Monotone finite volume schemes for diffusion equations on polygonal meshes ☆

Guangwei Yuan, Zhiqiang Sheng *

Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics,
P.O. Box 8009, Beijing 100088, China

Abstract

We construct a nonlinear finite volume (FV) scheme for diffusion equation on star-shaped polygonal meshes and prove that the scheme is monotone, i.e., it preserves positivity of analytical solutions for strongly anisotropic and heterogeneous full tensor coefficients. Our scheme has only cell-centered unknowns, and it treats material discontinuities rigorously and offers an explicit expression for the normal flux. Numerical results are presented to show how our scheme works for preserving positivity on various distorted meshes for both smooth and non-smooth highly anisotropic solutions. And numerical results show that our scheme appears to be approximate second-order accuracy for the solution and first-order accuracy for the flux.
© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Accurate and reliable discretization methods are very important for the numerical simulations of Lagrangian hydrodynamics. Development of a new discrete scheme for diffusion equation should satisfy some desirable properties [11], specifically the scheme must

– be locally conservative, or ensure the continuity of the normal flux through cell interfaces;
– be monotone, i.e., preserve positivity of the differential solution or satisfy the discrete maximum principle;
– be reliable on unstructured anisotropic meshes that may be severely distorted;
– allow heterogeneous full diffusion tensors;
– result in a sparse system with minimal number of non-zero entries;

– have the accuracy that is higher than the first order for smooth solutions;
– have only cell-centered unknowns.

Monotonicity is one of the key requirements to discretization schemes. In the context of anisotropic thermal conduction, the scheme without preserving monotonicity can lead to the violation of the entropy constraints of the second law of thermodynamics, causing heat to flow from regions of lower temperature to higher temperature. In regions of large temperature variations, this can cause the temperature to become negative. In order to avoid negative temperature, the scheme must be monotone. For the linear cases, the monotonicity is equivalent with the discrete maximum principle. However, for general cases, the discrete maximum principle is more restrictive than monotonicity.

It is well known that classical finite volume (FV) and finite element (FE) schemes fail to satisfy the discrete maximum principle for strong anisotropic diffusion tensors and on distorted meshes [7,9,15]. To our knowledge, a linear scheme satisfying all the above requirements is not known. There are several different linear schemes [1,2,4,5,8,10,13,14,16,18] satisfying one or more requirements above, but not all of them. For example, some schemes must impose severe restrictions on the geometry of meshes or diffusion coefficients in order to satisfy the monotonicity or the discrete maximum principle. Recently, based on repair technique and constrained optimization, two approaches have been suggested to enforce discrete maximum principle for linear finite element solutions of general elliptic equations with self-adjoint operator on triangular meshes in [12].

The criteria for the monotonicity of control volume methods on quadrilateral meshes was derived in [15], and it was shown that it is impossible to construct linear nine-point methods which unconditionally satisfy the monotonicity criteria when the discretization satisfies local conservation and exact reproduction of linear solution.

On the other hand, a few nonlinear schemes [6,11,17] have been proposed to guarantee monotonicity. A nonlinear stabilized Galerkin approximation of the Laplace operator has been analyzed in [6] and a nonlinear FV scheme for highly anisotropic diffusion operators on unstructured triangular meshes has been proposed in [17]. It was shown in [17] that the scheme is monotone only for parabolic equations and sufficiently small time steps. The nonlinear FV scheme suggested in [17] has been further developed and analyzed for elliptic problems in [11], which satisfies the above requirements on triangular meshes. They proved that the scheme is monotone on triangular meshes for strongly anisotropic and heterogeneous full tensor coefficients, with a special choice of collocation points (i.e., cell centers). Moreover, they did not propose a monotone scheme for anisotropic and heterogeneous full tensor coefficients on general polygonal meshes.

In this paper we will further develop the nonlinear monotone FV schemes, and construct a nonlinear FV scheme with monotonicity for anisotropic and heterogeneous full tensor coefficients on polygonal meshes. We will propose an adaptive strategy of constructing discrete flux to guarantee monotonicity on polygonal meshes. The basic idea is to choose appropriate cell-edge in the derivation of discrete flux expression according to mesh geometry. Compared with [11], our scheme is monotone for strongly anisotropic and heterogeneous full tensor coefficients on polygonal meshes, and need not use a specific definition of collocation points. For our scheme, we can simply take collocation points as the cell centers which are defined in Lagrangian hydrodynamics algorithm for polygonal meshes. It follows that our scheme avoids a remap from the values on collocation points to those on cell centers, and would be suitable for coupled radiation diffusion/hydrodynamics calculations on such meshes. Moreover an alternative interpolation technique is used and compared with other techniques in [11,21].

The remainder of this article is organized as follows. In Section 2 we describe the construction of the nonlinear FV scheme and then prove it is monotone. In Section 3 we extend this scheme to non-stationary diffusion equations. Then in Section 4 we present some numerical results to illustrate the features of the scheme. Finally some conclusions are given in Section 5.

2. Construction of monotone nonlinear scheme

2.1. Problem and notation

Consider the stationary diffusion problem for unknown $u = u(x)$:

$$-\nabla \cdot (\kappa \nabla u) = f \quad \text{in } \Omega, \tag{2.1}$$

$$u(x) = g \quad \text{on } \partial\Omega, \tag{2.2}$$

where Ω is an open bounded polygonal set of R^2 with boundary $\partial\Omega$, and κ is diffusion tensor (possibly anisotropic).

In this paper, we use a mesh on Ω made up of polygons and denote the cell by K or L. And with each cell K we associate one point (the so-called collocation point or cell center) denoted also by K: the centroid is a qualified candidate but other points can be chosen.

We assume that each polygon is star-shaped with respect to the collocation point, that is any ray emanating from the cell center K intersects the boundary of cell K at exactly one point. Note that any convex polygon satisfies the assumption.

Denote the cell vertex by A, B or $P_1, P_2, P_3, P_4, \ldots$, and the cell side by σ (see Fig. 2.1). If the cell side σ is a common edge of cells K and L, and its vertices are A and B, then we denote

$$\sigma = K|L = BA.$$

Let $\mathcal{J}$ be the set of all cells, $\mathcal{E}$ be the set of all cell side, and $\mathcal{E}_K$ be the set of all cell side of cell K. Denote $\mathcal{E}_{\text{int}} = \mathcal{E} \cap \Omega$, $\mathcal{E}_{\text{ext}} = \mathcal{E} \cap \partial\Omega$. Denote $h = (\sup_{K \in \mathcal{J}} m(K))^{1/2}$, where $m(K)$ is the area of cell K.

We adopt the following notations (see Fig. 2.1). $\vec{n}_{K\sigma}$ (resp. $\vec{n}_{L\sigma}$) is the unit outer normal on the edge σ of cell K (resp. L). There holds $\vec{n}_{K\sigma} = -\vec{n}_{L\sigma}$ for $\sigma = K \mid L$. $\vec{t}_{KP_i}$ and $\vec{t}_{LP_i}$ are the unit tangential vectors on the line KP_i and LP_i $(i = 1, 2, \ldots)$, respectively.

Let κ^{T} be the transpose of matrix κ. The ray originated in the point K along the direction $\kappa^{\text{T}}\vec{n}_{K\sigma}$ must intersect one of the cell-side of cell K, and this cell-side is denoted by P_1P_2, and the cross point is O_1. Similarly, the ray originated in the point L along the direction $\kappa^{\text{T}}\vec{n}_{L\sigma}$ must intersect one of the cell-side of cell L, and we denote this cell-side by P_3P_4, and the cross point is O_2. Let θ_{K_1} be the angle between KP_1 and KO_1, θ_{K_2} be the angle between KO_1 and KP_2, θ_{L_1} be the angle between LP_4 and LO_2, and θ_{L_2} be the angle between LO_2 and LP_3. Denote $\theta_K = \theta_{K_1} + \theta_{K_2}$, i.e., θ_K is the angle between KP_1 and KP_2, and $\theta_L = \theta_{L_1} + \theta_{L_2}$, i.e., θ_L is the angle between LP_3 and LP_4. Notice that the polygon is star-shaped, then the three point K, P_1 and P_2 can form a triangle, θ_K is an internal angle of the triangle KP_1P_2. Similar, θ_L is an internal angle of the triangle LP_3P_4. Hence, there are

$$0 \leqslant \theta_{K_1}, \theta_{K_2}, \theta_{L_1}, \theta_{L_2} < \pi,$$

and

$$0 < \theta_K, \theta_L < \pi.$$

2.2. Construction of scheme

Integrate (2.1) over the cell K, to obtain

$$\sum_{\sigma \in \mathcal{E}_K} \mathcal{F}_{K,\sigma} = \int_K f(x) \mathrm{d}x, \tag{2.3}$$
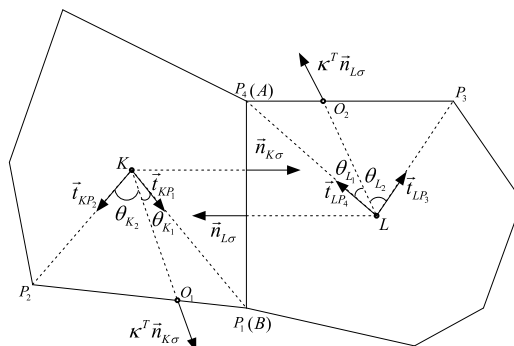

Fig. 2.1. Stencil and notation.

where the continuous flux on the edge σ is

$$\mathcal{F}_{K,\sigma} = -\int_\sigma \kappa(x)\nabla u(x) \cdot \vec{n}_{K\sigma} \mathrm{d}l. \tag{2.4}$$

Noticing that

$$(\kappa\nabla u) \cdot v = \nabla u \cdot (\kappa^{\mathrm{T}} v),$$

we have

$$\mathcal{F}_{K,\sigma} = -\int_\sigma \nabla u(x) \cdot \kappa(x)^{\mathrm{T}} \vec{n}_{K\sigma} \mathrm{d}l. \tag{2.5}$$

Since KP_1 and KP_2 are two edges of the triangle KP_1P_2, the two vectors $\vec{t}_{KP_1}$ and $\vec{t}_{KP_2}$ cannot be collinear (see Fig. 2.1). Then there is

$$\frac{\kappa^{\mathrm{T}}\vec{n}_{K\sigma}}{|\kappa^{\mathrm{T}}\vec{n}_{K\sigma}|} = \frac{\sin\theta_{K_2}}{\sin\theta_K}\vec{t}_{KP_1} + \frac{\sin\theta_{K_1}}{\sin\theta_K}\vec{t}_{KP_2}. \tag{2.6}$$

