Journal of Computational Mathematics
Vol.xx, No.x, 2023, 1–23.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2303-m2022-0139

MONOTONICITY CORRECTIONS FOR NINE-POINT SCHEME OF DIFFUSION EQUATIONS*

Wang Kong
School of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
Key Laboratory of Mathematical Modelling and High Performance Computing of Air Vehicles (NUAA), MIIT, Nanjing 211106, China
Email: wkong@nuaa.edu.cn
Zhenying Hong
Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, Beijing 100088, China
Email: zyhong@iapcm.ac.cn
Guangwei Yuan
Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, Beijing 100088, China
Email: yuan_guangwei@iapcm.ac.cn
Zhiqiang Sheng[1)]
Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, Beijing 100088, China
HEDPS, Center for Applied Physics and Technology, Peking University, Beijing 100871, China
Email: sheng_zhiqiang@iapcm.ac.cn

Abstract

In this paper, we present a nonlinear correction technique to modify the nine-point scheme proposed in [SIAM J. Sci. Comput., 30:3 (2008), 1341–1361] such that the resulted scheme preserves the positivity. We first express the flux by the cell-centered unknowns and edge unknowns based on the stencil of the nine-point scheme. Then, we use a nonlinear combination technique to get a monotone scheme. In order to obtain a cell-centered finite volume scheme, we need to use the cell-centered unknowns to locally approximate the auxiliary unknowns. We present a new method to approximate the auxiliary unknowns by using the idea of an improved multi-points flux approximation. The numerical results show that the new proposed scheme is robust, can handle some distorted grids that some existing finite volume schemes could not handle, and has higher numerical accuracy than some existing positivity-preserving finite volume schemes.

Mathematics subject classification: 52B10, 65D18, 68U05, 68U07.
Key words: Monotonicity corrections, Diffusion equation, Improved MPFA, Distorted meshes.

1. Introduction

In the numerical simulation of inertial confinement fusion, reservoir simulation and astrophysics, we often need to numerically solve the diffusion equation on the distorted meshes.

* Received May 26, 2022 / Revised version received February 22, 2023 / Accepted March 14, 2023 / Published online October 30, 2023 /

[1)] Corresponding author

To avoid non-physical oscillations in the numerical solution, we need to choose the positivity-preserving schemes. It was shown in [21] that there is no locally conservative, unconditionally positivity-preserving, linear nine-point scheme such that the discretization has a second-order accuracy and exactly reproduces the linear solution on the distorted meshes. To get a monotone scheme, some pre- and post-processing methods are proposed in [1, 4, 12, 17, 18, 20, 23, 30–32].

On the other hand, Le Potier presents a nonlinear monotone finite volume scheme for time-dependent anisotropic diffusion problems on unstructured triangular meshes [13]. As far as we know, there have many papers so far, e.g. [3, 5, 10, 15, 19, 22, 25–27, 34], devoted to positivity-preserving nonlinear finite volume schemes to solve diffusion equations on distorted meshes. Besides, the nonlinear finite volume schemes which satisfy the stricter requirement – the discrete maximum principle, have been discussed in [2, 6, 7, 9, 14, 16, 28].

Radiation diffusion calculation occupies an important position in solving actual radiation fluid mechanics problems. In the calculation of multi-medium Lagrange radiation fluid, the flow of fluid will cause the distortion of the grid. Triangular meshes have good adaptability to the complex calculation areas, so they are often used in Lagrange radiation fluid calculations. However, the previously proposed positive-preserving finite volume schemes [25–27, 34] cannot handle highly distorted triangular meshes well, such as the triangular Kershaw meshes showed in Section 4.1. In this paper, we will propose a new nonlinear positive-preserving finite volume scheme, which can handle highly distorted triangular meshes better.

The monotone schemes in [25, 27, 34] adaptively select discrete templates, which can adapt to various large deformed meshes. However, when a certain cell has a large degree of distortion, the expression of discrete flux on some edge of the cell may not include the physical quantity on the edge. In this case, although the discrete flux design is well adapted to the geometric deformation of the cell, it fails to directly reflect the change of physical quantities on some edge of the cell, which may affect the discrete accuracy of the discrete normal flow. Besides, the expression of the discrete flux proposed in [26] contains the unknown at the midpoint of the edge. Hence, the discrete flux design can directly reflect the change of physical quantities on the edge. However, the construction process of the scheme is relatively complicated, and involves the elimination of two types of auxiliary unknowns: the vertex unknowns and the edges unknowns.

In this paper, we construct a linear flux on each cell-edge as [24, 33], which contains the unknown at the midpoint of the edge. And then, we deal with the tangential difference along the edge in the discrete flux to get a new nonlinear expression of the discrete flux that includes the cell-centered unknown and some edge unknowns. The construction process of the new proposed scheme is relatively simple, and the new expression of the discrete flux is not only suitable for the distortion of the mesh, but also directly reflects the change of the physical quantity on the edge.

The auxiliary edge unknowns should be locally approximated with the surrounding cell-centered unknowns. For a mesh with a small degree of distortion, we can use the method in [25] to approximate the auxiliary edge unknowns. However, it is found through numerical experiments that the absolute values of the interpolation coefficients obtained by the method in [25] on some distorted triangular meshes are often large, resulting in an unstable scheme. We present a new method to approximate the auxiliary edge unknowns inspired by the idea of an improved multi-points flux approximation.

However, the approximate auxiliary unknowns obtained by this new method may be negative even if the surrounding cell-centered unknowns are non-negative. We use an idea similar to

[26] to assure the resulting nonlinear scheme is monotone by introducing two non-negative parameters when constructing the conservative flux. The new proposed scheme can deal with some distorted grids that the previous finite volume schemes could not handle well, such as the triangular Kershaw meshes.

The article is organized as follows. In Section 2, we introduce the nonlinear correction technique to modify the nine-point scheme. In Section 3, we give a new approach to eliminate the auxiliary unknowns. In Section 4, we present some numerical results to test the monotonicity and accuracy of the new scheme. Finally, we give some conclusions in Section 5.

2. Construction of Scheme

2.1. Notations

We consider the numerical solution of the diffusion equation on an open bounded polygonal domain Ω in $\mathbb{R}^2$

$$\begin{cases} -\nabla\cdot\big(\kappa(\mathrm{x})\nabla u(\mathrm{x})\big)=f(\mathrm{x}), & \mathrm{x}\in\Omega,\\ u(\mathrm{x})=g(\mathrm{x}), & \mathrm{x}\in\Gamma_1,\\ \alpha(\mathrm{x})\kappa(\mathrm{x})\dfrac{\partial u}{\partial\vec{n}}(\mathrm{x})+\beta(\mathrm{x})u(\mathrm{x})=h(\mathrm{x}), & \mathrm{x}\in\Gamma_2. \end{cases} \tag{2.1}$$

Here, the boundary $\partial\Omega$ is divided into two disjoint parts: Γ_1 with a Dirichlet boundary condition and Γ_2 with a Robin boundary condition where the non-negative parameters $\alpha(\mathrm{x})$ and $\beta(\mathrm{x})$ do not vanish at the same time. Moreover, the diffusion tensor $\kappa(\mathrm{x})$ is piecewise smooth and satisfies the following uniform ellipticity condition:

$$\exists\,\lambda_1,\lambda_2>0,\quad \lambda_1|\xi|^2\le\xi^T\kappa(\mathrm{x})\xi\le\lambda_2|\xi|^2,\quad \forall\,\mathrm{x}\in\Omega,\quad \xi\in\mathbb{R}^2.$$

We construct a second-order positivity-preserving finite volume scheme for the diffusion equation (2.1) on distorted meshes. On the discontinuity of the diffusion tensor, we request that the solution $u(\mathrm{x})$ and the normal flux $\kappa(\mathrm{x})(\partial u/\partial\vec{n})(\mathrm{x})$ are continuous, but the gradient $\nabla u(\mathrm{x})$ is discontinuous. We require the discontinuity of the diffusion tensor matches some mesh edges.

$\mathcal{T}$ denotes the set of all cells, and $\mathcal{E}$ is the set of all edges. We denote $\mathcal{P}_{int}$ as the set of the cell centers and $\mathcal{P}_{out}$ as the set of the midpoints for the boundary edges. Besides, we choose $h=\sup_{K\in\mathcal{T}}\mathrm{diam}(K)$, where $\mathrm{diam}(K)$ is the diameter of K.

For the selected cell K, its vertices are numbered counterclockwise by $\{P_k\}_{k=1}^m$. The cell center is still denoted K, the edges are denoted as $\{\sigma_k=\overline{P_kP_{k+1}}\}_{k=1}^m$ with $P_{m+1}=P_1$ and the midpoint of σ_k is M_k. Integrating the diffusion equation (2.1) on K, we can obtain

$$-\int_K\nabla\cdot\big(\kappa(\mathrm{x})\nabla u(\mathrm{x})\big)d\mathrm{x}=\int_K f(\mathrm{x})d\mathrm{x},$$

and then it follows from the divergence theorem to obtain

$$\sum_{k=1}^m\mathcal{F}_{K,\sigma_k}=\int_K f(\mathrm{x})d\mathrm{x},$$

where $\mathcal{F}_{K,\sigma_k}$ is the continuous normal flux on edge σ_k

$$\mathcal{F}_{K,\sigma_k}=-\int_{\sigma_k}\kappa(\mathrm{x})\nabla u(\mathrm{x})\cdot\vec{n}_{K,\sigma_k}d\Gamma=-\int_{\sigma_k}\nabla u(\mathrm{x})\cdot\kappa(\mathrm{x})^T\vec{n}_{K,\sigma_k}d\Gamma. \tag{2.2}$$

2.2. The non-conservative discrete flux

As shown in Fig. 2.1, we denote an edge of the cell K by σ, the endpoints of σ by A and B, the midpoint by I. If $\theta_{K,\sigma}$ is denoted as the angle between $\vec{\tau}_{KI}$ and $\vec{n}_{K,\sigma}$, then according to the construction of the discrete flux on σ in [24], we have

$$\mathcal{F}_{K,\sigma} = \alpha_{K,\sigma}\big[u(A) - u(B)\big] - \frac{|A-B|}{|I-K|}\beta_{K,\sigma}\big[u(I) - u(K)\big] + \mathcal{O}(h^2) \tag{2.3}$$

with the coefficients are defined by

$$\beta_{K,\sigma} = \frac{1}{\cos\theta_{K,\sigma}}\vec{n}_{K,\sigma}\cdot\Big(\kappa_K^T\vec{n}_{K,\sigma}\Big) > 0, \quad \alpha_{K,\sigma} = \frac{1}{\cos\theta_{K,\sigma}}\vec{\nu}_{KI}\cdot\Big(\kappa_K^T\vec{n}_{K,\sigma}\Big),$$

where we denote $\kappa_K = \kappa(K)$ and

$$\vec{\nu}_{KI} = \sin\theta_{K,\sigma}\vec{n}_{K,\sigma} - \cos\theta_{K,\sigma}\vec{\tau}_{BA}.$$

