Interpolation-based second-order monotone finite volume schemes for anisotropic diffusion equations on general grids

Jiming Wu, Zhiming Gao

Interpolation-based second-order monotone finite volume schemes for anisotropic diffusion equations on general grids[☆]

Jiming Wu, Zhiming Gao[*]

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

Abstract

We propose two interpolation-based monotone schemes for the anisotropic diffusion problems on unstructured polygonal meshes through the linearity-preserving approach. The new schemes are characterized by their nonlinear two-point flux approximation, which is different from the existing ones and has no constraint on the associated interpolation algorithm for auxiliary unknowns. Thanks to the new nonlinear two-point flux formulation, it is no longer required that the interpolation algorithm should be a positivity-preserving one. The first scheme employs vertex unknowns as the auxiliary ones, and a second-order but not positivity-preserving interpolation algorithm is utilized. The second scheme uses the so-called harmonic averaging points located on cell edges to define the auxiliary unknowns, and a second-order positivity-preserving interpolation method is employed. Both schemes have nearly the same convergence rates as compared with their related second-order linear schemes. Numerical results demonstrate that the new schemes are monotone, and have the second-order accuracy for the solution and first-order for its gradient on severely distorted meshes.

Keywords: diffusion equation, monotone finite volume scheme, linearity-preserving criterion, nonlinear two-point flux approximation

1. Introduction

The anisotropic diffusion problem is often coupled with some other physical processes such as the Lagrange approach in radiation hydrodynamics. In this case, the solution should be non-negative physically, the mesh may be severely distorted and even worse, the problem may be suffered by some moving discontinuity whose location is blind to the diffusion solver. In this paper we are concerned with the nonlinear cell-centered finite

[☆]This work was supported by the National Natural Science Fund of China (Nos. 91330205, 11271053).

[*]Corresponding author.

Email addresses: wu_jiming@iapcm.ac.cn (Jiming Wu), gao@iapcm.ac.cn (Zhiming Gao)

volume methods which not only preserve solution positivity, but also approximately have a second-order accuracy on severely distorted grids.

It is well known that the linear schemes, such as mixed finite element, multi-point flux approximation (MPFA) and the mimetic finite difference (MFD) schemes, are second-order accurate on unstructured grids, but do not always satisfy monotonicity for distorted meshes or with high anisotropy ratio [10]. MPFA [1] used built-in flexibility to increase the monotonicity regions. Two methods based on repair technique and constrained optimization have been introduced to enforce discrete extremum principle for linear finite element solutions on triangular meshes [15]. A linear scheme satisfying the maximum principle on distorted grids was developed in [18], but it is generally only first-order accurate for smooth solutions. The sufficient condition to ensure the monotonicity of MFD was analyzed in [12].

A few monotone schemes have been proposed with the nonlinear two-point flux approximation in recent years. The original idea belongs to C. Le Potier [17] for triangular meshes. The monotonicity of this method for steady state diffusion problems was proven in [13], and its 3D extension was analyzed in [11]. Further development of the method was made in [24, 20, 14]. A common property of the methods in [13, 24, 20] is that in addition to primary unknowns defined at cell centers, solution values at mesh vertices or edge midpoints are involved. These auxiliary unknowns are usually interpolated from primary cell-centered unknowns and the weights in the interpolation formula are required to be positive so that the monotonicity of the schemes can be proved. When negative interpolation weights occur, the authors in [20, 24] suggested that the interpolation formula should be replaced by some lower-order but positivity-preserving ones, such as the inverse distance weighting method used in [13]. Unfortunately, this technique usually leads to a great loss of accuracy on largely distorted meshes even in the case of a constant diffusion coefficient with a smooth solution. The direct use of a second-order interpolation algorithm in C. Le Potier's nonlinear two-point flux approximation without weights replacement may spoil the monotonicity, although it seems a theoretical problem instead of a numerical one in some cases [19]. In most cases, the construction of a second-order positivity-preserving interpolation algorithm is a more challenging task than the design of interpolation-based monotone schemes itself. By noting this fact, the authors in [14] suggested an interpolation-free monotone scheme, however, this alternative approach introduces a constraint on the choice of cell centers. A similar work can be found in [4] where a proper choice of the cell centers was used to construct a monotone scheme on triangular meshes. Recently, some monotone corrections for generic cell-centered finite volume schemes were suggested in [5], however, it is not clear why the accuracy of the original schemes decreases after performing these corrections.

In this article, we further develop C. Le Potier's idea to suggest a new nonlinear two-point flux approximation and then construct two interpolation-based second-order monotone finite volume schemes outlined below.

- In the first scheme, we take cell-vertex unknowns as the auxiliary ones, and a second-order interpolation algorithm for cell-vertex unknowns in [8] are utilized. Although this interpolation method is not positivity-preserving, the monotonicity of the scheme is still maintained theoretically due to the new nonlinear two-point flux approximation.

- In the second scheme, we take cell-edge unknowns as the auxiliary ones which are located at the so-called harmonic averaging point [3], and can be interpolated from the cell-centered unknowns defined on the two cells sharing this same edge. The use of harmonic averaging points not only simplifies the interpolation procedure, but also assures it to be a positivity-preserving one. In this case, our nonlinear two-point flux approximation reduces to the existing ones [17, 13, 24].

Our new schemes have the following characteristics:

- they are monotone;

- they are locally conservative and have only cell-centered unknowns;

- they allow heterogeneous full diffusion tensors;

- they are linearity-preserving except for some extreme cases. The linearity-preserving property can be found in many articles for instance in [1, 3, 7, 22].

- they have approximately second-order accuracy on severely distorted meshes in case that the diffusion tensor is taken to be anisotropic and discontinuous.

It should be emphasized that the last property is the main difference between the two new schemes and some existing interpolation-based monotone schemes.

The outline of this article is as follows. In section 2, we state the diffusion problem and give some necessary notations. In section 3, two nonlinear monotone finite volume schemes are constructed under a unified framework and a new nonlinear two-point flux approximation. Numerical experiments are then presented in section 4, showing the good numerical performance of the new schemes, especially on severely distorted grids. Some concluding remarks are given in the last section.

2. Steady diffusion problem and notations

We consider a diffusion problem on an open bounded connected polygonal domain $\Omega \subset \mathbb{R}^2$,

$$-\mathrm{div}(\Lambda \nabla u) = f \quad \text{in} \quad \Omega, \tag{1}$$

$$u = g_D \quad \text{on} \quad \Gamma_D, \tag{2}$$

$$-\Lambda \nabla u \cdot \boldsymbol{n} = g_N \quad \text{on} \quad \Gamma_N, \tag{3}$$

where $\Lambda(\boldsymbol{x})$ is the diffusion tensor, f is a source term, $\partial\Omega = \bar{\Gamma}_D \cup \bar{\Gamma}_N$ is the boundary of Ω, $\boldsymbol{n}$ denotes the exterior normal vector and g_D, g_N are given scalar functions which are almost everywhere defined on Γ_D, Γ_N, respectively. We employ the following notations and assumptions throughout this paper.

1. $\mathcal{M} = \{K\}$ is a finite family of disjoint open polygonal cells in Ω such that $\bar{\Omega} = \cup_{K\in\mathcal{M}} \bar{K}$.
2. $\mathcal{E}$= $\{\sigma\}$ is a finite family of disjoint edges in $\bar{\Omega}$ such that for $\sigma \in \mathcal{E}$, σ is a line segment with a strictly positive length denoted by $|\sigma|$. When $\sigma \in \mathcal{E}^{ext}$, σ is on either Γ_D or Γ_N, but not both of them. Let $\mathcal{E}^{int} = \mathcal{E} \cap \Omega$ and $\mathcal{E}^{ext} = \mathcal{E} \cap \partial\Omega$. For $K \in \mathcal{M}$, there exists a subset $\mathcal{E}_K$ of $\mathcal{E}$ such that $\partial K = \cup_{\sigma\in\mathcal{E}_K} \bar{\sigma}$ and the number of edges in $\mathcal{E}_K$ is n_K. For $\sigma \in \mathcal{E}_K$, notation σ may denote either an edge on ∂K or the local numbering of this same edge, depending on the context, and $\boldsymbol{n}_{K,\sigma}$ denotes the unit vector normal to σ outward to K.
3. $\mathcal{O} = \{\boldsymbol{x}_K, K \in \mathcal{M}\}$ is a set of points, known as cell centers, where $\boldsymbol{x}_K \in K$.
4. $\mathcal{P} = \cup_{K\in\mathcal{M}} \mathcal{P}_K$, where $\mathcal{P}_K$ is the set of interpolation points associated with cell K.
5. Approximation of the solution u at the cell center $\boldsymbol{x}_K$ is known as the primary variable and denoted as u_K. By contrast, the auxiliary variable $u_{K,i}$ (resp., u_σ) is the approximation of u at the interpolation point $\boldsymbol{x}_{K,i}$ (resp., $\boldsymbol{y}_\sigma$) defined in the later discussion. In addition, we assume that Λ is constant on each cell $K \in \mathcal{M}$ with Λ_K denoting the restriction of Λ on K.

3. Monotone nonlinear finite volume schemes

In this section, we will construct some nonlinear monotone finite volume schemes with a new two-point flux approximation, which is done under the same framework suggested in [23] and consists of five steps (see Section 3.1–Section 3.5). Throughout, all the derivations are subjected to the following fundamental principles for finite volume methods: the normal component of flux is continuous across all cell edges and the solution is continuous at any interpolation points.

3.1. Definitions of the primary and auxiliary variables

The primary variables are defined at the cell centers. In this article, the barycenter and the geometric center of K are the usual candidates for the cell center $\boldsymbol{x}_K$ but not the necessary ones. The auxiliary variables are defined at the interpolation points, for which we have two different choices in this paper:

1. The first choice for the set of interpolation points is $\mathcal{P}_K = \mathcal{P}_K^v := \{\boldsymbol{x}_{K,i}, i = 1, \cdots, n_K\}$ where $\boldsymbol{x}_{K,i}$ denotes a generic vertex of cell K.
2. The second choice for the set of interpolation points is $\mathcal{P}_K = \mathcal{P}_K^e := \{\boldsymbol{y}_\sigma, \sigma \in \mathcal{E}_K\}$ where $\boldsymbol{y}_\sigma$ is a specific point associated with edge $\sigma \in \mathcal{E}_K$ and defined as follows.

