh for Example 5.2.

$1/h$	E_2	Rate	E_∞	Rate
5	3.4988E-2	-	1.2793E-1	-
10	1.0346E-2	1.7578	2.8213E-2	2.1809
20	3.0516E-3	1.7614	7.8994E-3	1.8924
40	8.7475E-4	1.8026	2.7236E-3	1.4804
80	2.1482E-4	2.0257	9.5045E-4	1.5188
160	5.5175E-5	1.9610	2.2700E-4	2.0659

Example 5.3. In this example, we consider a discontinuous diffusion tensor to examine the performance of this new scheme. We still apply the same computational domain and meshes as those employed in above two examples. The diffusion tensor $\Lambda(\mathbf{x})$ and exact solution u are respectively defined as:

$$\Lambda(\mathbf{x}) = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} & \text{if } x < 0.5, \\ \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix} & \text{if } x \geq 0.5, \end{cases}$$

$$u(x, y) = \begin{cases} x^2 y^3 + \cos(xy), & \text{if } x < 0.5, \\ \left(\frac{2x+99}{200}\right)^2 y^3 + \cos\left(\frac{(2x+99)y}{200}\right), & \text{if } x \geq 0.5. \end{cases}$$

In Example 5.3, the two constants C' and C'' in scheme (4.5) are still taken as $\frac{1}{100}$ and h , respectively. The numerical results are reported in Tables 5-8. From these numerical results, we note that optimal convergence is observed in discrete L^2 norm on both uniform

Table 5: Standard FVE method on uniform mesh for Example 5.3.

$1/h$	E_2	Rate	E_∞	Rate
10	9.6548E-4	-	2.3174E-4	-
20	2.4215E-4	1.9953	6.0136E-5	1.9462
40	6.0588E-5	1.9988	1.5127E-5	1.9911
80	1.5150E-5	1.9997	3.7949E-6	1.9950
160	3.7877E-6	1.9999	9.4910E-7	1.9994

Table 6: Corrected FVE method on uniform mesh for Example 5.3.

$1/h$	E_2	Rate	E_∞	Rate
10	1.0721E-3	-	4.0566E-4	-
20	2.7066E-4	1.9859	1.0768E-4	1.9135
40	6.7844E-5	1.9962	2.7092E-5	1.9908
80	1.6972E-5	1.9991	6.8089E-6	1.9924
160	4.2439E-6	1.9997	1.7026E-6	1.9997

Table 7: Standard FVE method on random distorted mesh for Example 5.3.

$1/h$	E_2	Rate	E_∞	Rate
10	1.1778E-3	-	2.9499E-3	-
20	3.1770E-4	1.8904	8.8539E-4	1.7363
40	8.2298E-5	1.9487	2.9659E-4	1.5778
80	2.5972E-5	1.6639	1.0208E-4	1.5388
160	6.1177E-6	2.0859	3.0250E-5	1.7547

Table 8: Corrected FVE method on random distorted mesh for Example 5.3.

$1/h$	E_2	Rate	E_∞	Rate
10	1.2895E-3	-	3.0189E-3	-
20	3.4613E-4	1.8974	8.9063E-4	1.7611
40	8.9402E-5	1.9529	2.9328E-4	1.6025
80	2.6600E-5	1.7489	1.0116E-4	1.5356
160	6.5291E-6	2.0265	3.0304E-5	1.7391

and distorted meshes. The convergence in the discrete L^∞ norm still appear descendance in Tables 6 and 8, but the nonlinear technique nearly does not destroy the numerical accuracy holding by the original scheme.

6 Conclusion

In this paper, we have developed a corrected FVE scheme for anisotropic problems on distorted triangular grids. We verify this numerical scheme holds the DMP, simultaneously without enforcing anisotropic non-obtuse angle condition to the computational meshes. The numerical results provided in last section not only confirm this property, but also indicate this numerical scheme keeps nearly the same numerical accuracy as the original FVE method.

Acknowledgments

This work was partially supported by the Postdoctoral Science Foundation of China (No. 2017M620689), and the National Science Foundation of China (Nos. 11571048 and 11401034), the CAEP developing fund of science and technology (No. 2014A0202009).