Hence, we can obtain a discrete approximation of the continuous flux $\mathcal{F}_{K,\sigma}$ as follows:

$$F_{1,\sigma} = -\tau_{K,\sigma}\left[\big(u(I) - u(K)\big) - D_{K,\sigma}\big(u(A) - u(B)\big)\right], \tag{2.4}$$

where

$$\tau_{K,\sigma} = \frac{|A-B|}{|I-K|}\beta_{K,\sigma} > 0, \quad D_{K,\sigma} = \frac{|I-K|\alpha_{K,\sigma}}{|A-B|\beta_{K,\sigma}}.$$

Assume L is the neighbor cell of K with $\sigma = K\bigcap L$. Similarly, we can give a discrete expression of the other flux $\mathcal{F}_{L,\sigma}$ on σ

$$F_{2,\sigma} = -\tau_{L,\sigma}\left[\big(u(I) - u(L)\big) - D_{L,\sigma}\big(u(B) - u(A)\big)\right] \tag{2.5}$$

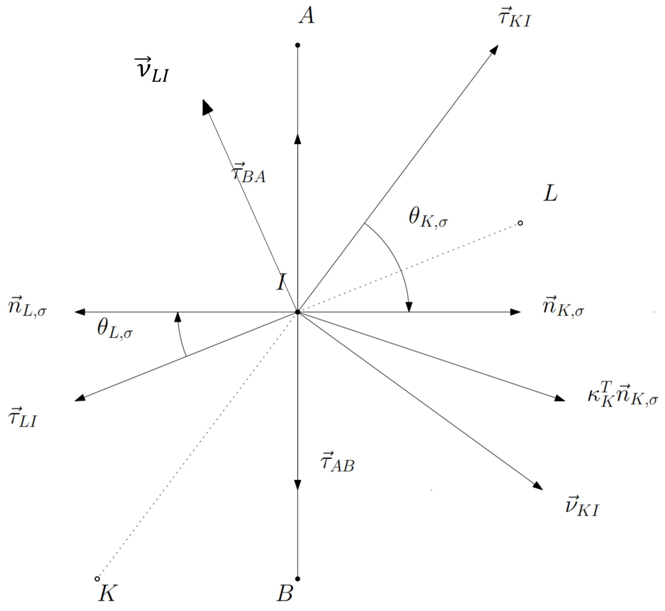

Fig. 2.1. The stencil of the discrete flux.

with the coefficients defined as follows:

$$\tau_{L,\sigma} = \frac{|A-B|\vec{n}_{L,\sigma}\cdot\left(\kappa_L^T\vec{n}_{L,\sigma}\right)}{|I-L|\cos\theta_{L,\sigma}} > 0, \quad D_{L,\sigma} = \frac{|I-L|\vec{\nu}_{LI}\cdot\left(\kappa_L^T\vec{n}_{L,\sigma}\right)}{|A-B|\vec{\tau}_{L,\sigma}\cos\theta_{L,\sigma}},$$

where $\theta_{L,\sigma}$ is the angle between $\vec{\tau}_{LI}$ and $\vec{n}_{L,\sigma}$, the diffusion tensor $\kappa_L = \kappa(L)$ and the vector

$$\vec{\nu}_{LI} = \sin\theta_{L,\sigma}\vec{n}_{L,\sigma} - \cos\theta_{L,\sigma}\vec{\tau}_{AB}.$$

In general, the coefficients in the above discrete fluxes (2.4) and (2.5) satisfy the following relationship:

$$\tau_{K,\sigma} \neq \tau_{L,\sigma}, \quad D_{K,\sigma} \neq D_{L,\sigma}.$$

Thus, the discrete flux defined above is non-conservative and satisfies

$$F_{1,\sigma} + F_{2,\sigma} = \mathcal{O}(h^2).$$

2.3. The conservative discrete flux on interior edges

It is found through numerical experiments that the accuracy of the existing positivity-preserving finite volume scheme is lower than that of the nine-point scheme. Hence, we will construct a positivity-preserving scheme starting from the discrete flux (2.4) and (2.5) of the nine-point scheme on σ.

In order to obtain a new discrete flux containing only the cell-centered unknowns and the edge unknowns, we can use an ideal from [33] to approximate the tangential difference $[u(A) - u(B)]$ contained in $F_{1,\sigma}$ (or $F_{2,\sigma}$) according to the sign of $D_{K,\sigma}$ (or $D_{L,\sigma}$). As shown in Fig. 2.2, the tangential difference $[u(A) - u(B)]$ will be approximated by the multiple of the directional derivative of $u(\mathrm{x})$ in a direction parallel to $\vec{\tau}_{AB}$ or at K (or L).

For the tangential difference $[u(A) - u(B)]$ in $F_{1,\sigma}$, we can find a point K' on ∂K such that $\vec{\tau}_{K'K} \parallel \mathrm{sgn}(D_{K,\sigma})\vec{\tau}_{BA}$, and then

$$\mathrm{sgn}(D_{K,\sigma})\frac{u(A)-u(B)}{|A-B|} = \mathrm{sgn}(D_{K,\sigma})\nabla u(I)\cdot\vec{\tau}_{BA} + \mathcal{O}(h) = \nabla u(K)\cdot\vec{\tau}_{K'K} + \mathcal{O}(h). \quad (2.6)$$

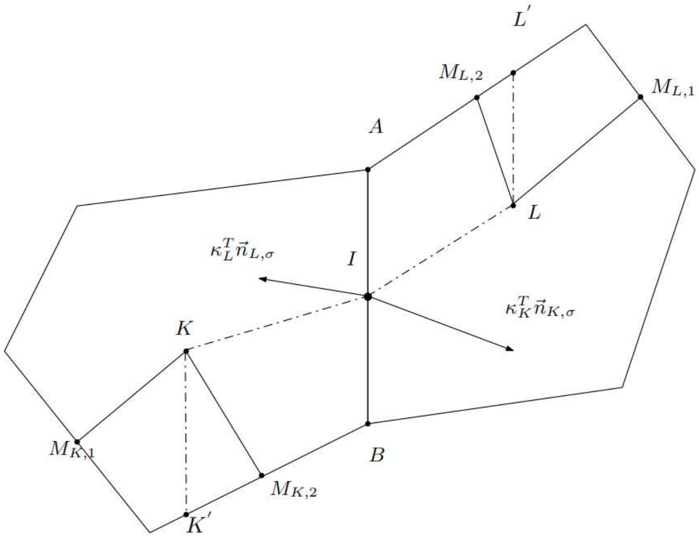
L'
$M_{L,2}$
$M_{L,1}$
A
L
$\kappa_L^T\vec{n}_{L,\sigma}$
I
$\kappa_K^T\vec{n}_{K,\sigma}$
K
$M_{K,1}$
B
$M_{K,2}$
K'

Fig. 2.2. The stencil of the conservative flux on interior edge.

There are two adjacent edge midpoints such that the ray originated at cell-center K along the direction $\vec{\tau}_{KK'}$ intersects the segment $\overline{M_{K,1}M_{K,2}}$. If we denote the angle between $\vec{\tau}_{K'K}$ and $\vec{\tau}_{M_{K,1}K}$ as $\theta_{K,1}$, and the one between $\vec{\tau}_{K'K}$ and $\vec{\tau}_{M_{K,2}K}$ as $\theta_{K,2}$, then $\vec{\tau}_{K'K}$ can be decomposed by

$$\vec{\tau}_{K'K} = \frac{\sin\theta_{K,2}}{\sin\theta_K}\vec{\tau}_{M_{K,1}K} + \frac{\sin\theta_{K,1}}{\sin\theta_K}\vec{\tau}_{M_{K,2}K} = \frac{1}{|A-B|}(\omega_{K,1}\vec{\tau}_{M_{K,1}K} + \omega_{K,2}\vec{\tau}_{M_{K,2}K}) \tag{2.7}$$

with

$$\theta_K = \theta_{K,1} + \theta_{K,2}, \quad 0 \le \theta_{K,1} < \pi, \quad 0 \le \theta_{K,2} < \pi.$$

Here, we assume that $\theta_{K,1}, \theta_{K,2}$ cannot be zero at the same time and $\theta_K < \pi$, and then we can get

$$\omega_{K,1} \ge 0, \quad \omega_{K,2} \ge 0.$$

Next, substituting (2.7) into (2.6) gives

$$F_{1,\sigma} = a_{K,\sigma}u(K) - \tilde{c}_{K,\sigma} + \mathcal{O}(h^2),$$

where

$$\begin{aligned} a_{K,\sigma} &= \tau_{K,\sigma} + \tau_{K,\sigma}|D_{K,\sigma}|\left(\frac{\omega_{K,1}}{|K-M_{K,1}|} + \frac{\omega_{K,2}}{|K-M_{K,2}|}\right), \\ \tilde{c}_{K,\sigma} &= \tau_{K,\sigma}\left[u(I) + \frac{|D_{K,\sigma}|\omega_{K,1}}{|K-M_{K,1}|}u(M_{K,1}) + \frac{|D_{K,\sigma}|\omega_{K,2}}{|K-M_{K,2}|}u(M_{K,2})\right]. \end{aligned}$$

If we denote the approximate value of $u(P)$ as u_P, we set

$$\bar{F}_{K,\sigma} \triangleq a_{K,\sigma}u_K - c_{K,\sigma}, \tag{2.8}$$

where

$$c_{K,\sigma} = \tau_{K,\sigma}\left(u_I + \frac{|D_{K,\sigma}|\omega_{K,1}}{|K-M_{K,1}|}u_{M_{K,1}} + \frac{|D_{K,\sigma}|\omega_{K,2}}{|K-M_{K,2}|}u_{M_{K,2}}\right).$$

Similarly, we can get the other discrete flux on σ as follows:

$$\bar{F}_{L,\sigma} = a_{L,\sigma}u_L - c_{L,\sigma} \tag{2.9}$$

with the coefficients $a_{L,\sigma}$ and $c_{L,\sigma}$ defined as above.