For a boundary edge $\sigma \in \mathcal{E}^{ext}$, $\boldsymbol{y}_\sigma$ is the midpoint of σ, and for an interior edge $\sigma \in \mathcal{E}_K \cap \mathcal{E}_L$,

$$\boldsymbol{y}_\sigma = \frac{d_{L,\sigma}\lambda_K^{(n)}\boldsymbol{x}_K + d_{K,\sigma}\lambda_L^{(n)}\boldsymbol{x}_L + d_{K,\sigma}d_{L,\sigma}\left(\Lambda_K^T - \Lambda_L^T\right)\boldsymbol{n}_{K,\sigma}}{d_{L,\sigma}\lambda_K^{(n)} + d_{K,\sigma}\lambda_L^{(n)}}, \tag{4}$$

where $\lambda_K^{(n)} = \boldsymbol{n}_{K,\sigma}^T\Lambda_K\boldsymbol{n}_{K,\sigma}, \lambda_L^{(n)} = \boldsymbol{n}_{L,\sigma}^T\Lambda_L\boldsymbol{n}_{L,\sigma}$ and $d_{K,\sigma}$ (resp., $d_{L,\sigma}$) denotes the orthogonal distance from $\boldsymbol{x}_K$ (resp., $\boldsymbol{x}_L$) to σ.

For the second choice, we introduce the following assumption:

(A1) For any $\sigma \in \mathcal{E}_K \cap \mathcal{E}_L \subset \mathcal{E}^{int}$, (i) K (resp., L) is a star-shaped polygonal cell with respect to $\boldsymbol{x}_K$ (resp., $\boldsymbol{x}_L$); (ii) $\boldsymbol{y}_\sigma \in \bar{\sigma}$.

When **(A1)** holds, $\boldsymbol{y}_\sigma$ coincides with the *harmonic averaging point* [3, 7, 23] and we have [9]

$$u(\boldsymbol{y}_\sigma) \simeq \frac{d_{K,\sigma}\lambda_L^{(n)}u(\boldsymbol{x}_L) + d_{L,\sigma}\lambda_K^{(n)}u(\boldsymbol{x}_K)}{d_{L,\sigma}\lambda_K^{(n)} + d_{K,\sigma}\lambda_L^{(n)}}, \tag{5}$$

where $\simeq$ indicates that (5) satisfies the so-called *linearity-preserving criterion*, i.e., the truncation error vanishes in the linear case where the solution u is linear and the diffusion coefficient is constant on any cell $K \in \mathcal{M}$.

As pointed in [9], when **(A1)** is violated, (5) may not hold, however, the harmonic averaging point $\boldsymbol{y}_\sigma$ defined by (4) can be still used as an interpolation point. A simple derivation of harmonic averaging point under the linearity-preserving criterion can be found in [9].

3.2. Construction of the one-sided flux

For $\sigma \in \mathcal{E}_K, K \in \mathcal{M}$, the one-sided flux $F_{K,\sigma}$ is the approximation of the flux $-\int_\sigma(\Lambda_K\nabla u)\cdot\boldsymbol{n}_{K,\sigma}\,\mathrm{d}s$, using only the information of K. Here we give an algorithm for a general $\mathcal{P}_K$, instead of confining ourselves to $\mathcal{P}_K^v$ or $\mathcal{P}_K^e$. As shown in Fig. 1, for $\sigma \in \mathcal{E}_K$,

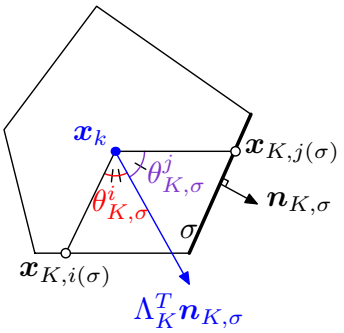


Fig. 1. Notations for the edge normal flux calculation.

we define the subset $\mathcal{P}_K^\sigma = \{\boldsymbol{x}_{K,i(\sigma)}, \boldsymbol{x}_{K,j(\sigma)}\}$ of $\mathcal{P}_K$ such that the co-normal $\Lambda_K^T\boldsymbol{n}_{K,\sigma}$ is located within $\boldsymbol{x}_{K,i(\sigma)} - \boldsymbol{x}_K$ and $\boldsymbol{x}_{K,j(\sigma)} - \boldsymbol{x}_K$. Sometimes we shall write $\boldsymbol{x}_{K,i}$, $\boldsymbol{x}_{K,j}$ instead of $\boldsymbol{x}_{K,i(\sigma)}$, $\boldsymbol{x}_{K,j(\sigma)}$ for simplicity. $\theta^i_{K,\sigma}$ (resp., $\theta^j_{K,\sigma}$) is the angle between $\boldsymbol{x}_{K,i} - \boldsymbol{x}_K$ (resp., $\boldsymbol{x}_{K,j} - \boldsymbol{x}_K$) and $\Lambda_K^T\boldsymbol{n}_{K,\sigma}$. We have the following identity [24, 14]

$$\frac{\Lambda_K^T\boldsymbol{n}_{K,\sigma}}{\|\Lambda_K^T\boldsymbol{n}_{K,\sigma}\|} = \frac{\sin\theta^j_{K,\sigma}}{\sin\theta_{K,\sigma}}\frac{\boldsymbol{x}_{K,i} - \boldsymbol{x}_K}{\|\boldsymbol{x}_{K,i} - \boldsymbol{x}_K\|} + \frac{\sin\theta^i_{K,\sigma}}{\sin\theta_{K,\sigma}}\frac{\boldsymbol{x}_{K,j} - \boldsymbol{x}_K}{\|\boldsymbol{x}_{K,j} - \boldsymbol{x}_K\|}, \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean vector norm and $\theta_{K,\sigma} = \theta^i_{K,\sigma} + \theta^j_{K,\sigma}$.

Multiplying both sides of (6) with $-\nabla u$ and integrating over σ, we obtain a linearity-preserving one-sided flux of the form

$$F_{K,\sigma} = \sum_{p\in\{i,j\}} \alpha^p_{K,\sigma}(u_K - u_{K,p}), \tag{7}$$

where $u_{K,i}$, $u_{K,j}$ are approximate solutions at the interpolation points $\boldsymbol{x}_{K,i}$, $\boldsymbol{x}_{K,j} \in \mathcal{P}_K^\sigma$ and the coefficients

$$\alpha^i_{K,\sigma} = \frac{\|\Lambda_K^T\boldsymbol{n}_{K,\sigma}\||\sigma|}{\|\boldsymbol{x}_{K,i} - \boldsymbol{x}_K\|}\frac{\sin\theta^j_{K,\sigma}}{\sin\theta_{K,\sigma}}, \quad \alpha^j_{K,\sigma} = \frac{\|\Lambda_K^T\boldsymbol{n}_{K,\sigma}\||\sigma|}{\|\boldsymbol{x}_{K,j} - \boldsymbol{x}_K\|}\frac{\sin\theta^i_{K,\sigma}}{\sin\theta_{K,\sigma}}.$$

For the construction of a monotone scheme using C. Le Potier's nonlinear two-point flux approximation, the coefficients in (7) are required to satisfy the following condition,

$$\alpha^i_{K,\sigma} \geq 0, \quad \alpha^j_{K,\sigma} \geq 0, \quad \forall\sigma \in \mathcal{E}_K, K \in \mathcal{M}. \tag{8}$$

A sufficient condition for (8) is that all angles $\theta^i_{K,\sigma}$, $\theta^j_{K,\sigma}$ and $\theta_{K,\sigma}$ are less than π.

Analogously, the one-sided flux constructed from cell L (that shares edge σ with cell K) can be obtained and given by

$$F_{L,\sigma} = \sum_{p\in\{i',j'\}} \alpha^p_{L,\sigma}(u_L - u_{L,p}), \tag{9}$$

also $\alpha^{i'}_{L,\sigma}$ and $\alpha^{j'}_{L,\sigma}$ are required to be non-negative.

3.3. A unique definition of the flux

This section is contributed to the nonlinear two-point flux approximation, and a unique definition of the flux is aimed to achieve local conservation. For $\sigma \in \mathcal{E}_K \cap \mathcal{E}_L$, we use two one-sided fluxes to define

$$\tilde{F}_{K,\sigma} = \mu_{K,\sigma} F_{K,\sigma} - \mu_{L,\sigma} F_{L,\sigma}, \ \tilde{F}_{L,\sigma} = \mu_{L,\sigma} F_{L,\sigma} - \mu_{K,\sigma} F_{K,\sigma}, \tag{10}$$

where $\mu_{K,\sigma}$ and $\mu_{L,\sigma}$ are two parameters. Obviously, for such defined fluxes, we have the local conservation condition

$$\tilde{F}_{K,\sigma} + \tilde{F}_{L,\sigma} = 0, \quad \forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_L \subset \mathcal{E}^{int}. \tag{11}$$

Substituting (7) and (9) into the first equation in (10), we have

$$\tilde{F}_{K,\sigma} = \mu_{K,\sigma} \sum_{p\in\{i,j\}} \alpha^p_{K,\sigma} u_K - \mu_{L,\sigma} \sum_{p\in\{i',j'\}} \alpha^p_{L,\sigma} u_L + \mu_{L,\sigma} a_{L,\sigma} - \mu_{K,\sigma} a_{K,\sigma}, \tag{12}$$

where

$$a_{K,\sigma} = \sum_{p\in\{i,j\}} \alpha^p_{K,\sigma} u_{K,p}, \quad a_{L,\sigma} = \sum_{p\in\{i',j'\}} \alpha^p_{L,\sigma} u_{L,p}. \tag{13}$$

