Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 196 (2006) 579–594
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

Improvements of the Mizukami–Hughes method for convection–diffusion equations

Petr Knobloch

Charles University, Faculty of Mathematics and Physics, Department of Numerical Mathematics, Sokolovská 83, 186 75 Praha 8, Czech Republic

Received 13 September 2005; received in revised form 30 May 2006; accepted 9 June 2006

Abstract

We consider the Mizukami–Hughes method for the numerical solution of scalar two-dimensional steady convection–diffusion equations using conforming triangular piecewise linear finite elements. We propose several modifications of this method to eliminate its shortcomings. The improved method still satisfies the discrete maximum principle and gives very accurate discrete solutions in convection-dominated regime, which is illustrated by several numerical experiments. In addition, we show how the Mizukami–Hughes method can be applied to convection–diffusion–reaction equations and to three-dimensional problems.

Keywords: Stabilized FEM; Convection–diffusion; Convection–diffusion–reaction; Petrov–Galerkin method; Discrete maximum principle

1. Introduction

In this paper we propose several improvements of the Mizukami–Hughes method introduced in [14] for solving the convection–diffusion equation

$$-\varepsilon\Delta u + \boldsymbol{b}\cdot\nabla u = f \quad \text{in } \Omega. \tag{1}$$

Here Ω is a bounded two-dimensional domain with a polygonal boundary $\partial\Omega$, f is a given outer source of the unknown scalar quantity u, $\varepsilon > 0$ is the diffusivity, which is assumed to be constant, and $\boldsymbol{b}$ is the flow velocity. Eq. (1) is equipped with boundary conditions

$$u = u_{\mathrm{b}} \quad \text{on } \Gamma^{\mathrm{D}}, \quad \varepsilon\frac{\partial u}{\partial \boldsymbol{n}} = g \quad \text{on } \Gamma^{\mathrm{N}}, \tag{2}$$

where Γ^{D} and Γ^{N} are disjoint and relatively open subsets of $\partial\Omega$ satisfying $\mathrm{meas}_1(\Gamma^{\mathrm{D}}) > 0$ and $\overline{\Gamma^{\mathrm{D}} \cup \Gamma^{\mathrm{N}}} = \partial\Omega$, $\boldsymbol{n}$ is the outward unit normal vector to $\partial\Omega$ and u_{b}, g are given functions.

Despite the apparent simplicity of problem (1) and (2), its numerical solution is by no means an easy task since convection often dominates diffusion and hence the solution of (1) and (2) typically contains narrow inner and boundary layers. It is well known that the application of the classical Galerkin finite element method is inappropriate in this case since the discrete solution is usually globally polluted by spurious oscillations.

To enhance the stability and accuracy of the Galerkin discretization of (1) and (2) in convection-dominated regime, various stabilization strategies have been developed during the last three decades. One of the most efficient procedures for solving convection-dominated equations is the streamline upwind/Petrov–Galerkin (SUPG) method [2] which consistently introduces numerical diffusion along streamlines. Although this method produces to a great extent accurate and oscillation-free solutions, it does not preclude small nonphysical oscillations localized in narrow regions along sharp layers. Since these oscillations are not permissible in many applications, various terms introducing artificial crosswind diffusion in the neighborhood of layers have been proposed to be added to the SUPG formulation in order to obtain a method which is monotone or which at least reduces the local oscillations (cf. e.g. [1,3–6,8,9,13,15] and the references there). This procedure

E-mail address: knobloch@karlin.mff.cuni.cz

0045-7825/$ - see front matter
doi:10.1016/j.cma.2006.06.004

is usually referred to as discontinuity capturing (or shock capturing). A basic problem of most of these methods is the design of appropriate stabilization parameters which lead to sufficiently small nonphysical oscillations without compromising accuracy.