Due to the error introduced in the process of approximating the tangential difference term, we can obtain that

$$\bar{F}_{K,\sigma} + \bar{F}_{L,\sigma} = \mathcal{O}(h^2),$$

which yields, the discrete fluxes $\bar{F}_{K,\sigma}$ and $\bar{F}_{L,\sigma}$ constructed above are non-conservative. To maintain local conservation, we can use the convex combination of the above non-conservative discrete fluxes to construct a discrete normal flux on σ

$$F_{K,\sigma} = \mu_{1,\sigma}\bar{F}_{K,\sigma} - \mu_{2,\sigma}\bar{F}_{L,\sigma}, \tag{2.10}$$

$$F_{L,\sigma} = -\mu_{1,\sigma}\bar{F}_{K,\sigma} + \mu_{2,\sigma}\bar{F}_{L,\sigma}, \tag{2.11}$$

where $\mu_{1,\sigma}, \mu_{2,\sigma}$ are coefficients satisfying $\mu_{1,\sigma} + \mu_{2,\sigma} = 1$ to be determined later. Substituting (2.8) and (2.9) into (2.10) gives

$$\begin{aligned} F_{K,\sigma} &= \mu_{1,\sigma}a_{K,\sigma}u_K - \mu_{2,\sigma}a_{L,\sigma}u_L - \mu_{1,\sigma}c_{K,\sigma} + \mu_{2,\sigma}c_{L,\sigma} \\ &= \mu_{1,\sigma}[a_{K,\sigma} + \mathrm{sgn}(u_K)\omega_{K,\sigma}]u_K - \mu_{2,\sigma}[a_{L,\sigma} + \mathrm{sgn}(u_L)\omega_{L,\sigma}]u_L \\ &\quad - \mu_{1,\sigma}(c_{K,\sigma} + \omega_{K,\sigma}|u_K|_\delta) + \mu_{2,\sigma}(c_{L,\sigma} + \omega_{L,\sigma}|u_L|_\delta) \\ &\quad + \mu_{1,\sigma}\omega_{K,\sigma}(|u_K|_\delta - |u_K|) - \mu_{2,\sigma}\omega_{L,\sigma}(|u_L|_\delta - |u_L|), \end{aligned} \tag{2.12}$$

where $\omega_{K,\sigma}$, $\omega_{L,\sigma}$ are non-negative parameters to be determined and $|\cdot|_\delta$ is defined by

$$|w|_\delta = \begin{cases} |w|, & |w| \geq \delta, \\ \delta, & |w| < \delta. \end{cases}$$

In order to maintain the precision of the numerical scheme, we choose the parameters δ as follows:

$$\delta = Ch^2,$$

where the constant C is generally a positive real number not greater than 10. If we truncate the last two terms in (2.12), we obtain an approximate flux

$$\begin{aligned} F_{K,\sigma}^{\delta} &= \mu_{1,\sigma}[a_{K,\sigma} + \mathrm{sgn}(u_K)\omega_{K,\sigma}]u_K - \mu_{2,\sigma}[a_{L,\sigma} + \mathrm{sgn}(u_L)\omega_{L,\sigma}]u_L \\ &\quad - \mu_{1,\sigma}(c_{K,\sigma} + \omega_{K,\sigma}|u_K|_\delta) + \mu_{2,\sigma}(c_{L,\sigma} + \omega_{L,\sigma}|u_L|_\delta). \end{aligned} \tag{2.13}$$

To get a two-point discrete flux on σ, the third and fourth term of the above expression must be offset, which yields

$$\begin{cases} \mu_{1,\sigma} + \mu_{2,\sigma} = 1, \\ \mu_{1,\sigma}(c_{K,\sigma} + \omega_{K,\sigma}|u_K|_\delta) - \mu_{2,\sigma}(c_{L,\sigma} + \omega_{L,\sigma}|u_L|_\delta) = 0. \end{cases} \tag{2.14}$$

If $c_{K,\sigma} + c_{L,\sigma} = 0$, we choose $\omega_{K,\sigma} = \omega_{L,\sigma} = 0$ and set

$$\mu_{1,\sigma} = \mu_{2,\sigma} = \frac{1}{2},$$

else if $c_{K,\sigma} + c_{L,\sigma} \neq 0$, we can choose

$$\begin{aligned} \mu_{1,\sigma} &= \frac{c_{L,\sigma} + \omega_{L,\sigma}|u_L|_\delta}{c_{K,\sigma} + c_{L,\sigma} + \omega_{K,\sigma}|u_K|_\delta + \omega_{L,\sigma}|u_L|_\delta}, \\ \mu_{2,\sigma} &= \frac{c_{K,\sigma} + \omega_{K,\sigma}|u_K|_\delta}{c_{K,\sigma} + c_{L,\sigma} + \omega_{K,\sigma}|u_K|_\delta + \omega_{L,\sigma}|u_L|_\delta}. \end{aligned}$$

To preserve positivity, the coefficients $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$ should be non-negative. When using the improved MPFA introduced in next section to approximate the edge unknown, even if the cell-centered unknowns used in the approximation are all non-negative, it is also possible to obtain a negative edge unknown, and then give a negative $c_{K,\sigma}$ or $c_{L,\sigma}$. When $c_{K,\sigma}$ ($c_{L,\sigma}$) is negative, we can find a positive parameter $\omega_{K,\sigma}$ ($\omega_{L,\sigma}$) such that

$$\omega_{K,\sigma}|u_K|_\delta + c_{K,\sigma} > 0, \quad \omega_{L,\sigma}|u_L|_\delta + c_{L,\sigma} > 0. \tag{2.15}$$

Here, we can choose

$$\omega_{K,\sigma} > \max\left\{0, -\frac{c_{K,\sigma}}{|u_K|_\delta}\right\}, \quad \omega_{L,\sigma} > \max\left\{0, -\frac{c_{L,\sigma}}{|u_L|_\delta}\right\}.$$

If σ is an interior edge, we can give the conservative discrete flux on σ

$$F_{K,\sigma}^{\delta} = A_{K,\sigma}u_K - B_{K,\sigma}u_L, \tag{2.16}$$

$$F_{L,\sigma}^{\delta} = A_{L,\sigma}u_L - B_{L,\sigma}u_K, \tag{2.17}$$

where the coefficients are defined by

$$\begin{aligned} A_{K,\sigma} &= B_{L,\sigma} = \mu_{1,\sigma}[a_{K,\sigma} + \mathrm{sgn}(u_K)\omega_{K,\sigma}], \\ A_{L,\sigma} &= B_{K,\sigma} = \mu_{2,\sigma}[a_{L,\sigma} + \mathrm{sgn}(u_L)\omega_{L,\sigma}]. \end{aligned}$$

If the unknowns $u_K, u_L \geq 0$, we can obtain that

$$A_{K,\sigma} = B_{L,\sigma} > 0, \quad A_{L,\sigma} = B_{K,\sigma} > 0.$$

2.4. The conservative discrete flux on boundary edges

As shown in Fig. 2.3, if σ is a boundary edge, we denote the midpoint of σ as K. The continuous normal flux on σ is defined as follows:

$$\mathcal{F}_{K,\sigma} = -\int_\sigma \nabla u(\mathrm{x}) \cdot \kappa(\mathrm{x})^T \vec{n}_{K,\sigma} d\Gamma \equiv \int_\sigma \nabla u(\mathrm{x}) \cdot \kappa^T(\mathrm{x}) \vec{n}_{L,\sigma} d\Gamma = -\mathcal{F}_{L,\sigma}, \tag{2.18}$$

where L is denoted as the center of cell that σ belongs to. If the endpoints of σ are denoted as A and B, we can discretize the continuous normal flux $\mathcal{F}_{L,\sigma}$ by

$$\mathcal{F}_{L,\sigma} = -\tau_\sigma \big[\big(u(K) - u(L)\big) - D_\sigma \big(u(B) - u(A)\big) \big] + \mathcal{O}(h^2)$$

with $\tau_\sigma \triangleq \tau_{L,\sigma}$ and $D_\sigma \triangleq D_{L,\sigma}$ defined as in [24].

According to (2.18), the flux $\mathcal{F}_{K,\sigma}$ defined above can be discretized by

$$\mathcal{F}_{K,\sigma} = -\tau_\sigma \big[\big(u(L) - u(K)\big) - D_\sigma \big(u(A) - u(B)\big) \big] + \mathcal{O}(h^2).$$

Then we define the discrete normal flux on σ as

$$F_{1,\sigma} = -\tau_\sigma [(u_L - u_K) - D_\sigma (u_A - u_B)], \tag{2.19}$$

$$F_{2,\sigma} = -\tau_\sigma [(u_K - u_L) - D_\sigma (u_B - u_A)]. \tag{2.20}$$

Since K is the midpoint of σ, we can handle the tangential difference $(u_A - u_B)$ in $F_{1,\sigma}$ by

$$\mathrm{sgn}(D_\sigma)(u_A - u_B) = 2(u_K - u_P) + \mathcal{O}(h^2),$$

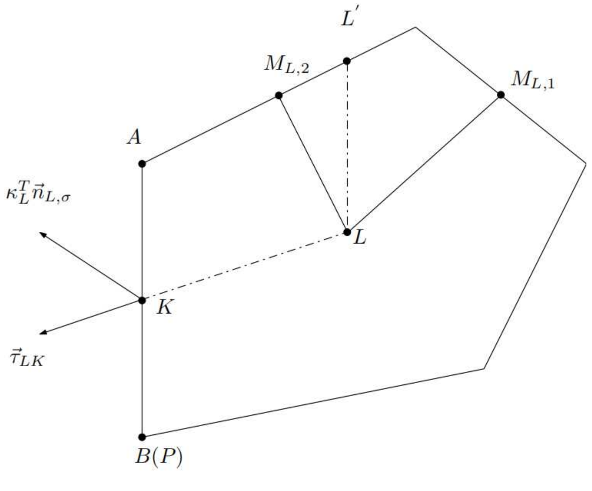

Fig. 2.3. The stencil of the conservative flux on boundary edge.

where

$$u_P = \begin{cases} u_B, & D_\sigma \geq 0, \\ u_A, & D_\sigma < 0. \end{cases}$$

Substituting it into (2.19) gives

$$F_{1,\sigma} = a_{K,\sigma} u_K - b_{K,\sigma} u_L - c_{K,\sigma} + \mathcal{O}(h^2) \triangleq \bar{F}_{K,\sigma} + \mathcal{O}(h^2) \tag{2.21}$$

with the coefficients defined as follows:

$$a_{K,\sigma} = \tau_\sigma (1 + 2|D_\sigma|), \quad b_\sigma = \tau_\sigma, \quad c_{K,\sigma} = 2\tau_\sigma |D_\sigma| u_P.$$