3.3.1. The previous nonlinear two-point flux approximation

In order to derive a two-point consistent flux approximation that leads to a final monotonic scheme, by following the ideas in [17, 13, 24], one sees that the undetermined parameters $\mu_{K,\sigma}$ and $\mu_{L,\sigma}$ should be selected in a way such that the following requirements are satisfied,

$$\begin{cases} \mu_{K,\sigma} \geq 0, \quad \mu_{L,\sigma} \geq 0, \\ \mu_{K,\sigma} + \mu_{L,\sigma} = 1, \\ \mu_{L,\sigma} a_{L,\sigma} - \mu_{K,\sigma} a_{K,\sigma} = 0, \end{cases} \tag{14}$$

which leads to the following algorithm,

$$\mu_{K,\sigma} = \begin{cases} 0.5, & a_{K,\sigma} + a_{L,\sigma} = 0, \\ \dfrac{a_{L,\sigma}}{a_{K,\sigma} + a_{L,\sigma}}, & a_{K,\sigma} + a_{L,\sigma} \neq 0 \end{cases} \quad \text{and} \quad \mu_{L,\sigma} = 1 - \mu_{K,\sigma}, \tag{15}$$

subjected to the constraint condition below

$$a_{K,\sigma} \geq 0, \quad a_{L,\sigma} \geq 0. \tag{16}$$

Then we obtain

$$\tilde{F}_{K,\sigma} = A_{K,\sigma}(u) u_K - A_{L,\sigma}(u) u_L, \quad \tilde{F}_{L,\sigma} = A_{L,\sigma}(u) u_L - A_{K,\sigma}(u) u_K, \tag{17}$$

where

$$A_{K,\sigma}(u) = \mu_{K,\sigma} \sum_{p\in\{i,j\}} \alpha^p_{K,\sigma}, \quad A_{L,\sigma}(u) = \mu_{L,\sigma} \sum_{p\in\{i',j'\}} \alpha^p_{L,\sigma}. \tag{18}$$

It can be easily seen from (13) that, the interpolation algorithm for auxiliary unknowns must be a positive-preserving one so that (16) is maintained. Usually, a second-order interpolation algorithm is one of the key ingredients for a second-order cell-centered scheme. When a second-order interpolation algorithm is not positive-preserving, the authors in [24] suggested that it should be replaced by some lower-order but positive-preserving ones, such as the inverse distance weights used in [13] and given by

$$u_0 = \sum_i w_i u_i, \quad w_i = \frac{\bar{w}_i}{\sum_i \bar{w}_i}, \quad \bar{w}_i = \frac{1}{\|\boldsymbol{x}_0 - \boldsymbol{x}_i\|}, \tag{19}$$

where u_0 (resp., u_i) denotes the auxiliary (resp., cell-centered) unknown defined at the interpolation point $\boldsymbol{x}_0$ (resp., cell-center $\boldsymbol{x}_i$). Unfortunately, in practical computation, especially for grids with large distortion, the positive-preserving property for nearly all second-order interpolation algorithms can be easily spoiled. As a consequence, one has to use lower-order interpolation algorithm at many points and the resulting monotone schemes have poor accuracy on severe distorted meshes.

3.3.2. The new nonlinear two-point flux approximation

To deal with this problem, we still start from (12) and suggest the following new algorithm

$$\mu_{K,\sigma} = \begin{cases} 0.5, & a_{K,\sigma} = a_{L,\sigma} = 0, \\ \dfrac{|a_{L,\sigma}|}{|a_{K,\sigma}| + |a_{L,\sigma}|}, & \text{otherwise} \end{cases} \quad \text{and} \quad \mu_{L,\sigma} = 1 - \mu_{K,\sigma}. \tag{20}$$

In this situation, the constraint condition (16) does not need to be satisfied, however, the third equation in (14) is spoiled generally. This problem can be remedied by employing the following new nonlinear two-point flux approximation. Set

$$B_\sigma = \begin{cases} \mu_{L,\sigma} a_{L,\sigma} - \mu_{K,\sigma} a_{K,\sigma}, & \sigma \in \mathcal{E}_K \cap \mathcal{E}_L \subset \mathcal{E}^{int}, \\ -a_{K,\sigma}, & \sigma \in \mathcal{E}_K \cap \mathcal{E}^{ext} \end{cases} \tag{21}$$

and

$$B^+_\sigma = \frac{|B_\sigma| + B_\sigma}{2}, \quad B^-_\sigma = \frac{|B_\sigma| - B_\sigma}{2}.$$

Then, we find that (12) can be rewritten as

$$\begin{aligned}\tilde{F}_{K,\sigma} &= \mu_{K,\sigma}\sum_{p\in\{i,j\}}\alpha^p_{K,\sigma}u_K-\mu_{L,\sigma}\sum_{p\in\{i',j'\}}\alpha^p_{L,\sigma}u_L+B^+_\sigma-B^-_\sigma\\ &= \tilde{A}_{K,\sigma}(u)u_K-\tilde{A}_{L,\sigma}(u)u_L+\frac{B^+_\sigma\varepsilon}{u_K+\varepsilon}-\frac{B^-_\sigma\varepsilon}{u_L+\varepsilon},\end{aligned}\tag{22}$$

where ε is a sufficient small positive parameter and

$$\tilde{A}_{K,\sigma}(u)=\mu_{K,\sigma}\sum_{p\in\{i,j\}}\alpha^p_{K,\sigma}+\frac{B^+_\sigma}{u_K+\varepsilon},\quad \tilde{A}_{L,\sigma}(u)=\mu_{L,\sigma}\sum_{p\in\{i',j'\}}\alpha^p_{L,\sigma}+\frac{B^-_\sigma}{u_L+\varepsilon}.\tag{23}$$

Finally, neglecting the last two terms in (22), we reach a new definition of the unique edge flux, given by

$$\tilde{F}_{K,\sigma}=\tilde{A}_{K,\sigma}(u)u_K-\tilde{A}_{L,\sigma}(u)u_L,\quad \tilde{F}_{L,\sigma}=\tilde{A}_{L,\sigma}(u)u_L-\tilde{A}_{K,\sigma}(u)u_K.\tag{24}$$

Obviously, these last fluxes are *nonlinear* two-point ones and the local conservation condition (11) is still maintained.

Remark 3.1. *Although the drop of the last two terms in (22) may run the risk of losing accuracy, we are much rewarded by a monotone scheme that has no constraints on the interpolation algorithm. The monotonicity of the new schemes can be proved only under condition (8), which can be seen in the later discussion. Moreover, by (20) and (21), we can easily conclude that*

$$\frac{B^+_\sigma\varepsilon}{u_K+\varepsilon}-\frac{B^-_\sigma\varepsilon}{u_L+\varepsilon}=\begin{cases}0, & a_{K,\sigma}a_{L,\sigma}\geq 0,\\ O(\varepsilon), & a_{K,\sigma}a_{L,\sigma}<0 \text{ and } u_Ku_L\neq 0,\\ O(B_\sigma), & a_{K,\sigma}a_{L,\sigma}<0 \text{ and } u_Ku_L=0.\end{cases}\tag{25}$$

Thus, when the interpolation algorithm is a positivity-preserving one, by (8) we reach (16), which is included in the first case of (25). In this case the nonlinear two-point flux approximation (24) reduces to that of C. Le Potier's (17)-(18). The third case in (25) is quite rare, and it does not occur in our numerical tests in this paper. Therefore, the neglecting of the last two terms in (22) is probably a theoretical issue instead of a numerical one.

As for a boundary edge $\sigma\in\mathcal{E}_K\cap\mathcal{E}^{ext}\subset\Gamma_D$, we simply set

$$\tilde{F}_{K,\sigma}=F_{K,\sigma}=\sum_{p\in\{i,j\}}\alpha^p_{K,\sigma}u_K+B^+_\sigma-B^-_\sigma,\tag{26}$$

where B^+_σ can be handled in a way analogous to that of an interior edge while B^-_σ can be moved to the right-hand side of the final finite volume equation.

3.4. Interpolation of the auxiliary variables

To make the finite volume scheme a cell-centered one, the auxiliary variables in the flux expressions have to be eliminated by a certain interpolation procedure. Since we have two types of the auxiliary variables, the corresponding interpolation techniques will be given separately.

3.4.1. Interpolation of the auxiliary variables at cell vertices

We first present an interpolation technique for vertex unknowns, which was originally suggested in [8] and named LPEW2 therein. Recently, LPEW2 has been used to construct a robust nonlinear scheme on triangle grids [19]. Here for completeness, we give an equivalent but simpler derivation, and the resulting formula is easier for coding.

Suppose that a vertex Q_0 is surrounded by the cell Ω_k centered at O_k and enclosed by edges Q_0P_k and Q_0P_{k+1} where $1 \le k \le n_0$, see Fig. 2(a) for an example of $n_0 = 5$. T_k is an interior point on the edge Q_0P_k, defined by

$$T_k = \tau Q_0 + (1-\tau)P_k, \quad k = 1, \cdots, n_0, \tag{27}$$

where $\tau \in (0,1)$. For simplicity, we assume that k is a periodic number so that $P_{n_0+1} = P_1$, $T_{n_0} = T_0$, etc. Let u_0, u_k, $\bar{u}_k$, Λ_k and $S_{k,i}$ denote the vertex unknown at Q_0, the cell-centered unknown at O_k, the edge unknown at T_k, the piecewise constant value of Λ on cell Ω_k and the area of triangle $\triangle O_kQ_0T_{k+i-1}$ $(i = 1, 2)$, respectively.

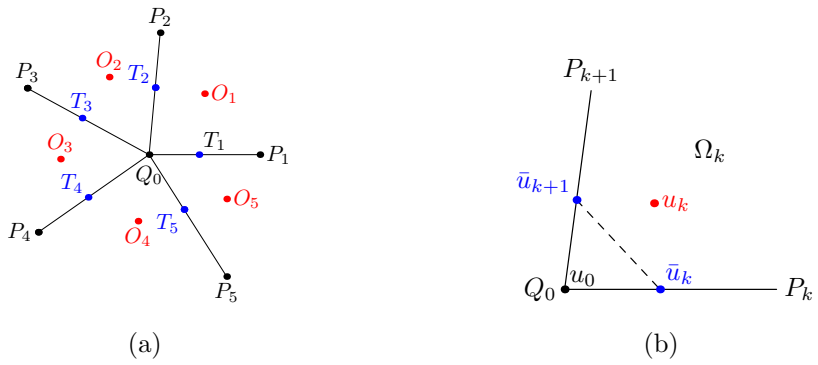


Fig. 2. The notations and local structure around a mesh vertex.