Similarly, there is

$$\frac{\kappa^{\mathrm{T}}\vec{n}_{L\sigma}}{|\kappa^{\mathrm{T}}\vec{n}_{L\sigma}|} = \frac{\sin\theta_{L_2}}{\sin\theta_L}\vec{t}_{LP_4} + \frac{\sin\theta_{L_1}}{\sin\theta_L}\vec{t}_{LP_3}. \tag{2.7}$$

Substituting (2.6) into (2.5), we obtain

$$\mathcal{F}_{K,\sigma} = -\int_\sigma |\kappa^{\mathrm{T}}\vec{n}_{K\sigma}|\left(\frac{\sin\theta_{K_2}}{\sin\theta_K}\nabla u(x)\cdot\vec{t}_{KP_1} + \frac{\sin\theta_{K_1}}{\sin\theta_K}\nabla u(x)\cdot\vec{t}_{KP_2}\right)\mathrm{d}l$$

$$= -|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|\left(\frac{\sin\theta_{K_2}}{\sin\theta_K}\frac{u(P_1)-u(K)}{|KP_1|} + \frac{\sin\theta_{K_1}}{\sin\theta_K}\frac{u(P_2)-u(K)}{|KP_2|}\right) + \mathrm{O}(h^2).$$

Similarly, we have

$$\mathcal{F}_{L,\sigma} = -\int_\sigma |\kappa^{\mathrm{T}}\vec{n}_{L\sigma}|\left(\frac{\sin\theta_{L_2}}{\sin\theta_L}\nabla u(x)\cdot\vec{t}_{LP_4} + \frac{\sin\theta_{L_1}}{\sin\theta_L}\nabla u(x)\cdot\vec{t}_{LP_3}\right)\mathrm{d}l$$

$$= -|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|\left(\frac{\sin\theta_{L_2}}{\sin\theta_L}\frac{u(P_4)-u(L)}{|LP_4|} + \frac{\sin\theta_{L_1}}{\sin\theta_L}\frac{u(P_3)-u(L)}{|LP_3|}\right) + \mathrm{O}(h^2).$$

Let $F_{K,\sigma}$ ($F_{L,\sigma}$) be the discrete normal flux on edge σ of cell K (L,resp.) defined as follows:

$$F_{K,\sigma} = -|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|\left(\frac{\sin\theta_{K_2}}{\sin\theta_K}\frac{u_{P_1}-u_K}{|KP_1|} + \frac{\sin\theta_{K_1}}{\sin\theta_K}\frac{u_{P_2}-u_K}{|KP_2|}\right),$$

$$F_{L,\sigma} = -|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|\left(\frac{\sin\theta_{L_2}}{\sin\theta_L}\frac{u_{P_4}-u_L}{|LP_4|} + \frac{\sin\theta_{L_1}}{\sin\theta_L}\frac{u_{P_3}-u_L}{|LP_3|}\right).$$

By continuity of the normal flux component $F_{K,\sigma} = -F_{L,\sigma}$, we have

$$F_{K,\sigma} = -\mu_1|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|\left(\frac{\sin\theta_{K_2}}{\sin\theta_K}\frac{u_{P_1}-u_K}{|KP_1|} + \frac{\sin\theta_{K_1}}{\sin\theta_K}\frac{u_{P_2}-u_K}{|KP_2|}\right)$$

$$+ \mu_2|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|\left(\frac{\sin\theta_{L_2}}{\sin\theta_L}\frac{u_{P_4}-u_L}{|LP_4|} + \frac{\sin\theta_{L_1}}{\sin\theta_L}\frac{u_{P_3}-u_L}{|LP_3|}\right),$$

where μ_1 and μ_2 are some coefficients satisfying $\mu_1 + \mu_2 = 1$, which will be determined later. The above equation can be rewritten to

$$F_{K,\sigma} = \mu_1\frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|} + \frac{\sin\theta_{K_1}}{|KP_2|}\right)u_K - \mu_2\frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|} + \frac{\sin\theta_{L_1}}{|LP_3|}\right)u_L$$

$$- \mu_1\frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|}u_{P_1} + \frac{\sin\theta_{K_1}}{|KP_2|}u_{P_2}\right) + \mu_2\frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|}u_{P_4} + \frac{\sin\theta_{L_1}}{|LP_3|}u_{P_3}\right). \tag{2.8}$$

In order to obtain the two-point flux approximation, the third and fourth term of the above expression should be vanished. Hence we choose μ_1 and μ_2 such that

$$\begin{cases} \mu_1 + \mu_2 = 1, \\ -a_1\mu_1 + a_2\mu_2 = 0, \end{cases} \tag{2.9}$$

where

$$a_1 = \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|}u_{P_1} + \frac{\sin\theta_{K_1}}{|KP_2|}u_{P_2}\right),$$

$$a_2 = \frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|}u_{P_4} + \frac{\sin\theta_{L_1}}{|LP_3|}u_{P_3}\right).$$

If $a_1 + a_2 \neq 0$, then we can obtain

$$\mu_1 = \frac{a_2}{a_1 + a_2}, \qquad \mu_2 = \frac{a_1}{a_1 + a_2}. \tag{2.10}$$

If $a_1 + a_2 = 0$, we can take

$$\mu_1 = \mu_2 = \frac{1}{2}.$$

From the definitions of $\theta_{K_1}, \theta_{K_2}, \theta_{L_1}, \theta_{L_2}, \theta_K$ and θ_L, we know that

$$\sin\theta_{K_1} \geqslant 0, \quad \sin\theta_{K_2} \geqslant 0, \quad \sin\theta_{L_1} \geqslant 0, \quad \sin\theta_{L_2} \geqslant 0,$$

and

$$\sin\theta_K > 0, \quad \sin\theta_L > 0.$$

Hence, there are

$$a_1 \geqslant 0, \quad a_2 \geqslant 0,$$

provided that

$$u_{P_i} \geqslant 0, \quad i = 1, 2, 3, 4, \ldots, \tag{2.11}$$

which imply that

$$\mu_1 \geqslant 0, \qquad \mu_2 \geqslant 0.$$

For $\sigma = K|L \in \mathcal{E}_{\mathrm{int}}$, by (2.8) and (2.9), we have

$$\begin{aligned} F_{K,\sigma} &= \mu_1\frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|} + \frac{\sin\theta_{K_1}}{|KP_2|}\right)u_K - \mu_2\frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|} + \frac{\sin\theta_{L_1}}{|LP_3|}\right)u_L \\ &= A_{K,\sigma}u_K - A_{L,\sigma}u_L, \end{aligned} \tag{2.12}$$

where

$$A_{K,\sigma} = \mu_1\frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|} + \frac{\sin\theta_{K_1}}{|KP_2|}\right) = \mu_1\frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{|KO_1|}, \tag{2.13}$$

and

$$A_{L,\sigma} = \mu_2\frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|} + \frac{\sin\theta_{L_1}}{|LP_3|}\right) = \mu_2\frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{|LO_2|}. \tag{2.14}$$

Under the condition (2.11), it is obvious that there are

$$A_{K,\sigma} \geqslant 0, \qquad A_{L,\sigma} \geqslant 0.$$

For $\sigma \subset \partial\Omega \cap \partial K$ (see Fig. 2.1), the ray originated in the point K along $\kappa^{\mathrm{T}}\overrightarrow{n}_{K\sigma}$ intersects an edge $\tilde{\sigma} = P_2P_1$ with the cross point $O_1 \in \tilde{\sigma}$, where $\tilde{\sigma}$ may be σ or not, and in this case we define

$$F_{K,\sigma} = -\frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin\theta_K}\left[\frac{\sin\theta_{K_2}}{|KP_1|}u_{P_1} + \frac{\sin\theta_{K_1}}{|KP_2|}u_{P_2} - \left(\frac{\sin\theta_{K_2}}{|KP_1|} + \frac{\sin\theta_{K_1}}{|KP_2|}\right)u_K\right] = A_{K,\sigma}u_K - a_{K,\sigma}, \tag{2.15}$$

where

$$A_{K,\sigma} = \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|} + \frac{\sin\theta_{K_1}}{|KP_2|}\right),$$

$$a_{K,\sigma} = \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|}u_{P_1} + \frac{\sin\theta_{K_1}}{|KP_2|}u_{P_2}\right).$$

In the formulae (2.12) and (2.15), if P_i lies on $\partial\Omega$, then we take $u_{P_i} = g_{P_i}$ in the corresponding formula.

In order to ensure the effect of boundary condition, we require there exists at least one edge $\sigma \subset \partial K$ ($K \cap \partial\Omega \subset \mathcal{E}_{\mathrm{ext}}$) such that the ray originated in the cell center K along $\kappa^{\mathrm{T}}\overrightarrow{n}_{K\sigma}$ intersects one edge $\tilde{\sigma}$ satisfying $\tilde{\sigma} \cap \partial\Omega \neq \emptyset$. If the above condition is not satisfied, we can obtain the expression of $F_{K,\sigma}$ on boundary similar to Section 2.3. Hence, we can always ensure the effect of boundary condition.

With the definition of $F_{K,\sigma}$ the finite volume scheme is constructed as follows:

$$\sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma} = f_K m(K), \quad \forall K \in \mathcal{J}, \tag{2.16}$$

$$u_{P_i} = g_{P_i}, \quad \forall P_i \in \partial\Omega, \tag{2.17}$$

where $f_K = f(K)$.

It is obvious that the cell center can be defined at any position of a cell in our scheme. The coefficients $A_{K,\sigma}$ and $A_{L,\sigma}$ depend on the cell vertex unknowns, hence, the scheme is nonlinear.

2.3. Robin boundary conditions

Consider the following robin boundary conditions:

$$\alpha\kappa\nabla u \cdot \vec{v} + \beta u = g, \tag{2.18}$$

where $\vec{v}$ is the outward unit normal vector of domain Ω.