As for the tangential difference $(u_B - u_A)$ in $F_{2,\sigma}$, using the method proposed in the previous subsection gives

$$F_{2,\sigma} = a_{L,\sigma} u_L - b_{L,\sigma} u_K - c_{L,\sigma} + \mathcal{O}(h^2), \tag{2.22}$$

where

$$\begin{aligned} a_{L,\sigma} &= \tau_{L,\sigma} + \tau_{L,\sigma} |D_{L,\sigma}| \left(\frac{\omega_{L,1}}{|L - M_{L,1}|} + \frac{\omega_{L,2}}{|L - M_{L,2}|} \right), \\ b_{L,\sigma} &= \tau_{L,\sigma}, \\ c_{L,\sigma} &= \tau_{L,\sigma} \left[\frac{|D_{L,\sigma}| \omega_{L,1}}{|L - M_{L,1}|} u_{M_{L,1}} + \frac{|D_{L,\sigma}| \omega_{L,2}}{|L - M_{L,2}|} u_{M_{L,2}} \right]. \end{aligned}$$

Next, we construct the conservative discrete flux on σ by a convex combination of the above non-conservative discrete fluxes

$$F_{K,\sigma} = \mu_{1,\sigma} \bar{F}_{K,\sigma} - \mu_{2,\sigma} \bar{F}_{L,\sigma}, \tag{2.23}$$

$$F_{L,\sigma} = -\mu_{1,\sigma} \bar{F}_{K,\sigma} + \mu_{2,\sigma} \bar{F}_{L,\sigma}. \tag{2.24}$$

Substituting $\bar{F}_{K,\sigma}$ and $\bar{F}_{L,\sigma}$ into the above expression gives

$$\begin{aligned} F_{K,\sigma} &= (\mu_{1,\sigma} a_{K,\sigma} + \mu_{2,\sigma} b_{L,\sigma}) u_K - (\mu_{1,\sigma} b_{K,\sigma} + \mu_{2,\sigma} a_{L,\sigma}) u_L \\ &\quad - \mu_{1,\sigma} c_{K,\sigma} + \mu_{2,\sigma} c_{L,\sigma} \\ &= \big[\mu_{1,\sigma} \big(a_{K,\sigma} + \mathrm{sgn}(u_K) \omega_{K,\sigma} \big) + \mu_{2,\sigma} b_{L,\sigma} \big] u_K \\ &\quad - \big[\mu_{1,\sigma} b_{K,\sigma} + \mu_{2,\sigma} \big(\mathrm{sgn}(u_L) \omega_{L,\sigma} + a_{L,\sigma} \big) \big] u_L \\ &\quad - \mu_{1,\sigma} (c_{K,\sigma} + \omega_{K,\sigma} |u_K|_\delta) + \mu_{2,\sigma} (c_{L,\sigma} + \omega_{L,\sigma} |u_L|_\delta) \\ &\quad + \mu_{1,\sigma} \omega_{K,\sigma} (|u_K|_\sigma - |u_K|) - \mu_{2,\sigma} \omega_{L,\sigma} (|u_L|_\sigma - |u_L|), \end{aligned} \tag{2.25}$$

where $\omega_{K,\sigma}$, $\omega_{L,\sigma}$ are non-negative parameters to be determined. Choosing $\delta = Ch^2$ and truncating the last two terms in (2.25) yields an approximate flux

$$\begin{aligned} F^\delta_{K,\sigma} &= \big[\mu_{1,\sigma} \big(a_{K,\sigma} + \mathrm{sgn}(u_K) \omega_{K,\sigma} \big) + \mu_{2,\sigma} b_{L,\sigma} \big] u_K \\ &\quad - \big[\mu_{1,\sigma} b_{K,\sigma} + \mu_{2,\sigma} \big(\mathrm{sgn}(u_L) \omega_{L,\sigma} + a_{L,\sigma} \big) \big] u_L \\ &\quad - \mu_{1,\sigma} (c_{K,\sigma} + \omega_{K,\sigma} |u_K|_\delta) + \mu_{2,\sigma} (c_{L,\sigma} + \omega_{L,\sigma} |u_L|_\delta). \end{aligned} \tag{2.26}$$

To get a two-point discrete flux on σ, the third and fourth terms of the above expression must be removed, which yields

$$\begin{cases} \mu_{1,\sigma} + \mu_{2,\sigma} = 1, \\ \mu_{1,\sigma} (c_{K,\sigma} + \omega_{K,\sigma} |u_K|_\delta) - \mu_{2,\sigma} (c_{L,\sigma} + \omega_{L,\sigma} |u_L|_\delta) = 0. \end{cases} \tag{2.27}$$

Using the method mentioned in the previous subsection, we can select the appropriate convex combination coefficients $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$.

In summary, we can give the conservative discrete flux on σ

$$F^{\delta}_{K,\sigma} = A_{K,\sigma}u_K - B_{K,\sigma}u_L, \tag{2.28}$$

$$F^{\delta}_{L,\sigma} = A_{L,\sigma}u_L - B_{L,\sigma}u_K, \tag{2.29}$$

where the coefficients are defined by

$$\begin{aligned} A_{K,\sigma} = B_{L,\sigma} &= \mu_{1,\sigma}[a_{K,\sigma} + \text{sgn}(u_K)\omega_{K,\sigma}] + \mu_{2,\sigma}b_{L,\sigma}, \\ A_{L,\sigma} = B_{K,\sigma} &= \mu_{1,\sigma}b_{K,\sigma} + \mu_{2,\sigma}[a_{L,\sigma} + \text{sgn}(u_L)\omega_{L,\sigma}]. \end{aligned}$$

If the unknowns $u_K, u_L \geq 0$, we can obtain that

$$A_{K,\sigma} = B_{L,\sigma} > 0, \quad A_{L,\sigma} = B_{K,\sigma} > 0.$$

In order to obtain a complete finite volume scheme, we need to discretize the boundary conditions on σ.

Dirichlet boundary condition. If Dirichlet boundary condition is given on σ, direct discretization gives

$$u_K = g(K),$$

and then the discrete flux on σ is given by

$$F^{\delta}_{L,\sigma} = A_{L,\sigma}u_L - B_{L,\sigma}g(K) = A_{L,\sigma}u_L - a_{L,\sigma}. \tag{2.30}$$

Robin boundary condition. As for the Robin boundary condition, integrating it on σ gives

$$\int_{\sigma} \alpha(\mathrm{x})\kappa(\mathrm{x})\nabla u(\mathrm{x}) \cdot \vec{n} d\Gamma + \int_{\sigma} \beta(\mathrm{x})u(\mathrm{x})d\Gamma = \int_{\sigma} h(\mathrm{x})d\Gamma,$$

and then we can discretize the above formula as follows:

$$\alpha_K F^{\delta}_{K,\sigma} + \beta_K|\sigma|u_K = |\sigma|h(K),$$

that is,

$$(\alpha_K A_{K,\sigma} + \beta_K|\sigma|)\, u_K - \alpha_K B_{K,\sigma}u_L = |\sigma|h(K). \tag{2.31}$$

2.5. The finite volume scheme

By discretizing the flux and the boundary value conditions as above, we can get the following finite volume scheme for solving the diffusion equation (2.1):

$$\sum_{\sigma\subset\partial K} F^{\delta}_{K,\sigma} = m(K)f_K, \qquad \forall K \in \mathcal{P}_{int}, \tag{2.32}$$

$$u_K = g_K, \qquad \forall K \in \mathcal{P}_{out} \bigcap \Gamma_1, \tag{2.33}$$

$$\alpha_K F^{\delta}_{K,\sigma} + \beta_K|\sigma|u_K = |\sigma|h_K, \quad \forall K \in \mathcal{P}_{out} \bigcap \Gamma_2, \tag{2.34}$$

where we denote $f_K = f(K)$, $g_K = g(K)$, $h_K = h(K)$ and $m(K)$ is the area of the cell K.

Let us denote $\mathbf{U_c}$ as the vector consisting of all cell-centered unknowns and the boundary unknowns on Γ_2, and $\mathbf{U_e}$ is the vector consisting of auxiliary unknowns. And then the finite volume scheme (2.32)-(2.34) can be rewritten as the following matrix form:

$$\mathbf{A}(\mathbf{U_e})\mathbf{U_c} = \mathbf{F} + \mathbf{G}, \tag{2.35}$$

where

$$\mathbf{A}(\mathbf{U_e}) = \sum_{\sigma\in\mathcal{E} \text{ or } \sigma\subset\Gamma_2} \mathbf{N}_\sigma \mathbf{A}_\sigma(\mathbf{U_e})\mathbf{N}_\sigma^T, \tag{2.36}$$

$$\mathbf{F} = \left(\sum_{\sigma\in\mathcal{E}} m(K) f_K + \sum_{\sigma\subset\Gamma_2} |\sigma| h_K\right)_{K\in\mathcal{P}_{int}\bigcup(\mathcal{P}_{out}\bigcap\Gamma_2)}, \tag{2.37}$$

$$\mathbf{G} = \left(\sum_{\sigma\subset K\bigcap\Gamma_1} a_{K,\sigma}\right)_{K\in\mathcal{P}_{int}\bigcup(\mathcal{P}_{out}\bigcap\Gamma_2)}. \tag{2.38}$$

Here, the matrices $\mathbf{A}_\sigma(\mathbf{U_e})$ are 2×2 matrices

$$\mathbf{A}_\sigma(\mathbf{U_e}) = \begin{pmatrix} A_{K,\sigma} & -A_{L,\sigma} \\ -A_{K,\sigma} & A_{L,\sigma} \end{pmatrix}, \qquad \sigma\in\mathcal{E}, \tag{2.39}$$

$$\mathbf{A}_\sigma(\mathbf{U_e}) = \begin{pmatrix} \alpha_K A_{K,\sigma} + \beta_K|\sigma| & -\alpha_K A_{L,\sigma} \\ -\alpha_K A_{K,\sigma} - \beta_K|\sigma| & \alpha_K A_{L,\sigma} \end{pmatrix}, \quad \sigma\subset\Gamma_2 \tag{2.40}$$

for $\sigma\in\mathcal{E}$ or $\sigma\subset\Gamma_2$ and are 1×1 matrices $\mathbf{A}_\sigma(\mathbf{U_e}) = A_{K,\sigma}$ for $\sigma\subset\Gamma_1$. Besides, the assembling matrices $\mathbf{N}_\sigma$ only consist of zeros and ones.

2.6. Picard iteration and monotonicity

We will use the Picard iterative method to solve the above system of nonlinear equations (2.35)

$$\mathbf{A}\left(\mathbf{U_e}^{(s)}\right)\mathbf{U_c}^{(s+1)} = \mathbf{F} + \mathbf{G}, \quad s = 0, 1, 2, \dots. \tag{2.41}$$

For the auxiliary unknowns $\mathbf{U_e}$, we need to use the cell-centered unknowns for locally interpolation approximation

$$\mathbf{U_e}^{(s)} \approx \mathbf{B}\mathbf{U_c}^{(s)}, \tag{2.42}$$

and then we can obtain

$$\mathbf{A}\left(\mathbf{B}\mathbf{U_c}^{(s)}\right)\mathbf{U_c}^{(s+1)} = \mathbf{F} + \mathbf{G}, \quad s = 0, 1, 2, \dots. \tag{2.43}$$

According to numerical evidence, this iteration always converge if the linear system can be solved with a small tolerance ε_{linear}. The slow convergence rate of Picard iteration can be accelerated by Anderson-mixing method [29].

