Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2007) 492–512
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Monotone finite volume schemes for diffusion equations on unstructured triangular and shape-regular polygonal meshes

K. Lipnikov [a], M. Shashkov [a], D. Svyatskiy [a,*], Yu. Vassilevski [b]

[a] *Los Alamos National Laboratory, Theoretical Division, MS B284, Los Alamos, NM 87545, USA*
[b] *Institute of Numerical Mathematics, Russian Academy of Sciences, 8, Gubkina, 117333 Moscow, Russia*

Received 3 May 2007; received in revised form 31 July 2007; accepted 1 August 2007
Available online 24 August 2007

Abstract

We consider a non-linear finite volume (FV) scheme for stationary diffusion equation. We prove that the scheme is monotone, i.e. it preserves positivity of analytical solutions on arbitrary triangular meshes for strongly anisotropic and heterogeneous full tensor coefficients. The scheme is extended to regular star-shaped polygonal meshes and isotropic heterogeneous coefficients.

1. Introduction

Predictive numerical simulations require not only more sophisticated physical models but also more accurate and reliable discretization methods for these models. In this article we consider a stationary diffusion problem with a full tensor coefficient. Development of a new discretization scheme for this problem should be based on a few practical requirements [3,4]. The scheme must

- be locally conservative;
- be monotone, i.e. preserve positivity of the differential solution;
- be reliable on unstructured anisotropic meshes that may be severely distorted;
- allow heterogeneous full diffusion tensors;
- result in a sparse system with minimal number of non-zero entries;
- have higher than the first order of accuracy for smooth solutions.

* Corresponding author. Tel.: +1 505 606 2124; fax: +1 505 665 5757.
E-mail addresses: lipnikov@lanl.gov (K. Lipnikov), shashkov@lanl.gov (M. Shashkov), dasvyat@lanl.gov (D. Svyatskiy), vasilevs@dodo.inm.ras.ru (Yu. Vassilevski).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2007.08.008

As far as we know, a *linear* scheme satisfying all the above requirements is not known. Several linear schemes satisfying one or more requirements have been proposed in [1,8,9,5]. In this article, we analyze a *non-linear* scheme that satisfies all six requirements.

Monotonicity is the most difficult requirement to satisfy. We distinguish two classes of monotone schemes. The larger class contains schemes which preserve positivity of a continuum solution. The smaller class contains schemes which satisfy the discrete maximum principle (DMP). Both classes are tightly connected to algebraic properties of the matrix of the discrete operator. A monotone matrix [20] guarantees that the solution of a system of linear algebraic equations will be non-negative for any non-negative right hand side. The discrete maximum principle requires the matrix to be monotone and to have weak diagonal dominance in rows [17].

Classical finite volume (FV) and finite element (FE) schemes violate the discrete maximum principle on general meshes and for full diffusion tensors [7,9]. The schemes which satisfy the DMP impose severe restrictions on both meshes and problem coefficients [14,15]. To enlarge the class of admissible problems and meshes, some schemes such as the multi-point flux approximation methods [1] use built-in flexibility to increase their monotonicity regions. The other schemes use the first physical principles such as the constrained minimization of the energy functional [12] to get the positive solution. In this article, we analyze a FV scheme which is monotone (i.e. preserves positivity of a continuum solution) and imposes no constraints on both the problem coefficients and mesh regularity.

Recently a few non-linear schemes [6,11] have been suggested to guarantee monotonicity on unstructured simplicial meshes. The Poisson equation in arbitrary space dimensions was analyzed in [6] and a general two-dimensional parabolic equation was considered in [11]. In this article, we further develop and analyze the non-linear FV scheme proposed in [11]. First, we rectify the scheme by giving correct positions of collocation points for the case of a full diffusion tensor and an unstructured triangular mesh. Second, we propose an alternative interpolation technique [16] to improve robustness of the scheme for problems with strong anisotropy and sharp gradients. Third, we *prove* monotonicity (in the sense of solution positivity) of the scheme for stationary diffusion equations. It was shown in [11] that the scheme is monotone only for parabolic equations and sufficiently small time steps. Fourth, we study numerically important features of the scheme such as violation of the DMP as well as impact of anisotropy of the diffusion tensor on the scheme convergence. Finally, we extend the scheme to shape-regular quadrilateral meshes and heterogeneous isotropic diffusion tensors. We also mention the recent extension of the scheme to tetrahedral meshes [21].

The outline of the article is as follows. In Section 2 we formulate the stationary diffusion equation and introduce the conformal simplicial mesh. In Section 3 we describe and analyze the non-linear FV scheme. In Section 4 we extend the scheme to polygonal meshes. In section 5 we present the numerical experiments which illustrate the basic features of the scheme.

2. Stationary diffusion equation

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

$$\begin{aligned} &-\operatorname{div}\mathbb{D}\nabla c = f \quad \text{in } \Omega \\ &c = g_D \quad \text{on } \Gamma_D \\ &-\mathbb{D}\frac{\partial c}{\partial \mathbf{n}} = g_N \quad \text{on } \Gamma_N \end{aligned} \tag{1}$$

where $\mathbb{D} = \mathbb{D}^T > 0$ is a piecewise constant (possibly anisotropic) diffusion tensor and $\mathbf{n}$ is the exterior normal vector.

Let $\mathcal{T}$ be a conformal triangulation composed of $N_\mathcal{T}$ triangular cells T. We assume that $\mathcal{T}$ is connected, i.e. it cannot be split into two sets having at most one common point (a mesh vertex). We assume that the tensor $\mathbb{D}$ is constant inside each cell and its jumps occur only along mesh edges of $\mathcal{T}$. Let $\mathbf{q} = -\mathbb{D}\nabla c$ denote the diffusion flux which satisfies the mass balance equation:

$$\operatorname{div}\mathbf{q} = f \quad \text{in } \Omega. \tag{2}$$

3. Monotone non-linear FV scheme on triangular meshes

In this section, we derive a non-linear FV scheme with 2-point flux approximation. Integrating the mass balance equation (2) over a cell T and using the Green formula we get

$$\int_{\partial T} \mathbf{q}\cdot\mathbf{n}\,\mathrm{d}s = \int_T f\,\mathrm{d}x, \quad \forall T\in\mathcal{T}, \tag{3}$$

where $\mathbf{n}$ denotes the outer unit normal to ∂T. Let e denote an edge of triangle T and $\mathbf{n}_e$ be the corresponding normal vector. For a single cell T, we shall always assume that $\mathbf{n}_e$ is the outward normal vector. We shall specify orientation of $\mathbf{n}_e$ in all other cases. Hereafter, it will be convenient to assume that $|\mathbf{n}_e| = |e|$ where $|e|$ denotes the length of edge e. Eq. (3) becomes

$$\sum_{e\in\partial T} \mathbf{q}_e\cdot\mathbf{n}_e = \int_T f\,\mathrm{d}x, \quad \forall T\in\mathcal{T}, \tag{4}$$

where $\mathbf{q}_e$ is the average flux density for edge e:

$$\mathbf{q}_e = \frac{1}{|e|}\int_e \mathbf{q}\,\mathrm{d}s.$$

The FV schemes differ by approximations for the fluxes $\mathbf{q}_e$. In this article we use a two-point flux approximation. For each cell T, we assign one degree of freedom C_T for concentration c. Let C be the vector of discrete unknowns. The two-point flux approximation uses only two degrees of freedom C_{T_+} and C_{T_-} corresponding to cells T_+ and T_- that share the edge e. Sometimes, we shall write C_+ instead of C_{T_+} for simplicity. The general form for the two-point flux is as follows:

$$\mathbf{q}_e^h\cdot\mathbf{n}_e = A_e^+ C_+ - A_e^- C_-,$$

where A_e^+ and A_e^- are some coefficients. For instance, $A_e^+ = A_e^-$ in some classical FV schemes. Substituting discrete approximation $\mathbf{q}_e^h$ for $\mathbf{q}_e$ in (4), we obtain a system of N_T equations with N_T unknowns C_T.

3.1. Non-linear two-point flux

In this section, we consider a non-linear two-point flux approximation where coefficients A_e^+ and A_e^- depend on concentration. We begin with the physical meaning of discrete unknowns. The discrete concentration C_T approximates the continuous concentration c at a point $\mathbf{x}_T$ inside triangle T. We shall refer to this point as the *collocation point*. Denoting the vertices of this triangle by $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$, we define the collocation point as follows:

$$\mathbf{x}_T = \sum_{i=1}^{3}\mathbf{v}_i\lambda_i, \quad \lambda_i = \frac{|\mathbf{n}_{\alpha(i)}|_{\mathbb{D}}}{\sum_{j=1}^{3}|\mathbf{n}_{\alpha(j)}|_{\mathbb{D}}}, \tag{5}$$

where $|\mathbf{n}|_{\mathbb{D}} = (\mathbb{D}\mathbf{n}\cdot\mathbf{n})^{1/2}$ is the length of vector $\mathbf{n}$ in metric $\mathbb{D}$ induced by the diffusion tensor in triangle T and $\alpha(i)$ denotes the edge opposite to vertex $\mathbf{v}_i$. The reason for such a choice of coordinates λ_i will be explained later.

Let us consider an interior mesh edge e with end points $\mathbf{v}_1$ and $\mathbf{v}_2$ shared by two triangles T_+ and T_-. Let $\mathbb{D}_+$ and $\mathbb{D}_-$ be the values of diffusion tensor in triangles T_+ and T_-, respectively. Similarly, we denote the collocation points for these triangles by $\mathbf{x}_+$ and $\mathbf{x}_-$ (see Fig. 1). We assume that the normal vector $\mathbf{n}_e$ is outward for triangle T_+.

Let T_i, $i = 1, 2$, be the triangle with vertices $\mathbf{x}_+$, $\mathbf{x}_-$, and $\mathbf{v}_i$. For triangle T_1, we denote the normal vectors to its edges by $\mathbf{n}_1^+$, $\mathbf{n}_1^-$ and $\mathbf{n}_M$ as shown in Fig. 1. We assume again that length of these vectors equals to length of the corresponding edge, i.e. $|\mathbf{n}_1^\pm| = |\mathbf{v}_1 - \mathbf{x}_\pm|$ and $|\mathbf{n}_M| = |\mathbf{x}_+ - \mathbf{x}_-|$. In a similar way we define normals $\mathbf{n}_2^\pm$ to edges of triangle T_2. The following identities hold:

$$\mathbf{n}_1^+ + \mathbf{n}_1^- + \mathbf{n}_M = 0 \quad \text{and} \quad \mathbf{n}_2^+ + \mathbf{n}_2^- - \mathbf{n}_M = 0. \tag{6}$$

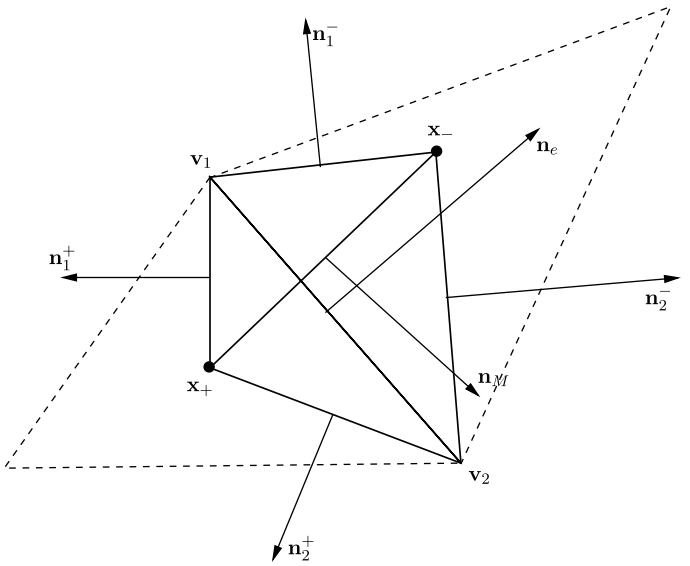

Fig. 1. Case I. Interior edge e with end points $\mathbf{v}_1$ and $\mathbf{v}_2$. The collocation points $\mathbf{x}_+$ and $\mathbf{x}_-$ are marked by solid balls. The triangles T_+ and T_- are marked by dashed lines.