Integrate (2.18) on cell-side $\sigma \in \partial\Omega$ to obtain

$$\int_\sigma \alpha\kappa\nabla u \cdot \vec{v} + \int_\sigma \beta u = \int_\sigma g. \tag{2.19}$$

Let K be the midpoint of σ, then we have

$$\alpha_K \mathcal{F}_{K,\sigma} + |\sigma|\beta_K u_K = |\sigma|g_K, \tag{2.20}$$

where

$$\mathcal{F}_{K,\sigma} = \int_\sigma \kappa\nabla u \cdot \vec{v} = -\int_\sigma \kappa\nabla u \cdot \vec{n}_{K,\sigma}.$$

Next, we discrete the above expression. As the vector $\kappa^{\mathrm{T}}\vec{v}$ is an outward vector of the domain Ω, which is always true due to the physics of the problem, the ray originated in the point K along the direction $\kappa^{\mathrm{T}}\vec{n}_{K\sigma}$ intersects one of the segments LA and LB (see Fig. 2.2), this segment is denoted by P_1P_2, and the cross point is O_1. In this figure, the points P_1 and L are the same point.
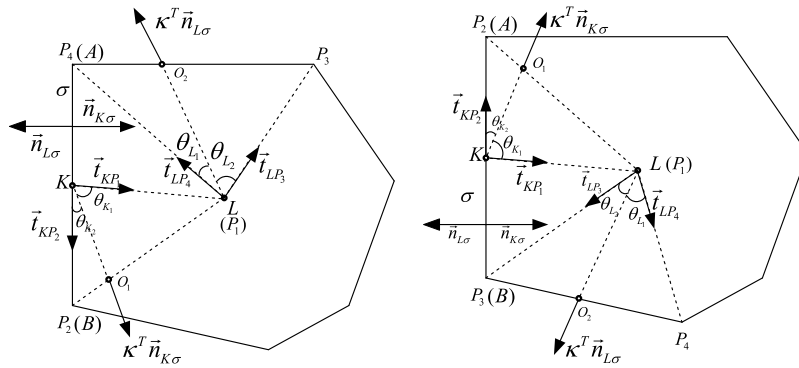
Fig. 2.2. Boundary stencil.

Similar to Section 2.2, we define

$$F_{K,\sigma} = -\mu_1 |\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma| \left(\frac{\sin \theta_{K_2}}{\sin \theta_K} \frac{u_{P_1} - u_K}{|KP_1|} + \frac{\sin \theta_{K_1}}{\sin \theta_K} \frac{u_{P_2} - u_K}{|KP_2|} \right)$$
$$+ \mu_2 |\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma| \left(\frac{\sin \theta_{L_2}}{\sin \theta_L} \frac{u_{P_4} - u_L}{|LP_4|} + \frac{\sin \theta_{L_1}}{\sin \theta_L} \frac{u_{P_3} - u_L}{|LP_3|} \right),$$

Noticing that the points P_1 and L are the same point, we can rewrite the above equation to

$$F_{K,\sigma} = \mu_1 \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin \theta_K} \left(\frac{\sin \theta_{K_2}}{|KP_1|} + \frac{\sin \theta_{K_1}}{|KP_2|} \right) u_K$$
$$- \left[\mu_2 \frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin \theta_L} \left(\frac{\sin \theta_{L_2}}{|LP_4|} + \frac{\sin \theta_{L_1}}{|LP_3|} \right) + \mu_1 \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin \theta_K} \frac{\sin \theta_{K_2}}{|KL|} \right] u_L$$
$$- \mu_1 \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin \theta_K} \frac{\sin \theta_{K_1}}{|KP_2|} u_{P_2} + \mu_2 \frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin \theta_L} \left(\frac{\sin \theta_{L_2}}{|LP_4|} u_{P_4} + \frac{\sin \theta_{L_1}}{|LP_3|} u_{P_3} \right). \quad (2.21)$$

In order to obtain the two-point flux approximation, these terms including the values of vertex in the expression (2.21) should be vanished. Hence, we choose μ_1 and μ_2 such that

$$\begin{cases} \mu_1 + \mu_2 = 1, \\ -a_1\mu_1 + a_2\mu_2 = 0, \end{cases} \quad (2.22)$$

where

$$a_1 = \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin \theta_K} \frac{\sin \theta_{K_1}}{|KP_2|} u_{P_2},$$
$$a_2 = \frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin \theta_L} \left(\frac{\sin \theta_{L_2}}{|LP_4|} u_{P_4} + \frac{\sin \theta_{L_1}}{|LP_3|} u_{P_3} \right).$$

If $a_1 + a_2 \neq 0$, then we can obtain

$$\mu_1 = \frac{a_2}{a_1 + a_2}, \qquad \mu_2 = \frac{a_1}{a_1 + a_2}. \quad (2.23)$$

If $a_1 + a_2 = 0$, we can take

$$\mu_1 = \mu_2 = \frac{1}{2}.$$

Similar to the Section 2.2, we can see that

$$a_1 \geqslant 0, \qquad a_2 \geqslant 0,$$

provided that

$$u_{P_i} \geqslant 0, \quad i = 2, 3, 4, \ldots \tag{2.24}$$

Hence, there are

$$\mu_1 \geqslant 0, \qquad \mu_2 \geqslant 0.$$

By (2.21) and (2.22), we get

$$F_{K,\sigma} = A_{K,\sigma} u_K - A_{L,\sigma} u_L, \tag{2.25}$$

where

$$A_{K,\sigma} = \mu_1 \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin \theta_K} \left(\frac{\sin \theta_{K_2}}{|KP_1|} + \frac{\sin \theta_{K_1}}{|KP_2|} \right), \tag{2.26}$$

and

$$A_{L,\sigma} = \mu_2 \frac{|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}||\sigma|}{\sin \theta_L} \left(\frac{\sin \theta_{L_2}}{|LP_4|} + \frac{\sin \theta_{L_1}}{|LP_3|} \right) + \mu_1 \frac{|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}||\sigma|}{\sin \theta_K} \frac{\sin \theta_{K_2}}{|KL|}. \tag{2.27}$$

Under the condition (2.24), it is obvious that there are

$$A_{K,\sigma} \geqslant 0, \quad A_{L,\sigma} \geqslant 0.$$

By (2.20) and (2.25), we have

$$(\alpha_K A_{K,\sigma} + |\sigma|\beta_K)u_K - \alpha_K A_{L,\sigma} u_L = |\sigma|g_K. \tag{2.28}$$

2.4. Special case

Next, we consider the special case of $\kappa = \lambda I$, where I is a unit matrix.

In this case, when we consider the normal flux through an edge $\sigma = K \mid L$ (the common side of cell K and L), the vertical line from the cell center K to the cell-side σ will always intersect one of cell-side $P_1 P_2$ of cell K (see Figs. 2.3, 2.4 and 2.5). The vertical line from the cell center L to the cell-side σ will always intersect one of cell-side $P_3 P_4$ of cell L.

Obviously the unit normal vectors $\vec{n}_{K\sigma}$ and $\vec{n}_{L\sigma}$ can be expressed by the linear combinations of two unit vectors with direction from the cell center to the cell vertex (see Fig. 2.3, 2.4 and 2.5), that is

$$\vec{n}_{K\sigma} = \frac{\sin \theta_{K_2}}{\sin \theta_K} \vec{t}_{KP_1} + \frac{\sin \theta_{K_1}}{\sin \theta_K} \vec{t}_{KP_2}, \tag{2.29}$$

$$\vec{n}_{L\sigma} = \frac{\sin \theta_{L_2}}{\sin \theta_L} \vec{t}_{LP_4} + \frac{\sin \theta_{L_1}}{\sin \theta_L} \vec{t}_{LP_3}. \tag{2.30}$$
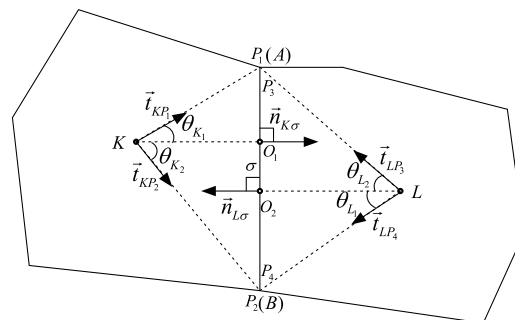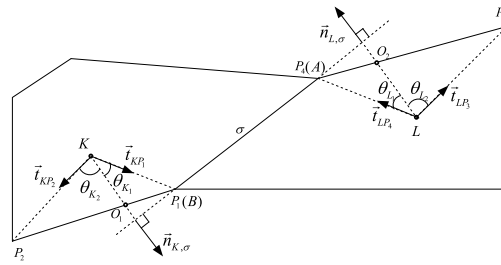

Fig. 2.3. Special stencil 1.
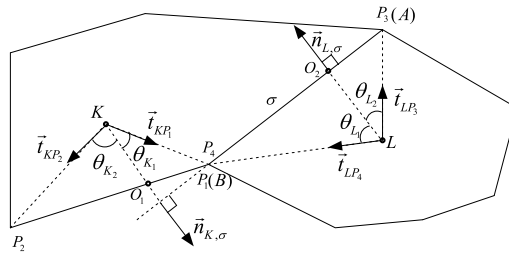
Fig. 2.4. Special stencil 2.

Fig. 2.5. Special stencil 3.

Similar to Section 2.2, we need only to set $|\kappa^{\mathrm{T}}(K)\vec{n}_{K\sigma}| = |\lambda(K)|$ and $|\kappa^{\mathrm{T}}(L)\vec{n}_{L\sigma}| = |\lambda(L)|$, then we can obtain the discrete normal flux on edge σ of cell K as follows:

$$F_{K,\sigma} = \mu_1 \frac{|\lambda(K)||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|} + \frac{\sin\theta_{K_1}}{|KP_2|}\right)u_K - \mu_2 \frac{|\lambda(L)||\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|} + \frac{\sin\theta_{L_1}}{|LP_3|}\right)u_L$$

$$- \mu_1 \frac{|\lambda(K)||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|}u_{P_1} + \frac{\sin\theta_{K_1}}{|KP_2|}u_{P_2}\right) + \mu_2 \frac{|\lambda(L)||\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|}u_{P_4} + \frac{\sin\theta_{L_1}}{|LP_3|}u_{P_3}\right). \tag{2.31}$$

Denote

$$a_1 = \frac{\lambda(K)|\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|}u_{P_1} + \frac{\sin\theta_{K_1}}{|KP_2|}u_{P_2}\right),$$

$$a_2 = \frac{\lambda(L)|\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|}u_{P_4} + \frac{\sin\theta_{L_1}}{|LP_3|}u_{P_3}\right),$$

and

$$\mu_1 = \frac{a_2}{a_1 + a_2}, \qquad \mu_2 = \frac{a_1}{a_1 + a_2}. \tag{2.32}$$

Then we can obtain the discrete normal flux on cell-side σ:

$$F_{K,\sigma} = A_{K,\sigma}u_K - A_{L,\sigma}u_L, \tag{2.33}$$

where

$$A_{K,\sigma} = \mu_1 \frac{|\lambda(K)||\sigma|}{\sin\theta_K}\left(\frac{\sin\theta_{K_2}}{|KP_1|} + \frac{\sin\theta_{K_1}}{|KP_2|}\right) = \mu_1 \frac{\lambda(K)|\sigma|}{|KO_1|}, \tag{2.34}$$

$$A_{L,\sigma} = \mu_2 \frac{|\lambda(L)||\sigma|}{\sin\theta_L}\left(\frac{\sin\theta_{L_2}}{|LP_4|} + \frac{\sin\theta_{L_1}}{|LP_3|}\right) = \mu_2 \frac{\lambda(L)|\sigma|}{|LO_2|}. \tag{2.35}$$

When $u_{P_i} \geqslant 0$ $(i = 1, 2, 3, 4, \ldots)$, there are

$$A_{K,\sigma} \geqslant 0, \quad A_{L,\sigma} \geqslant 0.$$

2.5. The expression of cell vertex unknowns

It is obvious that the coefficients $A_{K,\sigma}$ and $A_{L,\sigma}$ depend on the vertex unknowns, i.e., there are the vertex unknowns in addition to cell-centered unknowns in the expression of flux. Now we consider how to eliminate the vertex unknowns locally, or approximate the vertex unknowns by the cell-centered unknowns.