Referring to [34], we can prove that the nonlinear finite volume scheme (2.32)-(2.34) is monotone.

Theorem 2.1. *Assume the vectors $\mathbf{F}, \mathbf{G}, \mathbf{U_c}^0 \geq \mathbf{0}$ and linear systems in Picard iterations are solved exactly. Then all iterates $\mathbf{U_c}^{(s)}$ are non-negative vectors*

$$\mathbf{U_c}^{(s)} \geq \mathbf{0}, \quad s = 1, 2, 3, \dots.$$

If the Picard iterative method converges, we can further prove the nonlinear scheme (2.32)-(2.34) is strongly positivity-preserving.

Theorem 2.2. *Assume the vectors* $\mathbf{F}, \mathbf{G} \geq \mathbf{0}$ *and* $\mathbf{F} + \mathbf{G} \neq \mathbf{0}$. *Then the solution of the nonlinear finite volume scheme* (2.32)-(2.34) *is positive for the interior cell, that is*

$$u_K > 0, \quad \forall K \in \mathcal{P}_{int}. \tag{2.44}$$

Proof. Since $\mathbf{F}, \mathbf{G} \geq \mathbf{0}$, the solution of the nonlinear finite volume scheme (2.32)-(2.34) is non-negative according to the monotonicity. The matrix $\mathbf{A}(\mathbf{U_e})$ is non-symmetric and weak diagonal dominance in column. Hence, $\mathbf{A}(\mathbf{U_e})$ is an M-matrix and all entries of $\mathbf{A}^{-1}(\mathbf{U_e})$ are non-negative. And then the entries of $\mathbf{U_c}$ are not all zero since $\mathbf{F} + \mathbf{G} \neq \mathbf{0}$.

If the cell-centered unknowns are not all positive, we can find an interior cell K such that

$$u_K = 0, \qquad \sum_{K^\sigma \bigcap K \neq \emptyset} u_{K^\sigma} > 0, \tag{2.45}$$

and

$$\sum_{\sigma \subset \partial K} F^\delta_{K,\sigma} = m(K) f_K \geq 0. \tag{2.46}$$

According to the construction process of the conservative discrete flux $F^\delta_{K,\sigma}$ in the previous section, we have that

$$A_{K,\sigma} > 0, \quad B_{K,\sigma} > 0, \quad \forall \sigma \subset \partial K. \tag{2.47}$$

And then combining (2.45) and (2.47), we can get that

$$\sum_{\sigma \subset \partial K} F^\delta_{K,\sigma} = \sum_{\sigma \subset \partial K} (A_{K,\sigma} u_K - B_{K,\sigma} u_{K^\sigma}) < 0,$$

which contradicts (2.46). Thus, the cell-centered unknowns are all positive. □

Remark 2.1. When the vectors $\mathbf{F} \geq \mathbf{0}$, $\mathbf{G} \geq \mathbf{0}$ and $\mathbf{F} + \mathbf{G} \neq \mathbf{0}$, the cell-centered unknowns u_K are all positive due to Theorem 2.2. By the definition of $|\cdot|_\delta$, we have $|u_K|_\delta = u_K$. And then, we can obtain $F^\delta_{K,\sigma} = F_{K,\sigma}$. Hence, the nonlinear finite volume scheme (2.32)-(2.34) has a second-order truncation accuracy while maintaining the positivity.

3. Method Eliminating the Auxiliary Unknowns

In this section, we will introduce a new method eliminating the auxiliary unknowns.

We can use a method shown in [25] that is derived from the idea of MPFA (multi-point flux approximations) to express the auxiliary unknowns, and it shows in [25] that the algorithm has a relatively high numerical accuracy. However, this algorithm cannot converge on some deformed meshes, such as the triangular Kershaw meshes.

As shown in Fig. 3.1(a), for a given vertex P, we denote $\{\mathrm{x}^c_k\}^m_{k=1}$ as the centers of the cells around P and $\{\mathrm{x}^c_k\}^m_{k=1}$ is sorted counterclockwise, besides, we denote $\{\mathrm{x}^e_k\}^m_{k=1}$ as the midpoints of the edges around P and $\{\mathrm{x}^e_k\}^m_{k=1}$ is also sorted counterclockwise, moreover, x^c_k is adjacent to x^e_k and x^e_{k+1}.

Inspired by the work in [8], we consider the quadrilateral pressure support Ω_k with x^c_k, x^e_{k+1}, P and x^e_k as vertices, as shown in Fig. 3.1(b), instead of a triangular pressure support considered

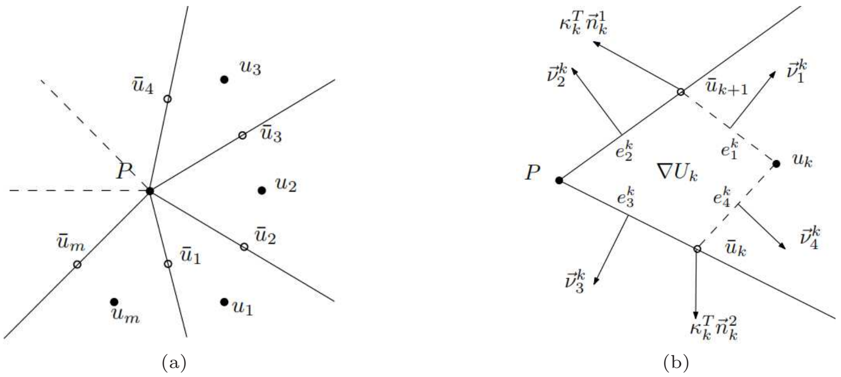

Fig. 3.1. The stencil of improved MPFA: (a) some notation around P; (b) some notation on Ω_k.

in [25]. Then, we propose a new method with full pressure support (FPS) to improve the above method expressing the auxiliary unknowns.

For simplicity, we will denote the above point as $\{\mathrm{x}_j\}_{j=1}^4$, and let e_j^k be the edge with two endpoints x_j, x_{j+1} and an outer normal vector $\vec{\nu}_j^k$. Integrating the gradient $\nabla u(\mathrm{x})$ on Ω_k and using the Green's formula in vector calculus, we have

$$\int_{\Omega_k} \nabla u(\mathrm{x}) d\mathrm{x} = \sum_{j=1}^4 \int_{e_j^k} u(\mathrm{x})\vec{\nu}_j^k d\Gamma.$$

If we approximate $u(\mathrm{x})$ linearly, then we can define the approximate gradient ∇U_k on Ω_k as follows:

$$\nabla U_k \triangleq \sum_{j=1}^4 \frac{|e_j^k|\vec{\nu}_j^k}{2|\Omega_k|}\big[u(\mathrm{x}_j) + u(\mathrm{x}_{j+1})\big],$$

thus, the continuous normal flux on e_2^k can be written as

$$\mathcal{F}_k^1 = \int_{e_2^k} \kappa(\mathrm{x})\nabla u(\mathrm{x}) \cdot \vec{n}_k^1 d\Gamma = \sum_{j=1}^4 \frac{|e_j^k||e_2^k|\vec{\nu}_j^k \cdot \kappa_k^T\vec{n}_k^1}{2|\Omega_k|}\big[u(\mathrm{x}_j) + u(\mathrm{x}_{j+1})\big] + \mathcal{O}(h^2).$$

Hence, we can give a second-order approximation of the continuous flux $\mathcal{F}_k^1$

$$F_k^1 = \sum_{j=1}^4 \frac{|e_j^k||e_2^k|\vec{\nu}_j^k \cdot \kappa_k^T\vec{n}_k^1}{2|\Omega_k|}\big[u(\mathrm{x}_j) + u(\mathrm{x}_{j+1})\big]. \tag{3.1}$$

Similarly, we can get a second-order approximation of the normal flux $\mathcal{F}_k^2$ on e_3^k

$$F_k^2 = \sum_{j=1}^4 \frac{|e_j^k||e_3^k|\vec{\nu}_j^k \cdot \kappa_k^T\vec{n}_k^2}{2|\Omega_k|}\big[u(\mathrm{x}_j) + u(\mathrm{x}_{j+1})\big]. \tag{3.2}$$

Then the continuity of the normal flux on e_2^k gives

$$F_k^1 + F_{k+1}^2 = 0.$$

If we denote $u_k = u(\mathrm{x}_k^c)$, $\bar{u}_k = u(\mathrm{x}_k^e)$ and $\bar{u}_{m+1} = \bar{u}_1$, $u_{m+1} = u_1$, we have

$$a_k^k\bar{u}_k + a_k^{k+1}\bar{u}_{k+1} + a_k^{k+2}\bar{u}_{k+2} + a_k^{m+1}u_P = b_k^k u_k + b_k^{k+1}u_{k+1}, \tag{3.3}$$

where the coefficients are defined by

$$
\begin{aligned}
a_k^k &= \frac{|e_2^k|}{2|\Omega_k|}\left[|e_3^k|\vec{\nu}_3^k \cdot \kappa_k^T \vec{n}_k^1 + |e_4^k|\vec{\nu}_4^k \cdot \kappa_k^T \vec{n}_k^1\right],\\
a_k^{k+1} &= \frac{|e_2^{k+1}|}{2|\Omega_{k+1}|}\left[|e_3^{k+1}|\vec{\nu}_3^{k+1} \cdot \kappa_{k+1}^T \vec{n}_{k+1}^2 + |e_4^{k+1}|\vec{\nu}_4^{k+1} \cdot \kappa_{k+1}^T \vec{n}_{k+1}^2\right]\\
&\quad + \frac{|e_1^k|}{2|\Omega_k|}\left[|e_1^k|\vec{\nu}_1^k \cdot \kappa_k^T \vec{n}_k^1 + |e_2^k|\vec{\nu}_2^k \cdot \kappa_k^T \vec{n}_k^1\right],\\
a_k^{k+2} &= \frac{|e_1^{k+1}|}{2|\Omega_{k+1}|}\left[|e_1^{k+1}|\vec{\nu}_1^{k+1} \cdot \kappa_{k+1}^T \vec{n}_{k+1}^2 + |e_2^{k+1}|\vec{\nu}_2^{k+1} \cdot \kappa_{k+1}^T \vec{n}_{k+1}^2\right],\\
a_k^{m+1} &= \frac{|e_2^{k+1}|}{2|\Omega_{k+1}|}\left[|e_3^{k+1}|\vec{\nu}_3^{k+1} \cdot \kappa_{k+1}^T \vec{n}_{k+1}^2 + |e_2^{k+1}|\vec{\nu}_2^{k+1} \cdot \kappa_{k+1}^T \vec{n}_{k+1}^2\right]\\
&\quad + \frac{|e_1^k|}{2|\Omega_k|}\left[|e_3^k|\vec{\nu}_3^k \cdot \kappa_k^T \vec{n}_k^1 + |e_2^k|\vec{\nu}_2^k \cdot \kappa_k^T \vec{n}_k^1\right],\\
b_k^k &= -\frac{|e_2^k|}{2|\Omega_k|}\left[|e_1^k|\vec{\nu}_1^k \cdot \kappa_k^T \vec{n}_k^1 + |e_4^k|\vec{\nu}_4^k \cdot \kappa_k^T \vec{n}_k^1\right],\\
b_k^{k+1} &= \frac{|e_1^{k+1}|}{2|\Omega_{k+1}|}\left[|e_1^{k+1}|\vec{\nu}_1^{k+1} \cdot \kappa_{k+1}^T \vec{n}_{k+1}^2 + |e_4^{k+1}|\vec{\nu}_4^{k+1} \cdot \kappa_{k+1}^T \vec{n}_{k+1}^2\right].
\end{aligned}
$$