On the one hand, by the linearity-preserving criterion and the Green formula, we obtain the approximate gradient on $\triangle O_kQ_0T_k$

$$\nabla u \simeq -\frac{1}{2S_{k,1}}\left[(u_k - u_0)\mathcal{R}\overrightarrow{Q_0T_k} + (\bar{u}_k - u_0)\mathcal{R}\overrightarrow{O_kQ_0}\right],$$

where $\mathcal{R}$ denotes an operator that rotates a vector clockwise to its normal direction. Then, the normal component of the flux $\mathbf{F} = -\Lambda\nabla u$ across the interface Q_0T_k can be expressed as

$$\mathbf{F} \cdot (\mathcal{R}\overrightarrow{Q_0T_k}) \simeq \xi_{k,1}\,(u_k - u_0) + \bar{\xi}_{k,1}(\bar{u}_k - u_0),$$

where for $k = 1, 2, \cdots, n_0$ and $i = 1, 2$,

$$\xi_{k,i} = \frac{(\mathcal{R}\overrightarrow{Q_0T_{k+i-1}})^T \Lambda_k (\mathcal{R}\overrightarrow{Q_0T_{k+i-1}})}{2S_{k,i}}, \ \bar{\xi}_{k,i} = \frac{(\mathcal{R}\overrightarrow{Q_0T_{k+i-1}})^T \Lambda_k (\mathcal{R}\overrightarrow{O_kQ_0})}{2S_{k,i}}.$$

Similarly, on the triangular domain $\triangle O_{k-1}T_kQ_0$ we obtain the normal flux

$$\mathbf{F} \cdot (\mathcal{R}\overrightarrow{T_kQ_0}) \simeq \xi_{k-1,2}\,(u_{k-1} - u_0) + \bar{\xi}_{k-1,2}(\bar{u}_k - u_0).$$

By employing the local conservation condition

$$\mathbf{F} \cdot (\mathcal{R}\overrightarrow{Q_0T_k}) + \mathbf{F} \cdot (\mathcal{R}\overrightarrow{T_kQ_0}) = 0, \tag{28}$$

we have

$$\bar{u}_k - u_0 = -\frac{\xi_{k,1}}{\bar{\xi}_{k,1} + \bar{\xi}_{k-1,2}}(u_k - u_0) - \frac{\xi_{k-1,2}}{\bar{\xi}_{k,1} + \bar{\xi}_{k-1,2}}(u_{k-1} - u_0). \tag{29}$$

On the other hand, under the linearity-preserving criterion, the contour integration of the normal flux along the boundaries of $\triangle Q_0T_kT_{k+1}(k = 1, \cdots, n_0)$ vanishes. Consequently, by using (28), we have

$$\sum_{k=1}^{n_0} \mathbf{F} \cdot \left(\mathcal{R}\overrightarrow{T_kT_{k+1}}\right) \simeq 0. \tag{30}$$

Employing the linearity-preserving criterion on the triangle $\triangle Q_0T_kT_{k+1}$, we have

$$\mathbf{F} \cdot \mathcal{R}\left(\overrightarrow{T_kT_{k+1}}\right) \simeq \eta_{k,1}(\bar{u}_{k+1} - u_0) - \eta_{k,2}(\bar{u}_k - u_0), \tag{31}$$

where in the cell $\Omega_k, k = 1, \cdots, n_0$,

$$\eta_{k,i} = \frac{(\mathcal{R}\overrightarrow{T_kT_{k+1}})^T \Lambda_k (\mathcal{R}\overrightarrow{Q_0T_{k+i-1}})}{2S_{\triangle Q_0T_kT_{k+1}}}, \quad i = 1, 2.$$

Inserting (29) and (31) into (30) gives the interpolation formula

$$u_0 = \sum_{i=1}^{n_0} w_i u_i, \tag{32}$$

where

$$w_i = \frac{\bar{w}_i}{\sum_{k=1}^{n_0} \bar{w}_k} \quad \text{and} \quad \bar{w}_k = \frac{\eta_{k-1,1} - \eta_{k,2}}{\bar{\xi}_{k-1,2} + \bar{\xi}_{k,1}}\xi_{k,1} + \frac{\eta_{k,1} - \eta_{k+1,2}}{\bar{\xi}_{k,2} + \bar{\xi}_{k+1,1}}\xi_{k,2}. \tag{33}$$

Here we remark that the above weights for cell-vertex unknowns are independent of τ [8], and generally speaking, the interpolation formula (32) is not positivity-preserving, i.e., $w_i \geq 0$ cannot be always guaranteed. As far as we know, there exists no second-order positivity-preserving interpolation algorithm for vertex unknowns. Thanks to the new nonlinear two-point flux approximation (24) and (23), the possible negative weights cause no problem to the monotonicity of the new scheme, at least it seems so theoretically.

3.4.2. Interpolation of the auxiliary variables at cell edges

For the cell-edge unknowns u_σ, we have the following interpolation formula

$$u_\sigma = \omega_{K,\sigma} u_K + \omega_{L,\sigma} u_L, \quad \forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_L, \tag{34}$$

where

$$\omega_{K,\sigma} = \frac{d_{L,\sigma}\lambda_K^{(n)}}{d_{L,\sigma}\lambda_K^{(n)} + d_{K,\sigma}\lambda_L^{(n)}}, \quad \omega_{L,\sigma} = 1 - \omega_{K,\sigma}. \tag{35}$$

By (5), (34) is linearity-preserving if **(A1)** holds. When **(A1)** is violated, (34) is still adopted, however, it is possible for the linearity-preserving property to be spoiled. Since $\omega_{K,\sigma}$ and $\omega_{L,\sigma}$ are non-negative, the positivity-preserving interpolation is always assured.

3.5. The finite volume scheme

With the definition of $\tilde{F}_{K,\sigma}$ and $\tilde{F}_{L,\sigma}$, we formulate the nonlinear monotone finite volume scheme as follows: find $\{u_K, K \in \mathcal{M}\}$ such that

$$\sum_{\sigma \in \mathcal{E}_K} \tilde{F}_{K,\sigma} = \int_K f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}, \quad \forall\, K \in \mathcal{M}. \tag{36}$$

From (36), we get a nonlinear algebraic system

$$\mathbb{M}(\mathbf{U})\mathbf{U} = \mathbf{F}(\mathbf{U}), \tag{37}$$

where $\mathbf{U}$ denotes the vector unknowns and $\mathbb{M}(\mathbf{U})$ the coefficient matrix. The right-hand side vector $\mathbf{F}(\mathbf{U})$ is generated by the source term, the boundary data and the possible movement of B_σ^- in (26) to the right-hand side of (36).

The nonlinear system can be solved by Picard iterations in this article: choose a small value $\varepsilon_{non} > 0$ and initial solution vector $\mathbf{U}^0$ with positive entries, and repeat for $k = 0, 1, 2, \cdots$,

1. Assemble the global matrix $\mathbb{M}(\mathbf{U}^k)$, and calculate the right-hand side vector $\mathbf{F}(\mathbf{U}^k)$.
2. Solve $\mathbb{M}(\mathbf{U}^k)\mathbf{U}^{k+1} = \mathbf{F}(\mathbf{U}^k)$ to obtain $\mathbf{U}^{k+1}$.
3. Stop if $\|\mathbb{M}(\mathbf{U}^{k+1})\mathbf{U}^{k+1} - \mathbf{F}(\mathbf{U}^{k+1})\| \le \varepsilon_{non}\|\mathbb{M}(\mathbf{U}^0)\mathbf{U}^0 - \mathbf{F}(\mathbf{U}^0)\|$.

The linear system in Step 2 with the non-symmetric matrix $\mathbb{M}(\mathbf{U}^k)$ is solved by GMRES method, and the GMRES iterations are terminated when the relative norm of the initial residual becomes smaller than a small positive parameter ε_{lin}.

Table 1. Three variants of the nonlinear monotone scheme.

Scheme	$\mathcal{P}_K$	$\mu_{K,\sigma}$	Interpolation method
LPS-TP1	$\mathcal{P}_K^v$	(15)	(32),(33) if $w_i \geq 0$; (19) otherwise
LPS-TP2	$\mathcal{P}_K^v$	(20)	(32),(33)
LPS-TP3	$\mathcal{P}_K^e$	(20)	(34),(35)

3.6. Variants of the nonlinear monotone scheme

From the above construction algorithm, it can be seen that we have different choices in Step 1, Step 3 and Step 4, which leads to some variants of the scheme summarized in Table 1. Here we point out that LPS-TP1 can be viewed as a direct extension of the one in [24], although the associated interpolation algorithms for vertex unknowns are totally different. We do not recommend it in this paper and it is given here just for the need of comparison.

Theorem 3.1. *Let $f \geq 0$, $g_D \geq 0$, $g_N \leq 0$ and $\Gamma_D \neq \emptyset$ in (1)–(3). Assume that (8) holds. If the initial solution vector $\mathbf{U}^0 \geq 0$ and linear systems in the Picard iterations are solved exactly, then for the new schemes LPS-TP2 and LPS-TP3, we have $\mathbf{U}^k \geq 0$ for $k \geq 1$.*

Proof. What we have to prove is that

$$\mathbb{M}^{-1}(\mathbf{U}^{k-1}) \geq 0, \quad \mathbf{F}(\mathbf{U}^{k-1}) \geq 0, \quad k = 1, 2, \cdots.$$

Since $f \geq 0$, $g_D \geq 0$ and $g_N \leq 0$, the second inequality follows immediately. As for the first one, the argument is similar to that in [24] so that only a sketch is given. Actually, under assumptions (8), one can easily see that $A_{K,\sigma}(u) > 0$ and $A_{L,\sigma}(u) > 0$ provided that $\mathbf{U}^{k-1} \geq 0$. Then, we deduce from (24) and (26) that: (i) All diagonal entries of matrix $\mathbb{M}(\mathbf{U}^{k-1})$ are positive; (ii) All off-diagonal entries of matrix $\mathbb{M}(\mathbf{U}^{k-1})$ are non-positive; (iii) All column sums in $\mathbb{M}(\mathbf{U}^{k-1})$ are non-negative and there exists at least one such sum that is positive. Thus, it follows that $\mathbb{M}^T(\mathbf{U}^{k-1})$ is an M-matrix, since it is obviously irreducible. As a consequence, $\mathbb{M}^{-T}(\mathbf{U}^{k-1})$ and in turn, $\mathbb{M}^{-1}(\mathbf{U}^{k-1})$, are non-negative, which completes the proof. □

The monotonicity proof of LPS-TP1 can be obtained analogously under the constraint conditions (8) and (16). Obviously, the requirements for the monotonicity of LPS-TP2 and LPS-TP3 are much weaker than those for LPS-TP1.