Two interpolation techniques have been considered in [11]. One is the linear interpolation by three unknown values of cell centers closest to the cell vertex [17]. Another is the inverse distance weighting [21] of the vertex value u_A for all cell $K \in \mathcal{J}$ sharing A as a vertex, i.e.,

$$u_A = \sum_{K \in \mathcal{U}(A)} u_K \omega_K, \quad \omega_K = \frac{|x_K - A|^{-1}}{\sum_{L \in \mathcal{U}(A)} |x_L - A|^{-1}},$$

where $\mathcal{U}(A)$ is the collection of these cells K that have A as a vertex.

We proposed some other methods of eliminating the vertex unknowns in [20]. Now we describe briefly one of the methods, which will be used in Section 4. Let

$$u_p = \sum_{j=1}^{n_p} \omega_j u_{q_j},$$ (2.36)

where q_j are the center of cell around the vertex p, n_p is the number of cell sharing the vertex p, and ω_j are some combination coefficients.

When the diffusion coefficient κ is continuous, we require that ω_j $(j = 1, \ldots, n_p)$ satisfy the following relation:

$$\begin{cases} \sum_{j=1}^{n_p} \omega_j = 1, \\ \sum_{j=1}^{n_p} x_{q_j p} \omega_j = 0, \\ \sum_{j=1}^{n_p} y_{q_j p} \omega_j = 0, \end{cases}$$ (2.37)

where $x_{q_j p} = x_{q_j} - x_p$ and $y_{q_j p} = y_{q_j} - y_p$ $(j = 1, \ldots, n_p)$. The linear system associated with this problem reduces to a under-determined system, we can solve this problem by using the least-squares method (see [20]).

When the coefficient κ is discontinuous, we determine the coefficients by the continuity of normal flux and tangential gradients at the cell vertex(see [20]).

However, the coefficients ω_j maybe negative, hence it maybe lead to $u_p < 0$ even if all u_{q_j} $(j = 1, \ldots, n_p)$ are non-negative. In this case we can use any interpolation of preserving positivity, e.g. the inverse distance weighting method, to guarantee $u_p \geqslant 0$.

2.6. Discrete system

Substituting (2.12) and (2.15) into (2.16), we get a nonlinear algebraic system. Let U be the vector discrete unknowns and $A(U)$ be the matrix of this system. The matrix $A(U)$ may be represented by assembling of 2×2 matrices

$$A_\sigma(U) = \begin{pmatrix} A_{K,\sigma}(U) & -A_{L,\sigma}(U) \\ -A_{K,\sigma}(U) & A_{L,\sigma}(U) \end{pmatrix}$$

for interior edges and 1×1 matrices $A_\sigma(U) = A_{K,\sigma}(U)$ for boundary edges. The global discrete nonlinear system reads as:

$$A(U)U = F,$$ (2.38)

where

$$A(U) = \sum_{\sigma \in \mathcal{E}} N_\sigma A_\sigma(U) N_\sigma^{\mathrm{T}}, \tag{2.39}$$

and N_σ are assembling matrices consisting of zeros and ones.

The matrix $A(U)$ is non-symmetric and has the following properties:

1. All diagonal entries of matrix $A(U)$ are positive.
2. All off-diagonal entries of $A(U)$ are non-positive.
3. Each column sum in $A(U)$ is non-negative and there exists a column with a positive sum.

These properties implies $A(U)$ is weak diagonal dominance in column.

The nonlinear system (2.38) may be solved by a number of different methods. Just as [11] we use the Picard iterations: Choose a small value $\varepsilon_{\mathrm{non}} > 0$ and initial vector $U^0 \geqslant 0$, and repeat for $k = 1, 2, \ldots,$

1. Solve $A(U^{k-1})U^k = F$,
2. Stop if $\|A(U^k)U^k - F\| \leqslant \varepsilon_{\mathrm{non}}\|A(U^0)U^0 - F\|$.

The linear system with non-symmetric matrix $A(U^{k-1})$ is solved by the Bi-Conjugate Gradient Stabilized (BiCGStab) method. The BiCGStab iterations are terminated when the relative norm of the initial residual becomes smaller than $\varepsilon_{\mathrm{lin}}$.

In our numerical experiments, the Picard iteration always converge, however, the number of nonlinear iteration is excessive. For some stationary problems, the number of nonlinear iteration is over 20. However, the main issue of this paper is the construction of monotone scheme, and the consideration for computational efficiency is our future plan.

2.7. Monotonicity

In order to show that our schemes are monotone, we introduce the following lemma [3].

Lemma 2.1. *For an irreducible matrix $A = (a_{ij})_{n \times n}$ satisfying $a_{ii} > 0$ $(1 \leqslant i \leqslant n)$ and $a_{ij} \leqslant 0$ $(1 \leqslant i, j \leqslant n, i \neq j)$, if A is weak diagonal dominance in rows, that is*

$$\sum_{j=1}^{n} a_{ij} \geqslant 0 \quad (i = 1, 2, \ldots, n), \tag{2.40}$$

with strict inequality for at least one of the Eq. (2.40). Then the matrix A is an M-matrix.

Now, we state that our scheme is monotone.

Theorem 2.2. *Let $F \geqslant 0$, $U^0 \geqslant 0$ and linear systems in Picard iterations are solved exactly. Then all iterates U^k are non-negative vectors:*

$$U^k \geqslant 0.$$

Proof. We first prove that the matrix $A(U)$ is monotone for any vector U with non-negative components. In the above section, we have state some properties of matrix $A(U)$. It is obvious that the matrix $A^{\mathrm{T}}(U)$ satisfies the conditions of Lemma 2.1, hence $A^{\mathrm{T}}(U)$ is an M-matrix, that is all entries of $(A^{\mathrm{T}}(U))^{-1}$ are non-negative. Since inverse and transpose operation commute, $(A^{\mathrm{T}}(U))^{-1} = (A^{-1}(U))^{\mathrm{T}}$, we conclude that all entries of $A^{-1}(U)$ are non-negative and $A(U)$ is monotone for any vector $U \geqslant 0$.

Noticing that $U^0 \geqslant 0$, we assume for some integer $k_0 > 0$,

$$U^{k_0-1} \geqslant 0.$$

Hence, the matrix $A(U^{k_0-1})$ is monotone, that is $A^{-1}(U^{k_0-1}) \geqslant 0$. Also notice $F \geqslant 0$, it follows that the solution U^{k_0} of $A(U^{k_0-1})U^{k_0} = F$ is a non-negative vector, that is

$$U^{k_0} \geqslant 0.$$

By induction argument, there are

$$U^k \geqslant 0, \quad \text{for all } k \geqslant 0. \qquad \square$$

3. Extension to non-stationary diffusion equations

Consider following non-stationary diffusion problem:

$$u_t - \nabla \cdot (\kappa(x,t)\nabla u) = f \quad \text{in } \Omega \times (0,T], \tag{3.1}$$
$$u(x,t) = g \quad \text{on } \partial\Omega \times (0,T], \tag{3.2}$$
$$u(x,0) = \varphi(x) \quad \text{on } \Omega, \tag{3.3}$$

where $\kappa = \kappa(x,t)$ is a diffusion tensor, $f = f(x,t)$, $g = g(x,t)$ and $\varphi(x)$ are given functions.

Integrate (3.1) over the cell K to obtain

$$\int_K u_t \, \mathrm{d}x + \sum_{\sigma \in \mathcal{E}_K} \mathcal{F}_{K,\sigma}^{n+1} = \int_K f(x,t^{n+1}) \, \mathrm{d}x, \tag{3.4}$$

where the continuous flux on edge σ is

$$\mathcal{F}_{K,\sigma}^{n+1} = -\int_\sigma \kappa(x,t^{n+1})\nabla u(x,t^{n+1}) \cdot \vec{n}_{K\sigma} \mathrm{d}l. \tag{3.5}$$

Using the similar process as for stationary diffusion problems, we obtain the finite volume scheme of the problem (3.1)–(3.3):

$$\frac{u_K^{n+1} - u_K^n}{\Delta t} m(K) + \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}^{n+1} = f_K^{n+1} m(K), \quad K \in \Omega, \tag{3.6}$$
$$u_{P_i}^{n+1} = g_{P_i}^{n+1}, \quad P_i \in \partial\Omega, \tag{3.7}$$
$$u_K^0 = \varphi(K), \quad K \in \Omega, \tag{3.8}$$

where

$$F_{K,\sigma}^{n+1} = A_{K,\sigma}^{n+1} u_K^{n+1} - A_{L,\sigma}^{n+1} u_L^{n+1}, \quad \sigma = K|L \in \mathcal{E}_{\mathrm{int}},$$

and $F_{K,\sigma}^{n+1}$ is similar to (2.15) for $\sigma \in \mathcal{E}_{\mathrm{ext}}$.