If we use (3.3) to express the auxiliary edge unknowns $\{u_k\}_{k=1}^m$, we need to eliminate the vertex unknown u_P. We will use the method mentioned in [27] to represent the vertex unknown as a convex combination of $\{\bar{u}_k\}_{k=1}^m$ and $\{u_k\}_{k=1}^m$

$$
a_{m+1}^{m+1} u_P + \sum_{k=1}^m a_{m+1}^k \bar{u}_k = \sum_{k=1}^m b_{m+1}^k u_k, \tag{3.4}
$$

where the interpolation coefficients satisfy

$$
a_{m+1}^k \le 0, \quad b_{m+1}^k \ge 0, \quad k = 1, 2, \ldots, m, \quad a_{m+1}^{m+1} = 1,
$$

and

$$
\sum_{k=1}^m \left(|b_{m+1}^k| + |a_{m+1}^k|\right) = 1.
$$

If we denote $\bar{\mathbf{u}} = (\bar{u}_1, \cdots, \bar{u}_m, u_P)^T$ and $\mathbf{u} = (u_1, \cdots, u_m)^T$, then (3.3) and (3.4) give the following system of linear equations:

$$
\bar{\mathrm{A}}\bar{\mathbf{u}} = \bar{\mathrm{B}}\mathbf{u},
$$

where the coefficient matrix $\bar{\mathrm{A}} = (a_k^j)_{(m+1)\times(m+1)}$ and $\bar{\mathrm{B}} = (b_k^j)_{(m+1)\times m}$. Denote $\mathrm{C} = \bar{\mathrm{A}}^{-1}\bar{\mathrm{B}}$, and then we can express the auxiliary edge unknowns as follows:

$$
\bar{u}_k = \sum_{j=1}^m c_k^j u_j, \quad k = 1, 2, \ldots, m. \tag{3.5}
$$

Remark 3.1. For each auxiliary edge unknown, the improved MPFA proposed above can give two different interpolation approximations. It can be found from numerical experiments that when the sum of the absolute values for the interpolation coefficients is large, some interpolation coefficients are more likely to be negative. Hence, we choose the one with the smaller sum of the absolute values for the interpolation coefficients.

It can be seen from the numerical experiments that the improved MPFA has wider applicability than MPFA and can handle more complex meshes.

4. Numerical Experiments

We give some numerical experiments to test the robustness and accuracy of the new proposed scheme. We define the discrete L_2-norm error of $u(\mathrm{x})$ as

$$\varepsilon_2^u = \Big[\sum_{K\in\mathcal{T}} \big(u_K - u(K)\big)^2 m(K) \Big]^{\frac{1}{2}},$$

and the discrete error for the normal flux $\mathcal{F}$ as

$$\varepsilon_h^F = \Big[\sum_{\sigma\in\mathcal{E}} (F_{K,\sigma} - \mathcal{F}_{K,\sigma})^2 \Big]^{\frac{1}{2}}.$$

Besides, the number of nonlinear iterations is recorded as $it_{non}^{\#}$.

For the sake of brevity, the new proposed positivity-preserving scheme with the auxiliary unknowns approximated by the improved MPFA is denoted as Scheme 1, and the new proposed scheme with the auxiliary unknowns approximated by the method proposed in [25] is denoted as Scheme 2.

4.1. Anisotropic diffusion problem

We will use the finite volume scheme (2.32)-(2.34) to solve the anisotropic diffusion problem on $\Omega = [0,1]^2$ with a diffusion tensor $\kappa = RDR^T$, where

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

Here, we choose $\theta = 5\pi/12, k_1 = 1+2x^2+y^2$ and $k_2 = 1+x^2+2y^2$. And the analytical solution is chosen as $u(x,y) = \sin(\pi x)\sin(\pi y)$.

At first, we use our scheme to solve the anisotropic diffusion problems on the random quadrilateral and the random triangular meshes. In addition, we use the scheme in [25, 26] for comparison. The numerical experiment results show that the new proposed scheme (2.32)-(2.34) gives calculation results similar to the scheme in [25, 26].

We also test our scheme on the quadrilateral Kershaw meshes shown in Fig. 4.1 and the triangular Kershaw meshes shown in Fig. 4.2.

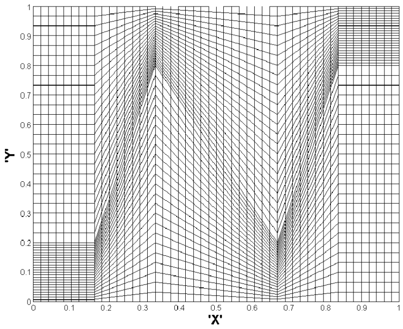

Fig. 4.1. The quadrilateral Kershaw mesh.

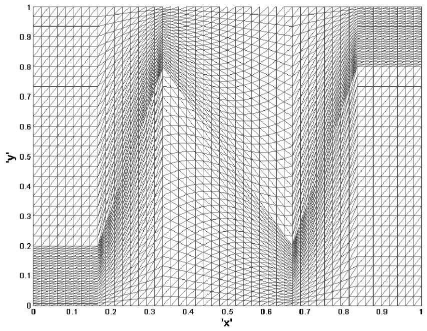

Fig. 4.2. The triangular Kershaw mesh for anisotropic diffusion problems.

Table 4.1 gives the numerical results of the new proposed scheme on the quadrilateral Kershaw meshes, and compares them with the numerical results in [25, 27]. The Kershaw meshes are highly distorted, but our positivity-preserving finite volume scheme can still have a convergence rate close to second order for the solution and a first-order convergence rate for the flux. On the coarse meshes, if improved MPFA is used to approximate the auxiliary edge unknowns, the accuracy for the solution is less than second order, moreover the accuracy of the new proposed scheme is better than that of the scheme in [25]. Besides, the scheme in [27] has lower accuracy on the quadrangular Kershaw meshes.

Table 4.2 shows the numerical accuracy of our positivity-preserving finite volume scheme on the triangular Kershaw meshes, and compares it with the calculation results in [27]. Here, we use improved MPFA for calculation.

From Table 4.2, it can be seen that our positivity-preserving finite volume scheme can give a higher convergence rate on the triangular Kershaw meshes than the scheme in [27].

The absolute values of the interpolation coefficients obtained by the method in [25] on the triangular Kershaw meshes are often large. In numerical experiments, no matter what initial value is selected, the nonlinear iteration to solve the finite volume scheme in [25] cannot converge. Hence, using MPFA to approximate auxiliary unknowns cannot obtain a convergent solution.

Combining the calculation results on random triangular, quadrilateral meshes and Kershaw meshes, we can see that our positivity-preserving finite volume scheme can deal with anisotropic

Table 4.1: Accuracy on the quadrilateral Kershaw meshes.

The number of cell		144	576	2304	9216	36864
Scheme 1	ε_2^u	2.50e-2	8.58e-3	1.71e-3	3.48e-4	7.99e-5
	rate	-	1.54	2.33	2.30	2.12
	ε_h^F	7.15e-1	2.73e-1	8.80e-2	2.88e-2	1.02e-2
	rate	-	1.39	1.63	1.61	1.50
	$it_{non}^{\#}$	64	122	171	211	242
Scheme 2	ε_2^u	1.90e-2	5.91e-3	1.27e-3	3.56e-4	9.72e-5
	rate	-	1.68	2.22	1.83	1.87
	ε_h^F	6.51e-1	2.64e-1	1.04e-1	3.86e-2	1.40e-2
	rate	-	1.31	1.34	1.43	1.47
	$it_{non}^{\#}$	64	133	186	230	262
The scheme in [25]	ε_2^u	1.96e-2	5.91e-3	1.24e-3	3.47e-4	9.49e-5
	rate	-	1.73	2.25	1.84	1.87
	ε_h^F	7.00e-1	2.66e-1	1.04e-1	3.85e-2	1.39e-2
	rate	-	1.40	1.35	1.43	1.47
	$it_{non}^{\#}$	64	134	187	231	262
The scheme in [27]	ε_2^u	3.14e-2	1.62e-2	6.44e-3	2.02e-3	5.48e-4
	rate	-	0.95	1.33	1.67	1.88
	ε_h^F	9.10e-1	4.88e-1	2.10e-1	7.57e-2	2.51e-2
	rate	-	0.90	1.22	1.47	1.59
	$it_{non}^{\#}$	70	129	212	287	382

Table 4.2: Accuracy on the triangular Kershaw meshes.

The number of cell		288	1152	4608	18432	73728
Scheme 1	ε_2^u	3.29e-2	1.18e-2	2.91e-3	7.03e-4	1.75e-4
	rate	-	1.48	2.02	2.05	2.01
	ε_h^F	1.34	5.33e-1	1.87e-1	6.17e-2	1.97e-2
	rate	-	1.32	1.51	1.60	1.65
	$it_{non}^{\#}$	107	235	397	517	690
Scheme 2	Not work!					
The scheme in [27]	ε_2^u	1.84e-2	1.25e-2	5.78e-3	1.93e-3	5.36e-4
	rate	-	0.56	1.11	1.58	1.85
	ε_h^F	1.31	6.96e-1	3.24e-1	1.30e-1	5.02e-2
	rate	-	0.91	1.10	1.32	1.37
	$it_{non}^{\#}$	104	131	234	336	534
The scheme in [25]	Not work!					

diffusion problems on highly distorted meshes and is more robust. Especially on Kershaw meshes, which are highly distorted meshes, our positivity-preserving finite volume scheme can give a higher convergence rate than some existing finite volume schemes.

