Russ. J. Numer. Anal. Math. Modelling, Vol. 24, No. 3, pp. 207–227 (2009)
DOI 10.1515/ RJNAMM.2009.014

A monotone nonlinear finite volume method for diffusion equations on conformal polyhedral meshes

A. A. DANILOV* and Yu. V. VASSILEVSKI*

Abstract — We have developed a new monotone cell-centered finite volume method for the discretization of diffusion equations on conformal polyhedral meshes. The proposed method is based on a nonlinear two-point flux approximation. For problems with smooth diffusion tensors and Dirichlet boundary conditions the method is interpolation-free. An adaptive interpolation is applied on faces where diffusion tensor jumps or Neumann boundary conditions are imposed. The interpolation is based on physical relationships such as continuity of the diffusion flux. The second-order convergence rate is verified with numerical experiments.

1. Introduction

The phenomenon of anisotropic diffusion plays a critical role in many physical models describing subsurface flows, heat conduction in structured materials and crystals, biological systems, and plasma physics. Accurate modelling of diffusive processes in these applications requires reliable discretization methods. Engineering 3D simulations use different types of meshes, such as tetrahedral, hexahedral, prismatic, octree, etc. All of them fall in the class of conformal meshes with polyhedral cells. The demand from the computational community for a simple and accurate conservative method applicable to general conformal meshes and anisotropic diffusion coefficients, is very distinct. In this article we present a new cell-centered finite volume method that preserves the solution positivity.

The conservative linear methods on unstructured meshes are well known: the multipoint flux approximation (MPFA), the mixed finite element (MFE) and the mimetic finite difference (MFD) methods. They are second-order accurate and are not monotone even when the diffusion coefficient is moderately (1:100) anisotropic. The cell-centered finite volume (FV) method with a linear two-point flux approximation is monotone, but not even first-order accurate for anisotropic problems or unstructured meshes. Nevertheless, this method is conventional in modelling flows in porous media due to its technological simplicity and monotonicity. The restrictions on monotonicity of the MPFA methods are analyzed in [1, 4, 13, 14]. The

*Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow 119333, Russia

This work has been supported in part by the RFBR grants 08-01-00159, 09-01-00115 and a grant from the Upstream Research Center, ExxonMobil corp.

conditions for the discrete maximum principle (DMP) to be satisfied by piecewise-linear finite element approximations generate restrictive mesh constraints [3, 8, 15]. Another class of monotone discretization methods for general meshes is formed by nonlinear methods. They guarantee solution positivity for the Poisson equation [2] and even for general diffusion equation [6, 9–12, 18, 19]

Following [6, 9, 11, 12, 18, 19] we propose a new monotone FV method based on a nonlinear two-point flux approximation scheme. The original idea belongs to C. LePotier [9] who proposed a monotone FV scheme for the discretization of parabolic equations on triangular meshes. Two years later the method and analysis of its monotonicity were extended to steady-state diffusion problems with full anisotropic tensors on triangulations or scalar diffusion coefficients on shape-regular polygonal meshes [11]. At the same time the method was extended to conformal tetrahedral meshes in [6, 18] where monotonicity was proved for the case of full anisotropic tensors. Recently the derivation of the nonlinear two-point flux stencil was modified on the basis of co-normal vector decomposition [19], which was originally suggested for a linear FV method in [7]. This approach extended the scheme to a much bigger class of polygonal meshes consisting of star-shaped cells and full tensor diffusion coefficients.

All these monotone nonlinear cell-centered FV methods use solution values at mesh cells (primary unknowns) and mesh nodes (auxiliary unknowns) for calculating discrete flux coefficients. Auxiliary unknowns are interpolated from primary cell-based unknowns. The choice of the interpolation method affects the accuracy of the nonlinear FV method [11, 19]. The particular interpolation method may be efficient for one problem and be inaccurate for another. Recently a new interpolation-free monotone cell-centered FV method with nonlinear two-point flux approximation was proposed for full diffusion tensors and unstructured conformal polygonal 2D meshes [12]. However, the method cannot be applied for *heterogeneous* diffusion tensor coefficients on an arbitrary mesh, since it may require partitions of certain cells.

In this paper, we extend the approach [12] to the case of 3D conformal polyhedral meshes and heterogeneous diffusion tensors. It is exact for linear and piecewise-linear solutions and thus has the second-order truncation error. However, the method may involve certain interpolation operations for a few auxiliary unknowns and thus is not interpolation-free formally. The important feature of the method is that most of auxiliary unknowns are interpolated from primary unknowns on the basis of a *physical* relationship. The latter expresses the continuity of the diffusion flux at the cell faces. The stencil of the interpolation operator at a cell face is two-point, and the coefficients depend on primary and auxiliary unknowns.

The main advantage of the proposed approach compared to other nonlinear FV methods using interpolation is that it is interpolation-free for diffusion tensors with smooth components and Dirichlet boundary conditions. In other cases, we apply physical interpolation on mesh faces where diffusion tensor jumps or Neumann boundary conditions are imposed. In rare pathological cases, the method requires auxiliary unknowns at the face edges. Interpolation at edges is arithmetic averaging the unknowns adjacent to the face. Such simple interpolation uses physically motivated data at cell faces and still provides an easy implementation. Although we provide monotonicity analysis only, our numerical experiments show the second-order convergence rate in the mesh-dependent L_2-norm.

The two-point flux approximation methods are technologically appealing due to the compact stencil even on polyhedral meshes. For cubic meshes and a diagonal diffusion tensor this stencil reduces to the conventional 7-point stencil. The major computational overhead in nonlinear FV methods is related to two nested iterations in the solution of a nonlinear algebraic problem. The outer iteration is the Picard method, which guarantees solution positivity in each iteration. The inner iteration is the Krylov subspace method for solving linearized problems.

The paper outline is as follows. In Section 2, we state the diffusion problem. In Section 3, we describe the nonlinear finite volume scheme. In Section 4, we prove the monotonicity of the proposed scheme. In Section 5, we present numerical analysis of the scheme using tetrahedral, hexahedral, and triangular prismatic meshes.

2. Diffusion equation

Let Ω be a three-dimensional polyhedral domain with a boundary $\Gamma = \Gamma_N \cup \Gamma_D$, where $\Gamma_D = \overline{\Gamma}_D$ and $\Gamma_D \neq \varnothing$. We consider a model diffusion problem for an unknown concentration c:

$$\begin{aligned} -\mathrm{div}(\mathbb{K}\nabla c) &= g && \text{in } \Omega \\ c &= g_D && \text{on } \Gamma_D \\ -\mathbb{K}\frac{\partial c}{\partial \mathbf{n}} &= g_N && \text{on } \Gamma_N \end{aligned} \tag{2.1}$$

where $\mathbb{K}(\mathbf{x}) = \mathbb{K}^T(\mathbf{x}) > 0$ is an anisotropic diffusion tensor, g is a source term, and $\mathbf{n}$ is the exterior normal vector.

Let $\mathscr{T}$ be a conformal polyhedral mesh composed of shape-regular cells with planar faces. We assume that each cell is a star-shaped 3D domain with respect to its barycenter, and each face is a star-shaped 2D domain with respect to the face's barycenter. Let $N_{\mathscr{T}}$ be the number of polyhedral cells and $N_{\mathscr{B}}$ be the number of boundary faces. We assume that $\mathscr{T}$ is face-connected, i.e. it cannot be split into two meshes having no common faces. We also assume that the tensor function $\mathbb{K}(\mathbf{x})$ varies slightly inside each cell; however it may jump across mesh faces, as well as may change the orientation of the principal directions.

We denote disjoint sets of interior and boundary faces by $\mathscr{F}_I$ and $\mathscr{F}_B$, respectively. The subset $\mathscr{F}_J$ of $\mathscr{F}_I$ collects faces with a jumping diffusion tensor. The set $\mathscr{F}_B$ is further split into subsets $\mathscr{F}_B^D$ and $\mathscr{F}_B^N$, where the Dirichlet and Neumann boundary conditions, respectively, are imposed. The cardinality of set $\mathscr{F}_*$ is denoted by $N_{\mathscr{F}_*}$. Finally, $\mathscr{F}_T$ and $\mathscr{E}_T$ denote the sets of faces and edges of the polyhedron T, respectively, whereas $\mathscr{E}_f$ denotes the set of edges of the face f.