Appendix

This section is devoted to proving the validity of (4.10) and (4.11).

Proof. For any $u_h, v_h \in U_h$, rewrite the bilinear form $a_h(u_h, \Pi_h^* v_h)$ as

$$\begin{aligned} a_h(u_h, \Pi_h^* v_h) &= \sum_{i=0}^N -v_h(P_i) \int_{\partial K_{P_i}^*} (\Lambda_h \nabla u_h) \cdot \mathbf{n} ds \\ &= \sum_{K \in \mathcal{T}_h} I_K(u_h, \Pi_h^* v_h). \end{aligned} \quad (\text{A.1})$$

Take the element shown in Fig. 2 as an example, suppose the normal vectors $\mathbf{l}_i, \mathbf{l}_j, \mathbf{l}_k$ and $\mathbf{n}_i, \mathbf{n}_j, \mathbf{n}_k$ have the same length as their corresponding edges. Then, the specific formulation of $I_K(u_h, \Pi_h^* v_h)$ is

$$\begin{aligned} I_K(u_h, \Pi_h^* v_h) &= (\Lambda(Q) \nabla u_h) \cdot \mathbf{l}_i (v_h(P_k) - v_h(P_j)) \\ &\quad + (\Lambda(Q) \nabla u_h) \cdot \mathbf{l}_j (v_h(P_i) - v_h(P_k)) + (\Lambda(Q) \nabla u_h) \cdot \mathbf{l}_k (v_h(P_j) - v_h(P_i)) \\ &= (\Lambda(Q) \nabla u_h) \cdot \mathbf{l}_i (v_h(P_k) - v_h(P_i)) + (\Lambda(Q) \nabla u_h) \cdot \mathbf{l}_i (v_h(P_i) - v_h(P_j)) \\ &\quad + (\Lambda(Q) \nabla u_h) \cdot \mathbf{l}_j (v_h(P_i) - v_h(P_k)) + (\Lambda(Q) \nabla u_h) \cdot \mathbf{l}_k (v_h(P_j) - v_h(P_i)) \\ &= (\Lambda(Q) \nabla u_h) \cdot \frac{\mathbf{n}_k}{2} (v_h(P_i) - v_h(P_k)) + (\Lambda(Q) \nabla u_h) \cdot \frac{\mathbf{n}_j}{2} (v_h(P_i) - v_h(P_j)) \\ &= S_Q (\Lambda(Q) \nabla u_h) \cdot \left[\frac{\mathbf{n}_k}{2S_Q} (v_h(P_i) - v_h(P_k)) + \frac{\mathbf{n}_j}{2S_Q} (v_h(P_i) - v_h(P_j)) \right] \\ &= S_Q (\Lambda(Q) \nabla u_h) \cdot \nabla v_h, \end{aligned}$$

where in the last two equalities we have used that $\mathbf{l}_j - \mathbf{l}_i = \frac{\mathbf{n}_k}{2}$, $\mathbf{l}_i - \mathbf{l}_k = \frac{\mathbf{n}_j}{2}$ and $\nabla v_h = \frac{\mathbf{n}_k}{2S_Q}(v_h(P_i) - v_h(P_k)) + \frac{\mathbf{n}_j}{2S_Q}(v_h(P_i) - v_h(P_j))$, respectively.

Because Λ is bounded, we first obtained that

$$I_K(u_h, \Pi_h^* v_h) \leq C |u_h|_{1,h,K} |v_h|_{1,h,K}. \quad (\text{A.2})$$

On the other hand, if taking $v_h = u_h$, it follows that from positive definiteness of Λ

$$I_K(u_h, \Pi_h^* u_h) \geq C |u_h|_{1,h,K}^2. \quad (\text{A.3})$$

Substituting (A.2) and (A.3) into (A.1) respectively, we could get

$$\begin{aligned} a_h(u_h, \Pi_h^* v_h) &\leq C |u_h|_{1,h} |v_h|_{1,h} \\ &\leq C |u_h|_1 |v_h|_1, \\ a_h(u_h, \Pi_h^* v_h) &\geq C |u_h|_{1,h}^2 \\ &\geq C |u_h|_1^2. \end{aligned}$$