4.2. The diffusion problem with point source

Consider the anisotropic diffusion problem on $\Omega = [0,1]^2$ with a point source. We choose an anisotropic diffusion tensor $\kappa = RDR^T$, where

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

and we choose $\theta = \pi/6, k_1 = 10000, k_2 = 1$. We put a point source at the center of Ω. To facilitate numerical calculations, we select the source term as

$$f(x,y) = \begin{cases} 101 \times 101, & (x,y) \in \left[\dfrac{50}{101}, \dfrac{51}{101}\right]^2, \\ 0, & \text{otherwise.} \end{cases}$$

Besides, the homogeneous Dirichlet boundary condition is imposed on $\partial\Omega$. We solve the strongly anisotropic diffusion problem with point source on a random quadrilateral mesh shown in Fig. 4.3. Although the analytical solution cannot be obtained, according to the maximum principle, we know that the solution should be positive.

Fig. 4.4(a) gives an image of the numerical solution for the point source problem on the random quadrilateral mesh given by our positivity-preserving finite volume scheme. It can be seen that all the numerical solutions on the interior cell are positive, so the scheme is positivity-preserving. And from the contour map in Fig. 4.4(b), it is obvious that our numerical solution captures the strong anisotropy of the diffusion tensor, which leads to the phenomenon that the solution is concentrated near the line $3y - x = 0$.

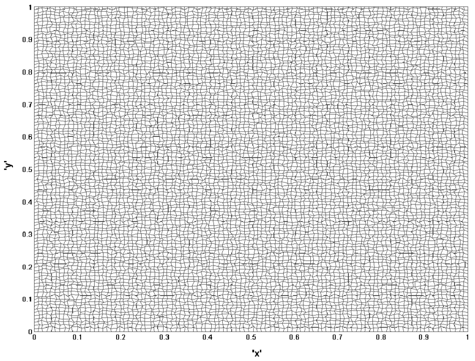

Fig. 4.3. The random quadrilateral mesh for the diffusion problem with point source (101×101).

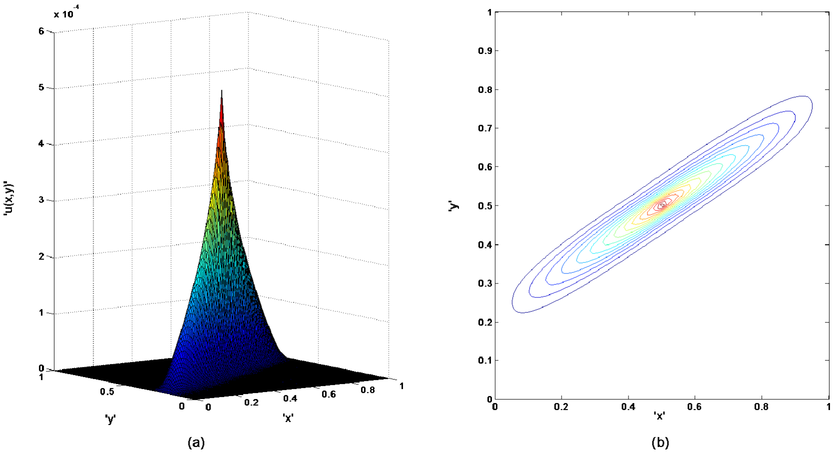

Fig. 4.4. Numerical results of the diffusion problem with point source: (a) numerical solution image; (b) contour map.

4.3. Vertical fault

Consider the vertical fault problem taken from [11]. We divide the calculation area $\Omega = [0,1]^2$ into two parts: the black area in Fig. 4.5 $\Omega_1 = \Omega_1^l \bigcup \Omega_1^r$, where

$$\Omega_1^l = (0.0, 0.5] \times \left(\bigcup_{k=0}^{4} \left[0.05 + 2k \times 0.1, 0.05 + (2k+1) \times 0.1 \right) \right),$$

$$\Omega_1^r = (0.5, 1.0) \times \left(\bigcup_{k=0}^{4} \left[2k \times 0.1, (2k+1) \times 0.1 \right) \right),$$

and the white area in Fig. 4.5 $\Omega_2 = \Omega \setminus \Omega_1$. Then, we choose the following layered anisotropic diffusion tensor:

$$\kappa = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix},$$

where $\alpha = 100, \beta = 10$ on Ω_1 and $\alpha = 10^{-2}, \beta = 10^{-3}$ on Ω_2. It can be seen that a vertical fault will appear at $x = 0.5$. We use a random quadrilateral mesh shown in Fig. 4.6 for calculation, where the mesh edges are divided along the discontinuity of the diffusion tensor.

Here, the Dirichlet boundary condition $u(x,y) = 1 - x$ are imposed on $\partial\Omega$, and we choose a zero force term. Then, according to the maximum principle, we know that the internal solution of the vertical fault problem should be between 0 and 1. Fig. 4.7 shows an approximate solution to the vertical fault problem on the random quadrilateral mesh shown in Fig. 4.6, which is obtained by using our positivity-preserving finite volume scheme. The maximum value of the numerical solution on the interior cell is 0.995, and the minimum value is 5.30e-3.

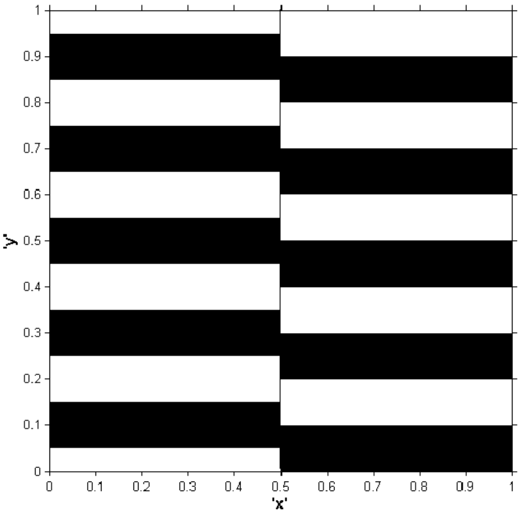

Fig. 4.5. Computation area for vertical fault problems.

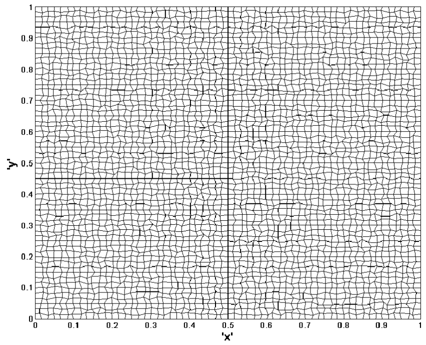

Fig. 4.6. The random quadrilateral mesh for vertical fault problems (60 × 60).

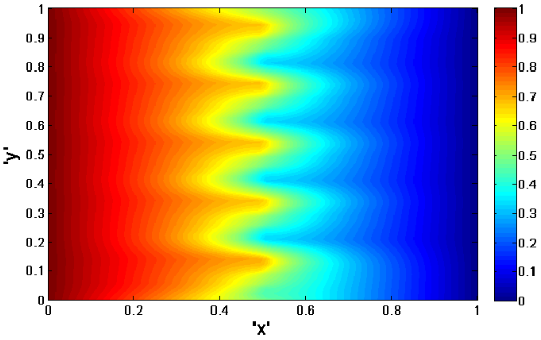

Fig. 4.7. Numerical results of vertical faults.

4.4. Heterogeneous diffusion tensor

Finally, we consider the heterogeneous diffusion tensor problem on $\Omega = [0,1]^2$ shown in Fig. 4.8. We choose a full diffusion tensor $\kappa(x,y) = RDR^T$, where

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

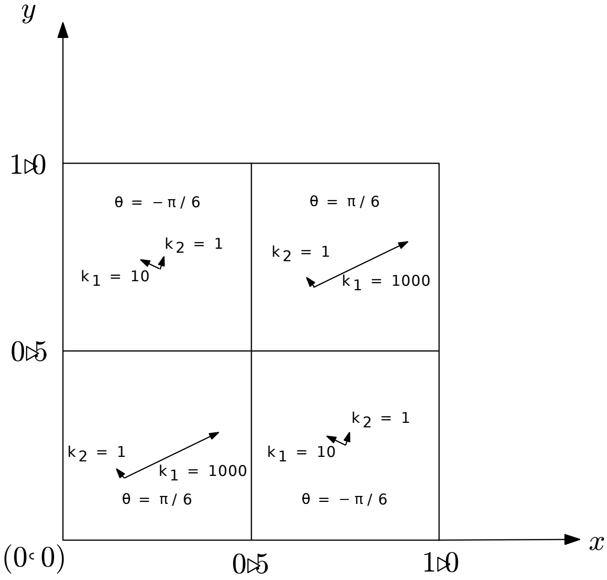

Fig. 4.8. The heterogeneous diffusion tensor.

When $0 \le x, y < 0.5$ or $0.5 < x, y \le 1$, we set

$$\theta = \frac{\pi}{6}, \quad k_1 = 1000, \quad k_2 = 1,$$

and on the rest of the calculation domain we choose

$$\theta = -\frac{\pi}{6}, \quad k_1 = 10, \quad k_2 = 1.$$

Thus, the diffusion tensor $\kappa(x, y)$ is anisotropic over Ω and is strongly discontinuous on $x = 0.5$ or $y = 0.5$. We examine the numerical solution on a random quadrilateral mesh shown in Fig. 4.9 by our positivity-preserving scheme. We choose the force term as

$$f(x, y) = \begin{cases} 10000, & \dfrac{7}{18} \le x, y \le \dfrac{11}{18}, \\ 0, & \text{otherwise}, \end{cases}$$

and the homogeneous Dirichlet boundary condition $g(x, y) = 0$ is imposed on $\partial\Omega$.

Fig. 4.10 shows the numerical solution for the problem with the heterogeneous diffusion tensor obtained by our positivity-preserving scheme on the random quadrilateral mesh. Fig. 4.10 gives the image of the numerical solution. It can be seen that even if the diffusion tensor is strongly anisotropic and discontinuous, the numerical solution given by our positivity-preserving scheme is still positive on the interior cell. Fig. 4.11(a) gives the contour map of the numerical solution, which clearly shows the influence of different diffusion coefficients in different regions. Fig. 4.11 shows the numerical solutions obtained by our schemes and the nine-point scheme in [24]. It can be seen from these figures that nine-point scheme produces negative values, however, our scheme preserves the positivity of the continuous solution.