3. Monotone nonlinear FV discretization

Let $\mathbf{q} = -\mathbb{K}\nabla c$ denote the flux which satisfies the mass balance equation:

$$\operatorname{div} \mathbf{q} = g \quad \text{in } \Omega. \tag{3.1}$$

We derive a FV scheme with a nonlinear two-point flux approximation. Integrating equation (3.1) over a polyhedron T and using the Green's formula we get:

$$\int_{\partial T} \mathbf{q} \cdot \mathbf{n}_T \, \mathrm{d}s = \int_T g \, \mathrm{d}x \tag{3.2}$$

where $\mathbf{n}_T$ denotes the outer unit normal to ∂T. Let f denote a face of the cell T and $\mathbf{n}_f$ be the corresponding normal vector. For a single cell T we always assume that $\mathbf{n}_f$ is the outward normal vector. In all other cases we specify the orientation of $\mathbf{n}_f$. It will be convenient to assume that $|\mathbf{n}_f| = |f|$, where $|f|$ denotes the area of the face f. The equation (3.2) becomes

$$\sum_{f \in \partial T} \mathbf{q}_f \cdot \mathbf{n}_f = \int_T g \, \mathrm{d}x \tag{3.3}$$

where $\mathbf{q}_f$ is the average flux density for face f.

For each cell T, we assign one degree of freedom, C_T, for the concentration c. Let C be the vector of all discrete concentrations. If two cells T_+ and T_- have a common face f, the two-point flux approximation is as follows:

$$\mathbf{q}_f^h \cdot \mathbf{n}_f = M_f^+ C_{T_+} - M_f^- C_{T_-} \tag{3.4}$$

where M_f^+ and M_f^- are some coefficients. In a linear FV method, these coefficients are equal and fixed. In the nonlinear FV method, they may be different and depend on concentrations in the surrounding cells. On the face $f \in \Gamma_D$, the flux has a form similar to (3.4) with an explicit value for one of the concentrations. For the Dirichlet boundary value problem, $\Gamma_D = \partial\Omega$, upon substituting (3.4) into (3.3), we obtain a system of $N_{\mathscr{T}}$ equations with $N_{\mathscr{T}}$ unknowns C_T. The Dirichlet and Neumann boundary conditions are considered in Subsection 3.3.

3.1. Notations

For every cell T in $\mathscr{T}$, we define the collocation point $\mathbf{x}_T$ at the barycenter of T. For every face $f \in \mathscr{F}_B \cup \mathscr{F}_I$, we denote the face barycenter by $\mathbf{x}_f$ and associate a collocation point with $\mathbf{x}_f$ for $f \in \mathscr{F}_B \cup \mathscr{F}_J$. We also define collocation points at the centers $\mathbf{x}_e$ of edges $e \in \mathscr{E}_f$, $f \in \mathscr{F}_B \cup \mathscr{F}_J$.

We shall refer to the collocation points on faces and edges as the *auxiliary* collocation points. They are introduced for mathematical convenience and will not contribute to the vector of unknowns in the final algebraic system, although will affect

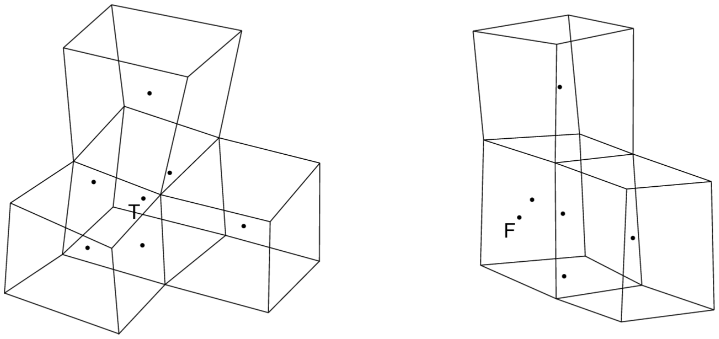

Figure 1. Examples of sets Σ_T (left) and $\Sigma_{f,T}$ (right).

the system coefficients. In contrast, we shall refer to the other collocation points as *primary* collocation points whose discrete concentrations form the unknown vector in the algebraic system.

For every cell T we define a set Σ_T of nearby collocation points as follows. First, we add to Σ_T the collocation point $\mathbf{x}_T$. Then, for every face $f \in \mathscr{F}_T \setminus (\mathscr{F}_J \cup \mathscr{F}_B)$, we add the collocation point $\mathbf{x}_{T'_f}$, where T'_f is a cell other than T, that has the face f. Finally, for any other face $f \in \mathscr{F}_T \cap (\mathscr{F}_B \cup \mathscr{F}_J)$, we add the collocation point $\mathbf{x}_f$ (see Fig. 1, left). Let $N(\Sigma_T)$ denote the cardinality of Σ_T.

Similarly, for every face $f \in \mathscr{F}_B \cup \mathscr{F}_J$ belonging to the cell T we define a set $\Sigma_{f,T}$ of nearby collocation points. We initialize $\Sigma_{f,T} = \{\mathbf{x}_f, \mathbf{x}_T\}$ and add to $\Sigma_{f,T}$ the points from Σ_T which are the barycenters of cells or faces adjacent to f (see Fig. 1, right). The cardinality of $\Sigma_{f,T}$ is denoted by $N(\Sigma_{f,T})$.

We assume that for every cell-face pair $T \to f$, $T \in \mathscr{T}$, $f \in \mathscr{F}_T$, there exist three points $\mathbf{x}_{f,1}$, $\mathbf{x}_{f,2}$, and $\mathbf{x}_{f,3}$ in the set Σ_T such that the following condition holds (see Fig. 2): The co-normal vector $\boldsymbol{\ell}_f = \mathbb{K}(\mathbf{x}_f)\mathbf{n}_f$ started from $\mathbf{x}_T$ belongs to the trihedral corner formed by the vectors

$$\mathbf{t}_{f,1} = \mathbf{x}_{f,1} - \mathbf{x}_T, \qquad \mathbf{t}_{f,2} = \mathbf{x}_{f,2} - \mathbf{x}_T, \qquad \mathbf{t}_{f,3} = \mathbf{x}_{f,3} - \mathbf{x}_T \tag{3.5}$$

and

$$\frac{1}{|\boldsymbol{\ell}_f|}\boldsymbol{\ell}_f = \frac{\alpha_f}{|\mathbf{t}_{f,1}|}\mathbf{t}_{f,1} + \frac{\beta_f}{|\mathbf{t}_{f,2}|}\mathbf{t}_{f,2} + \frac{\gamma_f}{|\mathbf{t}_{f,3}|}\mathbf{t}_{f,3} \tag{3.6}$$

where $\alpha_f \geqslant 0$, $\beta_f \geqslant 0$, $\gamma_f \geqslant 0$.

The coefficients α_f, β_f, γ_f are computed as follows:

$$\alpha_f = \frac{D_{f,1}}{D_f}, \qquad \beta_f = \frac{D_{f,2}}{D_f}, \qquad \gamma_f = \frac{D_{f,3}}{D_f} \tag{3.7}$$

where

$$D_f = \frac{\begin{vmatrix}\mathbf{t}_{f,1} & \mathbf{t}_{f,2} & \mathbf{t}_{f,3}\end{vmatrix}}{|\mathbf{t}_{f,1}||\mathbf{t}_{f,2}||\mathbf{t}_{f,3}|}, \qquad D_{f,1} = \frac{\begin{vmatrix}\boldsymbol{\ell}_f & \mathbf{t}_{f,2} & \mathbf{t}_{f,3}\end{vmatrix}}{|\boldsymbol{\ell}_f||\mathbf{t}_{f,2}||\mathbf{t}_{f,3}|}$$

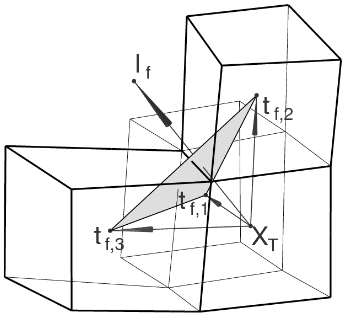

Figure 2. Co-normal vector and vector triplet.

$$D_{f,2} = \frac{|\mathbf{t}_{f,1} \quad \boldsymbol{\ell}_f \quad \mathbf{t}_{f,3}|}{|\mathbf{t}_{f,1}||\boldsymbol{\ell}_f||\mathbf{t}_{f,3}|}, \qquad D_{f,3} = \frac{|\mathbf{t}_{f,1} \quad \mathbf{t}_{f,2} \quad \boldsymbol{\ell}_f|}{|\mathbf{t}_{f,1}||\mathbf{t}_{f,2}||\boldsymbol{\ell}_f|}$$

and $|\mathbf{a}\ \mathbf{b}\ \mathbf{c}| = |(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}|$.