Case I. To illustrate the general idea of the method, we consider first the case $\mathbb{D}_+ = \mathbb{D}_- = \mathbb{D}$. The Green formula for triangle T_1 and definition of flux $\mathbf{q}$ yield:

$$\int_{T_1} \mathbb{D}^{-1}\mathbf{q}\,\mathrm{d}x = -\int_{\partial T_1} c\mathbf{n}\,\mathrm{d}s. \tag{7}$$

Applying the mid-point (second-order) quadrature rule for both integrals, we obtain

$$-|T_1|\mathbb{D}^{-1}\mathbf{q}^h_{e,1} = \frac{C_1 + C_+}{2}\mathbf{n}_1^+ + \frac{C_1 + C_-}{2}\mathbf{n}_1^- + \frac{C_+ + C_-}{2}\mathbf{n}_M,$$

where C_1, C_+ and C_- are the values of concentration c at points $\mathbf{v}_1$, $\mathbf{x}_+$, and $\mathbf{x}_-$, respectively. Only concentrations $C_\pm$ are our discrete unknowns. The concentration C_1 will be eliminated later. Using identity (6), we get

$$\mathbf{q}^h_{e,1} = \frac{1}{2|T_1|}\mathbb{D}\big(C_+\mathbf{n}_1^- + C_-\mathbf{n}_1^+ - C_1(\mathbf{n}_1^+ + \mathbf{n}_1^-)\big). \tag{8}$$

Now we apply the same derivations to triangle T_2 to obtain the second formula for the flux density:

$$\mathbf{q}^h_{e,2} = \frac{1}{2|T_2|}\mathbb{D}\big(C_+\mathbf{n}_2^- + C_-\mathbf{n}_2^+ - C_2(\mathbf{n}_2^+ + \mathbf{n}_2^-)\big). \tag{9}$$

Given two flux density approximations (8) and (9), we seek for the discrete flux $\mathbf{q}^h_e \cdot \mathbf{n}_e$ through edge e as their linear combination:

$$\mathbf{q}^h_e \cdot \mathbf{n}_e = \mu_1 \mathbf{q}^h_{e,1} \cdot \mathbf{n}_e + \mu_2 \mathbf{q}^h_{e,2} \cdot \mathbf{n}_e, \tag{10}$$

where μ_1 and μ_2 are positive unknown coefficients. The approximation of flux density yields

$$\mu_1 + \mu_2 = 1. \tag{11}$$

The second equation for these coefficients follows from the requirement that $\mathbf{q}^h_e \cdot \mathbf{n}_e$ is the *two-point* flux approximation. Substituting (8) and (9) into (10), we require that:

$$\mu_1 \frac{C_1\mathbb{D}(\mathbf{n}_1^+ + \mathbf{n}_1^-)\cdot\mathbf{n}_e}{|T_1|} + \mu_2 \frac{C_2\mathbb{D}(\mathbf{n}_2^+ + \mathbf{n}_2^-)\cdot\mathbf{n}_e}{|T_2|} = 0. \tag{12}$$

Substituting (6) into (12) we may rewrite it as follows:

$$\mathbb{D}\mathbf{n}_M \cdot \mathbf{n}_e \left(\mu_2 \frac{C_2}{|T_2|} - \mu_1 \frac{C_1}{|T_1|} \right) = 0. \tag{13}$$

If $\mathbb{D}\mathbf{n}_M \cdot \mathbf{n}_e = 0$, then requirements (12) and (13) are fulfilled for any μ_1 and μ_2. This is almost impossible to achieve for unstructured triangular meshes. Thus the last term in (13) should be zero. Together with (11), this gives

$$\mu_1 = \frac{C_2/|T_2|}{C_1/|T_1| + C_2/|T_2|} \quad \text{and} \quad \mu_2 = \frac{C_1/|T_1|}{C_1/|T_1| + C_2/|T_2|}. \tag{14}$$

Substituting (14) in (10) gives the discrete flux through the interior edge e:

$$\mathbf{q}_e^h \cdot \mathbf{n}_e = A_e^+(C) C_+ - A_e^-(C) C_-, \tag{15}$$

where

$$\begin{aligned} A_e^+(C) &= \frac{\mu_1}{2|T_1|} \mathbf{n}_1^- \cdot \mathbb{D}\mathbf{n}_e + \frac{\mu_2}{2|T_2|} \mathbf{n}_2^- \cdot \mathbb{D}\mathbf{n}_e, \\ A_e^-(C) &= -\frac{\mu_1}{2|T_1|} \mathbf{n}_1^+ \cdot \mathbb{D}\mathbf{n}_e - \frac{\mu_2}{2|T_2|} \mathbf{n}_2^+ \cdot \mathbb{D}\mathbf{n}_e. \end{aligned} \tag{16}$$

The coefficients A_e^+ and A_e^- depend on concentrations C_1, C_2, i.e. the flux (15) is *non-linear*. The unknown concentrations C_1 and C_2 must be approximated using the original degrees of freedom, i.e. concentrations at collocation points. The total number of collocation points is $N_{\mathcal{T}}$ which leave enough flexibility for accurate approximation of these concentrations. We consider two interpolation methods.

First interpolation method uses three collocation points closest to $\mathbf{v}_1$ that form a imaginary non-degenerate triangle $\widetilde{T}$ containing $\mathbf{v}_1$. We denote these points by $\mathbf{x}_{T_j}$, $j = 1, 2, 3$. The linear interpolation over this triangle gives a second-order approximation for C_1 [11]:

$$C_1 = \sum_{j=1}^{3} C(\mathbf{x}_{T_j}) \tilde{\lambda}_j, \tag{17}$$

where $\tilde{\lambda}_j$, $j = 1, 2, 3$, are the barycentric coordinates of point $\mathbf{v}_1$ in triangle $\widetilde{T}$. Note that $0 \leqslant \tilde{\lambda}_j \leqslant 1$. We found out that this interpolation method is not robust for problems with strong anisotropy and/or solutions with sharp gradients (see Section 5).

Second interpolation method uses the inverse distance weighting [16] of values $C(\mathbf{x}_T)$ for all triangles $T \in \mathcal{T}$ that have $\mathbf{v}_1$ as a vertex. Let $\mathcal{U}(\mathbf{v}_1)$ be the collection of these triangles. Then

$$C_1 = \sum_{T \in \mathcal{U}(\mathbf{v}_1)} C(\mathbf{x}_T) w_T, \quad w_T = \frac{|\mathbf{x}_T - \mathbf{v}_1|^{-1}}{\sum_{T' \in \mathcal{U}(\mathbf{v}_1)} |\mathbf{x}_{T'} - \mathbf{v}_1|^{-1}}. \tag{18}$$

Note that $0 \leqslant w_T \leqslant 1$. We shall use this fact later. The same interpolation methods are used for approximating C_2.

Case II. Now we proceed to the general case $\mathbb{D}_+ \neq \mathbb{D}_-$. In this case the interval $[\mathbf{x}_+, \mathbf{x}_-]$ may not intersect the edge e. Therefore we define $\mathbf{m}$ as the mid-point of edge e, see Fig. 2. The edge e and point $\mathbf{m}$ split the quadrilateral $\mathbf{v}_1\mathbf{x}_+\mathbf{x}_-\mathbf{v}_2$ into four triangles $T_i^{\pm}$, $i = 1, 2$. For example, triangle T_1^+ is defined by vertices $\mathbf{m}$, $\mathbf{x}_+$ and $\mathbf{v}_1$.

In addition to vectors introduced above (see Fig. 1), we define vectors $\mathbf{n}_M^{\pm}$, and $\mathbf{n}_{e,i}$, $i = 1, 2$, that are normal to intervals $[\mathbf{m}; \mathbf{x}_{\pm}]$ and $[\mathbf{m}; \mathbf{v}_i]$, $i = 1, 2$, respectively. The orientation of these normal vectors is shown in Fig. 2. We assume again that their length equals to the length of corresponding intervals; for example, $|\mathbf{n}_M^+| = |\mathbf{x}_+ - \mathbf{m}|$. Since $\mathbf{n}_{e,i} = \frac{1}{2}\mathbf{n}_e$, the following identities hold:

$$\mathbf{n}_1^{\pm} + \mathbf{n}_M^{\pm} \pm \frac{1}{2}\mathbf{n}_e = 0. \tag{19}$$

Applying the Green formula (7) for triangle T_1^+ and using the mid-point (second-order) quadrature rules for both integrals, we get

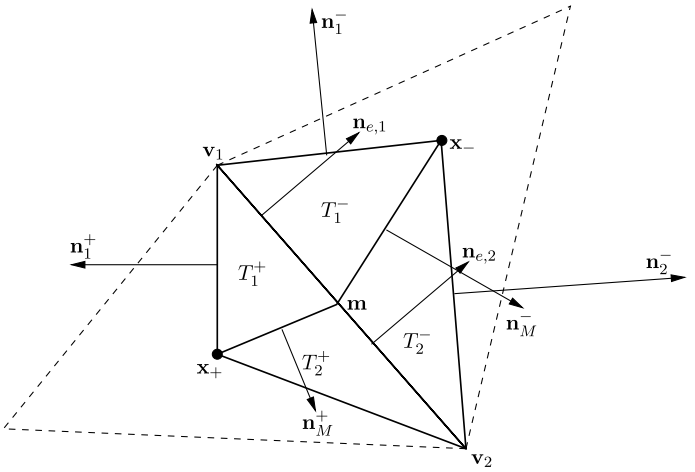

Fig. 2. Case II. Interior edge e with end points $\mathbf{v}_1$ and $\mathbf{v}_2$. The collocation points $\mathbf{x}_+$ and $\mathbf{x}_-$ are marked by solid balls. The triangles $T_1^\pm$ and $T_2^\pm$ are marked by thick lines. The triangles of $\mathcal{T}$ sharing the edge e are marked with dashed lines.

$$-2|T_1^+|\mathbb{D}_+^{-1}\mathbf{q}_{e,1}^{h,+} = (C_1 + C_+)\mathbf{n}_1^+ + \frac{1}{2}(C_1 + C_m)\mathbf{n}_e + (C_+ + C_m)\mathbf{n}_M^+, \tag{20}$$

where C_m is the concentration value at point $\mathbf{m}$. A similar formula holds for triangle T_1^-:

$$-2|T_1^-|\mathbb{D}_-^{-1}\mathbf{q}_{e,1}^{h,-} = (C_1 + C_-)\mathbf{n}_1^- - \frac{1}{2}(C_1 + C_m)\mathbf{n}_e + (C_- + C_m)\mathbf{n}_M^-. \tag{21}$$