4. Numerical Examples

We use discrete L_2-norm to evaluate discretization errors for the solution:

$$E_u = \left(\sum_{K \in \mathcal{M}} |K| (u(\boldsymbol{x}_K) - u_K)^2 \right)^{\frac{1}{2}}.$$

Discrete L_2-norm of the error on the solution gradient can be defined similarly and is denoted by E_q. The rate of convergence R_α ($\alpha = u, q$) is obtained by a least squares fit on the ones computed on each two successive meshes by the following formula

$$\frac{\log[E_\alpha(h_2)/E_\alpha(h_1)]}{\log(h_2/h_1)},$$

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

All tests are performed in double precision, and we use GMRES for solving linear systems with stopping tolerance ε_{lin}=1.0E-15. The nonlinear iterations are terminated when the reduction of the initial residual norm becomes smaller than ε_{non}=1.0E-07. In addition, we use the following notations for the numerical tests:

- `umin`: minimal value of the approximate solution;
- `umax`: maximal value of the approximate solution;
- `nitn`: number of nonlinear iterations.

The first three sections Section 4.1–Section 4.3 are contributed to verify the monotonicity of the schemes LPS-TP1, LPS-TP2 and LPS-TP3 numerically on different kind of models and different meshes. The linearity-preserving property for our schemes is confirmed in Section 4.4. Finally in Section 4.5–Section 4.7, the convergence analysis for three schemes LPS-TP1, LPS-TP2 and LPS-TP3 are investigated.

4.1. Monotonicity test I: oblique flow

In this test, we consider test 3 described in FVCA V[10]. The computational domain is $\Omega = [0, 1]^2$, and the anisotropic tensor is

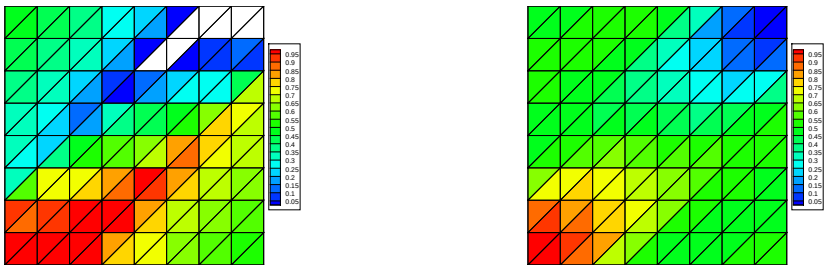
$$\Lambda = R_\theta \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} R_\theta^{-1}, \tag{38}$$

where R_θ is the rotation of angle $\theta = 40°$, $k_1 = 1$, $k_2 = 10^{-4}$ and the source term $f = 0$. Dirichlet boundary conditions are considered, $u = g_D$ on $\partial\Omega$, with g_D a continuous and

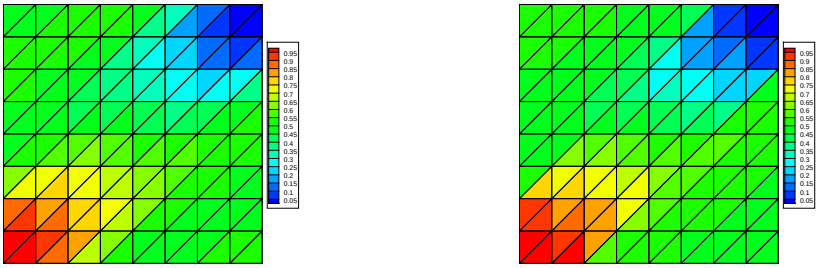
piecewise linear function defined by

$$g_D(x,y) = \begin{cases} 1 & \text{on} \quad (0,0.2)\times\{0\}\cup\{0\}\times(0,0.2), \\ 0 & \text{on} \quad (0.8,1)\times\{1\}\cup\{1\}\times(0.8,1), \\ 0.5 & \text{on} \quad (0.3,1)\times\{0\}\cup\{0\}\times(0.3,1), \\ 0.5 & \text{on} \quad (0,0.7)\times\{1\}\cup\{1\}\times(0,0.7). \end{cases}$$

The solution features a Z across the $y = x$ axis and the monotonicity is not always easy to verify for such a solution (see Table 4 in [10]).



(a) MPFA (b) LPS-TP1



(c) LPS-TP2 (d) LPS-TP3

Fig. 3. Solution profiles with MPFA and our schemes on the uniform triangle mesh. (White color denotes position of negative values).

We give the results on the uniform triangular mesh with 128 triangles, and Fig. 3 shows the solution behavior of classical MFPA-O method [1] and the schemes LPS-TP1, LPS-TP2 and LPS-TP3. We observe white areas where the discrete solutions are negative for the MPFA-O method, and solutions obtained with our three schemes are non-negative everywhere in Ω. Note that if we choose a uniform rectangular mesh, no oscillations appear with MPFA-O method.

4.2. Monotonicity test II: heterogeneous diffusion tensor

This problem is given in [13]. We solve the problem (1)–(2) in the domain $\Omega = [0,1]^2$ with the source term

$$f(\boldsymbol{x}) = \begin{cases} \frac{81}{4} & \text{if } \boldsymbol{x} \in \left[\frac{7}{18}, \frac{11}{18}\right]^2, \\ 0 & \text{otherwise.} \end{cases}$$

As shown in Fig. 5, the domain Ω is partitioned into four square subdomains Ω_i, $i = 1, \cdots, 4$. The diffusion tensor Λ is given by (38).

We give the numerical results on the randomly distorted quadrilateral mesh which is constructed from the uniform square mesh with the mesh size h by random distortion of its nodes:

$$x := x + \alpha\xi_x h, \qquad y := y + \alpha\xi_y h. \tag{39}$$

Here ξ_x and ξ_y are random variables with values located in $[-0.5, 0.5]$ and $\alpha \in [0, 1]$ is the degree of distortion. As shown in Fig. 4, we choose $\alpha = 0.8$ and all the mesh nodes located on the line $x = 0.5$ (resp., $y = 0.5$) were distorted only in the $y-$direction (resp., $x-$direction) in this test.

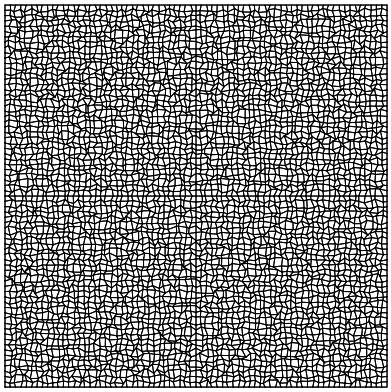
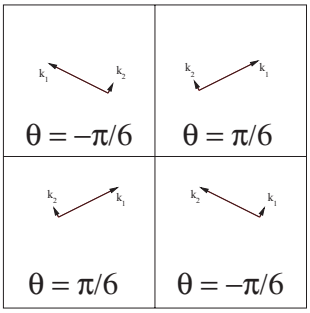
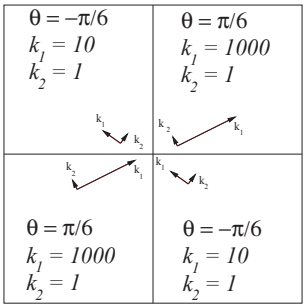


Fig. 4. 72×72 distorted quadrilateral mesh with distortion parameter $\alpha = 0.8$.



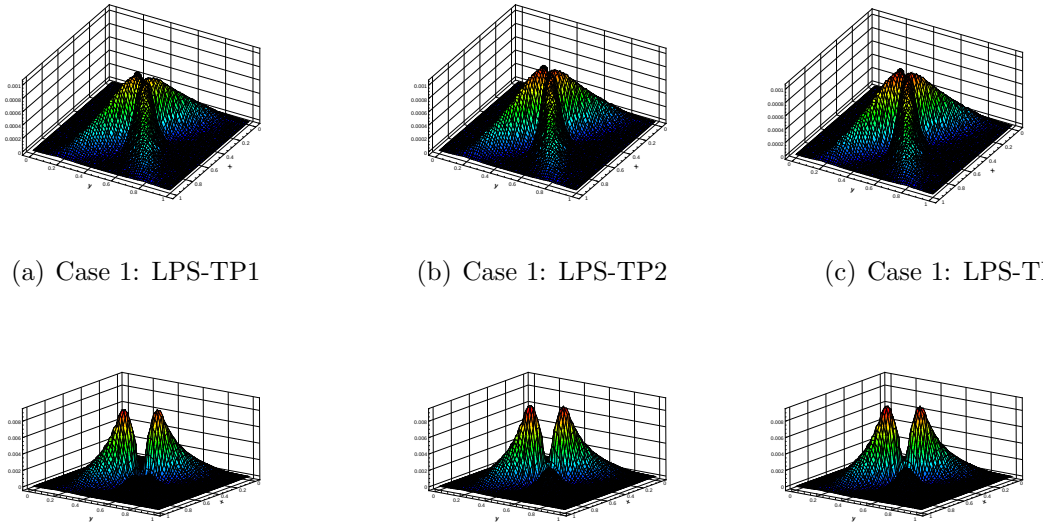
(a) Case 1: $k_1 = 1000, k_2 = 1$



(b) Case 2

Fig. 5. Principle directions and eigenvalues of the diffusion tensor.

Firstly, as illustrated in Fig. 5(a), we set $k_1 = 1000$, $k_2 = 1$ and vary only the angle θ, the solution profiles are depicted in the top three figures of Fig. 6. Secondly, we use various θ and the chess board distribution of k_1 and k_2 (see Fig. 5(b)), and the solution profiles are given in the bottom three figures of Fig. 6. In both cases the non-negative solutions are obtained for the three schemes.



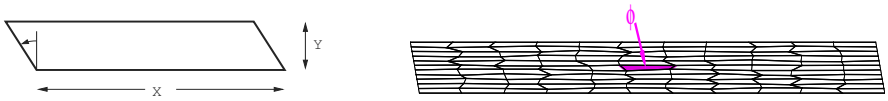
(a) Case 1: LPS-TP1 (b) Case 1: LPS-TP2 (c) Case 1: LPS-TP3

(d) Case 2: LPS-TP1 (e) Case 2: LPS-TP2 (f) Case 2: LPS-TP3

Fig. 6. Profiles of the discrete solutions for three schemes in Case 1 and Case 2.