Similarly, we assume that for every face-cell pair $f \to T$, $f \in \mathscr{F}_B \cup \mathscr{F}_J$, $T : f \in \mathscr{F}_T$ there exist three points $\mathbf{x}_{f,1}$, $\mathbf{x}_{f,2}$, and $\mathbf{x}_{f,3}$ in set $\Sigma_{f,T}$ such that the vector $\boldsymbol{\ell}_{f,T} = -\mathbb{K}_T(\mathbf{x}_f)\mathbf{n}_f$ started from $\mathbf{x}_f$ belongs to the trihedral corner formed by the vectors

$$\mathbf{t}_{f,1} = \mathbf{x}_{f,1} - \mathbf{x}_f, \quad \mathbf{t}_{f,2} = \mathbf{x}_{f,2} - \mathbf{x}_f, \quad \mathbf{t}_{f,3} = \mathbf{x}_{f,3} - \mathbf{x}_f \tag{3.8}$$

and (3.6), (3.7) hold true.

We suggest a simple and efficient algorithm for searching a triplet satisfying (3.6) with non-negative coefficients. We present here the algorithm for a cell-face pair $T \to f$. The algorithm for a face-cell pair $f \to T$ is obtained from Algorithm 3.1 by substitution of $\mathbf{x}_T$ and Σ_T for $\mathbf{x}_f$ and $\Sigma_{f,T}$, respectively.

For a general conformal polyhedral mesh this algorithm may fail to find an appropriate triplet. In this case, we extend the sets of nearby collocation points and repeat Algorithm 3.1. The reuse of Algorithm 3.1 guarantees detecting a triplet. For the case of the cell-face pair $T \to f$ we proceed as follows. First, for every $f \in \mathscr{F}_T \cap (\mathscr{F}_J \cup \mathscr{F}_B)$ we add to Σ_T the collocation points $\mathbf{x}_e$, $e \in \mathscr{E}_f$. Second, for $f \in \mathscr{F}_T \setminus (\mathscr{F}_J \cup \mathscr{F}_B)$, we add to Σ_T the collocation points $\mathbf{x}_{T'_e}$, where T'_e is the cell, not belonging to Σ_T, that has an edge $e \in \mathscr{E}_f$ and may be connected to T with a polylinear path $\{\mathbf{x}_{T'_e}, \dots, \mathbf{x}_T\}$ through the face barycenters. The path should belong to the cells sharing the edge e and should not intersect a face from $\mathscr{F}_J$. For the case of the face-cell pair $f \to T$ we add to $\Sigma_{f,T}$ the collocation points $\mathbf{x}_e$, for $e \in \mathscr{E}_f$.

3.2. Nonlinear two-point flux approximation for an interior face

Let f be an interior face. We denote by T_+ and T_- the cells that share f and assume that $\mathbf{n}_f$ is outward for T_+. Let $\mathbf{x}_\pm$ (or $\mathbf{x}_{T_\pm}$) be the collocation point of $T_\pm$. Let $C_\pm$ (or $C_{T_\pm}$) be the discrete concentrations in $T_\pm$.

We begin with the case $f \notin \mathscr{F}_J$ and introduce $\mathbb{K}_f = \mathbb{K}(\mathbf{x}_f)$. Let $T = T_+$. Using

Algorithm 3.1. Algorithm for searching vector triplet for a pair $T \to f$.

```
 1: Define on the unit sphere points $\bar{\mathbf{x}}_{\boldsymbol{\ell}_f} = \mathbf{x}_T + |\boldsymbol{\ell}_f|^{-1}\boldsymbol{\ell}_f$, $\bar{\mathbf{x}}_i = \mathbf{x}_T + \bar{\mathbf{t}}_i$, $\bar{\mathbf{t}}_i = \mathbf{t}_i/|\mathbf{t}_i|$,
    $i = 1,\dots,N(\Sigma_T)$.
 2: Reorder $\bar{\mathbf{x}}_i$ according to increase of the distance $|\bar{\mathbf{x}}_i\bar{\mathbf{x}}_{\boldsymbol{\ell}_f}|$.
 3: for $i = 1, N(\Sigma_T) - 2$ do
 4:     for $j = i+1, N(\Sigma_T) - 1$ do
 5:         for $k = j+1, N(\Sigma_T)$ do
 6:             Calculate coefficients (3.7) for vectors $\bar{\mathbf{t}}_i$, $\bar{\mathbf{t}}_j$, $\bar{\mathbf{t}}_k$.
 7:             if all coefficients are non-negative then
 8:                 if all coefficients are less or equal to 1 then go to 16
 9:                 else put triplet $\{i,j,k\}$ to the set $\Sigma_T^*$
10:                 end if
11:             end if
12:         end for
13:     end for
14: end for
15: Pick from $\Sigma_T^*$ triplet $\{i,j,k\}$ with minimal value of $\max\{\alpha_f,\beta_f,\gamma_f\}$.
16: Set $\mathbf{t}_{f,1} = |\mathbf{x}_i - \mathbf{x}_T|\bar{\mathbf{t}}_i$, $\mathbf{t}_{f,2} = |\mathbf{x}_j - \mathbf{x}_T|\bar{\mathbf{t}}_j$, $\mathbf{t}_{f,3} = |\mathbf{x}_k - \mathbf{x}_T|\bar{\mathbf{t}}_k$.
```

the above notations, the definition of the directional derivative,

$$\frac{\partial c}{\partial \boldsymbol{\ell}_f}|\boldsymbol{\ell}_f| = \nabla c \cdot (\mathbb{K}_f\,\mathbf{n}_f)$$

and assumption (3.6), we write

$$\mathbf{q}_f \cdot \mathbf{n}_f = -\frac{|\boldsymbol{\ell}_f|}{|f|}\int_f \frac{\partial c}{\partial \boldsymbol{\ell}_f}\,\mathrm{d}s = -\frac{|\boldsymbol{\ell}_f|}{|f|}\int_f \left(\alpha_f \frac{\partial c}{\partial \mathbf{t}_{f,1}} + \beta_f \frac{\partial c}{\partial \mathbf{t}_{f,2}} + \gamma_f \frac{\partial c}{\partial \mathbf{t}_{f,3}}\right)\mathrm{d}s. \tag{3.9}$$

Replacing the directional derivatives by finite differences, we get

$$\int_f \frac{\partial c}{\partial \mathbf{t}_{f,i}}\,\mathrm{d}s = \frac{C_{f,i} - C_T}{|\mathbf{x}_{f,i} - \mathbf{x}_T|}\,|f| + O(h_T^3), \quad i = 1,2,3 \tag{3.10}$$

where h_T is the diameter of the cell T. Using the finite difference approximations (3.10), we transform formula (3.9) to

$$\mathbf{q}_f^h \cdot \mathbf{n}_f = -|\boldsymbol{\ell}_f|\left(\frac{\alpha_f}{|\mathbf{t}_{f,1}|}(C_{f,1} - C_T) + \frac{\beta_f}{|\mathbf{t}_{f,2}|}(C_{f,2} - C_T) + \frac{\gamma_f}{|\mathbf{t}_{f,3}|}(C_{f,3} - C_T)\right). \tag{3.11}$$

At the moment, this flux involves four rather than two concentrations. To derive a two-point flux approximation, we consider the cell T_- and derive another approximation of the flux through the face f. To distinguish between T_+ and T_-, we add subscripts ± and omit the subscript f. Since $\mathbf{n}_f$ is the internal normal vector for T_-, we have to change the sign of the right-hand side:

$$\mathbf{q}^h_\pm \cdot \mathbf{n}_f = \mp|\boldsymbol{\ell}_f| \left(\frac{\alpha_\pm}{|\mathbf{t}_{\pm,1}|}(C_{\pm,1} - C_\pm) + \frac{\beta_\pm}{|\mathbf{t}_{\pm,2}|}(C_{\pm,2} - C_\pm) + \frac{\gamma_\pm}{|\mathbf{t}_{\pm,3}|}(C_{\pm,3} - C_\pm) \right) \tag{3.12}$$

where $\alpha_\pm$, $\beta_\pm$, and $\gamma_\pm$ are given by (3.7) and $C_{\pm,i}$ denote concentrations at points $\mathbf{x}_{\pm,i}$ from $\Sigma_{T_\pm}$.