Let U^{n+1} be the vector discrete unknowns and $B(U^{n+1})$ be the matrix of this system. The matrix $B(U^{n+1})$ can be represented by the combination of two matrices $B_1(U^{n+1})$ and $B_2(U^{n+1})$, that is

$$B(U^{n+1}) = B_1(U^{n+1}) + B_2(U^{n+1}).$$

The matrix $B_1(U^{n+1})$ may be represented by assembling of 2×2 matrices

$$B_{1,\sigma}(U^{n+1}) = \begin{pmatrix} \Delta t A_{K,\sigma}^{n+1}(U) & -\Delta t A_{L,\sigma}^{n+1}(U) \\ -\Delta t A_{K,\sigma}^{n+1}(U) & \Delta t A_{L,\sigma}^{n+1}(U) \end{pmatrix}.$$

The matrix $B_2(U^{n+1})$ is diagonal matrix, and may be represented by assembling of 1×1 matrices

$$B_{2,K}(U^{n+1}) = (\, m(K)\,).$$

The global discrete nonlinear system reads as:

$$B(U^{n+1})U^{n+1} = F^{n+1}. \tag{3.9}$$

We use the Picard iteration method in Section 2.6 to solve the above system. It is easy to see that the solution satisfies $(U^{n+1})^k \geqslant 0$ for k and n, provided that $f(x,t) \geqslant 0, g(x,t) \geqslant 0$ and $\varphi(x) \geqslant 0$.

There has no stability constraint for time step due to the implicit time discretion, and our scheme is monotone for any time step $\Delta t > 0$.

4. Numerical experiments

We use several numerical experiments to demonstrate that the discretization scheme satisfies the practical requirements mentioned in the introduction. The convergence rate is studied and the positivity of discrete solution is verified on different types of meshes.

We use discrete L_2-norms to evaluate approximation errors. For the solution u, we use the following L_2-norm:

$$\varepsilon_2^u = \left[\sum_{k \in \mathcal{J}} (u_K - u(K))^2 m(K) \right]^{1/2}.$$

For the flux F, we use the following L_2-norm (which is different from that defined in [11])

$$\varepsilon_2^F = \left[\sum_{\sigma \in \mathcal{E}} (F_{K,\sigma} - \mathcal{F}_{K,\sigma})^2 \right]^{1/2}.$$

For the stationary diffusion problems, we take $\varepsilon_{\text{non}} = 1.0e - 5$ and $\varepsilon_{\text{lin}} = 1.0e - 10$. For the non-stationary diffusion problems, we take $\varepsilon_{\text{non}} = 1.0e - 5$ and $\varepsilon_{\text{lin}} = 1.0e - 15$.

The random quadrilateral meshes over the physical domain $\Omega = [0,1] \times [0,1]$ is defined by $x_{ij} = \frac{i}{I} + \frac{\sigma}{I}(R_x - 0.5)$, $y_{ij} = \frac{i}{J} + \frac{\sigma}{J}(R_y - 0.5)$, where $\sigma \in [0,1]$ is a parameter, R_x and R_y are two normalized random variables. In this paper, we let $\sigma = 0.7$.

In the following Sections 4.1–4.5, we will use our method of eliminating the cell vertex unknowns mentioned in Section 2.5, moreover two other methods are considered and compared with ours in the Section 4.6.

4.1. Positivity of numerical solutions

Let us consider the problem (2.1), (2.2) in the unit square $\Omega = (0,1)^2$ and set

$$\kappa = \begin{pmatrix} y^2 + \varepsilon x^2 & -(1-\varepsilon)xy \\ -(1-\varepsilon)xy & \varepsilon y^2 + x^2 \end{pmatrix}, \quad \varepsilon = 5 \times 10^{-3}, \tag{4.1}$$

and

$$f(x,y) = \begin{cases} 1 & \text{if } (x,y) \in [3/8, 5/8]^2, \\ 0 & \text{otherwise.} \end{cases}$$

We impose the homogeneous Dirichlet boundary condition on $\partial\Omega$.

First, we test our nonlinear FV scheme on rectangular meshes and random quadrilateral meshes (see Fig. 4.1). The exact solution $u(x,y)$ is unknown but the maximum principle states that it is non-negative. The numerical solutions obtained by the MPFA (MPFA-O method in [1]) and our scheme on rectangular meshes and random quadrilateral meshes are shown in Figs. 4.3 and 4.4, respectively. From these figures, we see that MPFA produces negative values, however, our scheme preserves the positivity of the continuous solution. For the rectangular meshes, there are about 13% of all cells on which the numerical solution obtained by the MPFA is negative. Moreover, for the random quadrilateral meshes, the numerical solutions obtained by the MPFA has non-physical oscillations.

For the rectangular meshes, the number of nonlinear iteration is 35. For the random quadrilateral meshes, the number of nonlinear iteration is 40. The computational efficient will be considered in the future.
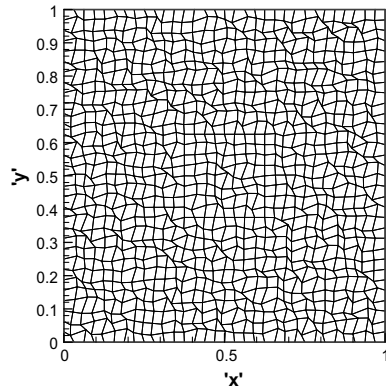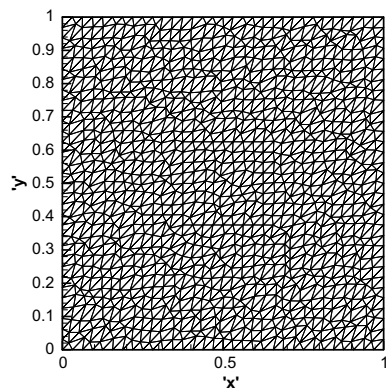
Fig. 4.1. Random quadrilateral meshes.

Fig. 4.2. Random triangular meshes.

$u_{min} = -3.8E - 5, \ u_{max} = 0.144$
MPFA

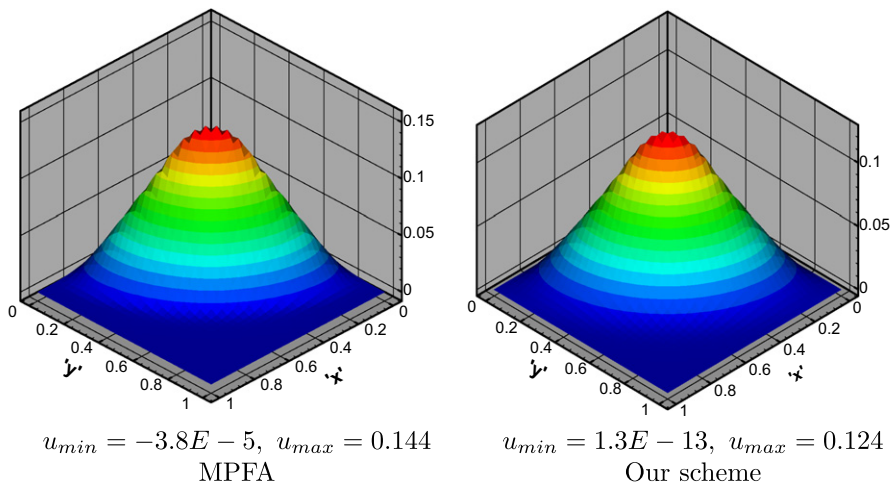$u_{min} = 1.3E - 13, \ u_{max} = 0.124$
Our scheme

Fig. 4.3. Comparison of MPFA and our scheme on rectangular meshes.

Then, we test our scheme on uniform triangular meshes and random triangular meshes (see Fig. 4.2). The numerical solutions obtained on uniform triangular meshes and random triangular meshes are shown in Fig. 4.5, which demonstrates that our scheme preserves positivity of the solution.

$u_{min} = -0.11, \ u_{max} = 0.17$
MPFA

$u_{min} = 2.3E - 11, \ u_{max} = 0.124$
Our scheme
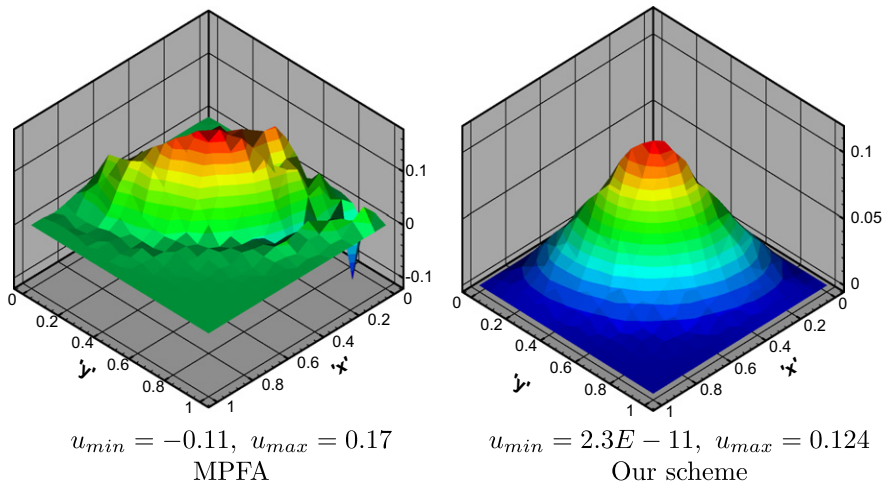
Fig. 4.4. Comparison of MPFA and our scheme on random quadrilateral meshes.

$u_{min} = 2.7E - 12, \ u_{max} = 0.117$

$u_{min} = 3.3E - 10, \ u_{max} = 0.116$
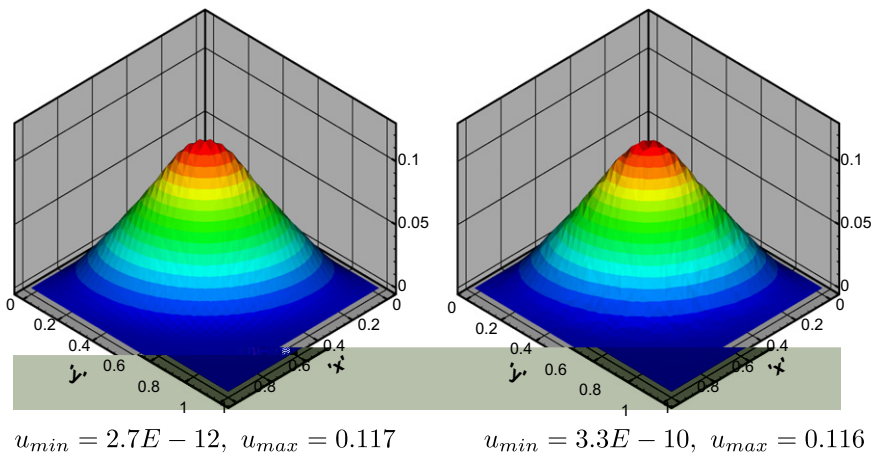
Fig. 4.5. Solution profile on uniform triangular (left) and random triangular meshes (right).