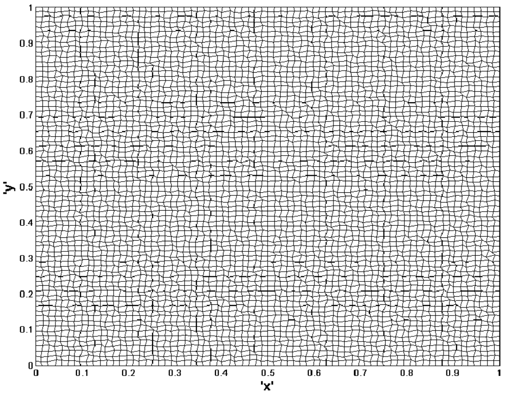

Fig. 4.9. The random quadrilateral mesh for problem with heterogeneous diffusion tensor (72 × 72).

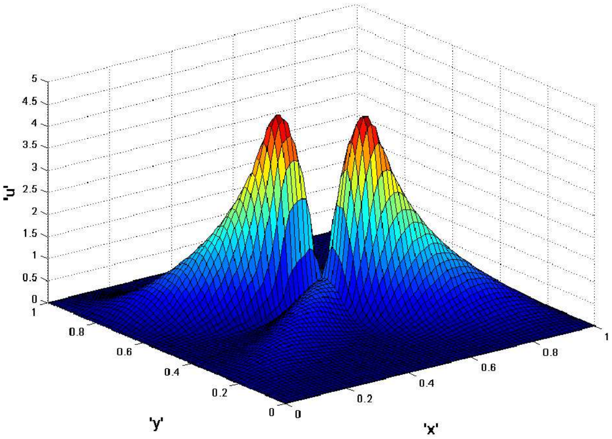

Fig. 4.10. Numerical solution of the problem with heterogeneous diffusion tensor.

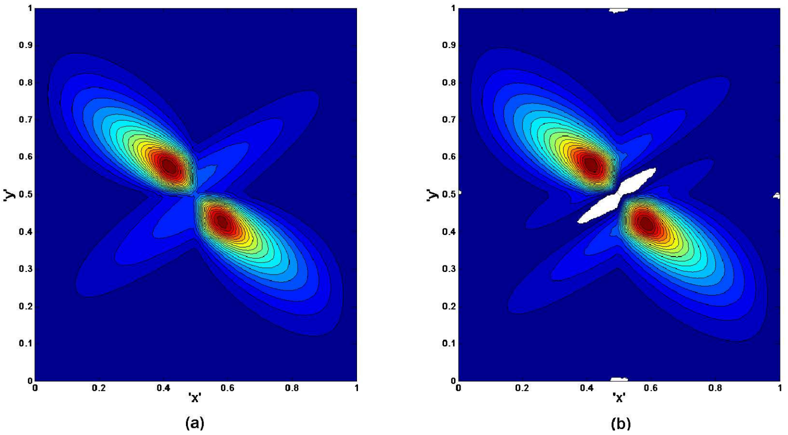

Fig. 4.11. Comparison of our scheme and the nine-point scheme in [24] on random quadrilateral meshes: (a) our scheme; (b) the nine-point scheme ($u_{\min} = -0.5901$, the negative part is shown in white).

5. Conclusion

We introduce a nonlinear combination technique to correct the nine-point scheme proposed in [24] for solving the diffusion equation. The auxiliary edge unknowns are eliminated locally by the improved MPFA, which results in a cell-centered scheme. The new proposed scheme is strongly positivity-preserving and can deal with some distorted grids, such as the triangular Kershaw meshes, but some existing positivity-preserving finite volume schemes could not handle well. The numerical tests show that the new proposed scheme positivity-preserving scheme is more robust than some existing positivity-preserving finite volume schemes.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 12071045, 12201299), by the Natural Science Foundation of Jiangsu Province (Grant No. BK20220870) and by the Youth Foundation of Laboratory of Computational Physics (Grant No. HX02021-37).

References

[1] O. Angélini, C. Chavant, E. Chénier, and R. Eymard, A finite volume scheme for diffusion problems on general meshes applying monotony constraints, *SIAM J. Numer. Anal.*, **47**:6 (2010), 4193–4213.

[2] E. Bertolazzi and G. Manzini, A second-order maximum principle preserving finite volume method for steady convection-diffusion problems, *SIAM J. Numer. Anal.*, **43**:5 (2005), 2172–2199.

[3] X. Blanc and E. Labourasse, A positive scheme for diffusion problems on deformed meshes, *ZAMM Z. Angew. Math. Mech.*, **96**:6 (2016), 660–680.

[4] O. Burdakov, I. Kapyrin, and Y. Vassilevski, Monotonicity recovering and accuracy preserving optimization methods for postprocessing finite element solutions, *J. Comput. Phys.*, **231**:8 (2012), 3126–3142.

[5] J.S. Camier and F. Hermeline, A monotone nonlinear finite volume method for approximating diffusion operators on general meshes, *Internat. J. Numer. Methods Engrg.*, **107**:6 (2016), 496–519.

[6] C. Cancés, M. Cathala, and C. Le Potier, Monotone corrections for generic cell-centered finite volume approximations of anisotropic diffusion equations, *Numer. Math.*, **125**:3 (2016), 387–417.

[7] J. Droniou and C. Le Potier, Construction and convergence study of schemes preserving the elliptic local maximum principle, *SIAM J. Numer. Anal.*, **49**:2 (2011), 459–490.

[8] M.G Edwards and H. Zheng, A quasi-positive family of continuous Darcy-flux finite-volume schemes with full pressure support, *J. Comput. Phys.*, **227**:22 (2008), 9333–9364.

[9] Z. Gao and J. Wu, A small stencil and extremum-preserving scheme for anisotropic diffusion problems on arbitrary 2D and 3D meshes, *J. Comput. Phys.*, **250** (2013), 308–331.

[10] Z. Gao and J. Wu, A second-order positivity-preserving finite volume scheme for diffusion equations on general meshes, *SIAM J. Sci. Comput.*, **37**:1 (2015), A420–A438.

[11] R. Herbin and F. Hubert, *Benchmark on Discretization Schemes for Anisotropic Diffusion Problems on Ggeneral Grids*, in: *Finite Volumes for Complex Applications V*, (2008), 659–692.

[12] W. Huang, Discrete maximum principle and a Delaunay-type mesh condition for linear finite element approximations of two-dimensional anisotropic diffusion problems, *Numer. Math. Theory Methods Appl.*, **4**:3 (2011), 319–334.

[13] C. Le Potier, Finite volume monotone scheme for highly anisotropic diffusion operators on unstructured triangular meshes, *C. R. Math. Acad. Sci. Paris*, **341**:12 (2005), 787–792.

[14] C. Le Potier, A nonlinear finite volume scheme satisfying maximum and minimum principles for diffusion operators, *Int. J. Finite Vol.*, **6** (2009), 1–20.

[15] K. Lipnikov, M. Shashkov, D. Svyatskiy, and Y. Vassilevski, Monotone finite volume schemes for diffusion equations on unstructured triangular and shape-regular polygonal meshes, *J. Comput. Phys.*, **227**:1 (2007), 492–512.

[16] K. Lipnikov, M. Shashkov, and Y. Vassilevski, Minimal stencil finite volume scheme with the discrete maximum principle, *Russian J. Numer. Anal. Math. Modelling*, **27**:4 (2012), 369–386.

[17] R. Liska and M. Shashkov, Enforcing the discrete maximum principle for linear finite element solutions of second-order elliptic problems, *Commun. Comput. Phys.*, **3**:4 (2008), 852–877.

[18] C. Lu, W. Huang, and E.S.V. 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** (2013), 24–36.

[19] S. Miao and J. Wu, A nonlinear correction scheme for the heterogeneous and anisotropic diffusion problems on polygonal meshes, *J. Comput. Phys.*, **448** (2022), 110729.

[20] C. Ngo and W. Huang, Monotone finite difference schemes for anisotropic diffusion problems via nonnegative directional splittings, *Commun. Comput. Phys.*, **19**:2 (2016), 473–495.

[21] J. Nordbotten, I. Aavatsmark, and G. Eigestad, Monotonicity of control volume methods, *Numer. Math.*, **106**:2 (2007), 255–288.

[22] M. Schneider, B. Flemisch, R. Helmig, K. Terekhov, and H. Tchelepi, Monotone nonlinear finite-volume method for challenging grids, *Comput. Geosci.*, **22**:2 (2018), 565–586.

[23] M. Sheng, D. Yang, and Z. Gao, A virtual element method-based positivity-preserving conservative scheme for convection-diffusion problems on polygonal meshes, *Numer. Methods Partial Differ. Equ.*, **39**:2 (2023), 1398–1424.

[24] Z. Sheng and G. Yuan, A nine point scheme for the approximation of diffusion operators on distorted quadrilateral meshes, *SIAM J. Sci. Comput.*, **30**:3 (2008), 1341–1361.

[25] Z. Sheng and G. Yuan, An improved monotone finite volume scheme for diffusion equation on polygonal meshes, *J. Comput. Phys.*, **231**:9 (2012), 3739–3754.

[26] Z. Sheng and G. Yuan, A new nonlinear finite volume scheme preserving positivity for diffusion equations, *J. Comput. Phys.*, **315** (2016), 182–193.

[27] Z. Sheng and G. Yuan, A cell-centered nonlinear finite volume scheme preserving fully positivity for diffusion equation, *J. Sci. Comput.*, **68**:2 (2016), 521–545.

[28] Z. Sheng and G. Yuan, Construction of nonlinear weighted method for finite volume schemes preserving maximum principle, *SIAM J. Sci. Comput.*, **40**:1 (2018), A607–A628.

[29] H.F. Walker and P. Ni, Anderson acceleration for fixed-point iterations, *SIAM J. Numer. Anal.*, **49**:4 (2011), 1715–1735.

[30] J. Wang and R. Zhang, Maximum principles for P1-conforming finite element approximations of quasi-linear second order elliptic equations, *SIAM J. Numer. Anal.*, **50**:2 (2012), 626–642.

[31] D. Yang, M. Sheng, Z. Gao, and G. Ni, The VEM-based positivity-preserving conservative scheme for radiation diffusion problems on generalized polyhedral meshes, *Comput. Fluids*, **239** (2022), 105356.

[32] B. Yu, H. Yang, Y. Li, and G. Yuan, Monotonicity correction for the finite element method of anisotropic diffusion problems, *Commun. Comput. Phys.*, **31**:5 (2022), 1489–1524.

[33] G. Yuan, Construction and analysis of nine-point scheme for diffusion equations on distored meshes, *Annual Report of Laboratory in Computational Physics*, (2005), 530–575.

[34] G. Yuan and Z. Sheng, Monotone finite volume schemes for diffusion equations on polygonal meshes, *J. Comput. Phys.*, **227**:12 (2008), 6288–6312.