4.3. Monotonicity test III: problem with point source on perturbed parallelograms

This test was given in [2, 10]. Problem (1) is considered with the diffusion tensor $\Lambda = \mathrm{Id}$ and the full Dirichlet boundary condition (2) on the parallelogram (see Fig. 7(a)). As shown in Fig. 7(b), a perturbed parallelogram mesh is employed and the source term f is equal to zero in all cells except cell $\phi = (6, 6)$ where $\int_\phi f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$. Note that the solution should be a function with a maximum in the cell ϕ. If the solution shows internal oscillations or is negative, Hopf's first lemma is violated.



(a) $X = 1$, $Y = \frac{1}{30}$ and angle 30° (b) 11×11 disturbed mesh

Fig. 7. The parallelogram-shaped domain (left) with associated perturbed mesh (right).

In Table 2, the minimum and maximum solutions of various schemes are given. The results on the fine grid are obtained with the scheme SUSHI-P [6] on a 201×201 uniform grid which was chosen parallel to the axes, and directly copied from Table 12 in [10]. There are only nonliner schemes which remain positive, namely SSEPS[9], LPS-TP1, LPS-TP2 and LPS-TP3.

Table 2. Minimum and maximum solutions of various schemes on the parallelogram mesh.

Scheme	`umin`	`umax`
Fine grid[10]	1.07E-24	4.10E-01
SLPS[23]	-7.51E-04	7.19E-02
MPFA-O[1]	-7.61E-02	1.87E-01
LPEW2[8]	-1.20E-03	7.85E-02
SSEPS[9]	1.58E-09	8.03E-02
LPS-TP1	3.92E-07	5.00E-02
LPS-TP2	3.92E-07	5.00E-02
LPS-TP3	1.41E-09	7.82E-02

4.4. Linearity-preserving verification

We investigate (1)-(2) and the domain Ω is composed of three subdomains

$$\begin{aligned}\Omega_1 &= \{(x,y)\in\Omega|\phi_1(x,y)<0\},\\ \Omega_2 &= \{(x,y)\in\Omega|\phi_1(x,y)>0,\phi_2(x,y)<0\},\\ \Omega_3 &= \{(x,y)\in\Omega|\phi_2(x,y)>0\},\end{aligned}$$

with $\phi_1(x,y)=y-\delta(x-0.5)-0.475$ and $\phi_2(x,y)=\phi_1(x,y)-0.05$. We take $\delta=0.2$ and define the exact solution $u(x,y)=-x-\delta y$. The permeability tensor Λ is chosen to be (38) with $\theta=\arctan\delta$, $k_1=100, k_2=10$ on Ω_2 and $k_1=1, k_2=0.1$ on $\Omega_1\cup\Omega_3$. This test can be found in [10].

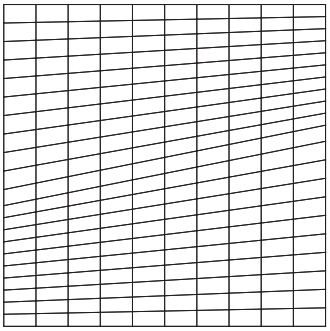


Fig. 8. Oblique mesh with mesh size h =1.25E-01.

We discrete this test using the oblique mesh (see Fig. 8), and set the parameters ε_{lin} and ε_{non} to be equal to the machine precision. The minimal and maximal values of exact solution `uemin` and `uemax`, and the minimal and maximal values of approximated solution `umin` and `umax` are given in the second–fifth columns of Table 3. The solutions are located in the interval $(-1.2, 0.0)$. The relative discrete L_2-norm of the solution errors E'_u and the errors on solution gradient E'_q are also given in the last two columns of Table 3. We can

conclude that our schemes are *linearity-preserving* [8], which means LPS-TP1, LPS-TP2 and LPS-TP3 exactly reproduce piecewise linear solutions as the exact solution is affine on the whole domain.

Table 3. Computational results on the oblique mesh.

scheme	`uemin`	`uemax`	`umin`	`umax`	E'_u	E'_q
LPS-TP1	-1.2	0.0	-1.14615	-0.05385	5.35E-16	6.70E-15
LPS-TP2	-1.2	0.0	-1.14615	-0.05385	5.50E-16	8.59E-15
LPS-TP3	-1.2	0.0	-1.14615	-0.05385	8.51E-16	8.85E-15

4.5. Convergence analysis I: smooth solution

In this section we investigate the convergence of our three schemes for an isotropic problem on largely distorted meshes. We consider two types of smooth solution problems.

In the fist problem, we consider a homogeneous isotropic medium with the exact solution $u = \cosh(\pi x)\cos(\pi y)$ on the distorted parallelogram mesh (see Fig. 9 as an example with aspect ratio 0.1), and the source term equals to zero. This problem is described in [2] where the authors demonstrate that the convergence of MPFA O-method is lost when aspect ratios deviate strongly from unity.

We employ the distorted parallelogram mesh with aspect ratio 0.001 which is a highly anisotropic mesh. The convergence rates for the solution and its gradient errors are graphically depicted in Fig. 10 as log-log plots of the discrete L_2 norm errors versus the characteristic mesh size h. The actual order is reflected by the slopes of error curves, and can be approximately evaluated by comparison with the "theoretical" first- and second-order slopes represented as dash-dot-dot lines in each figure. The convergence rates of LPS-TP1 and LPS-TP2 are h^2 for the solution and $h^{1.5}$ for the solution gradient, and the convergence rates of LPS-TP3 are approximately $h^{1.4}$ for the solution and $h^{0.9}$ for the solution gradient. The discrete L_2 errors for LPS-TP3 are much smaller than that for LPS-TP1 and LPS-TP2.

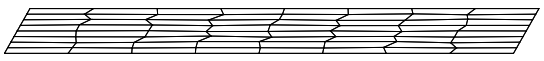
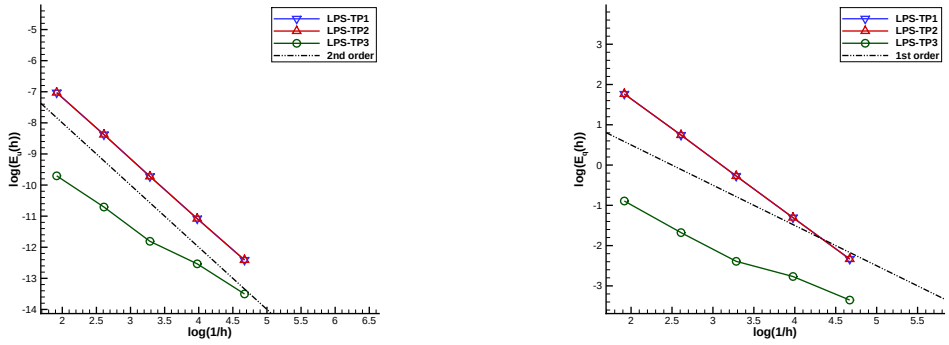


Fig. 9. Disturbed parallelogram mesh with aspect ratio 0.1 and angle 30°.

In the second problem, we consider (1)–(2) in $\Omega = [0,1]^2$ with the constant diffusion tensor $\Lambda = \mathrm{Id}$ and exact solution $u(x,y) = \sin(\pi x)\sin(\pi y)$. We study the convergence behaviours of the three schemes on a series of distorted meshes.

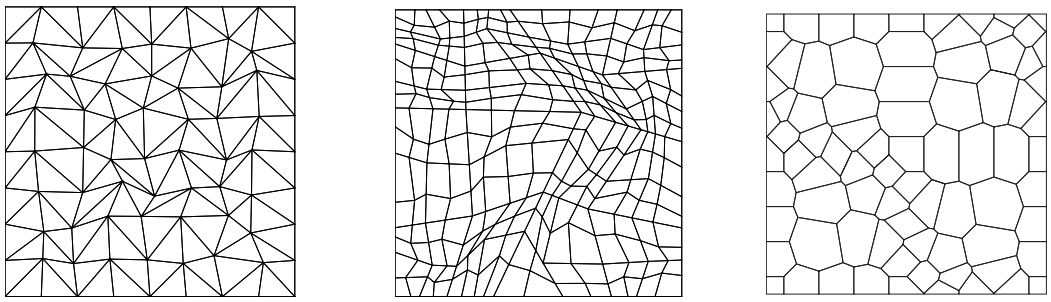
Firstly, we use a sequence of distorted triangular mesh, Shestakov mesh and polygonal mesh in this test (see Fig. 11), and the mesh refinement levels are also given. The distorted



(a) solution (b) gradient

Fig. 10. L_2 errors versus mesh size h for the three schemes on the disturbed parallelogram mesh.

triangular mesh is constructed from the uniform triangle mesh with the mesh size h by random distortion of internal nodes which are defined by (39). We set the distortion parameter $\alpha = 0.5$ for the distorted triangular mesh. The Shestakov mesh [21] is highly skewed and highly distorted.



(a) distorted triangular mesh (2.38E-01,1.61E-02) (b) Shestakov mesh (3.05E-01,4.42E-02) (c) polygonal mesh (2.30E-01,1.51E-02)

Fig. 11. Samples of the meshes: each mesh was used with 5 successive mesh levels, and the range of mesh size is shown in the caption of each mesh.

Averaged numbers of nonlinear iterations `itn`, convergence rates of solution and its gradient for three schemes are given in Table 4. On the distorted triangle mesh, we observe the second-order convergence rate for the solution and first-order convergence rate for its gradient. On the polygonal mesh, the convergence rates of LPS-TP2 and LPS-TP3 are h^2 for the solution errors and $h^{1.6}$ for its gradient errors. But the convergence rates of LPS-TP1 are only $h^{1.6}$ for the solution errors and $h^{0.6}$ for its gradient errors. Similar results are obtained on the Shestakov mesh, schemes LPS-TP2 and LPS-TP3 results in nearly second-order convergence rates for the solution errors, but the convergence rate of LPS-TP1 is only first order for the solution errors.