We define a new discrete flux as a linear combination of $\mathbf{q}^h_\pm \cdot \mathbf{n}_f$ with non-negative weights $\mu_\pm$:

$$\begin{aligned} \mathbf{q}^h_f \cdot \mathbf{n}_f &= \mu_+ \mathbf{q}^h_+ \cdot \mathbf{n}_f + \mu_- \mathbf{q}^h_- \cdot \mathbf{n}_f \\ &= \mu_+ |\boldsymbol{\ell}_f| \left(\frac{\alpha_+}{|\mathbf{t}_{+,1}|} + \frac{\beta_+}{|\mathbf{t}_{+,2}|} + \frac{\gamma_+}{|\mathbf{t}_{+,3}|} \right) C_+ \\ &\quad - \mu_- |\boldsymbol{\ell}_f| \left(\frac{\alpha_-}{|\mathbf{t}_{-,1}|} + \frac{\beta_-}{|\mathbf{t}_{-,2}|} + \frac{\gamma_-}{|\mathbf{t}_{-,3}|} \right) C_- \\ &\quad - \mu_+ |\boldsymbol{\ell}_f| \left(\frac{\alpha_+}{|\mathbf{t}_{+,1}|} C_{+,1} + \frac{\beta_+}{|\mathbf{t}_{+,2}|} C_{+,2} + \frac{\gamma_+}{|\mathbf{t}_{+,3}|} C_{+,3} \right) \\ &\quad + \mu_- |\boldsymbol{\ell}_f| \left(\frac{\alpha_-}{|\mathbf{t}_{-,1}|} C_{-,1} + \frac{\beta_-}{|\mathbf{t}_{-,2}|} C_{-,2} + \frac{\gamma_-}{|\mathbf{t}_{-,3}|} C_{-,3} \right). \end{aligned} \tag{3.13}$$

The obvious requirement for the weights is to cancel the terms in the last two rows of (3.13), which results in a two-point flux formula. The second requirement is to approximate the true flux. These requirements lead us to the following system

$$\begin{cases} -\mu_+ d_+ + \mu_- d_- = 0 \\ \mu_+ + \mu_- = 1 \end{cases} \tag{3.14}$$

where

$$d_\pm = |\boldsymbol{\ell}_f| \left(\frac{\alpha_\pm}{|\mathbf{t}_{\pm,1}|} C_{\pm,1} + \frac{\beta_\pm}{|\mathbf{t}_{\pm,2}|} C_{\pm,2} + \frac{\gamma_\pm}{|\mathbf{t}_{\pm,3}|} C_{\pm,3} \right).$$

Since the coefficients $d_\pm$ depend both on geometry and concentration, so do the weights $\mu_\pm$ as well. Thus, the resulting two-point flux approximation is *nonlinear*.

It may happen that the concentration $C_{+,i}$ ($C_{-,i}$), $i = 1,2,3$, is defined at the same collocation point as C_- (C_+). In this case the terms to be canceled are changed so that they do not incorporate $C_\pm$. By doing so, for the Laplace operator we recover the classical linear scheme with the $\{-1,-1,-1,6,-1,-1,-1\}$ stencil on uniform cubic meshes.

The solution of (3.14) can be written explicitly. In all cases $d_\pm \geqslant 0$ if $C \geqslant 0$. If $d_\pm = 0$, we set $\mu_+ = \mu_- = 1/2$. Otherwise,

$$\mu_+ = \frac{d_-}{d_- + d_+}, \qquad \mu_- = \frac{d_+}{d_- + d_+}.$$

This implies that the weights $\mu_{\pm}$ are non-negative. Substituting this into (3.13), we get the two-point flux formula (3.4) with coefficients

$$M_f^{\pm} = \mu_{\pm}|\boldsymbol{\ell}_f|(\alpha_{\pm}/|\mathbf{t}_{\pm,1}| + \beta_{\pm}/|\mathbf{t}_{\pm,2}| + \gamma_{\pm}/|\mathbf{t}_{\pm,3}|). \tag{3.15}$$

Now we consider the case $f \in \mathscr{F}_J$ when $\mathbb{K}_+(\mathbf{x}_f)$ and $\mathbb{K}_-(\mathbf{x}_f)$ differ. We derive two-point flux approximations in the cells T_+ and T_- independently:

$$(\mathbf{q}_f^h \cdot \mathbf{n}_f)_+ = N^+ C_+ - N_f^+ C_f \tag{3.16}$$

$$-(\mathbf{q}_f^h \cdot \mathbf{n}_f)_- = N^- C_- - N_f^- C_f. \tag{3.17}$$

Non-negative coefficients N^+, N_f^+, N^-, N_f^- are derived similarly to coefficients (3.15) on the basis of discrete concentrations at collocation points from $\Sigma_{T_\pm}$, $\Sigma_{f,T_\pm}$ and $\boldsymbol{\ell}_{\pm} = \mp\mathbb{K}_{\pm}(\mathbf{x}_f)\,\mathbf{n}_f$, the co-normal vectors to face f outward with respect to $T_{\pm}$. The continuity of the diffusive flux allows us to eliminate C_f from (3.16), (3.17)

$$C_f = (N^+ C_+ + N^- C_-)/(N_f^+ + N_f^-) \tag{3.18}$$

and derive the two-point flux approximation (3.4) with coefficients

$$M_f^{\pm} = N^{\pm} N_f^{\mp}/(N_f^+ + N_f^-). \tag{3.19}$$

If $N_f^{\pm} = 0$, we set $M_f^{\pm} = N^{\pm}/2$ and $C_f = (C_+ + C_-)/2$.

3.3. Flux approximation for a boundary face

First, we consider the case of the Dirichlet boundary face $f \in \mathscr{F}_B^D$ where we define

$$C_f = \bar{g}_{D,f} = \frac{1}{|f|}\int_f g_D \,\mathrm{d}s \tag{3.20}$$

and for every edge $e \in \mathscr{E}_f$

$$C_e = \bar{g}_{D,e} = \frac{1}{|e|}\int_e g_D \,\mathrm{d}x. \tag{3.21}$$

It may be convenient to think about f as the ghost cell with zero volume. Let T be the cell with the face f. Replacing C_+ and C_- with C_T and C_f, and Σ_{T_+}, Σ_{T_-} with Σ_T, $\Sigma_{f,T}$, respectively, we get

$$\mathbf{q}_f^h \cdot \mathbf{n}_f = M_f^+ C_T - M_f^- C_f \tag{3.22}$$

where coefficients $M_f^{\pm}$ are given by (3.15).

Now consider the case of a Neumann boundary face $f \in \mathscr{F}_B^N$. The flux through this face is

$$\mathbf{q}_f^h \cdot \mathbf{n}_f = \bar{g}_{N,f}|f| \tag{3.23}$$

where $\bar{g}_{N,f}$ is the mean value of g_N on face f.

3.4. Recovery of discrete solution at auxiliary collocation points

The coefficients $M_f^{\pm}$ in (3.15), (3.19), (3.22) may depend on discrete solutions C_f and C_e at auxiliary collocation points $\mathbf{x}_f$, $f \in \mathscr{F}_B \cup \mathscr{F}_J$, and $\mathbf{x}_e$, $e \in \mathscr{E}_f$. On the other hand, the discrete FV system is formulated only for concentrations C_T at the primary collocation points. The values C_f, C_e, $f \in \mathscr{F}_B^D$, $e \in \mathscr{E}_f$, are computed by (3.20), (3.21) from the Dirichlet data. The values C_f, $f \in \mathscr{F}_J$, are recalculated from (3.18). However, values C_f, C_e, $f \in \mathscr{F}_B^N$, $e \in \mathscr{E}_f$, $e \notin \Gamma_D$ and the values C_e, $e \in \mathscr{E}_f$, $e \notin \Gamma_D$, $f \in \mathscr{F}_J$ have to be recovered from available data.

We recover the concentrations at Neumann faces from C_T using (3.22) and (3.23). The coefficients $M_f^{\pm}$ can depend on the values at primary collocation points and C_f, $f \in \mathscr{F}_J \cup \mathscr{F}_B$, and C_e, $e \in \mathscr{E}_f$. Therefore, concentrations C_f at mesh faces f, $f \in F_J \cup \mathscr{F}_B$, are interpolated from the cell data on the basis of *physical* relationships, such as the diffusion flux continuity or a given diffusion flux. The coefficients of interpolation can depend on concentrations C_e to be found at $\mathbf{x}_e$, $e \in \mathscr{E}_f$, $f \in \mathscr{F}_J \cup \mathscr{F}_B^N$, $e \notin \Gamma_D$. For every such edge, we suggest to compute C_e by arithmetic averaging of C_f for all faces $f \in \mathscr{F}_B^N \cup \mathscr{F}_J$ sharing e.