Taking into account identities (19) and continuity of the normal flux across edge e,

$$\mathbf{q}_{e,1}^{h,+}\cdot\mathbf{n}_e = \mathbf{q}_{e,1}^{h,-}\cdot\mathbf{n}_e \equiv \mathbf{q}_{e,1}^{h}\cdot\mathbf{n}_e,$$

we eliminate C_m from (20) and (21). To simplify formula, we introduce the following numbers:

$$k_\pm^{(i)} = \mathbb{D}_\pm\mathbf{n}_i^\pm\cdot\mathbf{n}_e, \quad i = 1,2, \quad \text{and} \quad d_\pm = \frac{1}{2}\mathbb{D}_\pm\mathbf{n}_e\cdot\mathbf{n}_e.$$

Then,

$$\mathbf{q}_{e,1}^{h}\cdot\mathbf{n}_e = \frac{C_+(d_+k_-^{(1)}) + C_-(d_-k_+^{(1)}) - C_1(d_+k_-^{(1)} + d_-k_+^{(1)})}{2(|T_1^+|k_-^{(1)} - |T_1^-|k_+^{(1)})}. \tag{22}$$

Repeating the above derivations for triangles T_2^- and T_2^+, we obtain a similar formula:

$$\mathbf{q}_{e,2}^{h}\cdot\mathbf{n}_e = \frac{C_+(d_+k_-^{(2)}) + C_-(d_-k_+^{(2)}) - C_2(d_+k_-^{(2)} + d_-k_+^{(2)})}{2(|T_2^+|k_-^{(2)} - |T_2^-|k_+^{(1)})}. \tag{23}$$

Now we proceed as in Case I. Given two flux densities, we seek for their linear combination:

$$\mathbf{q}_e^h\cdot\mathbf{n}_e = \mu_1\mathbf{q}_{e,1}^h\cdot\mathbf{n}_e + \mu_2\mathbf{q}_{e,2}^h\cdot\mathbf{n}_e, \tag{24}$$

where μ_1 and μ_2 are positive unknowns. The approximation of flux density yields

$$\mu_1 + \mu_2 = 1. \tag{25}$$

The second equation for these coefficients follows from requirement of two-point flux approximation. Substituting (22) and (23) into (24), we require that

$$\mu_1\gamma_1 + \mu_2\gamma_2 = 0, \quad \gamma_i = \frac{C_i(d_+k_-^{(i)} + d_-k_+^{(i)})}{2(|T_i^+|k_-^{(i)} - |T_i^-|k_+^{(i)})}. \tag{26}$$

The solution of system (25) and (26) gives

$$\mu_1 = \frac{\gamma_2}{\gamma_2 - \gamma_1} \quad \text{and} \quad \mu_2 = \frac{-\gamma_1}{\gamma_2 - \gamma_1}. \tag{27}$$

Therefore, the non-linear flux through an interior edge e is

$$\mathbf{q}_e^h \cdot \mathbf{n}_e = A_e^+(C)C_+ - A_e^-(C)C_-, \tag{28}$$

where

$$\begin{aligned} A_e^+(C) &= \mu_1 \frac{d_+ k_-^{(1)}}{2(|T_1^+|k_-^{(1)} - |T_1^-|k_+^{(1)})} + \mu_2 \frac{d_+ k_-^{(2)}}{2(|T_2^+|k_-^{(2)} - |T_2^-|k_+^{(2)})}, \\ A_e^-(C) &= -\mu_1 \frac{d_- k_+^{(1)}}{2(|T_1^+|k_-^{(1)} - |T_1^-|k_+^{(1)})} - \mu_2 \frac{d_- k_+^{(2)}}{2(|T_2^+|k_-^{(2)} - |T_2^-|k_+^{(2)})}. \end{aligned} \tag{29}$$

Boundary edge. We consider separately the case of Dirichlet and Neumann boundary edge e. If $e \in \Gamma_N$, we simply set

$$\mathbf{q}_e^h \cdot \mathbf{n}_e = \bar{g}_N |\mathbf{n}_e|, \tag{30}$$

where $\bar{g}_N$ is the mean value of g_N on edge e. If $e \in \Gamma_D$, there exists a triangle $T_e \in \mathcal{T}$ such that $T_e \cap \Gamma_D = e$. Let T_e be the triangle containing the collocation point $\mathbf{x}_+$ as shown in Fig. 1. Using notations introduced there, we define triangle T^+ with vertices $\mathbf{x}_+$, $\mathbf{v}_1$ and $\mathbf{v}_2$. The Green formula (7) for triangle T^+, mid-point quadrature rules for both integrals, and the identity (6) yield:

$$-|T^+|\mathbb{D}_+^{-1}\mathbf{q}_e^h = \frac{C_1 + C_+}{2}\mathbf{n}_1^+ + \frac{C_2 + C_+}{2}\mathbf{n}_2^+ - \frac{C_1 + C_2}{2}(\mathbf{n}_1^+ + \mathbf{n}_2^+). \tag{31}$$

Since C_1 and C_2 are end points of the Dirichlet edge, $C_i = g_D(\mathbf{v}_i)$. From (31) we derive the linear approximation of flux through edge e:

$$\mathbf{q}_e^h \cdot \mathbf{n}_e = \frac{1}{2|T^+|}(g_D(\mathbf{v}_1)\mathbf{n}_2^+ + g_D(\mathbf{v}_2)\mathbf{n}_1^+) \cdot \mathbb{D}_+\mathbf{n}_e - \frac{1}{2|T^+|}C_+(\mathbf{n}_1^+ + \mathbf{n}_2^+) \cdot \mathbb{D}_+\mathbf{n}_e$$

or in a compact form:

$$\mathbf{q}_e^h \cdot \mathbf{n}_e = A_e^+ C_+ + A_e^-, \tag{32}$$

where

$$A_e^+ = -\frac{1}{2|T^+|}(\mathbf{n}_1^+ + \mathbf{n}_2^+) \cdot \mathbb{D}_+\mathbf{n}_e, \quad A_e^- = \frac{1}{2|T^+|}(g_D(\mathbf{v}_1)\mathbf{n}_2^+ + g_D(\mathbf{v}_2)\mathbf{n}_1^+) \cdot \mathbb{D}_+\mathbf{n}_e. \tag{33}$$

In Section 3.3, we show that the coefficients $A_e^\pm$ appeared in (15), (28), and (32) are positive.

3.2. Discrete system and its iterative solution

Let $\mathcal{E}_I$ and $\mathcal{E}_B$ denote the sets of interior and boundary edges of $\mathcal{T}$, respectively. We split the set $\mathcal{E}_B$ into subsets $\mathcal{E}_B^D$ and $\mathcal{E}_B^N$ of Dirichlet and Neumann edges, respectively. The normal vector $\mathbf{n}_e$ to edge e is defined according to the following rules. If $e \in \mathcal{E}_B$, we choose the outward normal vector to Ω. If $e \in \mathcal{E}_I$, we denote by T_{e+} and T_{e-} the two triangles that share edge e and assume that $\mathbf{n}_e$ is outward for T_{e+}. Eq. (4) may be rewritten as

$$\sum_{e \in \partial T} \chi(T,e)\mathbf{q}_e^h \cdot \mathbf{n}_e = \int_T f \, dx, \quad \forall T \in \mathcal{T}, \tag{34}$$

where $\chi(T,e) = 1$ for $e \in \mathcal{E}_B$ and

$$\chi(T,e) = \begin{cases} 1, & \text{if } T = T_{e+}, \\ -1, & \text{if } T = T_{e-} \end{cases}$$

otherwise.

Substituting (15), (28), and (32) into (34), we get a system of $N_{\mathcal{T}}$ equations in $N_{\mathcal{T}}$ unknowns C_T. Let C be the vector discrete unknowns and $\mathbb{A}(C)$ be the matrix of this system. The matrix $\mathbb{A}(C)$ may be represented by assembling of 2×2 matrices

$$\mathbb{A}_e(C) = \begin{pmatrix} A_e^+(C) & -A_e^-(C) \\ -A_e^+(C) & A_e^-(C) \end{pmatrix}$$

for interior edges and 1×1 matrices $\mathbb{A}_e(C) = A_e^+$ for Dirichlet edges. The coefficients $A_e^{\pm}(C)$ are defined in (16), (29), and (33). The global discrete non-linear system reads as

$$\mathbb{A}(C)C = F, \tag{35}$$

where

$$\mathbb{A}(C) = \sum_{e\in\mathcal{T}} \mathbb{N}_e \mathbb{A}_e(C) \mathbb{N}_e^{\mathrm{T}}, \tag{36}$$

$$F_T = \int_T f\,\mathrm{d}x - \sum_{e\in\mathcal{E}_{\mathrm{B}}^{\mathrm{D}}\cap\partial T} A_e^- - \sum_{e\in\mathcal{E}_{\mathrm{B}}^{\mathrm{N}}\cap\partial T} \int_e g_{\mathrm{N}}\,\mathrm{d}s, \tag{37}$$

A_e^- is defined in (33) and $\mathbb{N}_e$ are assembling matrices consisting of zeros and ones.

The non-linear system (35) may be solved by a number of different methods. We use the Picard iterations: Choose a small value $\varepsilon_{\mathrm{non}} > 0$ and initial vector $C^0 \in \Re^{N_{\mathcal{T}}}$ with positive entries, $C_i^0 \geqslant 0$, $i = 1,\ldots,N_{\mathcal{T}}$, and repeat for $k = 1,2,\ldots,$

1. solve $\mathbb{A}(C^{k-1})C^k = F$,
2. stop if $\|\mathbb{A}(C^k)C^k - F\| \leqslant \varepsilon_{\mathrm{non}}\|\mathbb{A}(C^0)C^0 - F\|$.

The linear system with non-symmetric matrix $\mathbb{A}(C^{k-1})$ is solved by the Bi-Conjugate Gradient Stabilized (BCGStab) method [18] with the second-order ILU preconditioner [10]. The BCGStab iterations are terminated when the relative norm of the initial residual becomes smaller than $\varepsilon_{\mathrm{lin}}$.

According to numerical evidence, the Picard iterations always converge provided that the linear systems are solved with very low tolerance $\varepsilon_{\mathrm{lin}}$.

3.3. Monotonicity

The main result of this section is the following theorem.

Theorem 3.1. Let $F_{T_i} \geqslant 0$, $C_{T_i}^0 \geqslant 0$ for $i = 1,\ldots,N_{\mathcal{T}}$ and linear systems in Picard iterations are solved exactly. Then all iterates $\mathrm{C^k}$ are non-negative vectors:

$$C_{T_i}^k \geqslant 0, \quad i = 1,\ldots,N_{\mathcal{T}}.$$

Proof. Assume for a moment that the matrix $\mathbb{A}(C^{k-1})$ is monotone for any non-negative vector $\mathrm{C^{k-1}}$. Then the solution $\mathrm{C^k}$ of $\mathbb{A}(C^{k-1})C^k = F$ is a non-negative vector and the next matrix $\mathbb{A}(C^k)$ is again monotone. Therefore, $C_{T_i}^k \geqslant 0$ for all i and k.