4.2. Non-smooth anisotropic solution

Let us now consider the problem (2.1), (2.2) with a non-smooth anisotropic solution. The computational domain is the unit square with a hole, $\Omega = (0, 1)^2 \setminus [4/9, 5/9]^2$, the boundary $\partial\Omega$ is composed of two disjoint parts Γ_1 and Γ_0 as shown in Figs. 4.6 and 4.7, where Γ_1 is the interior boundary and Γ_0 is the exterior boundary.

We set $f = 0$, $g = 0$ on Γ_0, $g = 2$ on Γ_1, and take the anisotropic diffusion tensor κ as follows

$$\kappa = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{4.2}$$

where $k_1 = 1$, $k_2 = 100$ and $\theta = \pi/6$.

We test this problem on two different meshes. One is the random triangular meshes with a hole (see Fig. 4.6), the other is the random quadrilateral meshes with a hole (see Fig. 4.7), and the scale of mesh is 72×72. The numerical solutions on random triangular meshes are shown in Fig. 4.8, the minimum value is close to zero and the maximum value is 1.988. The numerical solutions on random quadrilateral meshes are shown in Fig. 4.9, the minimum value is also close to zero and the maximum value is 1.981. Hence, our method obtains the non-negative discrete solutions.
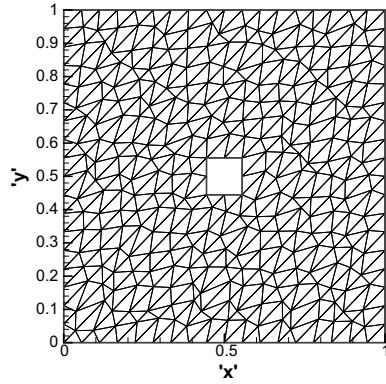
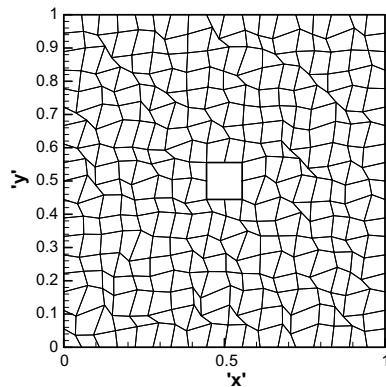Fig. 4.6. Random triangular meshes with a hole.

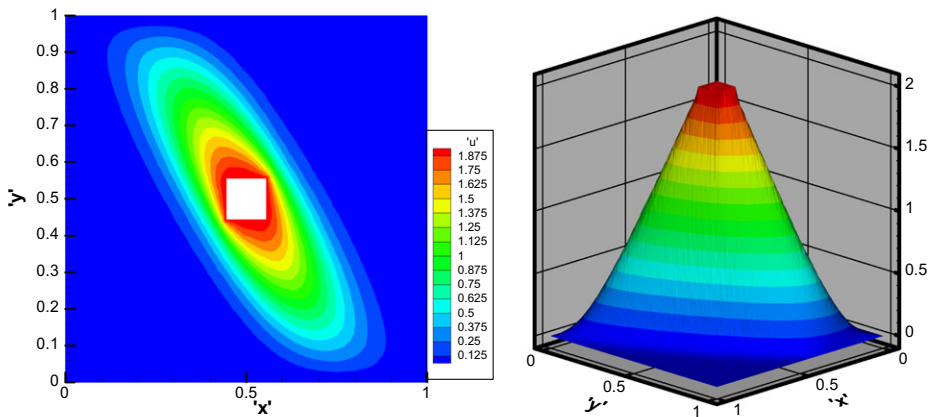Fig. 4.7. Random quadrilateral meshes with a hole.

Fig. 4.8. Problem with non-smooth anisotropic solution on random triangular meshes with a hole: (a) colormap of numerical solution; (b) solution profile: ($u_{\min} = 1.2\mathrm{E} - 14$, $u_{\max} = 1.988$). (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Heterogeneous diffusion tensor

In this section we demonstrate that our scheme can handle strong jumps of full diffusion tensor across mesh edges. Consider the problem (2.1), (2.2) in the unit square $\Omega = (0,1)^2$ with the source term
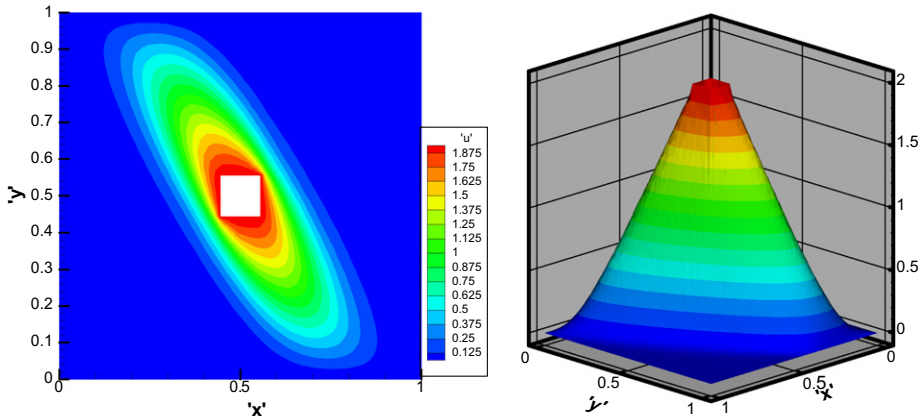
Fig. 4.9. Problem with non-smooth anisotropic solution with on random quadrilateral meshes with a hole: (a) colormap of numerical solution; (b) solution profile: ($u_{\min} = 1.2E - 14$, $u_{\max} = 1.981$). (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

$$f(x, y) = \begin{cases} 10000 & \text{if } (x, y) \in [7/18, 11/18]^2, \\ 0 & \text{otherwise.} \end{cases}$$

and the homogeneous Dirichlet boundary condition $g = 0$.

The domain Ω is partitioned into four square subdomains Ω_i, $i = 1, \ldots, 4$, as shown in Fig. 4.12a. The diffusion tensor is given by formula (4.2) with different parameters k_1, k_2 and θ in subdomains Ω_i. First, we fix the anisotropy ratio by setting $k_1 = 10^3$ and $k_2 = 1$ and vary only parameter θ (see Fig. 4.12a). Second, we use different parameters k_1, k_2 and θ on different subdomains(see Fig. 4.13a). We compute these problems on random quadrilateral meshes (see Fig. 4.10), and the scale of mesh is 72×72. In both cases we get the non-negative discrete solutions (see Fig. 4.12b and Fig. 4.13b).

4.4. Results on polygonal meshes

In this subsection, we test our scheme on polygonal meshes. Consider the problem in Section 4.1 and use the polygon meshes shown in Fig. 4.11. The polygon meshes is generated by Voronoi tessellation, and the site points are centers of random meshes. The contour of discrete solution on rectangular meshes and polygonal meshes are shown in Figs. 4.14 and 4.15, respectively. We see that the contour on polygonal meshes accord with that on rectangular meshes. Moreover, the numerical solution obtained by our scheme is non-negative in Ω.
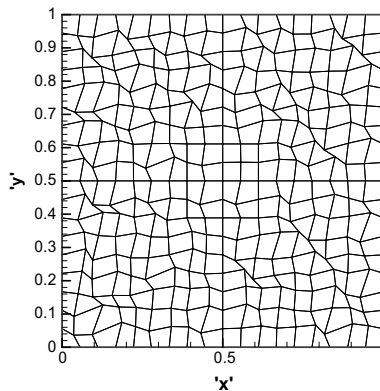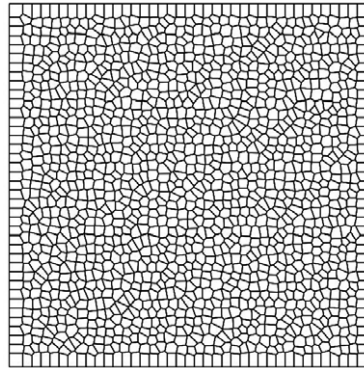

Fig. 4.10. Random quadrilateral meshes.
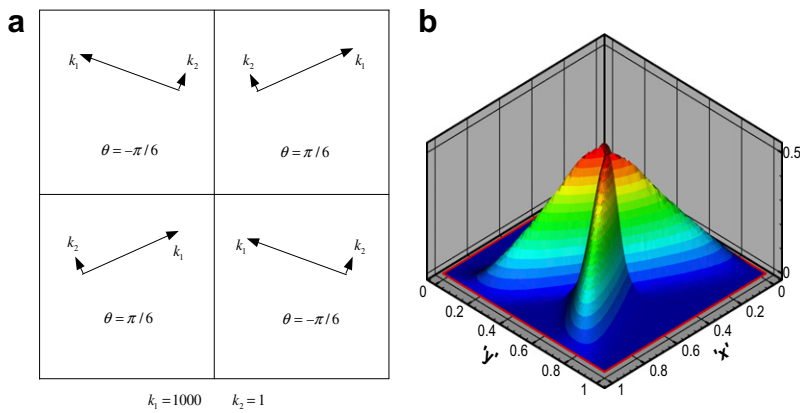
Fig. 4.11. The polygonal meshes.

Fig. 4.12. Principle directions of the anisotropic diffusion tensor with fixed eigenvalues k_1 and k_2 (left picture) and profile of discrete solution on random quadrilateral meshes (right picture).
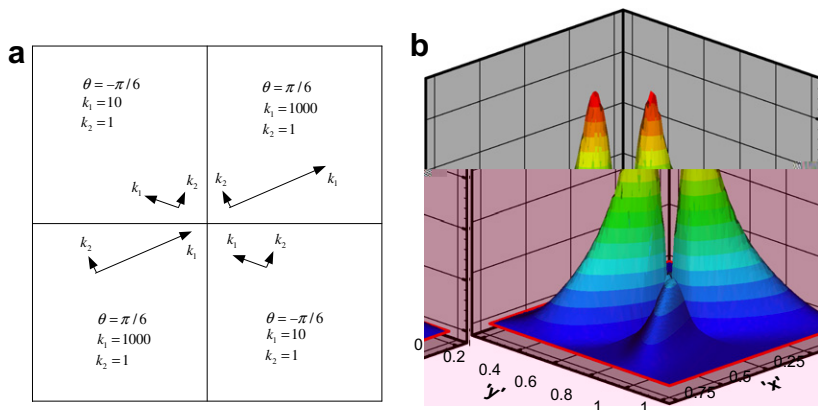

Fig. 4.13. Principle directions and eigenvalues of heterogeneous anisotropic diffusion tensor (left picture) and profile of the discrete solution on random quadrilateral meshes (right picture).