By the equivalence of the semi-norm $|\cdot|$ and norm $\|\cdot\|_1$ in the space H_0^1 , we could obtain the desired conclusion. \square

References

- [1] L. Agélas, G. Enchéry, B. Flemisch, M. Schneider, Convergence of nonlinear finite volume schemes for heterogeneous anisotropic diffusion on general meshes, *J. Comput. Phys.*, **351**, 80-107 (2017).
- [2] G.R. Barrenechea, V. John, P. Knobloch, An algebraic flux correction scheme satisfying the discrete maximum principle and linearity preservation on general meshes, *Math. Models Methods Appl. Sci.*, **27**(3), 525-548 (2017).
- [3] E. Burman, A. Ern, Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection-diffusion-reaction equation, *Comput. Methods Appl. Mech. Engrg.*, **191**(35), 3833-3855 (2002).
- [4] E. Burman, A. Ern, Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes, *Comptes Rendus Mathématique*, **338**(8), 641-646 (2004).
- [5] E. Burman, A. Ern, Stabilized galerkin approximation of convection-diffusion-reaction equation: discrete maximum principle and convergence, *Math. Comp.*, **74**(252), 1637-1652 (2005).
- [6] E. Burman, P. Hansbo, Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems, *Comput. Methods Appl. Mech. Engrg.*, **193**, 1437-1453 (2004).
- [7] E. Burman, Stabilised finite element methods for ill-posed problems with conditional stability, *Building Bridges: Connections and Challenges in Modern Approaches to Numerical Partial Differential Equations*, **14**, 93-127 (2016).
- [8] G.R. Barrenechea, E. Burman, F. Karakatsani, Edge-based nonlinear diffusion for finite element approximations of convection-diffusion equations and its relation to algebraic flux-correction schemes, *Numer. Math.*, **135**, 521-545 (2017).
- [9] L. Chen, A new class of high order finite volume methods for second order elliptic equations. *SIAM Numer. Anal.* **47**, 4021-4043 (2010).