It remains to prove that matrix $\mathbb{A}(C)$ is monotone for any vector C with non-negative components. We begin by showing that for any conformal triangulation $\mathcal{T}$ and any piecewise constant diffusion tensor $\mathbb{D}$, the following inequalities hold:

$$\begin{aligned} &A_e^{\pm}(C) \geqslant 0, \quad \forall e \in \mathcal{E}_{\mathrm{I}}, \\ &A_e^+ > 0, \quad \forall e \in \mathcal{E}_{\mathrm{B}}^{\mathrm{D}}. \end{aligned} \tag{38}$$

Let us show that

$$k_+^{(1)} = \mathbb{D}_+ \mathbf{n}_1^+ \cdot \mathbf{n}_e < 0. \tag{39}$$

To this end we consider a triangle $T_+ \in \mathcal{T}$ with vertices $\mathbf{v}_i$, $i = 1,2,3$ (see Fig. 3). We use the same notations as in Figs. 1 and 2. We denote the normals to the triangle edges by $\mathbf{n}_{13}$, $\mathbf{n}_{23}$ and $\mathbf{n}_e$. As before, the length of these normals equal to the length of corresponding edges. For example, $|\mathbf{n}_{13}| = |\mathbf{v}_1 - \mathbf{v}_3|$. Let $\alpha_{\mathbb{D}}(\mathbf{n},\mathbf{m})$ denote the angle in metric $\mathbb{D}$ between vectors $\mathbf{n}$ and $\mathbf{m}$,

$$\alpha_{\mathbb{D}}(\mathbf{n},\mathbf{m}) = \arccos\left(\frac{\mathbf{n}\cdot\mathbb{D}\mathbf{m}}{|\mathbf{n}|_{\mathbb{D}}|\mathbf{m}|_{\mathbb{D}}}\right).$$

Without loss of generality, we put the origin of the coordinate system in vertex $\mathbf{v}_1$. Eq. (5) gives the following formula for the collocation point $\mathbf{x}_{T_+}$:

$$\mathbf{x}_{T_+} = \frac{\mathbf{v}_2|\mathbf{n}_{13}|_{\mathbb{D}_+} + \mathbf{v}_3|\mathbf{n}_e|_{\mathbb{D}_+}}{|\mathbf{n}_e|_{\mathbb{D}_+} + |\mathbf{n}_{13}|_{\mathbb{D}_+} + |\mathbf{n}_{23}|_{\mathbb{D}_+}}.$$

Note that $\mathbf{n}_1^+$ is orthogonal to $\mathbf{x}_{T_+}$, $\mathbf{n}_e$ and $\mathbf{n}_{13}$ are orthogonal to vectors $\mathbf{v}_2$ and $\mathbf{v}_3$, respectively. We search $\mathbf{n}_1^+$ as a linear combination of vectors $\mathbf{n}_e$ and $\mathbf{n}_{13}$. The direct substitution verifies that

$$\mathbf{n}_1^+ = -\frac{\mathbf{n}_e|\mathbf{n}_{13}|_{\mathbb{D}_+} - \mathbf{n}_{13}|\mathbf{n}_e|_{\mathbb{D}_+}}{|\mathbf{n}_e|_{\mathbb{D}_+} + |\mathbf{n}_{13}|_{\mathbb{D}_+} + |\mathbf{n}_{23}|_{\mathbb{D}_+}}$$

and

$$\frac{\mathbb{D}_+\mathbf{n}_e\cdot\mathbf{n}_1^+}{|\mathbf{n}_e|_{D_+}} + \frac{\mathbb{D}_+\mathbf{n}_{13}\cdot\mathbf{n}_1^+}{|\mathbf{n}_{13}|_{D_+}} = 0. \tag{40}$$

Identity (40) implies that angles between $\mathbf{n}_e$ and $\mathbf{n}_1^+$ and between $\mathbf{n}_{13}$ and $-\mathbf{n}_1^+$ are equal in metric $\mathbb{D}_+$. We shall refer to the line which passes through a triangle vertex and cuts angles with the above properties as the angle $\mathbb{D}_+$-bisectors. From the mutual orientation of vectors shown in Fig. 4, we conclude that

$$\alpha_{\mathbb{D}_+}(\mathbf{n}_e,\mathbf{n}_1^+) = \alpha_{\mathbb{D}_+}(\mathbf{n}_1^+,\mathbf{n}_{13}) + \alpha_{\mathbb{D}_+}(\mathbf{n}_{13},\mathbf{n}_e)$$

and

$$\alpha_{\mathbb{D}_+}(-\mathbf{n}_1^+,\mathbf{n}_{13}) = \pi - \alpha_{\mathbb{D}_+}(\mathbf{n}_1^+,\mathbf{n}_{13}).$$

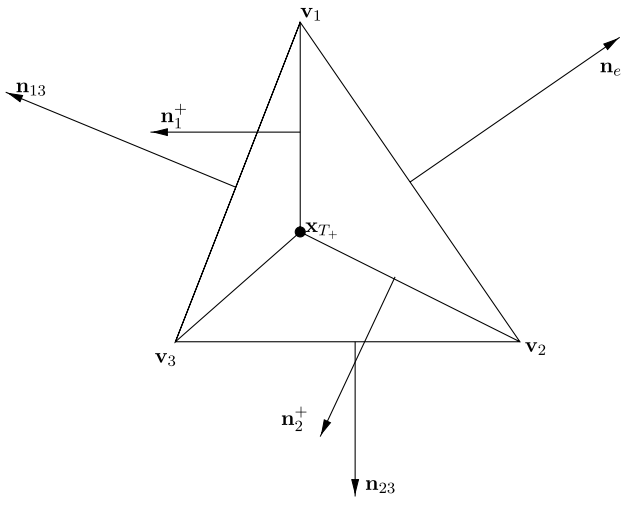

Fig. 3. Notations for triangle T_+. The collocation point $\mathbf{x}_{T_+}$ is marked by a solid bullet.

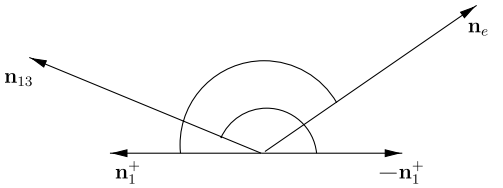

Fig. 4. Normals emanating from a common point. The marked angles are equal in metric $\mathbb{D}_+$.

Since $\alpha_{\mathbb{D}_+}(\mathbf{n}_e, \mathbf{n}_1^+) = \alpha_{\mathbb{D}_+}(-\mathbf{n}_1^+, \mathbf{n}_{13})$, we get that

$$\alpha_{\mathbb{D}_+}(\mathbf{n}_e, \mathbf{n}_1^+) = \frac{\pi}{2} + \frac{1}{2}\alpha_{\mathbb{D}_+}(\mathbf{n}_{13}, \mathbf{n}_e),$$

which in turn implies that the angle between $\mathbf{n}_e$ and $\mathbf{n}_1^+$ is obtuse in metric $\mathbb{D}_+$. Therefore $k_+^{(1)} < 0$. Using similar arguments we show that

$$k_+^{(2)} \equiv \mathbb{D}_+\mathbf{n}_2^+ \cdot \mathbf{n}_e < 0 \quad \text{and} \quad k_-^{(i)} \equiv \mathbb{D}_-\mathbf{n}_i^- \cdot \mathbf{n}_e > 0, \quad i = 1, 2. \tag{41}$$

The positive-definiteness of the diffusion tensor implies that the coefficients $\mathrm{d}_\pm$ are positive.

Now, we show that for non-negative C_{T_i}, $i = 1, \ldots, N_{\mathcal{T}}$, the coefficients μ_1 and μ_2 in (14) and (27) are non-negative. For μ's in formula (14) this follows from non-negativity of C_1, C_2 and positivity-preserving interpolation methods (17) and (18). For μ's in formula (27) we need to show that γ_1 and γ_2 have opposite signs. Since the denominators in definition of γ's are positive, we have to analyze signs of the nominators. Introducing a 2×2 matrix $\mathbb{U} = \frac{1}{2}\mathbb{D}_-(\mathbf{n}_e\mathbf{n}_e^{\mathrm{T}})\mathbb{D}_+$ and using identity $\mathbf{n}_1^+ + \mathbf{n}_1^- + \mathbf{n}_2^+ + \mathbf{n}_2^- = 0$, we get

$$\begin{aligned} d_+k_-^{(1)} + d_-k_+^{(1)} &= \frac{1}{2}\mathbf{n}_e^{\mathrm{T}}\mathbb{D}_+\mathbf{n}_e\mathbf{n}_e^{\mathrm{T}}\mathbb{D}_-\mathbf{n}_1^- + \frac{1}{2}\mathbf{n}_e^{\mathrm{T}}\mathbb{D}_-\mathbf{n}_e\mathbf{n}_e^{\mathrm{T}}\mathbb{D}_+\mathbf{n}_1^+ = \mathbf{n}_1^- \cdot \mathbb{U}\mathbf{n}_e + \mathbf{n}_1^+ \cdot \mathbb{U}^{\mathrm{T}}\mathbf{n}_e \\ &= -\mathbf{n}_2^- \cdot \mathbb{U}\mathbf{n}_e - \mathbf{n}_2^+ \cdot \mathbb{U}\mathbf{n}_e + \mathbf{n}_1^+ \cdot (\mathbb{U}^{\mathrm{T}} - \mathbb{U})\mathbf{n}_e \\ &= -(\mathbf{n}_2^- \cdot \mathbb{U}\mathbf{n}_e + \mathbf{n}_2^+ \cdot \mathbb{U}^{\mathrm{T}}\mathbf{n}_e) + \mathbf{n}_2^+ \cdot \mathbb{U}^{\mathrm{T}}\mathbf{n}_e - \mathbf{n}_2^+ \cdot \mathbb{U}\mathbf{n}_e + \mathbf{n}_1^+ \cdot (\mathbb{U}^{\mathrm{T}} - \mathbb{U})\mathbf{n}_e \\ &= -(d_+k_-^{(2)} + d_-k_+^{(2)}) + \mathbf{n}_1^+ \cdot (\mathbb{U}^{\mathrm{T}} - \mathbb{U})\mathbf{n}_e + \mathbf{n}_2^+ \cdot (\mathbb{U}^{\mathrm{T}} - \mathbb{U})\mathbf{n}_e. \end{aligned} \tag{42}$$

Based on identity $\mathbf{n}_1^+ + \mathbf{n}_2^+ + \mathbf{n}_e = 0$ and skew-symmetry of matrix $\mathbb{U}^{\mathrm{T}} - \mathbb{U}$ we conclude that sum of the last two terms in (42) is zero. Thus γ's in (27) have opposite signs and therefore μ's are non-negative.

Using (39), (41), and non-negativity of μ_1 and μ_2, we get that the first inequality in (38) holds for any non-negative vector $C \in \Re^{N_{\mathcal{T}}}$. Similarly, from (33), (39), and (41) we get the second inequality in (38). Summarizing, we have proved three important statements.

1. All diagonal entries of matrix $\mathbb{A}(C)$ are positive.
2. All off-diagonal entries of $\mathbb{A}(C)$ are non-positive.
3. Each column sum in $\mathbb{A}(C)$ is non-negative and there exists a column with a positive sum ($\mathcal{E}_{\mathrm{B}}^{\mathrm{D}} \neq \emptyset$). Moreover, by construction,
4. Matrices $\mathbb{A}(C)$ and $\mathbb{A}^{\mathrm{T}}(C)$ are irreducible since their directed graphs are strongly connected. The graphs are strongly connected because $A_e^{\pm}(C) \neq 0$ and the mesh is assumed to be connected (see [19] for more detail).