4.5. Problem with mixed boundary conditions

In this subsection, we consider the diffusion problem with mixed boundary conditions. The equation is the same as Section 4.1, but with the following mixed boundary condition:
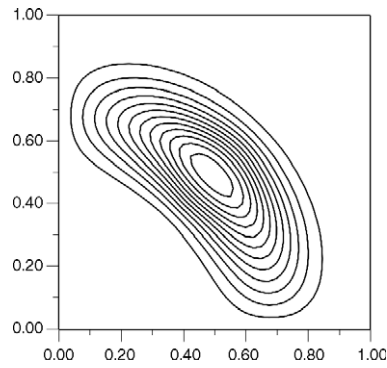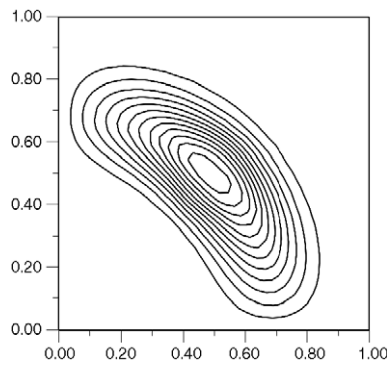
Fig. 4.14. The contour on rectangular meshes.

Fig. 4.15. The contour on polygonal meshes.

$$u + \kappa \nabla u \cdot \vec{v} = 0 \quad \text{at } y = 0, \qquad \kappa \nabla u \cdot \vec{v} = 0 \quad \text{at } y = 1,$$
$$u + \kappa \nabla u \cdot \vec{v} = 0 \quad \text{at } x = 0, \qquad \kappa \nabla u \cdot \vec{v} = 0 \quad \text{at } x = 1.$$

We test this problem on random quadrilateral meshes (see Fig. 4.1). The numerical solutions obtained by our scheme on random quadrilateral meshes are shown on Fig. 4.16. From this figure, we see that our scheme preserves the positivity for the problem with mixed boundary conditions.

4.6. The accuracy of our scheme

Now we consider the accuracy of our nonlinear FV scheme with three methods (I)–(III) of eliminating the cell vertex unknowns. The first method (I) is our method described in Section 2.5. The second method (II) is
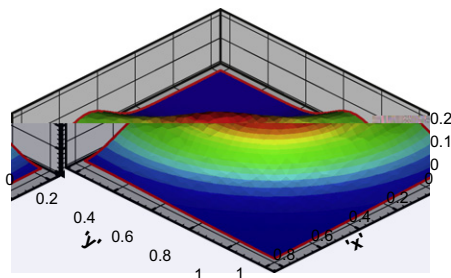

Fig. 4.16. Solution profile on random quadrilateral meshes for the problem with mixed boundary conditions ($u_{\min} = 2.13\text{E} - 9$, $u_{\max} = 0.2141$).

the inverse distance weighting method (see [11]) mentioned in Section 2.5. The third method (III) is the simple weighting method, that is,

$$\omega_i = 1/n_p,$$

where n_p is the number of cell sharing the cell vertex p.

4.6.1. The elliptic problem

Consider the problem (2.1), (2.2) with Dirichlet boundary condition in the unit square $\Omega = (0, 1)^2$. Let $\kappa = RDR^T$, and

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad D = \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix},$$

where $\theta = \frac{5\pi}{12}$, $k_1 = 1 + 2x^2 + y^2$, $k_2 = 1 + x^2 + 2y^2$. The solution is chosen to be $u(x, y) = \sin(\pi x)\sin(\pi y)$.

We test our method (I) on random quadrilateral meshes (see Fig. 4.17) and random triangular meshes (see Fig. 4.18). Table 4.1 gives L_2-norm of the error between exact solution and numerical solution and L_2-norm of the error between exact flux and numerical flux on random quadrilateral meshes. From this table, one can know that our method gives second-order convergence rate for the solution and the first-order convergence rate for the flux.
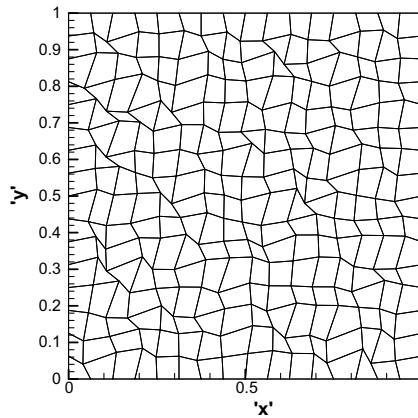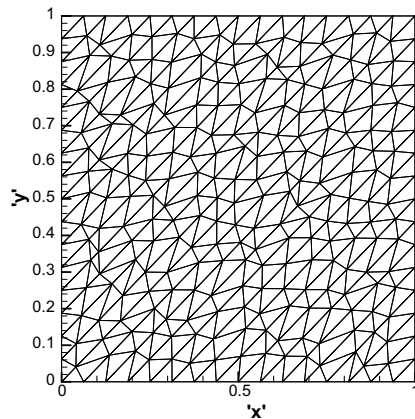

Fig. 4.17. Random quadrilateral meshes.

Fig. 4.18. Random triangular meshes.

Table 4.1
Convergence results for the elliptic problem on random quadrilateral meshes (method I)

The number of cell	64	256	1024	4096	16384
ε_2^u	1.37E−2	2.54E−3	6.61E−4	1.64E−4	3.83E−5
Rate	–	2.43	1.94	2.01	2.10
ε_2^F	1.76E−1	7.11E−2	3.28E−2	1.59E−2	7.91E−3
Rate	–	1.31	1.12	1.04	1.01

Tables 4.2 and 4.3 give the numerical results of the methods (II) and (III). We can see that the error of these methods are not remarkably decreased as the number of cells is increased. Hence, the methods of inverse distance weighting and simple weighting fail to convergence as the number of cells is increased, which demonstrate that the methods of eliminating the cell vertex unknowns will remarkably affect the accuracy of scheme, and in some cases they lead to convergence failure.

Table 4.4 gives L_2-norm of the error on random triangular meshes. It shows that our method (I) also obtains second-order convergence rate for the solution and first-order convergence rate for the flux on random triangular meshes.

Tables 4.5 and 4.6 give the numerical results of the methods (II) and (III) on random triangular meshes, which show that the error of these methods are not remarkably decreased as the number of cells is increased. Hence, these methods also fail to convergence as the number of cells is increased.

From these experiments, we conclude that our method (I) of eliminating the cell vertex unknowns mentioned in Section 2.5 is robust.

Table 4.2
Convergence results for the elliptic problem on random quadrilateral meshes (method II)

The number of cell	64	256	1024	4096	16384
ε_2^u	1.22E−2	1.46E−2	1.67E−2	1.72E−2	1.74E−2
ε_2^F	3.66E−1	2.27E−1	2.79E−1	2.71E−1	2.79E−1

Table 4.3
Convergence results for the elliptic problem on random quadrilateral meshes (method III)

The number of cell	64	256	1024	4096	16384
ε_2^u	3.48E−2	3.26E−2	3.85E−2	3.87E−2	3.88E−2
ε_2^F	8.09E−1	4.67E−1	6.26E−1	5.95E−1	6.10E−1

Table 4.4
Convergence results for the elliptic problem on random triangular meshes (method I)

The number of cell	128	512	2048	8192	32768
ε_2^u	1.06E−2	2.23E−3	6.43E−4	1.70E−4	4.48E−5
Rate	–	2.25	1.79	1.92	1.92
ε_2^F	1.27E−1	5.91E−2	2.56E−2	9.36E−3	4.16E−3
Rate	–	1.10	1.21	1.45	1.17

Table 4.5
Convergence results for the elliptic problem on random triangular meshes (method II)

The number of cell	128	512	2048	8192	32768
ε_2^u	1.52E−2	9.55E−3	1.19E−2	1.19E−2	1.20E−2
ε_2^F	1.35E−1	5.17E−2	4.73E−2	3.18E−2	2.30E−2

Table 4.6
Convergence results for the elliptic problem on random triangular meshes (method III)

The number of cell	128	512	2048	8192	32768
ε_2^u	3.51E−2	2.08E−2	2.67E−2	2.64E−2	2.67E−2
ε_2^F	7.14E−1	3.66E−1	5.20E−1	4.95E−1	5.05E−1

4.6.2. The parabolic problems

Then consider the problem (3.1)–(3.3) with Dirichlet boundary condition in the unit square $\Omega = (0,1)^2$. Let $\kappa(x,t) = 1$, $f = 0$, $g = 0$ and $\varphi = \sin(\pi x)\sin(\pi y)$. The exact solution is $u = e^{-2\pi^2 t}\sin(\pi x)\sin(\pi y)$.

We test our method (I) on random quadrilateral meshes and random triangular meshes, respectively. Table 4.7 and 4.8 give L_2-norm of the error on random quadrilateral meshes and random triangular meshes, respectively. In these computations, we take $T = 0.1$ and $\Delta t = 1/N$, where N is the number of cell. We take this small time step in order not to affect the spatial accuracy. From these tables, we can see that our method gives second-order convergence rate for both the solution and the flux.

4.6.3. Discontinuous coefficient problem

Next, we consider a discontinuous coefficient problem (see [4]). The conductivity κ is discontinuous and given by

$$\kappa = \begin{cases} \kappa_1, & x \leqslant \frac{1}{2}, \\ \kappa_2, & x > \frac{1}{2}. \end{cases}$$

We set $f = 0$ and the exact solution is

$$u(x,y) = \begin{cases} a + bx + cy, & x \leqslant \frac{1}{2}, \\ a + b\frac{\kappa_2 - \kappa_1}{2\kappa_2} + b\frac{\kappa_1}{\kappa_2}x + cy, & x > \frac{1}{2}. \end{cases}$$

This solution and its normal component of flux are continuous at $x = \frac{1}{2}$, while tangential component of flux is $\kappa_1 c$ on the left side of the interface and $\kappa_2 c$ on the right side of the interface.

The numerical experiments use $\kappa_1 = 4$, $\kappa_2 = 1$, $a = b = c = 1$ and the random meshes shown in Fig. 4.19. We apply Dirichlet boundary conditions which are directly deduced from the analytical solution.

The calculated isolines of the numerical solution on random meshes are shown in Fig. 4.20. The L_2-norm of error is 9.27E−16. For this test problem, the L_2-norm of the scheme in [4] is 2.03E−3 on random meshes (10×10), while the asymptotic errors obtained by our scheme are close to zero. Hence, our scheme reproduces exactly the linear solution.