4. Discrete system and monotonicity analysis

For every T in $\mathscr{T}$, the cell equation (3.3) is

$$\sum_{f \in \mathscr{F}_T} \chi(T,f)\, \mathbf{q}_f^h \cdot \mathbf{n}_f = \int_T f \, \mathrm{d}x \tag{4.1}$$

where $\chi(T,f) = \mathrm{sign}(\mathbf{n}_f \cdot \mathbf{n}_T(\mathbf{x}_f))$. Substituting two-point flux formula (3.4) with non-negative coefficients given by (3.15) and (3.19) into (4.1), and using equations (3.20), (3.21) and (3.22), (3.23) to eliminate concentrations at boundary faces, and using arithmetic averaging of recovered face concentrations at non-Dirichlet edges $e \in \mathscr{E}_f$, $f \in \mathscr{F}_B^N \cup \mathscr{F}_J$, we get a nonlinear system of $N_{\mathscr{T}}$ equations

$$\mathbb{M}(C)C = G(C). \tag{4.2}$$

The matrix $\mathbb{M}(C)$ may be represented by assembling 2×2 matrices

$$\mathbb{M}_f(C) = \begin{pmatrix} M_f^+(C) & -M_f^-(C) \\ -M_f^+(C) & M_f^-(C) \end{pmatrix} \tag{4.3}$$

for the interior faces and 1×1 matrices $\mathbb{M}_f(C) = M_f^+(C)$ for Dirichlet faces (see Algorithm 4.1 for more detail). The right-hand side vector $G(C)$ is generated by the source and the boundary data:

$$G_T(C) = \int_T g \, \mathrm{d}x + \sum_{f \in \mathscr{F}_B^D \cap \mathscr{F}_T} M_f^-(C)\bar{g}_{D,f} - \sum_{f \in \mathscr{F}_B^N \cap \mathscr{F}_T} |f| \bar{g}_{N,f} \qquad \forall T \in \mathscr{T}. \tag{4.4}$$

Algorithm 4.1. Generation and solution of nonlinear system (4.2).

```
1: For each cell-face pair $T \to f$, $f \in \mathscr{F}_T$, and each face-cell pair $f \to T$, $f \in F_J \cup \mathscr{F}_B$ find vectors $\mathbf{t}_{f,1}$, $\mathbf{t}_{f,2}$, $\mathbf{t}_{f,3}$, satisfying conditions (3.5), (3.6) and (3.8), (3.6), respectively.
2: Select initial vectors $C^0 \in \Re^{N_{\mathscr{T}}}$ and $C_f^0 \in \Re^{N_{\mathscr{F}_J}+N_{\mathscr{F}_B^N}}$ with non-negative entries and a small value $\varepsilon_{\text{non}} > 0$.
3: Calculate concentrations $C_e^0$ at the auxiliary collocation points on edges using (3.21) or arithmetic averaging of neighboring data $C_f^0$.
4: for $k = 0, \ldots,$ do
5:     Assemble the global matrix $\mathbb{M}_k = \mathbb{M}(C^k, C_f^k, C_e^k)$ from the face-based matrices $\mathbb{M}_f(C^k, C_f^k, C_e^k)$. To form $\mathbb{M}_f(C^k, C_f^k, C_e^k)$, use (3.15) for $f \in \mathscr{F}_B^D \cup \mathscr{F}_I \setminus \mathscr{F}_J$ and (3.19) for $f \in \mathscr{F}_J$.
6:     Calculate the right-hand side vector $G^k = G(C^k, C_f^k, C_e^k)$ using (4.4).
7:     Stop if $\|\mathbb{M}_k C^k - G^k\| \leqslant \varepsilon_{\text{non}} \|\mathbb{M}_0 C^0 - G^0\|$.
8:     Solve $\mathbb{M}_k C^{k+1} = G^k$.
9:     Calculate concentrations $C_f^{k+1}$ at the auxiliary collocation points on faces $f \in \mathscr{F}_J \cup \mathscr{F}_B$ using (3.18), (3.20), (3.22), (3.23), and data $C^{k+1}$, $C_f^k$, $C_e^k$.
10:    Calculate concentrations $C_e^{k+1}$ at the auxiliary collocation points on edges using (3.21) or arithmetic averaging of neighboring data $C_f^{k+1}$.
11: end for
```

For data functions $g \geqslant 0$, $g_D \geqslant 0$ and $g_N \leqslant 0$ the components of the vector G are non-negative. We use the Picard iterations to solve the nonlinear system (4.2) (see Algorithm 4.1).

The linear system in Step 8 with the non-symmetric matrix $\mathbb{M}_k = M(C^k, C_f^k, C_e^k)$ and the right-hand side $G^k = G(C^k, C_f^k, C_e^k)$ is solved by the Bi-Conjugate Gradient Stabilized (BiCGStab) method [16] with the second-order ILU preconditioner [5]. The BiCGStab iterations are terminated when the relative norm of the residual of the linear system becomes smaller than ε_{lin}.

Now we demonstrate that the matrix $\mathbb{M}_k$ is an M-matrix provided that $C^k \geqslant 0$. Our derivation shows that auxiliary unknowns $C_f^k \geqslant 0$, $C_e^k \geqslant 0$, and coefficients $M_f^{\pm}(C^k)$ are positive. Thus, all diagonal entries of the matrix $\mathbb{M}_k$ are positive and all off-diagonal entries of $\mathbb{M}_k$ are non-positive. The structure of face-based matrices (4.3) implies that each column sum in $\mathbb{M}_k$ is non-negative. Moreover, for the elements having Dirichlet faces, the corresponding column sum is strictly positive. For a connected mesh, the matrices $\mathbb{M}_k$ and $\mathbb{M}_k^T$ are irreducible, since their directed graphs are strongly connected. Under the above conditions, the well known linear algebra result [17] implies that matrix $\mathbb{M}_k^T$ is an M-matrix and all entries of $(\mathbb{M}_k^T)^{-1}$ are positive. Since the inverse and transpose operations commute,

$(\mathbb{M}_k^T)^{-1} = (\mathbb{M}_k^{-1})^T$, we conclude that $\mathbb{M}_k$ is monotone. Due to the signs of diagonal and off-diagonal entries $\mathbb{M}_k$ is an M-matrix as well. Therefore, we have proved the following theorem.

Theorem 4.1. *Let $g \geqslant 0$, $g_D \geqslant 0$, $g_N \leqslant 0$ and $\Gamma_D \neq \varnothing$ in* (2.1)*. If $C^0 \geqslant 0$ and linear systems in the Picard method are solved exactly, then $C^k \geqslant 0$ for $k \geqslant 1$.*

The considered FV method is exact for piecewise linear concentrations and has the second-order truncation error. Therefore, we may expect the second order of convergence for the scalar variable C and at least the first order of convergence for the flux degrees of freedom.

5. Numerical experiments

We use discrete L_2-norms to evaluate discretization errors for the concentration c and the flux $\mathbf{q}$:

$$\varepsilon_2^c = \left[\frac{\sum\limits_{T\in\mathscr{T}} (c(\mathbf{x}_T) - C_T)^2 |T|}{\sum\limits_{T\in\mathscr{T}} (c(\mathbf{x}_T))^2 |T|}\right]^{1/2}, \qquad \varepsilon_2^q = \left[\frac{\sum\limits_{f\in\mathscr{F}_I\cup\mathscr{F}_B} \Big((\mathbf{q}_f - \mathbf{q}_f^h)\cdot\mathbf{n}_f\Big)^2 |V_f|}{\sum\limits_{f\in\mathscr{F}_I\cup\mathscr{F}_B} (\mathbf{q}_f\cdot\mathbf{n}_f)^2 |V_f|}\right]^{1/2}$$

where $|V_f|$ is a representative volume for the face f. More precisely, $|V_f|$ is the arithmetic mean of the volumes of mesh cells sharing the face. The nonlinear iterations are terminated when the relative norm of the residual norm becomes smaller than $\varepsilon_{\text{non}} = 10^{-9}$. The convergence tolerance for the linear solver is set to $\varepsilon_{\text{lin}} = 10^{-12}$.

We consider three classes of polyhedral meshes for the unit cube $[0,1]^3$. All meshes are considered to be quasiuniform.

Hexahedral meshes are constructed from uniform cubic meshes by the distortion of internal nodes. In each plane $x = 0.5$, $y = 0.5$, and $z = 0.5$ the nodes are randomly shifted along the planes. The position of other nodes is determined by the requirement of planarity of the faces. The distance and direction in which the nodes are shifted from the original position are chosen randomly. The shifts of all nodes do not exceed $0.3h$, where h is the cubic mesh size.

Prismatic meshes are constructed as a tensor product of a quasiuniform unstructured triangular xy-mesh and 1D z-mesh, both meshes having the size h. Additionally, z-planes are slightly tilted in such a way, that they do not intersect each other and the distance between them is at least $0.75h$. The height of each cell in these meshes is between $0.75h$ and $1.25h$.

Tetrahedral meshes are quasiuniform unstructured tetrahedral meshes with a mesh size h. There is no hierarchical relation between the tetrahedral meshes.