Therefore, matrix $\mathbb{A}^{\mathrm{T}}(C)$ is the M-matrix and all entries of $(\mathbb{A}^{\mathrm{T}}(C))^{-1}$ are positive, see [19]. Since inverse and transpose operation commute, $(\mathbb{A}^{\mathrm{T}}(C))^{-1} = (\mathbb{A}^{-1}(C))^{\mathrm{T}}$, we conclude that all entries of $\mathbb{A}^{-1}(C)$ are positive and $\mathbb{A}(C)$ is the monotone matrix. □

Corollary 3.1. For any tensor $\mathbb{D}$ the angle $\mathbb{D}$-bisectors of triangle T are concurrent and intersect at the collocation point $\mathbf{x}_{\mathrm{T}}$ defined by (5).

Corollary 3.2. Let $g_N \leqslant 0$ on Γ_N, $f \geqslant 0$ in Ω, $g_D \geqslant 0$ on Γ_D. Then $A_e^- \leqslant 0$ in (37) and therefore $F_{T_i} \geqslant 0$, $i = 1, N_T$.

Remark 3.1. The original version of the method [11] gives the wrong position of the collocation point $\mathbf{x}_T$ in the case of a full diffusion tensor. For the triangle with vertices $(1,0)$, $(0,1)$, and $(0.25, 0.25)$ and for the diagonal tensor $\mathbb{D} = \mathrm{diag}\{16, 1\}$ the method in [11] results in a non-monotone scheme.

4. Monotone non-linear FV scheme on polygonal meshes

Construction of a non-linear FV scheme on a polygonal mesh is similar to that on a triangular mesh. The main difficulty is to determine a position of collocation point inside each mesh cell such that the resulting system is monotone. For the triangular case it is proved that such points exist for any diffusion tensor and any geometry. For general polygonal meshes such points exist only for a restricted class of meshes and diffusion tensors. We modify the scheme to relax some of the restrictions.

Let $\mathbb{D}$ be an isotropic heterogeneous diffusion tensor and $\mathcal{Q}$ be a conformal polygonal mesh composed of $N_\mathcal{Q}$ cells. We assume that the mesh is composed of *shape-regular and star-shaped* cells in the following sense:

1. For each polygonal cell $Q \in \mathcal{Q}$, we have
$$\frac{d(Q)}{\rho(Q)} \leqslant R_*,$$
where $d(Q)$ is the diameter of Q, $\rho(Q)$ is radius of maximal inscribed circle, and R_* is a constant independent of the mesh.
2. Each cell Q is star-shaped with respect to an interior point $\mathbf{x}_Q$, i.e. any ray emanating from this point intersects the boundary ∂Q at exactly one point. If geometry allows, we shall always place $\mathbf{x}_Q$ at the center of mass of Q.

Let $\mathcal{E}_I$ and $\mathcal{E}_B$ denote again the sets of interior and boundary edges of $\mathcal{Q}$, respectively. We split $\mathcal{E}_B$ into two subsets of Dirichlet, $\mathcal{E}_B^D$, and Neumann, $\mathcal{E}_B^N$, edges. To each edge e we assign a normal vector $\mathbf{n}_e$ such that $|\mathbf{n}_e| = |e|$. If $e \in \mathcal{E}_B$, we choose the outward normal to Ω. For $e \in \mathcal{E}_I$ we denote by Q_{e+} and Q_{e-} the two polygons that share edge e and assume that $\mathbf{n}_e$ is outward for Q_{e+}. Eq. (4) may be rewritten as

$$\sum_{e \in \partial Q} \chi(Q, e) \mathbf{q}_e^h \cdot \mathbf{n}_e = \int_Q f \, \mathrm{d}x, \quad \forall Q \in \mathcal{Q}, \tag{43}$$

where $\chi(Q, e)$ is defined in the same way as the function $\chi(T, e)$ in Section 3.2.

Given a two-point flux formula (28) we may follow the path described in the previous section to get a non-linear system (35). In order to guarantee positivity of coefficients in formula (28), we propose the following method. For an edge $e \in \mathcal{E}_I$ with end points $\mathbf{v}_1$ and $\mathbf{v}_2$, we define a minimal interval $e' = [\mathbf{v}_1'; \mathbf{v}_2']$ containing e such that

$$\mathbb{D}_- \mathbf{n}_i^- \cdot \mathbf{n}_e \geqslant 0 \quad \text{and} \quad \mathbb{D}_+ \mathbf{n}_i^+ \cdot \mathbf{n}_e \leqslant 0, \quad i = 1, 2, \tag{44}$$

where $\mathbf{n}_i^\pm$ are outward normals to edges of polygon $\mathbf{v}_1' \mathbf{x}_+ \mathbf{v}_2 \mathbf{x}_-$ as shown in Fig. 5. Formally extending coefficients $D_\pm$ to the respective half planes of e', we may use formula (28) to calculate the flux density through e' and associate this flux density with the mesh edge e. The accuracy of such a modification depends on the ratio $|e'|/|e|$ which is bounded for shape-regular polygonal meshes and isotropic heterogeneous tensors. The monotonicity of the matrix $\mathbb{A}(C)$ for any non-negative vector C follows from (44) and arguments used in Section 3.3.

5. Numerical experiments

We consider several numerical tests to demonstrate that the discretization scheme satisfies the practical requirements mentioned in the introduction. The convergence rate is studied for both smooth and non-smooth

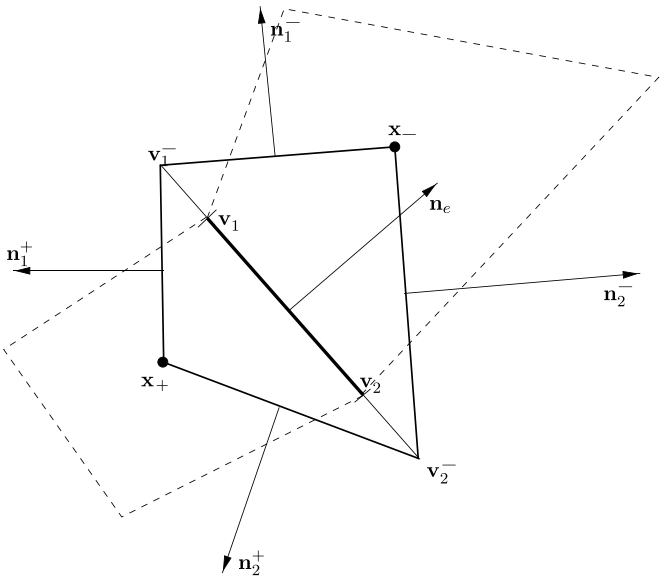

Fig. 5. Interval $[\mathbf{v}_1'; \mathbf{v}_2']$ containing the interior mesh edge e with end points $\mathbf{v}_1$ and $\mathbf{v}_2$. The collocation points $\mathbf{x}_+$ and $\mathbf{x}_-$ are marked by solid balls. The quadrilaterals Q_+ and Q_- are marked by dashed lines.

highly anisotropic solutions. The positivity of a discrete solution is verified on different types of meshes. We show that the discretization scheme is applicable to unstructured and strongly distorted meshes and can accommodate full heterogeneous and anisotropic diffusion tensor.

5.1. Implementation issues

Since the FV scheme uses the collocation points $\mathbf{x}_Q$ ($\mathbf{x}_T$ for triangular meshes) to approximate the solution, it is appropriate to use discrete L_2-norms to evaluate approximation errors. For the concentration c, we use the following norm:

$$\varepsilon_2^c = \left[\sum_{i=1}^{N_\mathcal{Q}} (c(\mathbf{x}_{Q_i}) - C_{Q_i})^2 |Q_i|\right]^{1/2}.$$

For the flux $\mathbf{q}$, we use the following norm:

$$\varepsilon_2^q = \left[\sum_{i=1}^{M_\mathcal{Q}} ((\mathbf{q} - \mathbf{q}_{e_i}^h) \cdot \mathbf{n}_{e_i})^2 |S_{e_i}|\right]^{1/2},$$

where $M_\mathcal{Q}$ is a total number of mesh edges, $\mathbf{n}_{e_i}$ is the unit normal vector to edge e_i, and $|S_{e_i}|$ is a representative area for that edge. More precisely, $|S_{e_i}|$ is the arithmetic average of areas of mesh cells sharing edge e_i.

Two interpolation methods were described in Section 3.1. The linear interpolation method is used in Sections 5.3.1 and 5.6. The inverse weighting interpolation method is used in the other sections. The numerical results presented in Section 5.4 demonstrate that the linear interpolation method is not robust for problems with strong anisotropy and/or solutions with sharp gradients.

To visualize a solution, we use the MATLAB tool which constructs the Delaunay triangulation from the set of collocation points and draws a solution on this triangular mesh.

5.2. Triangular meshes: positivity of solution

In this section, we consider several test problems illustrating Theorem 3.1. We also to compare the non-linear FV method with the mixed finite element (MFE) method and the multi-point flux approximation (MPFA)

method. Recall that the MFE method always results in an algebraic problem with a symmetric positive definite matrix. The MPFA method results in a non-symmetric matrix whose positivity was not proved in general.

5.2.1. Comparison with linear methods

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

$$\mathbb{D} = \begin{pmatrix} y^2 + \varepsilon x^2 & -(1-\varepsilon)xy \\ -(1-\varepsilon)xy & \varepsilon y^2 + x^2 \end{pmatrix}, \quad \varepsilon = 5 \times 10^{-2} \tag{45}$$

and

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

We impose the homogeneous Dirichlet boundary conditions on $\partial\Omega$. Let $\mathcal{T}$ be the triangular partition of Ω shown in Fig. 6.

The exact solution is unknown but the maximum principle states that $c(x,y)$ is non-negative. The numerical solutions obtained with the MFE, MPFA, and non-linear FV methods are shown in Fig. 7. Only the FV method preserves positivity of the continuum solution. Both linear methods produce negative values in large subdomains. The largest negative values appear in the vicinity of the source term area where the solution has sharp gradients. The MPFA solution has more non-physical oscillations than the MFE solution. As parameter ε decreases, the oscillations grow. This behavior of linear methods has been also observed by other researchers [13]. The special technique which improves monotone properties of MFPA methods has been proposed in [2].

5.2.2. Different type of meshes

Quality of the solution produced with a linear method is improved when the mesh is aligned with the solution. The numerical results presented in this section demonstrate that the non-linear FV method preserves positivity of a continuum solution on different triangulations and produces solutions of the same quality. We consider the diffusion problem described in the previous section and the following triangular partitions: the regular structured mesh (Fig. 8(a)), the regular unstructured mesh (Fig. 8(b)), and the anisotropic mesh (Fig. 8(c)). In all cases the discrete solution is non-negative.