Table 4.7
Convergence results for the parabolic problem on random quadrilateral meshes

The number of cell	64	256	1024	4096
ε_2^u	2.30E−2	5.63E−3	1.40E−3	3.53E−4
Rate	–	2.03	2.01	1.99
ε_2^F	2.06E−1	5.20E−2	1.31E−2	3.39E−3
Rate	–	1.99	1.99	1.95

Table 4.8
Convergence results for the parabolic problem on random triangular meshes

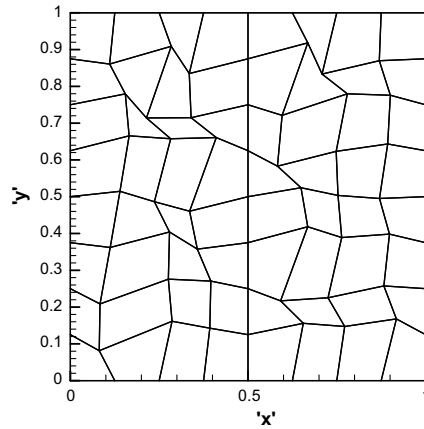| The number of cell	128	512	2048	8192
ε_2^u	2.18E−2	5.32E−3	1.31E−3	3.31E−4
Rate	–	2.03	2.02	1.98
ε_2^F	2.82E−1	7.12E−2	1.77E−2	4.50E−3
Rate	–	1.99	2.01	1.98

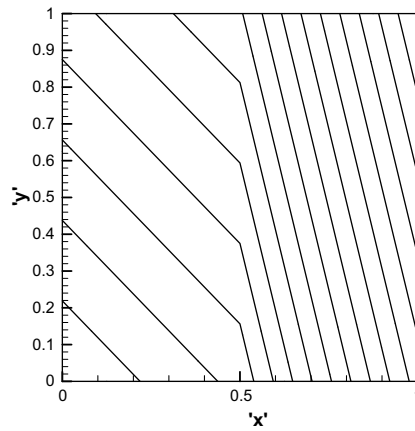Fig. 4.19. The random meshes with a discontinuity at $x = \frac{1}{2}$ (8×8).

Fig. 4.20. Isolines for the discontinuous coefficient problem.

4.6.4. Highly anisotropic tensors problem

In this subsection, we give the convergence analysis of the method for highly anisotropic tensors.

Consider the problem (2.1), (2.2) with Dirichlet boundary condition in the unit square $\Omega = (0,1)^2$. Let

$$\kappa = \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix},$$

where $\kappa_1 = 10$ and $\kappa_2 = 0.01$. The solution is chosen to be $u(x,y) = \sin(\pi x)\sin(\pi y)$.

We test our scheme on random quadrilateral meshes (see Fig. 4.17). Table 4.9 gives L_2-norm of the error between exact solution and numerical solution and L_2-norm of the error between exact flux and numerical flux

Table 4.9
Convergence results for highly anisotropic tensors problem on random quadrilateral meshes

The number of cell	64	256	1024	4096	16384
ε_2^u	1.20E−2	3.51E−3	1.21E−3	3.12E−4	1.02E−4
Rate	–	1.77	1.54	1.96	1.61
ε_2^F	8.90E−1	3.13E−1	1.50E−1	6.07E−2	2.77E−2
Rate	–	1.51	1.06	1.31	1.13

on random quadrilateral meshes. From this table, one can know that our method obtains the convergence rate larger than one and a half-order for the solution and the first-order convergence rate for the flux.

4.6.5. Problem with mixed boundary conditions

Consider the following diffusion problem (see [19]):

$$\frac{1}{v}\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left[D\frac{\partial u}{\partial x}\right] + \frac{\partial}{\partial y}\left[D\frac{\partial u}{\partial y}\right] + f.$$

The boundary conditions are that there is zero flux through the top and bottom boundaries and mixed or Robin boundary conditions on the left and right boundaries:

$$D\frac{\partial u}{\partial y} = 0 \quad \text{at } y = 0, \qquad D\frac{\partial u}{\partial y} = 0 \quad \text{at } y = 1,$$

$$u - 2D\frac{\partial u}{\partial x} = 0 \quad \text{at } x = 0, \qquad u + 2D\frac{\partial u}{\partial x} = 1 \quad \text{at } x = 1.$$

The initial condition is $u(x, y, 0) = 0$.

We consider the problem with $v = 300$, $D = \frac{1}{30}$, and $f = 0$, which has $1D$ linear steady-state solution $u = (x + 2D)/(1 + 4D)$. We take $\Delta t = 1.0E - 2$, $T = 1$, and compute this problem on the random meshes (see Fig. 4.21). The calculated isolines of the numerical solution on random mesh are shown in Fig. 4.22.
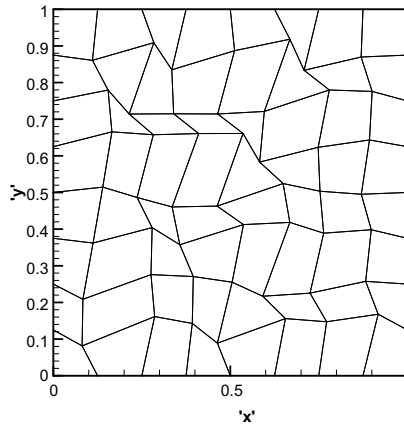

Fig. 4.21. The random meshes (8×8).
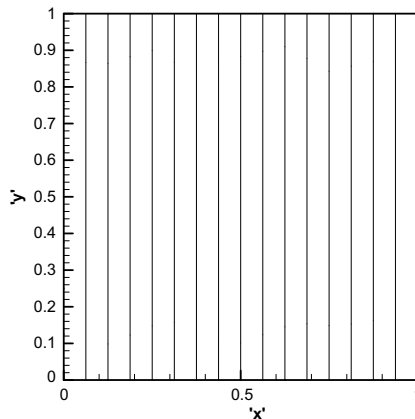

Fig. 4.22. Isolines on the random meshes.

The L_2-norm of error is 1.51e−15. It obviously that our scheme reproduces the steady-state solution exactly on this random meshes.

5. Conclusions

The new nonlinear FV scheme that we have constructed for solving diffusion equations is monotone on star-shaped polygonal meshes. This FV scheme is a development of the scheme proposed in [11]. The construction of the FV scheme on unstructured polygonal meshes is based on an adaptive strategy of discretization of flux. The resulting scheme satisfies the practical requirements mentioned in Section 1, without severe restrictions on meshes and diffusion coefficients and without a specific definition of collocation points. It follows that our scheme would be suitable for coupled radiation diffusion/hydrodynamics calculations on polygonal meshes. Numerical experiments demonstrate the ability of preserving positivity of the new nonlinear scheme and also show that the convergence rate of the scheme is about the same as that of some known linear FV schemes.

Acknowledgments

References

[1] I. Aavatsmark, An introduction to multipoint flux approximations for quadrilateral grids, Comput. Geosci. 6 (2002) 405–432.
[2] I. Aavatsmark, T. Barkve, O. Boe, T. Mannseth, Discretization on unstructured grids for inhomogeneous, anisotropic media, Part I: Derivation of the methods, SIAM. J. Sci. Comput. 19 (5) (1998) 1700–1716.
[3] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, New York, 1979.
[4] J. Breil, P.-H. Maire, A cell-centered diffusion scheme on two-dimensional unstructured meshes, J. Comput. Phys. 224 (2007) 785–823.
[5] F. Brezzi, K. Lipnikov, M. Shashkov, V. Simoncini, A new discretization methodology for diffusion problems on generalized polyhedral meshes, Comput. Meth. Appl. Mech. Eng. 196 (2007) 3682–3692.
[6] E. Burman, A. Ern, Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes, C. R. Acad. Sci. Paris, Ser. I 338 (2004) 641–646.
[7] A. Draganescu, T.F. Dupont, L.R. Scott, Failure of the discrete maximum principle for an elliptic finite element problem, Math. Comput. 74 (249) (2004) 1–23.
[8] J. Droniou, R. Eymard, A mixed finite volume scheme for anisotropic diffusion problems on any grid, Numer. Math. 105 (2006) 35–71.
[9] H. Hoteit, R. Mose, B. Philippe, Ph. Ackerer, J. Erhel, The maximum principle violations of the mixed-hybrid finite-element method applied to diffusion equations, Numer. Meth. Eng. 55 (12) (2002) 1373–1390.
[10] S. Korotov, M. Krizek, P. Neittaanmäki, Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle, Math. Comput. 70 (2000) 107–119.
[11] K. Lipnikov, M. Shashkov, D. Svyatskiy, Yu. Vassilevski, Monotone finite volume schemes for diffusion equations on unstructured triangular and shape-regular polygonal meshes, J. Comput. Phys. 227 (2007) 492–512.
[12] R. Liska, M. Shashkov, Enforcing the discrete maximum principle for linear finite element solutions of second-order elliptic problems, Commun. Comput. Phys. 3 (2008) 852–877.
[13] I.D. Mishev, Finite volume methods on voronoi meshes, Numer. Meth. Part. Differ. Equat. 12 (2) (1998) 193–212.
[14] J.M. Nordbotten, I. Aavatsmark, Monotonicity conditions for control volume methods on uniform parallelogram grids in homogeneous media, Comput. Geosci. 9 (2005) 61–72.
[15] J.M. Nordbotten, I. Aavatsmark, G.T. Eigestad, Monotonicity of control volume methods, Numer. Math. 106 (2007) 255–288.
[16] G.J. Pert, Physical constraints in numerical calculations of diffusion, J. Comput. Phys. 42 (1981) 20–52.
[17] C. Le Potier, Finite volume monotone scheme for highly anisotropic diffusion operators on unstructured triangular meshes, C. R. Acad. Sci. Paris, Ser. I 341 (2005) 787–792.
[18] Prateek Sharma, Gregory W. Hammett, Preserving monotonicity in anisotropic diffusion, J. Comput. Phys. 227 (2007) 123–142.
[19] M. Shashkov, S. Steinberg, Solving diffusion equations with rough coefficients in rough grids, J. Comput. Phys. 129 (1996) 383–405.
[20] Zhiqiang Sheng, Guangwei Yuan, A nine point scheme for the approximation of diffusion operators on distorted quadrilateral meshes, SIAM J. Sci. Comput. 30 (3) (2008) 1341–1361.
[21] D. Shepard, A two-dimensional interpolation function for irregularly spaced data, in: Proceedings of the 23d ACM National Conference, 517–524, ACM, NY, 1968.