Representative examples of all three mesh classes are shown in Fig. 3.

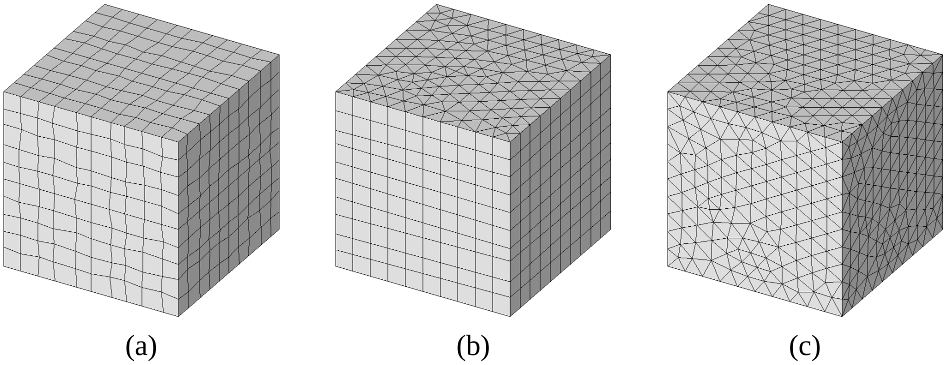

Figure 3. Examples of hexahedral (a), triangular prismatic (b), and tetrahedral (c) meshes.

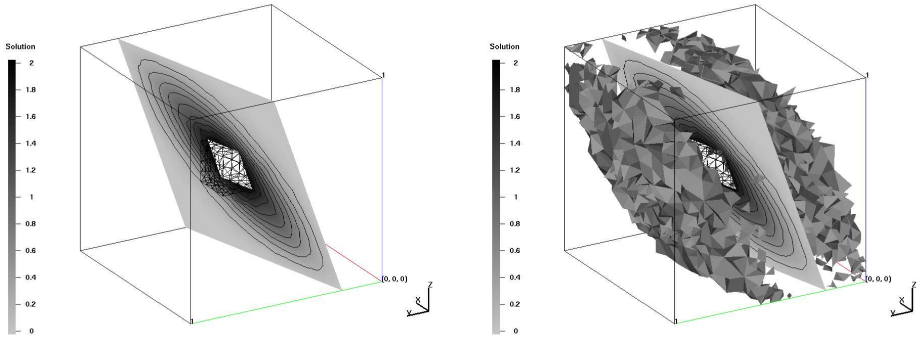

Figure 4. Cutplane of the solution calculated with the nonlinear FV method (left) and with the MFE method (right) for a problem with the Dirichlet boundary conditions. Elements with solution values less than -10^{-3} are also shown. Tetrahedral mesh with $h = 1/20$.

5.1. Monotonicity test

Numerical results of this section verify the assertion of Theorem 4.1. We consider two benchmark problems with a highly anisotropic diffusion tensor and demonstrate numerically that the discrete solution is always non-negative, although it can violate the discrete maximum principle (DMP).

5.1.1. Dirichlet boundary conditions. Here we consider problem (2.1) defined in a unit cube with a cubic hole, $\Omega = (0,1)^3/[0.4,0.6]^3$. The boundary of Ω consists of two disjoint parts, interior Γ_0 and outer Γ_1. We set $\Gamma_N = \varnothing$, $f = 0$, $g_D = 2$ on Γ_0, $g_D = 0$ on Γ_1, and take the anisotropic diffusion tensor $\mathbb{K}$,

$$\mathbb{K} = R_z(-\theta_z)R_y(-\theta_y)R_x(-\theta_x)\mathrm{diag}(k_1,k_2,k_3)R_x(\theta_x)R_y(\theta_y)R_z(\theta_z) \tag{5.1}$$

where $k_1 = 100$, $k_2 = 10$, $k_3 = 1$, $\theta_x = \pi/3$, $\theta_y = \pi/4$, $\theta_z = \pi/6$, and $R_a(\alpha)$ is the rotation matrix in the plane orthogonal to Oa with angle α. According to the maximum principle for elliptic PDEs, the exact solution should be between 0 and 2.

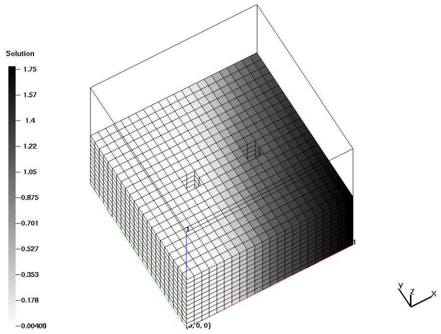

Figure 5. Solution calculated with the nonlinear FV method for the problem with no-flow boundary conditions, $h = 1/22$.

Table 1.
The maximum concentration values for the problem with no-flow boundary conditions.

h	$1/11$	$1/22$	$1/44$	$1/88$
$\max C$	2.163	1.765	1.136	1.024

Discrete solutions computed with the nonlinear FV method on all types of the considered meshes are non-negative everywhere in Ω (see Fig. 4, left). The computed solution does not violate the DMP on the considered meshes either. We remark that discretizations with the lowest-order Raviart–Thomas MFE on simplicial meshes computed using public libraries [20, 21] generate extensive areas of negative solutions (see Fig. 4, right). Similar observations [11, 12] have been made for MFE and MPFA discretizations of the 2D analogue of (2.1).

5.1.2. No-flow boundary conditions. Now we demonstrate that the FV discrete solution can violate the DMP even on cubic meshes.

We consider the 3D analogue of the problem described and investigated in [1]. We consider a unit cube with two vertical holes P_1, P_2, $\Omega = (0,1)^3 \setminus (P_1 \cup P_2)$, $P_i = S_i \times (0,1)$, $i = 1,2$, $S_1 = [3/11, 4/11] \times [5/11, 6/11]$, $S_2 = [7/11, 8/11] \times [5/11, 6/11]$. The domain boundary is split into the outer part Γ_N where the homogeneous Neumann (no-flow) boundary condition is set, and two inner parts $\Gamma_{D,1}$, $\Gamma_{D,2}$ where the Dirichlet boundary conditions are set: $g_D(\mathbf{x}) = 0$, $\mathbf{x} \in \Gamma_{D,1}$, $g_D(\mathbf{x}) = 1$, $\mathbf{x} \in \Gamma_{D,2}$. The anisotropic diffusion tensor is defined by (5.1) with $k_1 = k_3 = 1$, $k_2 = 10^{-3}$, $\theta_x = \theta_y = 0$, $\theta_z = 67.5°$. According to the maximum principle for elliptic PDEs the exact solution should be between 0 and 1 and have no extrema on the no-flow boundary Γ_N.

The FV discrete solution is shown in Fig. 5. It is non-negative in agreement with Theorem 4.1, but demonstrates overshoots near the no-flow boundary. These overshoots are decreased rapidly as we refine the mesh, see Table 1.

Table 2.
Number of Picard iterations for different types of meshes and tolerances ε_{non}, $10^{-3}/10^{-6}/10^{-9}$.

$h\setminus$Mesh	hexahedral	prismatic	tetrahedral
$1/10$	6/18/30	7/25/43	6/23/42
$1/20$	7/30/54	8/37/69	9/39/74
$1/40$	7/48/95	9/61/127	8/67/137

5.2. Picard method

The iterative solution of the nonlinear algebraic problem is the major computational overhead in the nonlinear FV method. The Picard method guarantees the solution positivity on each nonlinear iteration.

We consider the problem described in Subsection 5.1.1 and measure the number of Picard iterations required for 10^3-, 10^6-, and 10^9-fold reduction of the initial nonlinear residual due to the initial vector composed of ones. Each linear system is solved with 10^{12}-fold reduction of the initial residual. In Table 2 we present the numbers of Picard iterations N_{it} for different types of meshes and tolerances. We observe a fast convergence of the first iterations, a moderate growth of N_{it} as $h \to 0$ for $\varepsilon_{\text{non}} = 10^{-3}$, proportionality of N_{it} to h^{-1} for $\varepsilon_{\text{non}} \to 0$, and slight sensitivity of N_{it} to the mesh type.

5.3. Convergence study: smooth solution

In this section we study the convergence of the method for problem (2.1) with a smooth solution. Let $\Omega = (0,1)^3$, $\Gamma_D = \partial\Omega$, and f be obtained by substitution of the exact solution

$$c(x,y,z) = \frac{1}{3\pi^2}\sin(\pi x)\sin(\pi y)\sin(\pi z) \tag{5.2}$$

in (2.1). The Dirichlet data g_D are equal to the trace of $c(x,y,z)$ on Γ_D. We consider two cases of the anisotropic diffusion tensor

$$\mathbb{K} = k(x,y,z)\cdot \operatorname{diag}(1,10,100)$$

represented by constant and smooth functions k. The first and the second cases are defined by relations $k(x,y,z) = 1$ and $k(x,y,z) = 1 + 0.25\cos(x+y-z)$, respectively. In both cases the diffusion tensor is smooth and $\mathscr{F}_J = \varnothing$.