5.3. Triangular meshes: convergence study

The next group of tests addresses the convergence rate of the non-linear FV scheme on randomly distorted triangular meshes. To construct such a mesh, we take a uniform square partition of Ω with a mesh size h, split each cell into four triangles, and distort randomly the positions of mesh nodes:

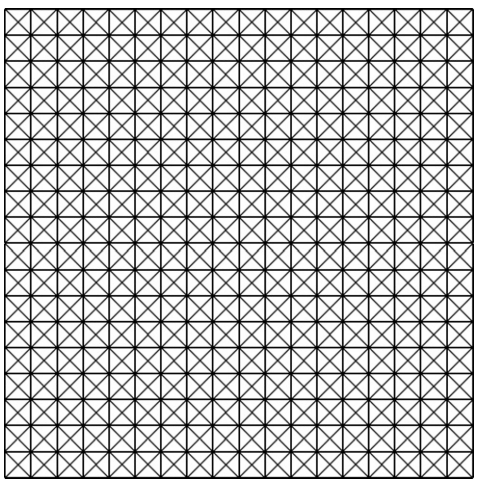

Fig. 6. Uniform triangular partion of Ω.

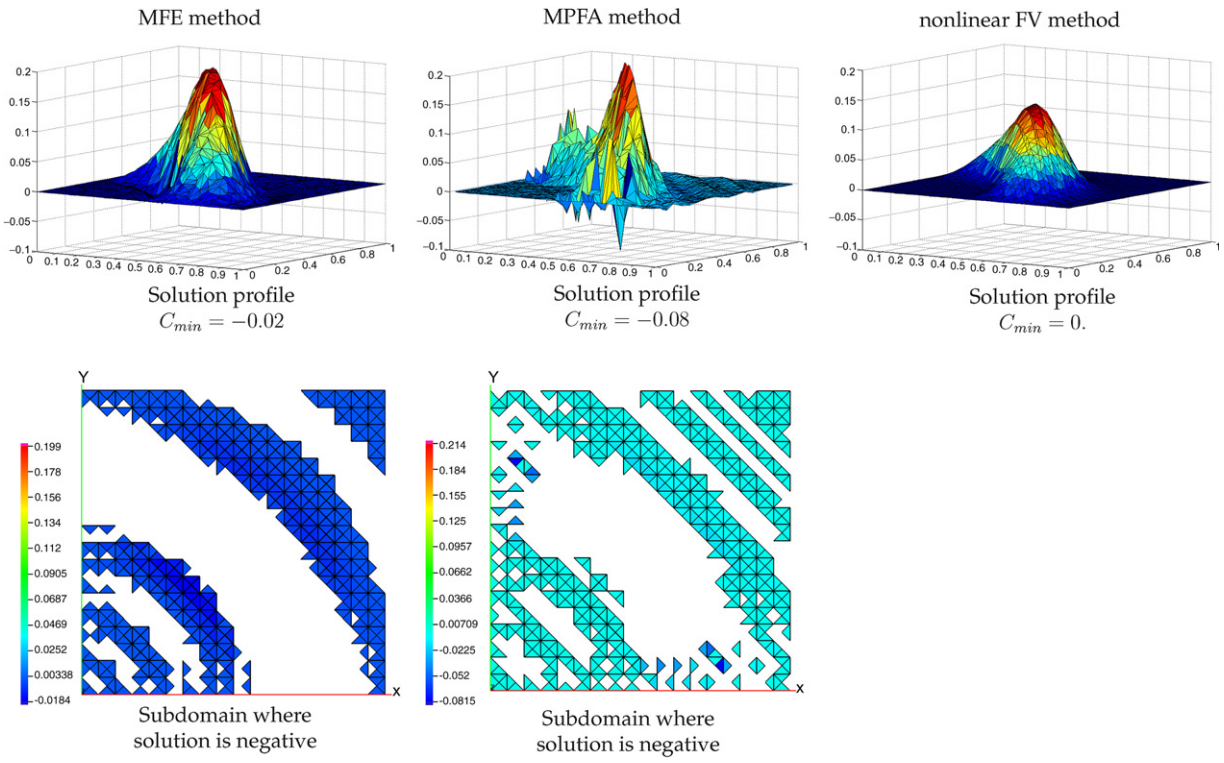

Fig. 7. Comparison of the MFE, MPFA, and non-linear FV methods.

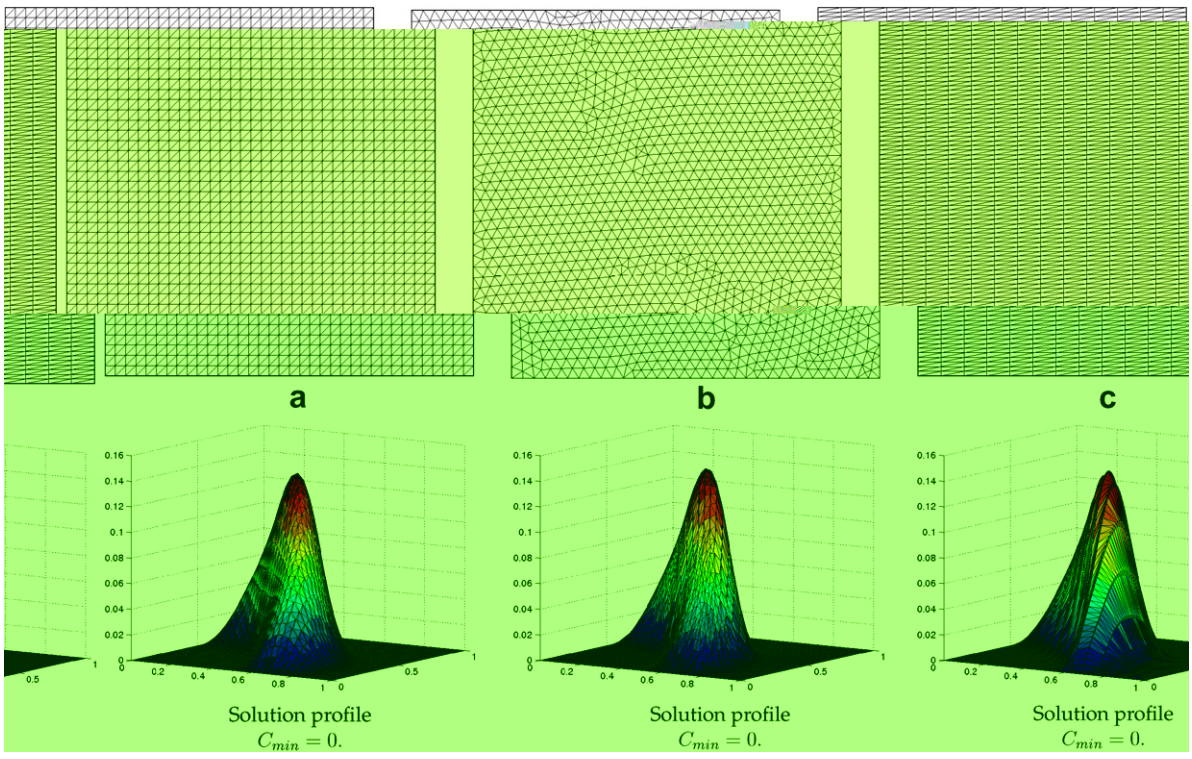

Fig. 8. Solution profile on different type of meshes.

$$x := x + \xi_x h, \quad y := y + \xi_y h,$$

where ξ_x and ξ_y are random variables with values between -0.15 and 0.15. It is pertinent to note that showing convergence of a scheme on a sequence of true random meshes is a more difficult task than that on a sequence of uniformly refined meshes.

5.3.1. Smooth solution

We consider problem (1) in the unit square $\Omega = (0,1)^2$ with the exact solution

$$c(x,y) = 2\cos(\pi x)\sin(2\pi y) + 2. \tag{46}$$

We set $\mathbb{D} = \mathbb{I}$ and impose the Dirichlet boundary condition of $\partial\Omega$.

The convergence results are presented in Table 1. The linear regression analysis shows that error ε_2^c approaches the second-order convergence rate. The convergence rate for the flux $\mathbf{q}$ is greater than the first-order. Note that in linear methods, the superconvergence of the flux is usually observed on smooth meshes.

5.3.2. Non-smooth anisotropic solution

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

We set $f = 0$, $g_D = 0$ on Γ_0, $g_D = 2$ on Γ_1, and take the anisotropic diffusion tensor $\mathbb{D}$,

$$\mathbb{D} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \tag{47}$$

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

Table 1
Convergence analysis for the smooth solution on randomly distorted triangular meshes

h	ε_2^c	ε_2^q
1/16	1.23e−3	1.04e−1
1/32	3.0e−3	4.05e−2
1/64	7.79e−4	1.75e−2
1/128	1.97e−4	8.47e−3
Rate	1.99	1.2

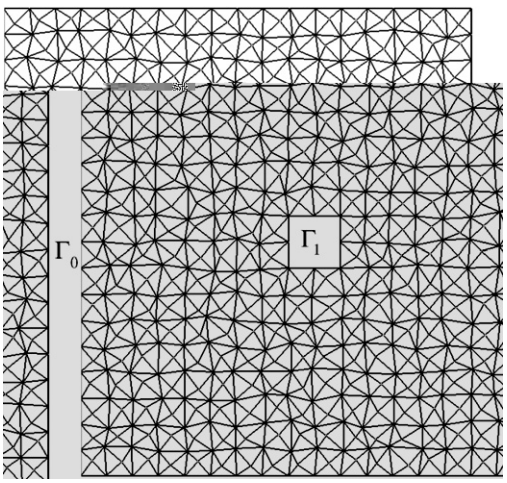

Fig. 9. Computational domain Ω and randomly distorted triangular partition.

Since the exact solution is unknown, we replace it with the discrete solution computed on a very fine mesh with $h = 1/576$ (Fig. 10). The numerical results shown in Table 2 indicate the first-order convergence rate for concentration c.

5.4. Triangular meshes: violation of discrete maximum principle

The non-linear FV scheme may not satisfy the DMP. In the absence of a source term, the discrete solution may have a few maxima inside the computational domain. We refer to this feature of the scheme as ''overshoots''. Numerical experiments presented below show that an appearance and values of overshoots depend on the mutual orientation of the solution and mesh edges. Moreover, the overshoots are sensitive to the interpolation method implemented in the scheme.

Let us consider the problem from Section 5.3.2 discretized on the uniform triangular partition shown in Fig. 11. The maximal value of the continuum solution is attained on the boundary and equals to *2*.

We have tested tensors (47) for different ratio k_1/k_2 and orientation θ of principal axes. The solution profiles are shown in Fig. 12. Maximum values of the discrete solutions are collected in Table 3. The inverse distance weighting interpolation method reduces overshoots and makes the scheme more robust. Moreover, no overshoots are observed when sharp gradients of the solution are aligned with mesh edges. The same observations are held for the MFE and MPFA schemes.