- [10] Z.Y. Chen, Y.S. Xu, Y.Y. Zhang, A construction of higher-order finite volume methods, *Math. Comp.*, **84**(292), 599-628 (2015).
- [11] Z.Y. Chen, Y.S. Xu, J.H. Zhang, A second-order hybrid finite volume method for solving the Stokes equation, *Appl. Numer. Math.*, **119**, 213-224 (2017).
- [12] Z.Y. Chen, R.H. Li, A.H. Zhou, A note on the optimal L^2 -estimate of the finite volume element method, *Adv. Comput. Math.*, **16**(4), 291-303 (2002).
- [13] Z.Q. Cai, On the finite volume element method, *Numer. Math.*, **58**, 713-735 (1991).
- [14] Z. Chen, J. Wu, Y. Xu, Higher-order finite volume methods for elliptic boundary value problems, *Adv. Comput. Math.*, **37**(2), 191-253 (2013).
- [15] A. Draganescu, T.F. Dupont, L.R. Scott, Failure of the discrete maximum principle for an elliptic finite element problem, *Math. Comp.*, **74**, 1-23 (2005).
- [16] J. Droniou, C. Le Potier, Construction and convergence study of schemes preserving the elliptic local maximum principle, *SIAM J. Numer. Anal.*, **49**(2), 459-490 (2011).
- [17] T. Ertekin, J.H. Abou-Kassem, G.R. King, *Basic Applied Reservoir Simulation*, SPE textbook series, Vol. 7, Richardson, Texas, 2001.
- [18] Z.M. Gao, J.M. Wu, A small stencil and extremum-preserving scheme for anisotropic diffusion problems on arbitrary 2D and 3D meshes, *J. Comput. Phys.*, **250**, 308-331 (2013).
- [19] S. Gunter, K. Lackner, A mixed implicit-explicit finite difference scheme for heat transport in magnetised plasmas, *J. Comput. Phys.*, **228**, 282-293 (2009).
- [20] J.L. Guermond, M. Nazarov, A maximum-principle preserving C^0 finite element method for scalar conservation equations, *Comput. Methods Appl. Mech. Eng.*, **272**, 198-213 (2014).
- [21] W.M. He, Z.M. Zhang, Q.S. Zou, Maximum-norms error estimates for high-order finite volume schemes over quadrilateral meshes, *Numer. Math.*, **138**(2), 473-500 (2018).
- [22] C. Kreuzer, A note on why enforcing discrete maximum principles by a simple posteriori cutoff is a good idea, *Numer. Meth. Part. Differ. Equ.*, **30**(3), 994-1002 (2014).
- [23] D. Kuzmin, M.J. Shashkov, D. Svyatskiy, A constrained finite element method satisfying the discrete maximum principle for anisotropic diffusion problems on arbitrary meshes, *J. Comput. Phys.*, **228**(9), 3448-3463 (2009).
- [24] D. Kuzmin, M. Moller, Algebraic flux correction I. Scalar conservation laws, in: D. Kuzmin, R. Lohner, S. Turek (Eds.), *Flux-Corrected Transport: Principles, Algorithms, and Applications*, Springer, Berlin, 155-206 (2005).
- [25] C.N. Lu, W.Z. Huang, E.S. Vleck, The cutoff method for the numerical computation of nonnegative solutions of parabolic PDEs with application to anisotropic diffusion and lubrication-type equations, *J. Comput. Phys.*, **242**, 24-36 (2013).
- [26] R. Liska, M. Shashkov, Enforcing the discrete maximum principle for linear finite element solutions of second-order elliptic problems, *Comm. Comput. Phys.*, **3**, 852-877 (2008).
- [27] X.P. Li, W.Z. Huang, An anisotropic mesh adaptation method for the finite element solution of heterogeneous anisotropic diffusion problems, *J. Comput. Phys.*, **229**, 8072-8094 (2010).
- [28] C. Le Potier, A nonlinear finite volume scheme satisfying maximum and minimum principles for diffusion operators, *Int. J. Finite Vol.*, **6**(20), 2009.
- [29] R.H. Li, Z.Y. Chen, W. Wu, *Generalized difference methods for differential equations: numerical analysis of finite volume methods*, Marcel Dekker, New York, 2000.
- [30] J.L. Lv, Y.H. Li, Optimal biquadratic finite volume element methods on quadrilateral meshes, *SIAM J. Numer. Anal.*, **50**, 2379-2399 (2012).
- [31] J.L. Lv, Y.H. Li, L^2 error estimate of the finite volume element methods on quadrilateral meshes, *Adv. Comput. Math.*, **33**, 129-148 (2010).
- [32] R.A. Nicolaides, T.A. Porsching, C.A. Hall, *Covolume methods in computational fluid dy-*

- namics. In: Hafez, M., Oshima, K. (eds.) Computational Fluid Dynamics Review, Wiley, New York, 279-299 (1995).
- [33] K. Nishikawa, M. Wakatani, Plasma Physics, Springer-Verlag Berlin Heidelberg, New York, 2000.
- [34] Z.Q. Sheng, G.W. Yuan, The finite volume scheme preserving extremum principle for diffusion equations on polygonal meshes, J. Comput. Phys., **230**(7), 2588-2604 (2011).
- [35] G. Stoyan, On maximum principles for monotone matrices, Lin. Alg. Appl., **78**, 147-161 (1986).
- [36] R.S. Varga, On a discrete maximum principle, SIAM J. Numer. Anal., **3**, 355-359 (1966).
- [37] H.F. Walker, P. Ni, Anderson acceleration for fixed-point iterations, SIAM J. Numer. Anal., **49**(4), 1715-1735 (2011).
- [38] J. Weickert, Anisotropic diffusion in image processing, Teubner-Verlag, Stuttgart, Germany, 1998.
- [39] Y.L. Yu, G.W. Yuan, A conservative finite volume scheme preserving maximum principle for diffusion equations, International Conference of Numerical Analysis and Applied Mathematics, 2017.
- [40] G.W. Yuan, Y.L. Yu, Existence of solution of a finite volume scheme preserving maximum principle for diffusion equations, Numer. Meth. Part. Differ. Equ., **34**(1), 80-96 (2018).
- [41] Z.M. Zhang, Q.S. Zou, Vertex-centered finite volume schemes of any order over quadrilateral meshes for elliptic boundary value problems, Numer. Math., **130**, 363-393 (2015).