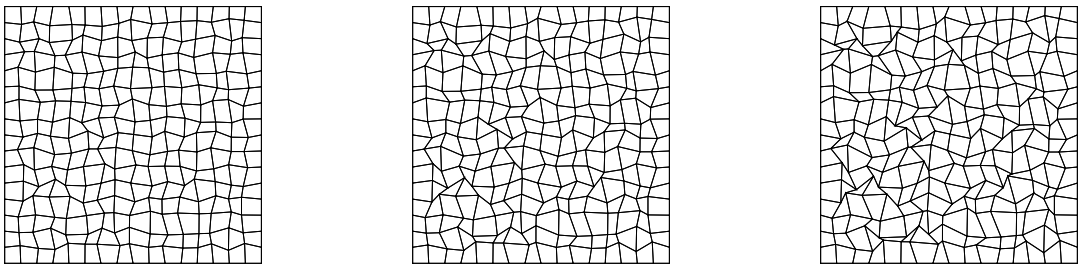
Secondly, we consider a set of randomly distorted quadrilateral meshes which were

Table 4. Solution behaviors on various meshes.

Scheme	distorted triangle mesh			polygonal mesh			Shestakov mesh		
	`nitn`	R_u	R_q	`nitn`	R_u	R_q	`nitn`	R_u	R_q
LPS-TP1	10.4	2.036	1.028	5.4	1.613	0.595	33.8	1.148	0.277
LPS-TP2	10.4	2.035	1.038	6.2	2.038	1.590	34.4	2.157	1.108
LPS-TP3	17.6	1.974	0.992	8.8	2.029	1.581	51.2	1.862	0.608

defined in Section 4.2 and the mesh nodes are defined by (39). As shown in Fig. 12, the larger α is, the more distorted mesh is generated. If $\alpha > 0.5$, mesh cells may be non-convex. The distortion is performed on each refinement level. We give the numerical results for $\alpha = 0.5$, 0.7 and 0.9 in Table 5, Table 6 and Table 7 respectively.

For various distortion, our schemes LPS-TP2 and LPS-TP3 always show a second-order convergence rate for the solution errors and first-order convergence rate fort its gradient errors. For the scheme LPS-TP1, the convergence rates on the mesh with $\alpha = 0.5$ are second order for the solution errors and first order for its gradient errors, on the mesh with $\alpha = 0.7$ they reduce to one-half, and on the mesh with $\alpha = 0.9$ the convergence is lost. The reason is that when the distortion grows, the numbers of negative weights in the interpolation formula (33) increase, so that we have to use (19) instead and the accuracy of cell-vertex unknowns decreases.



(a) $\alpha = 0.5$ (b) $\alpha = 0.7$ (c) $\alpha = 0.9$

Fig. 12. 16×16 randomly distorted quadrilateral meshes with various distortion parameter α.

Finally, we give some remarks on LPS-TP1 and LPS-TP2, the two-point schemes based on the vertex interpolation. It can be seen that in this test the scheme LPS-TP1 performs badly on largely distorted meshes, but scheme LPS-TP2 results in optimal convergence. Similar results can be observed if other second-order vertex interpolation methods are used. For example, we can replace our interpolation method (32) by the least square interpolation method [16], and the convergence rates for the solution and its gradient errors on Shestakov mesh and randomly distorted quadrilateral meshes with $\alpha = 0.7, 0.9$ are graphically depicted in Fig. 13 as log-log plots of the discrete L_2 norm

Table 5. Behavior on randomly distorted quadrilateral mesh with distortion $\alpha = 0.5$.

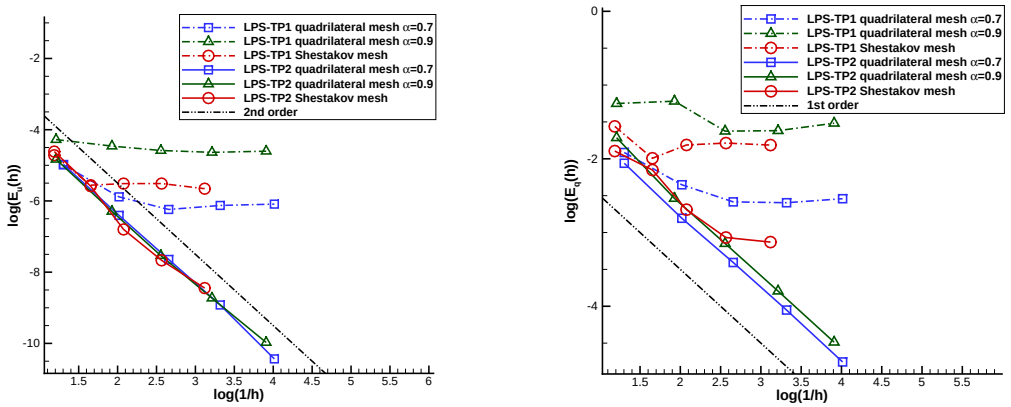
h	LPS-TP1			LPS-TP2			LPS-TP3		
	`nitn`	E_u	E_q	`nitn`	E_u	E_q	`nitn`	E_u	E_q
1/8	8	6.14E-3	9.08E-2	8	6.14E-3	9.08E-2	10	4.16E-3	8.65E-2
1/16	7	1.64E-3	4.41E-2	7	1.64E-3	4.41E-2	10	1.13E-3	3.86E-2
1/32	7	4.30E-4	2.37E-2	7	4.30E-4	2.37E-2	10	3.09E-4	2.09E-2
1/64	7	1.18E-4	1.25E-2	7	1.18E-4	1.25E-2	9	8.86E-5	1.14E-2
1/128	7	2.74E-5	6.20E-3	7	2.74E-5	6.20E-3	10	2.02E-5	5.66E-3
Rates		1.988	0.985		1.988	0.985		1.956	0.999

Table 6. Behavior on randomly distorted quadrilateral mesh with distortion $\alpha = 0.7$.

h	LPS-TP1			LPS-TP2			LPS-TP3		
	`nitn`	E_u	E_q	`nitn`	E_u	E_q	`nitn`	E_u	E_q
1/8	10	7.06E-3	1.34E-1	10	7.06E-3	1.34E-1	13	5.74E-3	1.31E-1
1/16	10	1.61E-3	6.86E-2	10	1.65E-3	6.27E-2	13	1.28E-3	5.52E-2
1/32	9	4.72E-4	3.61E-2	10	4.73E-4	3.40E-2	12	3.55E-4	2.94E-2
1/64	9	2.21E-4	2.93E-2	9	1.31E-4	1.80E-2	12	1.07E-4	1.63E-2
1/128	9	2.65E-4	3.16E-2	9	3.06E-5	8.97E-3	12	2.48E-5	8.12E-3
Rates		1.219	0.536		2.006	0.995		2.007	1.022

Table 7. Behavior on randomly distorted quadrilateral mesh with distortion $\alpha = 0.9$.

h	LPS-TP1			LPS-TP2			LPS-TP3		
	`nitn`	E_u	E_q	`nitn`	E_u	E_q	`nitn`	E_u	E_q
1/8	16	9.92E-3	2.60E-1	21	8.34E-3	2.00E-1	20	7.88E-3	1.97E-1
1/16	14	6.80E-3	2.38E-1	15	2.12E-3	8.72E-2	25	1.64E-3	7.95E-2
1/32	13	3.65E-3	1.36E-1	18	5.44E-4	4.55E-2	21	4.58E-4	4.04E-2
1/64	12	3.11E-3	1.28E-1	18	1.57E-4	2.43E-2	23	2.24E-4	3.64E-2
1/128	14	3.16E-3	1.49E-1	20	5.13E-5	1.27E-2	29	4.84E-5	1.92E-2
Rates		0.434	0.223		1.892	1.019		1.876	0.854



(a) solution (b) gradient

Fig. 13. L_2 errors versus mesh size h for schemes LPS-TP1 and LPS-TP2 with the least square interpolation method[16].

errors versus the characteristic mesh size h. We find that the convergence of LPS-TP1 is lost on three highly distorted meshes, but scheme LPS-TP2 represents approximately second-order convergence rate for the solution errors and first-order convergence rate for its gradient errors.

4.6. Convergence analysis II: discontinuous diffusion tensor

We solve the problem (1)-(2) on $\Omega = [0,1]^2$, and let the diffusion tensor Λ change the eigenvalues and orientation of eigenvectors across the line $x = 0.5$,

$$\Lambda = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & x \leq 0.5, \\ \begin{pmatrix} 10 & 3 \\ 3 & 1 \end{pmatrix}, & x > 0.5. \end{cases}$$

We choose the following exact solution

$$u(x,y) = \begin{cases} 1 - 2y^2 + 4xy + 6x + 2y, & x \leq 0.5, \\ -2y^2 + 1.6xy - 0.6x + 3.2y + 4.3, & x > 0.5. \end{cases}$$

This test is inspired by a numerical experiment in [14], and the condition number of the diffusion tensor is $\mathrm{cond}(\Lambda) = 1$ for $x \leq 0.5$ and $\mathrm{cond}(\Lambda) = 118.99$ otherwise. The numerical test is performed on the randomly distorted triangular mesh, randomly distorted quadrilateral mesh and Shestakov mesh (see Fig. 11(b)). The definitions of distorted meshes can be found in the previous sections and in this test all the mesh nodes located on the line $x = 0.5$ were distorted only in the $y-$direction. For each space resolution, the randomly distorted quadrilateral mesh and Shestakov mesh have the same numbers of cells, and the corresponding distorted triangular mesh has twice the cell numbers.

Numbers of nonlinear iterations, discrete L_2-errors of solution and its gradient for three schemes on each refinement level are given in Table 8, Table 9 and Table 10, respectively. On the distorted triangular mesh, the scheme LPS-TP1 is not convergent, but LPS-TP2 and LPS-TP3 obtain optimal convergence. On the distorted quadrilateral mesh and Shestakov mesh, the convergence rates for the schemes LPS-TP2 and LPS-TP3 are about $h^{1.8\sim2.0}$ for the solution errors and $h^{0.9\sim1.1}$ for its gradient errors, but the approximate solution for LPS-TP1 is only linear convergent in the L_2-norm.

Finally we can confirm that the convergence of scheme LPS-TP1 is lost in case of discontinuous model on distorted grids, but schemes LPS-TP2 and LPS-TP3 perform approximately quadratical convergence for the solution and linear convergence for its gradient. The results are in good agreement with those presented in the previous section.