The convergence results are presented in Tables 3 and 4. The convergence rate for the scalar variable C demonstrates the second-order reduction as $h \to 0$, whereas for the normal component of the flux it decreases at least linearly as $h \to 0$.

Now we proceed to problem (2.1) with the identity diffusion tensor discretized on highly anisotropic meshes. Following the benchmark test [1], we consider the distortion of the unit cube by k-fold shrinking the cube in z-direction, $k = 10, 100$, and tilting its yz-faces on $\pi/6$. Thus we produce meshes with different aspect ratios, 0.1 and 0.01 (see Fig. 6). We consider two types of quasiuniform hexahedral

Table 3.
The convergence results for the problem with the smooth solution and a constant tensor.

	Hexahedral meshes		Prismatic meshes		Tetrahedral meshes	
h	ε_2^C	ε_2^q	ε_2^C	ε_2^q	ε_2^C	ε_2^q
$1/10$	6.79e-3	1.39e-2	8.31e-3	6.85e-3	2.42e-2	5.72e-2
$1/20$	1.69e-3	4.45e-3	2.05e-3	2.26e-3	5.45e-3	3.05e-2
$1/40$	3.77e-4	1.51e-3	5.34e-4	8.09e-4	1.44e-3	1.47e-2

Table 4.
The convergence results for the problem with the smooth solution and a smooth variable tensor.

	Hexahedral meshes		Prismatic meshes		Tetrahedral meshes	
h	ε_2^C	ε_2^q	ε_2^C	ε_2^q	ε_2^C	ε_2^q
$1/10$	6.91e-3	1.38e-2	8.43e-3	6.86e-3	2.43e-2	5.73e-2
$1/20$	1.72e-3	4.42e-3	2.08e-3	2.26e-3	5.46e-3	3.04e-2
$1/40$	3.84e-4	1.49e-3	5.42e-4	8.02e-4	1.45e-3	1.45e-2

meshes in the original unit cube, the undistorted cubic meshes, and the distorted cubic meshes presented in the beginning of the section.

We define the exact solution

$$c(x,y,z) = \cosh(\pi x)\cos(\pi z)$$

generating a zero source term and non-homogeneous Dirichlet boundary conditions. The convergence results are presented in Table 5. The asymptotic second-order convergence for the scalar variable is observed for aspect ratio 0.1. For the small aspect ratio, the asymptotic convergence rate demonstrates the dependence h^β, $1 \leqslant \beta < 2$, $h \to 0$. The convergence for the flux variable is higher than the first order for both aspect ratios.

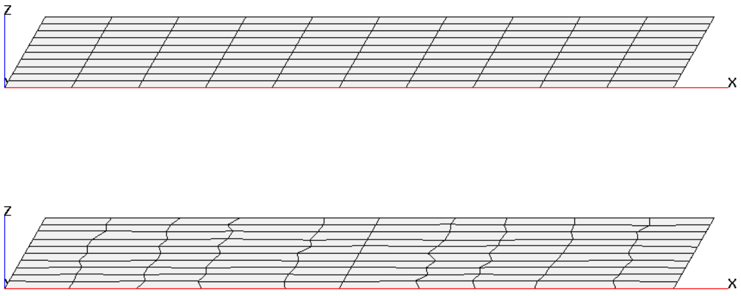

Figure 6. Anisotropic meshes with $h = 1/10$, aspect ratio 0.1 and tilt 30°. Above: Undistorted. Below: Distorted.

Table 5.
The convergence results for the problem defined on the anisotropic hexahedral meshes.

	Undistorted meshes				Distorted meshes			
	Aspect ratio 0.1		Aspect ratio 0.01		Aspect ratio 0.1		Aspect ratio 0.01	
h	ε_2^C	ε_2^q	ε_2^C	ε_2^q	ε_2^C	ε_2^q	ε_2^C	ε_2^q
1/10	1.02e-3	1.70e-1	1.55e-5	1.77e-0	1.15e-3	1.47e-1	4.76e-4	2.03e-0
1/20	4.35e-4	6.11e-2	1.17e-5	6.32e-1	5.00e-4	5.45e-2	2.04e-4	1.14e-0
1/40	1.35e-4	2.03e-2	8.10e-6	2.24e-1	1.48e-4	1.97e-2	6.89e-5	4.60e-1
1/80	3.60e-5	6.58e-3	4.96e-6	8.08e-2	3.82e-5	6.40e-3	2.41e-5	1.97e-1

5.4. Convergence study: solutions with sharp gradients

In this section we study the convergence of the nonlinear FV method for a problem with a known highly anisotropic solution. We consider problem (2.1) defined in a unit cube $\Omega = (0,1)^3$ which is divided into three subdomains $\Omega_1 = (0,1)\times(0,Y_1)\times(0,1)$, $\Omega_2 = (0,1)\times[Y_1,Y_2]\times(0,1)$, $\Omega_3 = (0,1)\times(Y_2,1)\times(0,1)$. We impose the homogeneous Dirichlet boundary condition on the non-horizontal parts Γ_D of $\partial\Omega$ and set the source function and the diffusion tensor as follows:

$$f(x,y,z) = \begin{cases} 0, & (x,y,z) \in \Omega_1 \cup \Omega_3 \\ f_0(y)\sin(\pi x), & (x,y,z) \in \Omega_2 \end{cases}, \qquad \mathbb{K} = \mathrm{diag}\{k,1,1\}.$$

In our experiments $f_0(y) = 50$, $k = 50$, $Y_1 = 0.3$, and $Y_2 = 0.7$. The exact solution to this problem can be calculated using the separation of variables. Taking

$$C(x,y,z) = \varphi(y)\sin(\pi x)$$

and substituting it into equation (2.1), we get the following equation for $\varphi(y)$:

$$-\frac{\partial^2\varphi}{\partial y^2} + \pi^2 k\varphi(y) = \hat{f}(y), \qquad \hat{f}(y) = \begin{cases} 0, & y \in [0,Y_1]\cup[Y_2,1] \\ f_0(y), & y \in (Y_1,Y_2) \end{cases}$$

that can be solved analytically. We seek the solution in the form

$$\varphi(y) = a\exp(\pi\sqrt{k}y) + b\exp(-\pi\sqrt{k}y) + \frac{1}{\pi^2 k}\hat{f}(y)$$

where the coefficients a and b are constant in each of the three intervals. The continuity and boundary conditions result in a system of order 6 for these coefficients.

We present the computed errors in Table 6. The dominant error is observed in the areas of large solution gradients (see Fig. 7). We observe a slow convergence for the scalar unknown C on the coarse meshes. On the fine hexahedral and prismatic meshes the convergence rate increases and becomes close to the second order. On the tetrahedral meshes the convergence rate increases as $h \to 0$ but just indicates to the second order asymptotics. The flux unknown q_h shows the first-order convergence on finer meshes.

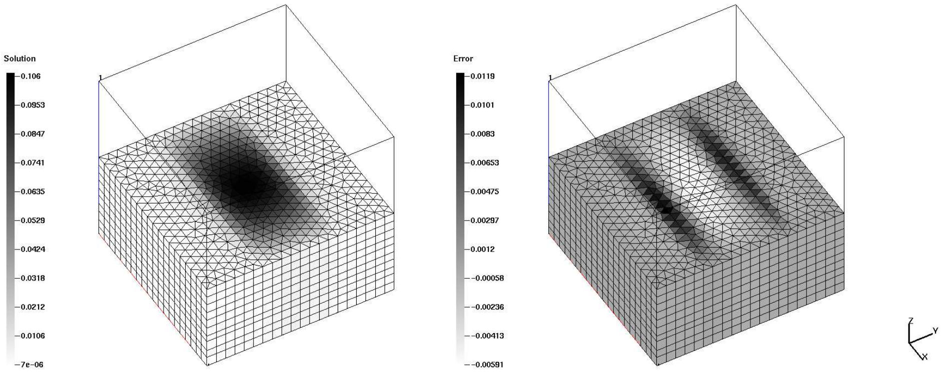

Figure 7. Solution and error distribution for the problem with sharp gradients. Prismatic mesh, $h = 1/20$.

Table 6.
The convergence results for the problem with sharp gradients.