An interesting monotone method for solving (1) and (2) was introduced by Mizukami and Hughes [14] for linear triangular finite elements. Although it is not clear how to generalize this method to other types of finite elements, it deserves some attention since it seems to give very accurate solutions and possesses many nice properties. First of all, in contrast to the most discontinuity-capturing methods, the solutions always satisfy the discrete maximum principle, which ensures that no spurious oscillations will appear, not even in the vicinity of sharp layers. Further, as a method of upwind type, it does not contain any stabilization parameters, which also is a great advantage in comparison with the most other stabilized methods. Moreover, it is conservative and since it is a Petrov–Galerkin method, it is consistent. Last but not least, the Mizukami–Hughes method is based on a clear and simple idea whereas many discontinuity-capturing methods are derived using heuristic ad hoc arguments. Like many discontinuity-capturing methods for solving (1) and (2), the Mizukami–Hughes method depends on the unknown discrete solution and hence it is nonlinear.

Although the Mizukami–Hughes discrete solutions are often very accurate, we observed that, in some cases, they are not correct. Moreover, sometimes it was very difficult to solve the nonlinear problem with a prescribed accuracy. Therefore, in this paper, we propose some improvements of the method which correct the mentioned shortcomings and keep its quality in cases in which it works well. We will be interested in the strongly convection-dominated case characterized by the condition $\varepsilon \ll |\boldsymbol{b}|$, where $|\boldsymbol{b}|$ is the Euclidean norm of $\boldsymbol{b}$.

A drawback of both the original and the improved versions of the Mizukami–Hughes method is that no existence, uniqueness and convergence results are available. Moreover, it seems to be rather difficult to generalize the method to more complicated problems. Nevertheless, we shall show that the method can be extended to convection–diffusion–reaction equations and to the three-dimensional case.

The plan of the paper is as follows. First, in the next section, we describe and comment the original Mizukami–Hughes method published in [14]. Then, in Sections 3–5, we discuss shortcomings of this method and propose some modifications to eliminate them. Since this will take several pages, we briefly summarize the improved method in Section 6. Section 7 contains our numerical results which illustrate the high accuracy of the improved method. In Section 8, we deal with a generalization of the Mizukami–Hughes method to convection–diffusion–reaction equations and, finally, in Section 9, we discuss the application of the Mizukami–Hughes method to the three-dimensional case.

2. The Mizukami–Hughes method

Let $\mathscr{T}_h$ be a triangulation of Ω consisting of a finite number of open triangular elements K. The discretization parameter h in the notation $\mathscr{T}_h$ is a positive real number satisfying $\operatorname{diam}(K) \leqslant h$ for any $K \in \mathscr{T}_h$. We assume that $\overline{\Omega} = \bigcup_{K\in\mathscr{T}_h} \overline{K}$ and that the closures of any two different elements K, $\widetilde{K} \in \mathscr{T}_h$ are either disjoint or possess either a common vertex or a common edge. Further, we assume that any edge of an element $K \in \mathscr{T}_h$ which lies on $\partial\Omega$ is contained either in $\overline{\Gamma^{\mathrm{D}}}$ or in $\overline{\Gamma^{\mathrm{N}}}$. Finally, we assume that the triangulation $\mathscr{T}_h$ is of weakly acute type, i.e., the magnitude of all angles of elements $K \in \mathscr{T}_h$ is less than or equal to $\pi/2$. This property will be used for proving the discrete maximum principle.

The solution u of (1) and (2) will be approximated by a continuous piecewise linear function u_h from the space

$$V_h = \{v \in C(\overline{\Omega}); v|_K \in P_1(K)\ \forall K \in \mathscr{T}_h\}.$$