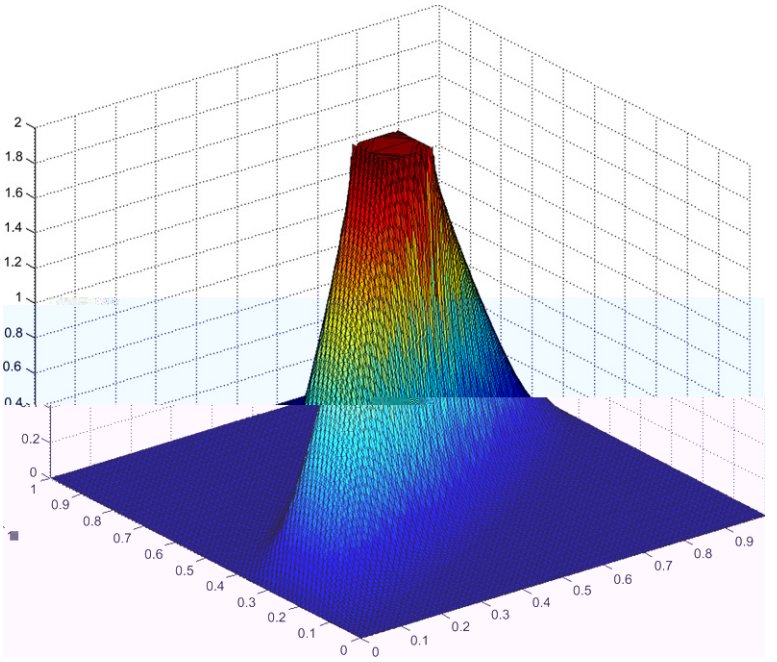

Fig. 10. Solution profile for the problem with the diffusion tensor defined by (47).

Table 2
Convergence analysis for the non-smooth solution on randomly distorted meshes

h	ε_2^c
1/18	8.69e−2
1/36	4.60e−2
1/72	2.34e−2
1/144	1.37e−2
1/288	6.72e−3
Rate	0.9

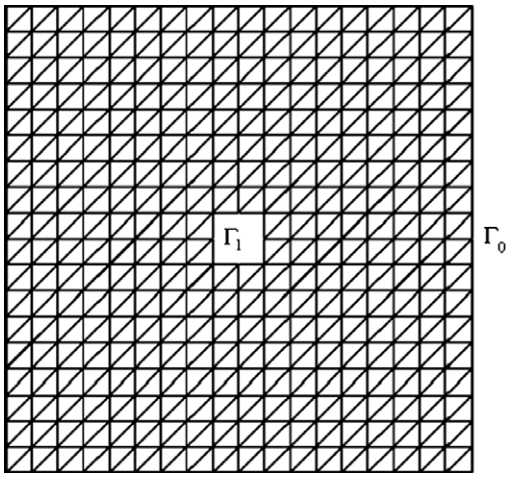

Fig. 11. Uniform triangular partion of Ω.

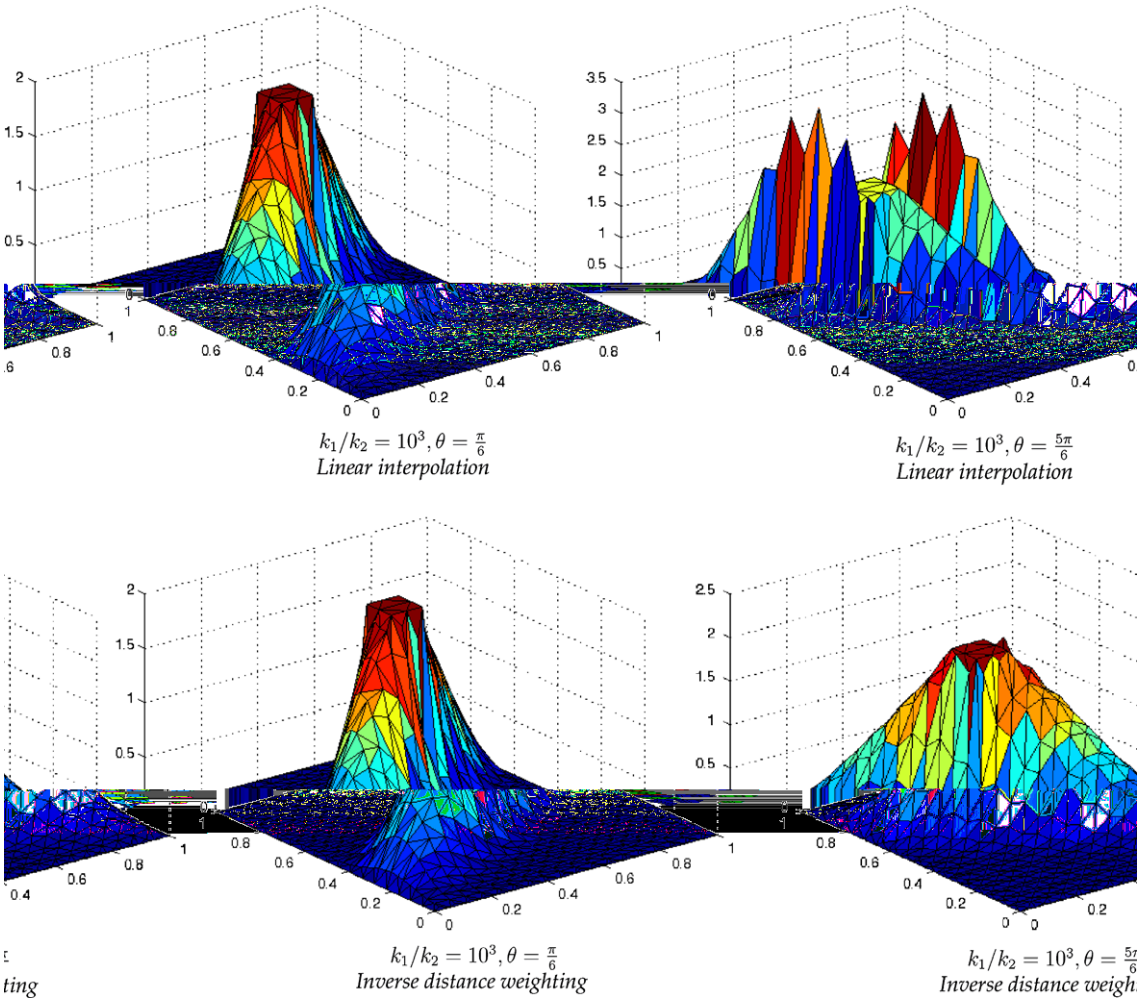

Fig. 12. Solution profiles for different diffusion tensors and different interpolation techniques.

Table 3
Maximum value of the discrete solution for different diffusion tensors and interpolation techniques

C_{max}	Interpolation method	
k_1/k_2	Linear	Inverse distance weighting
$\theta = \frac{\pi}{6}$		
10^1	1.82	1.82
10^2	1.90	1.90
10^3	1.98	1.98
$\theta = \frac{5\pi}{6}$		
10^1	1.89	1.89
10^2	2.39	2.00
10^3	3.41	2.05

Table 4
Reduction of the overshoot error ε_{over} for $k_1/k_2 = 10^3$ and $\theta = 5\pi/6$

h	ε_{over}
1/18	2.48e−3
1/36	1.40e−3
1/72	5.89e−4
1/144	2.24e−4

Table 4 demonstrates that the discrete L_2-norm of the overshoot error

$$\varepsilon_{over} = \left[\sum_{i=1}^{N_Q} (\max\{0, C_{Q_i} - 2\})^2 |Q_i|\right]^{1/2}$$

goes to zero linearly with h.

5.5. Triangular meshes: heterogeneous diffusion tensor

In this section, we demonstrate that the non-linear FV scheme can handle strong jumps of full diffusion tensor across mesh edges. We consider problem (1) in the unit square $\Omega = (0,1)^2$ with the source term

$$f(\mathbf{x}) = \begin{cases} \frac{1}{|\omega|} & \text{if } x \in \omega, \\ 0 & \text{otherwise,} \end{cases} \quad \text{where } \omega = [7/18, 11/18]^2$$

and the homogeneous Dirichlet boundary condition $g_D = 0$ on $\Gamma_D = \partial\Omega$.

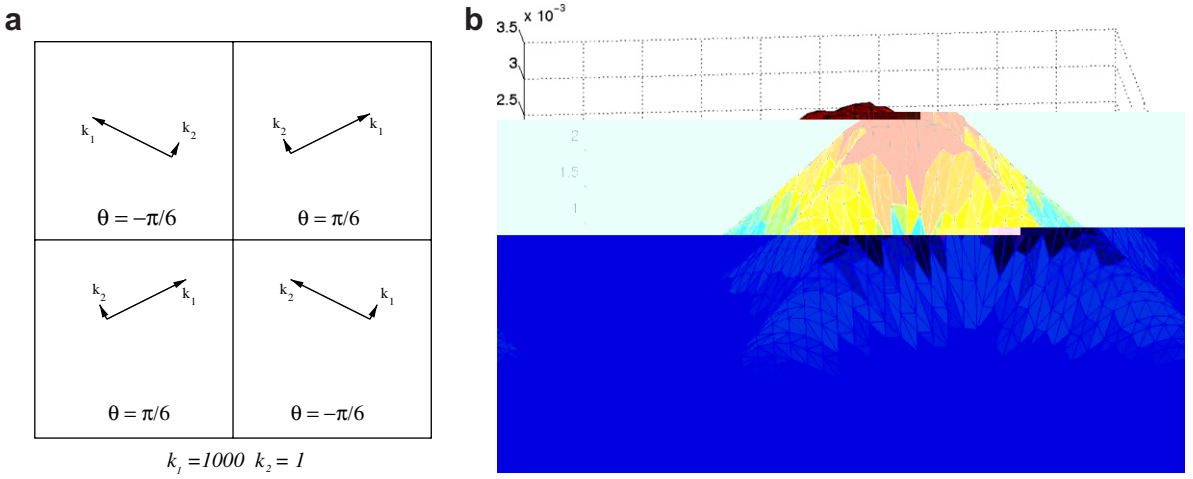

Fig. 13. Principle directions of the anisotropic diffusion tensor with fixed eigenvalues k_1 and k_2 (left picture) and profile of the discrete solution (right picture).

The domain Ω is partitioned into four square subdomains Ω_i, $i = 1, \ldots, 4$, as shown in Fig. 13(a). The diffusion tensor is given by formula (47) with different parameters k_1, k_2, and θ in subdomains Ω_i. First, we fix the anisotropy ratio by setting $k_1 = 10^3$ and $k_2 = 1$ and vary only parameter θ (see Fig. 13(a)). Second, we use the same values of θ and the chess board distribution of k_1 and k_2 (see Fig. 14(a)). In both cases we get the non-negative discrete solution (see Figs. 13(b) and 14(b)). Both discrete solutions have a good eye-ball quality.

5.6. Quadrilateral meshes: convergence study

The next group of tests addresses the convergence rate of the non-linear FV scheme on polygonal meshes in the case of isotropic diffusion tensors. We consider a set of randomly distorted quadrilateral meshes. The quadrilateral mesh is constructed from the uniform square mesh with the mesh size h by random distortion of its nodes:

$$x := x + \alpha\xi_x h, \quad y := y + \alpha\xi_y h.$$

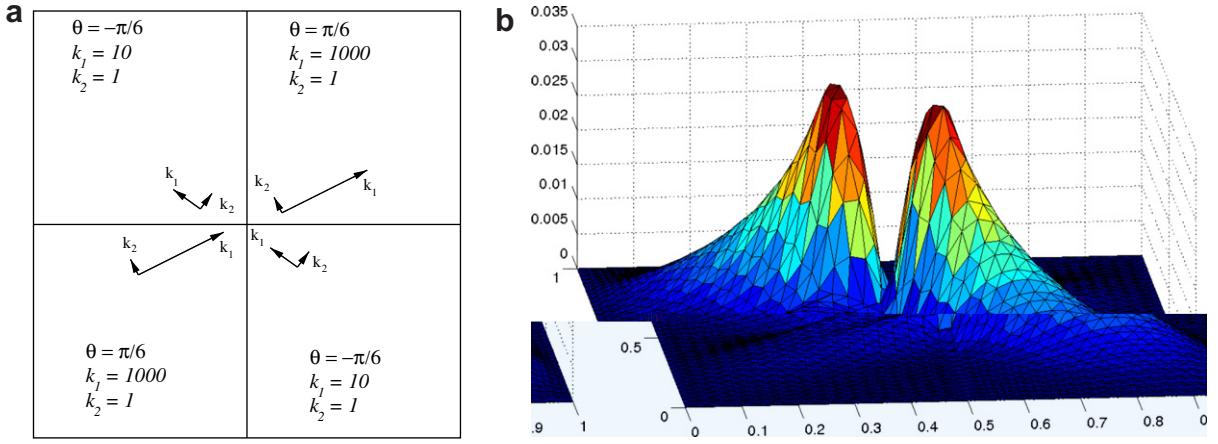

Fig. 14. Principle directions and eigenvalues of the heterogeneous anisotropic diffusion tensor (left picture) and profile of the discrete solution (right picture).