	Hexahedral meshes		Prismatic meshes		Tetrahedral meshes	
h	ε_2^C	ε_2^q	ε_2^C	ε_2^q	ε_2^C	ε_2^q
$1/10$	4.56e-2	5.07e-2	8.01e-2	6.19e-2	9.22e-2	1.15e-1
$1/20$	2.49e-2	2.84e-2	7.02e-2	6.63e-2	7.15e-2	7.72e-2
$1/40$	7.92e-3	1.07e-2	1.80e-2	1.93e-2	3.54e-2	4.75e-2
$1/80$	2.10e-3	3.94e-3	4.28e-3	5.47e-3	1.29e-2	2.52e-2

5.5. Convergence study: discontinuous diffusion tensor

In this section we consider the convergence towards a solution for a problem with a jumping diffusion tensor. Let $\Omega = (0,1)^3$ be split into two non-overlapping subdomains $\Omega^{(1)} = \Omega \cap \{x < 0.5\}$, $\Omega^{(2)} = \Omega \cap \{x > 0.5\}$, with the interface defined by the plane $x = 0.5$, and tensor $\mathbb{K}$ jump across the interface. Let $\mathbb{K}(\mathbf{x}) = \mathbb{K}^{(i)}$ for $\mathbf{x} \in \Omega^{(i)}$ where

$$\mathbb{K}^{(1)} = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathbb{K}^{(2)} = \begin{pmatrix} 10 & 3 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The spectral decomposition $\mathbb{K}^{(i)} = (W^{(i)})^T \Lambda^{(i)} W^{(i)}$ demonstrates a significant jump of the eigenvalues and orientation of the eigenvectors of $\mathbb{K}(\mathbf{x})$:

$$\Lambda^{(1)} = \mathrm{diag}\{4,2,1\}, \qquad \Lambda^{(2)} \approx \mathrm{diag}\{10.908, 0.092, 1\}$$

$$W^{(1)} \approx \begin{pmatrix} 0.707 & 0.707 & 0 \\ -0.707 & 0.707 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad W^{(2)} \approx \begin{pmatrix} 0.957 & 0.290 & 0 \\ -0.290 & 0.957 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

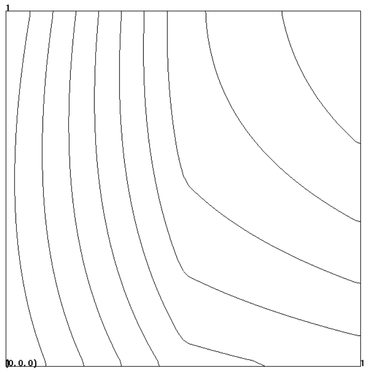

Figure 8. The solution isolines in *xy*-plane for the problem with the jumping diffusion tensor.

Table 7.
The convergence results for the problem with the jumping diffusion tensor.

	Hexahedral meshes		Prismatic meshes		Tetrahedral meshes	
h	ε_2^C	ε_2^q	ε_2^C	ε_2^q	ε_2^C	ε_2^q
$1/10$	3.38e-4	5.02e-3	2.99e-4	2.73e-3	4.68e-4	7.54e-3
$1/20$	8.64e-5	2.09e-3	1.37e-4	2.09e-3	1.67e-4	4.09e-3
$1/40$	2.18e-5	7.85e-4	3.45e-5	6.35e-4	4.62e-5	1.95e-3

We define the following exact solution of (2.1) with $\Gamma_D = \partial\Omega$:

$$c(\mathbf{x}) = \begin{cases} 1 - 2y^2 + 4xy + 2y + 6x, & \mathbf{x} \in \Omega^{(1)} \\ 3.5 - 2y^2 + 2xy + x + 3y, & \mathbf{x} \in \Omega^{(2)}. \end{cases}$$

The numerical tests were performed on the hexahedral, prismatic and tetrahedral meshes defined above. The meshes were generated so that the interface $x = 0.5$ is approximated by the mesh faces exactly. The solution isolines on the *xy*-plane are shown in Fig. 8. The convergence results presented in Table 7 demonstrate that the discontinuity of the diffusion tensor does not affect the convergence rate for all the considered meshes.

Conclusion

We have proposed and analyzed the new monotone finite volume method for the discretization of the anisotropic diffusion equation on conformal polyhedral meshes. We have proved the non-negativity of the numerical solution if the source term and the initial guess are non-negative. The method is applicable to full anisotropic heterogeneous diffusion tensors. The numerical experiments demonstrate the second-order convergence for the scalar unknown and the first-order convergence for the flux variable (a) on unstructured polyhedral quasiuniform meshes and meshes with moderately small aspect ratios and (b) for problems with highly anisotropic coefficients.

Acknowledgments

The authors are thankful to K. Nikitin, K. Lipnikov, and D. Svyatskiy for fruitful discussions and technical help and to Yu. Kuznetsov for proposing the test case studied in Subsection 5.4. The authors are also grateful to R. Garimella for the permission to use the MSTK library.

References

1. I. Aavatsmark, G. Eigestad, B. Mallison, and J. Nordbotten, A compact multipoint flux approximation method with improved robustness. *Numer. Meth. Partial Diff. Equations* (2008) **24**, No. 5, 1329–1360.
2. E. Burman and A. Ern, Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes. *Comptes Rendus Mathematique* (2004) **338**, No. 8, 641–646.
3. P. G. Ciarlet and P. -A. Raviart, Maximum principle and uniform convergence for the finite element method. *Comput. Meth. Appl. Mech. Engrg.* (1973) **2**, 17–31.
4. M. Edwards and H. Zheng, A quasi-positive family of continuous Darcy-flux finite-volume schemes with full pressure support. *J. Comp. Phys.* (2008) **227**, No. 22, 9333–9364.
5. I. E. Kaporin, High quality preconditioning of a general symmetric positive definite matrix based on its $U^TU+U^TR+R^TU$-decomposition. *Numer. Linear Algebra Appl.* (1998) **5**, No. 6, 483 – 509.
6. I. Kapyrin, A family of monotone methods for the numerical solution of three-dimensional diffusion problems on unstructured tetrahedral meshes. *Doklady Math.* (2007) **76**, No. 2, 734–738.
7. A. Koldoba, Yu. Poveschenko, and Yu. Popov, Algorithm for the solution of heat equation on non-orthogonal grids. *Diff. Equations* (1985) **21**, No. 7, 1273–1276 (in Russian).
8. S. Korotov, M. Křížek, and P. Neittaanmäki, Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle. *Math. Comp.* (2001) **233**, No. 70, 107–119 (electronic).
9. C. LePotier, Schema volumes finis monotone pour des operateurs de diffusion fortement anisotropes sur des maillages de triangle non structures. *C. C. Acad. Sci. Paris* (2005) **341**, 787–792.
10. C. LePotier, Finite volume scheme satisfying maximum and minimum principles for anisotropic diffusion operators. In: *Finite Volumes for Complex Applications, Vol. V* (Eds. R. Eymard, J.-M. Hérard), 2008, pp. 103–118.
11. K. Lipnikov, D. Svyatskiy, M. Shashkov, and Y. Vassilevski, Monotone finite volume schemes for diffusion equations on unstructured triangular and shape-regular polygonal meshes. *J. Comp. Phys.* (2007) **227**, 492–512.
12. K. Lipnikov, D. Svyatskiy, and Y. Vassilevski, Interpolation-free monotone finite volume method for diffusion equations on polygonal meshes. *J. Comp. Phys.* (2009) **228**, No. 3, 703–716.
13. J. M. Nordbotten, I. Aavatsmark, and G. T. Eigestad, Monotonicity of control volume methods. *Numer. Math.* (2007) **106**, No. 2, 255–288.
14. M. Pal and M. Edwards, Continuous Darcy-flux approximation for general 3-D grids of any element type, paper SPE 106486. In: *SPE Reservoir Simulation Symposium*. Houston, Texas, USA, 26–28 February 2007.
15. V. Ruas Santos, On the strong maximum principle for some piecewise linear finite element approximate problems of nonpositive type. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* (1982) **29**, No. 2,

473–491.

16. H. van der Vorst, Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear systems. *SIAM J. Sci. Stat. Comp.* (1992) **13**, No. 2, 631–644.

17. R. S. Varga, *Matrix Iterative Analysis*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1962.

18. Yu. Vassilevski and I. Kapyrin, Two splitting schemes for nonstationary convection-diffusion problems on tetrahedral meshes. *Comp. Math. Math. Phys.* (2008) **48**, No. 8, 1349–1366.

19. A. Yuan and Z. Sheng, Monotone finite volume schemes for diffusion equations on polygonal meshes. *J. Comp. Phys.* (2008) **227**, No. 12, 6288–6312.

20. http://sourceforge.net/projects/ani2d

21. http://sourceforge.net/projects/ani3d