Table 8. Convergence results on randomly distorted triangular meshes with $\alpha = 0.5$.

h	LPS-TP1			LPS-TP2			LPS-TP3		
	nitn	E_u	E_q	nitn	E_u	E_q	nitn	E_u	E_q
1/8	15	2.36E-2	5.78E-1	20	6.41E-3	1.13E-1	47	7.36E-3	9.89E-2
1/16	21	3.20E-2	7.76E-1	37	1.77E-3	5.58E-2	74	1.99E-3	4.91E-2
1/32	27	2.98E-2	7.24E-1	50	4.61E-4	2.63E-2	110	5.l0E-4	2.30E-2
1/64	30	3.40E-2	7.16E-1	69	1.11E-4	1.27E-2	143	1.27E-4	1.14E-2
1/128	32	3.36E-2	7.59E-1	97	3.01E-5	6.34E-3	189	3.45E-5	5.73E-3
Rates		-0.128	-0.100		1.989	1.068		1.990	1.057

Table 9. Convergence results on randomly distorted quadrilateral meshes with $\alpha = 0.8$.

h	LPS-TP1			LPS-TP2			LPS-TP3		
	nitn	E_u	E_q	nitn	E_u	E_q	nitn	E_u	E_q
1/8	25	2.06E-2	5.01E-1	26	8.70E-3	1.32E-1	44	7.04E-3	1.13E-1
1/16	33	1.43E-2	4.39E-1	33	2.21E-3	6.47E-2	61	1.96E-3	6.04E-2
1/32	50	2.66E-3	1.74E-1	51	5.72E-4	2.96E-2	83	4.51E-4	2.65E-2
1/64	74	1.68E-3	1.99E-1	74	1.61E-4	1.49E-2	119	1.22E-4	1.31E-2
1/128	102	1.15E-3	2.23E-1	135	4.26E-5	7.33E-3	153	3.23E-5	6.68E-3
Rates		1.086	0.307		1.969	1.072		1.997	1.051

Table 10. Convergence results on Shestakov mesh.

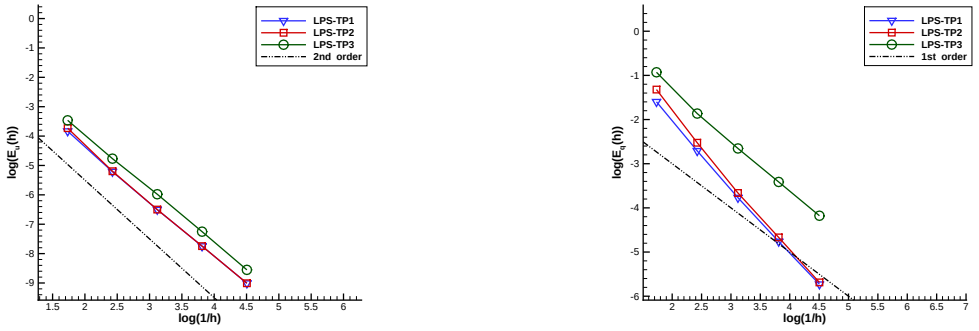
h	LPS-TP1			LPS-TP2			LPS-TP3		
	nitn	E_u	E_q	nitn	E_u	E_q	nitn	E_u	E_q
1/8	19	3.34E-2	6.99E-1	23	1.07E-2	1.29E-1	38	1.09E-2	1.24E-1
1/16	38	4.15E-3	8.83E-2	38	4.15E-3	8.83E-2	60	4.07E-3	7.59E-2
1/32	66	9.17E-3	2.95E-1	69	1.86E-3	5.86E-2	101	1.52E-3	4.72E-2
1/64	107	5.32E-3	2.92E-1	110	8.21E-4	3.59E-2	160	7.25E-4	3.34E-2
1/128	164	6.36E-3	3.38E-1	182	3.40E-4	2.15E-2	273	3.16E-4	2.18E-2
Rates		0.907	0.414		1.767	0.907		1.839	0.899

4.7. Convergence analysis III: heterogeneous rotating anisotropy

Problem (1)–(2) is defined in $\Omega = [0,1]^2$ with a rotating anisotropic diffusion tensor:

$$\Lambda = \frac{1}{x^2+y^2}\begin{pmatrix} \alpha x^2 + y^2 & (\alpha-1)xy \\ (\alpha-1)xy & x^2 + \alpha y^2 \end{pmatrix}, \qquad \alpha = 10^{-6},$$

where α characterizes the level of anisotropy, and the anisotropy ratio is 10^6 in this case. We consider the smooth exact solution $u(x,y) = \sin(\pi x)\sin(\pi y)$ and use the uniform square mesh in this test. The minimum solutions for three schemes on the coarsest mesh are 3.3E-2, 3.3E-2 and 2.8E-2 respectively. The convergence rates for the solution and its gradient errors are graphically depicted in Fig. 14 as log-log plots of the discrete L_2 norm errors versus the characteristic mesh size h. The optimal convergence rates are observed for high anisotropy ratio in this test.



(a) solution (b) gradient

Fig. 14. L_2 errors versus mesh size h for schemes on the uniform square mesh.

5. Conclusion

What we have seen from the above is the approximately second-order accuracy of two interpolation-based nonlinear finite volume schemes (LPS-TP2 and LPS-TP3) for

anisotropic diffusion equations on severely distorted grids. The two schemes are monotone, use two types of auxiliary unknowns, have compact stencil and are applicable to general unstructured meshes and full anisotropic heterogeneous diffusion tensors. The main feature of the scheme LPS-TP2 is that we do not need to replace the interpolation weights for auxiliary cell-vertex unknowns when the negative weights occur, and the main feature of LPS-TP3 is that a positivity-preserving and accurate interpolation method is utilized for the auxiliary cell-edge unknowns thanks to the harmonic averaging point.

Many dedicate experiments demonstrate that our new schemes are not only monotone, but also have asymptotically quadratic convergence rate for the approximate solution and first-order accuracy for its gradient on meshes with severely distortion and for problems with highly anisotropic diffusion tensors. Compared with LPS-TP1, LPS-TP2 and LPS-TP3 have better flexibility on mesh distortion, for example, desirable results can still be expected for randomly distorted quadrilateral meshes with distortion degree up to $\alpha = 0.9$.

References

[1] I. Aavatsmark, T. Barkve, Ø. Bøe, 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.

[2] I. Aavatsmark, G. Eigestad, B. Mallison, J. Nordbotten, A compact multipoint flux approximation method with improved robustness, Numer. Meth. Part. D. E. 24 (5) (2008) 1329–1360.

[3] L. Agelas, R. Eymard, R. Herbin, A nine-point finite volume scheme for the simulation of diffusion in heterogeneous media, C. R. Acad. Sci. Paris, Ser. I 347 (2009) 673–676.

[4] C. Aricò, T. Tucciarelli, Monotonic solution of heterogeneous anisotropic diffusion problems, J. Comput. Phys. 252 (2013) 219–249.

[5] C. Cancès, M. Cathala, C. L. Potier, Monotone corrections for generic cell-centered finite volume approximations of anisotropic diffusion equations, Numer. Math. 125 (2013), 387–417.

[6] R. Eymard, T. Gallouet, R. Herbin, Benchmark for anisotropic problems. SUSHI: A scheme using stabilization and hybrid interfaces for anisotropic heterogeneous diffusion problems, in: R. Eymard, J.-M. Herard (eds.), Finite Volumes for Complex Applications V-Problems and Perspectives, Wiley, 2008.

[7] R. Eymard, R. Herbin, C. Guichard, Small-stencil 3D schemes for diffusive flows in porous media, M2AN Math. Model. Numer. Anal. 46 (2012) 265–290.

[8] Z. Gao, J. Wu, A linearity-preserving cell-centered scheme for the heterogeneous and anisotropic diffusion equations on general meshes, Int. J. for Numer. Meth. Fluids 67 (2011) 2157–2183.

[9] Z. Gao, 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] R. Herbin, F. Hubert, Benchmark on discretization schemes for anisotropic diffusion problems on general grids, in: R. Eymard, J.-M. Herard (eds.), Finite Volumes for Complex Applications V-Problems and Perspectives, Wiley press, 2008.

[11] I. Kapyrin, A family of monotone methods for the numerical solution of three-dimensional diffusion problems on unstructured tetrahedral meshes, Dokl. Math. 76 (2007) 734–738.

[12] K. Lipnikov, G. Manzini, D. Svyatskiy, Analysis of the monotonicity conditions in the mimetic finite difference method for elliptic problems, J. Comput. Phys. 230 (2011) 2620–2642.

[13] K. Lipnikov, M. Shashkov, D. Svyatskiy, 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.

[14] K. Lipnikov, D. Svyatskiy, Y. Vassilevski, Interpolation-free monotone finite volume method for diffusion equations on polygonal meshes, J. Comput. Phys. 228 (2009) 703–716.

[15] 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.

[16] G. Manzini, M. Putti, Mesh locking effects in the finite volume solution of 2-D anisotropic diffusion equations, J. Comput. Phys. 220 (2007) 751–771.

[17] C. Le Potier, schema volumes finis monotone pour des operateurs de diffusion fortement anisotropes sur des maillages de triangle non structures, C. R. Math. Acad. Sci. Paris 341 (2005) 787–792.

[18] C. Le Potier, A linear scheme satisfying a maximum principle for anisotropic diffusion operators on distorted grids, C. R. Acad. Sci. Paris, Ser. I 347 (2009) 105–110.

[19] L. E. S. Queiroz, M. R. A. Souza, F. R. L. Contreras, P. R. M. Lyra and D. K. E. de Carvalho, On the accuracy of a nonlinear finite volume method for the solution of diffusion problems using different interpolations strategies, Int. J. Numer. Meth. Fluids, published online, 2013.

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

[21] A. I. Shestakov, J. A. Harte, D. S. Kershaw, solution of the diffusion equation by the finite elements in Lagrangian hydrodynamic codes, J. Comput. Phys. 76 (1988) 385–413.

[22] J. Wu, Z. Dai, Z. Gao, G. Yuan, Linearity preserving nine-point schemes for diffusion equation on distorted quadrilateral meshes, J. Comput. Phys. 229 (9) (2010) 3382–3401.

[23] J. Wu, Z. Gao, Z. Dai, A stabilized linearity-preserving scheme for the heterogeneous and anisotropic diffusion problems on polygonal meshes, J. Comput. Phys. 231 (2012) 7152–7169.

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