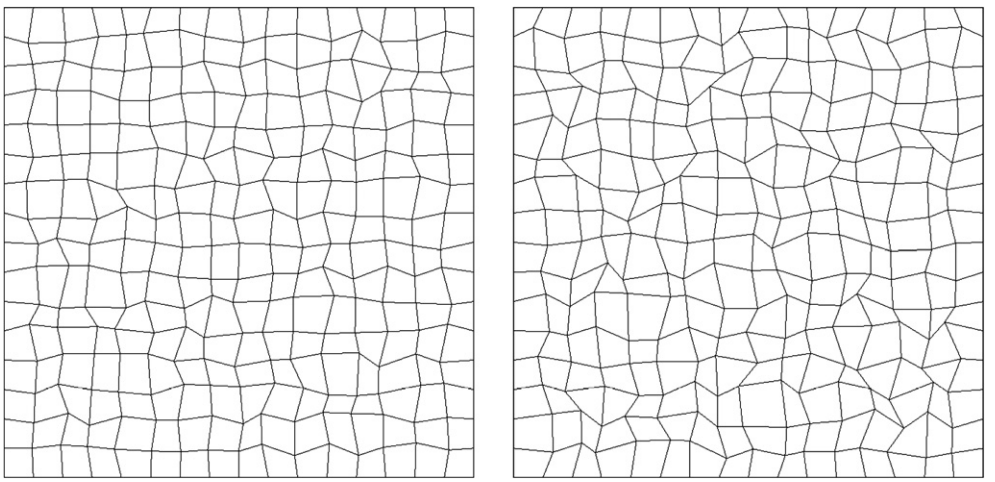

Fig. 15. Two randomly distorted quadrilateral meshes with $\alpha = 0.5$ (left picture) and $\alpha = 0.7$ (right picture).

Table 5
Convergence results for different distortion parameters

h	ε_2^c			ε_2^q		
	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$
1/16	9.06e−3	9.56e−3	1.04e−3	1.31e−1	1.49e−2	1.72e−2
1/32	2.24e−3	2.31e−3	2.63e−3	5.02e−2	5.95e−2	7.27e−2
1/64	5.46e−4	5.91e−4	6.37e−4	2.25e−2	2.73e−2	3.22e−2
1/128	1.38e−4	1.48e−4	1.59e−4	1.1e−2	1.33e−2	1.61e−2
Rate	2.01	1.97	2.01	1.18	1.15	1.14

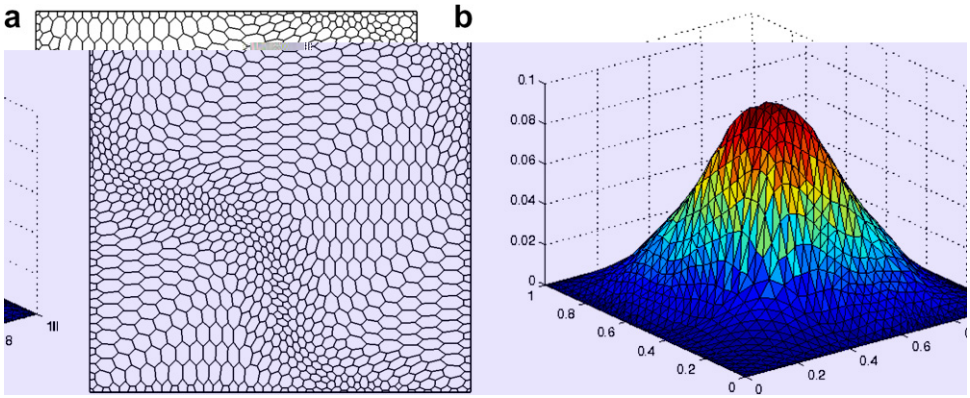

Fig. 16. The polygonal mesh (left picture) and the solution profile (right picture).

Here ξ_x and ξ_y are random variables with values between -0.5 and 0.5 and $\alpha \in [0,1]$ is the degree of distortion. We consider $\alpha \in [0.5, 0.7]$. The larger α is, the more distorted mesh is produced (see Fig. 15). If $\alpha > 0.5$, mesh cells may be non-convex. For each quadrilateral cell Q the collocation point $\mathbf{x}_Q$ is defined to be the mass center.

We consider the Dirichlet boundary value problem (1) in the unit square $\Omega = (0,1)^2$ with the isotropic diffusion tensor $\mathbb{D} = \mathbb{I}$ and the smooth exact solution

$$c(x,y) = 2\cos(\pi x)\sin(2\pi y) + 2. \tag{48}$$

In all experiments the edge extention factor $\frac{|e'|}{|e|}$ was bounded by 1.5. The numerical results presented in Table 5 show that the convergence rate of the non-linear FV scheme is not affected by the distortion parameter α. For the considered degrees of distortion we observe the second-order convergence rate for concentration c and greater than the first-order convergence rate for flux $\mathbf{q}$.

5.7. Polygonal meshes: positivity of solution

We return to the problem discussed in Section 5.2.1 and discretize it on the polygonal partition Ω_h of $\Omega = (0,1)^2$ shown in Fig. 16(a). Since the polygonal extention of the non-linear FV scheme is restricted to the case of isotropic or slightly anisotropic diffusion tensors, we pick a larger parameter $\varepsilon = 0.1$ in the formula (45) for the diffusion tensor.

The exact solution $c(x,y)$ is unknown but according to the maximum principle it is positive in Ω. The discrete solution profile shown in Fig. 16(b) demonstrates that the discretization scheme preserves solution positivity.

6. Conclusion

In this article, we further developed the non-linear finite volume method proposed by Le Potier in [11]. First, we rectified the method by providing the correct formula for positions of collocation points. Second,

we proposed the alternative interpolation technique which improves robustness of the method for problems with strong anisotropy and sharp gradients. Third, we proved monotonicity of the method for the *stationary* diffusion equation. Fourth, we studied numerically important properties of the method such as the convergence rate and violation of the discrete maximum principle. Fifth, we extended the method to regular star-shaped polygonal meshes and heterogeneous isotropic diffusion tensors.

The non-linear FV method is *monotone* and *conservative* for arbitrary triangular meshes and arbitrary full tensor diffusion coefficients. It has the *four-point* stencil for triangular meshes and the *five-point* stencil for quadrilateral meshes. It gives the *second-order* convergence rate for the scalar unknown and the *first-order* convergence rate for the flux unknown. The price for these appealing features is the method non-linearity.

Acknowledgments

This work was carried out under the auspices of the National Nuclear Security Administration of the US Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396 and the DOE Office of Science Advanced Scientific Computing Research (ASCR) Program in Applied Mathematics Research.

The authors thank C. Le Potier (CEA, France) and I. Kapyrin (INM, Russia) for fruitful discussions.

References

[1] I. Aavatsmark, T. Barkve, O. Boe, T. Mannseth, Discretization on unstructured grids for inhomogeneous, anisotropic media. Part I: derivation of the methods, SIAM J. Sci. Comput. 19 (5) (1998) 1700–1716.
[2] I. Aavatsmark, G.T. Eigestad, J.M. Nordbotten, Monotonicity of control volume methods, Numer. Math., APR 106 (2) (2007) 255–288.
[3] G. Bernard-Michel, C. Le Potier, A. Beccantini, S. Gounand, M. Chraibi, The Andra Couplex 1 test case: comparisons between finite element, mixed hybrid finite element and finite volume discretizations, Comput. Geosci. 8 (2004) 83–98.
[4] A. Bourgeat, M. Kern, S. Schumacher, J. Talandier, The COUPLEX test cases: nuclear waste disposal simulation, Comput. Geosci. 8 (2004) 83–98.
[5] F. Brezzi, K. Lipnikov, M. Shashkov, V. Simoncini, A new discretization methodology for diffusion problems on generalized polyhedral meshes, Comput. Methods Appl. Mech. Eng. 196 (37–40 SPEC. ISS.) (2007) 3682–3692.
[6] E. Burman, A. Ern, Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes, C. R. Math. 338 (8) (2004) 641–646.
[7] A. Draganescu, T.F. Dupont, L.R. Scott, Failure of the discrete maximum principle for an elliptic finite element problem, Math. Comp. 74 (249) (2004) 1–23.
[8] J. Droniou, R. Eymard, A mixed finite volume scheme for anisotropic diffusion problems on any grid, Numer. Math. 105 (1) (2006) 35–71.
[9] H. Hoteit, R. Mose, B. Philippe, Ph. Ackerer, J. Erhel, The maximum principle violations of the mixed-hybrid finite-element method applied to diffusion equations, Numer. Meth. Eng. 55 (12) (2002) 1373–1390.
[10] I. 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. 5 (1998) 483–509.
[11] C. Le Potier, Schema volumes finis monotone pour des operateurs de diffusion fortement anisotropes sur des maillages de triangle nonstructures, C. C. Acad. Sci. Paris, Ser. I 341 (2005) 787–792.
[12] R. Liska, M. Shashkov, Enforcing the discrete maximum prinicple for linear finite element solutions of elliptic problems, Commun. Comput. Phys. 2007, accepted for publication.
[13] M. Mlacnik, L. Durlofsky, R. Juanes, H. Tchelepi, Multi-point flux approximations for reservoir simulation, 12th Annual SUPRI-HW Meeting, Stanford University, November 18–19, 2004.
[14] I.D. Mishev, Finite volume methods on Voronoi meshes, Numer. Methods Partial Diff. Eq. 12 (2) (1998) 193–212.
[15] S. Korotov, M. Krizek, P. Neittaanmäki, Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle, Math. Comp. 70 (2000) 107–119.
[16] D. Shepard, A two-dimensional interpolation function for irregularly spaced data, in: Proceedings of the 23d ACM National Conference, ACM, NY, 1968, pp. 517–524.
[17] G. Stoyan, On maximum principles for monotone matrices, Linear Algebra Appl. 78 (1986) 147–161.
[18] 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. Statist. Comput. 13 (1992) 631–644.
[19] R. Varga, Matrix Iterative Analysis, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1962.
[20] R. Varga, On a discrete maximum principle, J. SIAM Numer. Anal. 3 (1966) 355–359.
[21] Yu. Vassilevski, I. Kapyrin, A family of monotone methods for the numerical solution of three-dimensional diffusion problems on unstructured tetrahedral meshes, Dokl. Math. 2007, accepted for publication.