Let $a_1, \ldots, a_{M_h}$ be the vertices of $\mathscr{T}_h$ lying in $\Omega \cup \Gamma^{\mathrm{N}}$ and let $a_{M_h+1}, \ldots, a_{N_h}$ be the vertices of $\mathscr{T}_h$ lying on $\overline{\Gamma^{\mathrm{D}}}$. For any $i \in \{1, \ldots, N_h\}$, let $\varphi_i \in V_h$ be the function satisfying $\varphi_i(a_j) = \delta_{ij}$ for $j = 1, \ldots, N_h$, where δ_{ij} is the Kronecker symbol. Then $V_h = \operatorname{span}\{\varphi_i\}_{i=1}^{N_h}$. The Mizukami–Hughes method is a Petrov–Galerkin method with weighting functions

$$\tilde{\varphi}_i = \varphi_i + \sum_{\substack{K\in\mathscr{T}_h,\\ a_i\in\overline{K}}} C_i^K \chi_K, \quad i = 1, \ldots, M_h,$$

where C_i^K are constants to be determined later and χ_K is the characteristic function of K (i.e., $\chi_K = 1$ in K and $\chi_K = 0$ elsewhere). The discrete solution u_h of (1) and (2) is defined by

$$\begin{aligned} &u_h \in V_h,\\ &\varepsilon(\nabla u_h, \nabla\varphi_i) + (\boldsymbol{b}\cdot\nabla u_h, \tilde{\varphi}_i) = (f, \tilde{\varphi}_i) + (g, \varphi_i)_{\Gamma^{\mathrm{N}}},\\ &\quad i = 1, \ldots, M_h,\\ &u_h(a_i) = u_{\mathrm{b}}(a_i), \quad i = M_h + 1, \ldots, N_h, \end{aligned}$$

where $(\cdot,\cdot)$ denotes the inner product in $L^2(\Omega)$ and $(\cdot,\cdot)_{\Gamma^{\mathrm{N}}}$ is the inner product in $L^2(\Gamma^{\mathrm{N}})$. Moreover, here and in the following, the flow velocity $\boldsymbol{b}$ is considered to be piecewise constant (equal to the original function $\boldsymbol{b}$ at barycentres of elements of $\mathscr{T}_h$).

It remains to define the constants C_i^K, which is the key point of the method. Mizukami and Hughes require for any $K \in \mathscr{T}_h$ that

$$C_i^K \geqslant -\tfrac{1}{3}\ \forall i \in \{1, \ldots, N_h\},\ a_i \in \overline{K}, \quad \sum_{\substack{i=1\\ a_i\in\overline{K}}}^{N_h} C_i^K = 0 \tag{3}$$

and that the local convection matrix A^K with entries

$$a_{ij}^K = (\boldsymbol{b}\cdot\nabla\varphi_j, \tilde{\varphi}_i)_K, \quad i = 1, \ldots, M_h,\ j = 1, \ldots, N_h,\ a_i, a_j \in \overline{K}$$

is of nonnegative type (i.e., off-diagonal entries of A^K are nonpositive and the sum of the entries in each row of A^K is nonnegative, cf. [7]). As usual, $(\cdot,\cdot)_K$ denotes the inner product in $L^2(K)$. The latter condition in (3) implies that u_h satisfies a discrete mass conservation law if the data in (1) and (2) satisfy $\Gamma^{\mathrm{N}} = \partial\Omega$, $g = 0$ and $\boldsymbol{b} = \text{const.}$, cf. [11].

The matrix A^K has three columns and at most three rows and it will be of nonnegative type as soon as $a_{ij}^K \leqslant 0$ for $i \neq j$. Note that

$$a_{ij}^K = (\boldsymbol{b}\cdot\nabla\varphi_j)|_K \int_K \tilde{\varphi}_i \,\mathrm{d}x = (\boldsymbol{b}\cdot\nabla\varphi_j)|_K \mathrm{meas}_2(K)\left(\tfrac{1}{3} + C_i^K\right).$$

Let K be any element of the triangulation $\mathscr{T}_h$ and let the vertices of K be a_1, a_2 and a_3. For each vertex a_i, $i = 1,2,3$, we define a vertex zone VZ_i and an edge zone EZ_i whose boundaries consist of lines intersecting the barycentre of K which are parallel to the two edges of K possessing the vertex a_i, see Fig. 1. The common part of the boundaries of two adjacent zones is included in the respective vertex zone. To avoid misunderstandings, we shall later also use the notation EZ_i^K instead of EZ_i.

Without loss of generality, we may assume that the vertices of K are numbered in such a way that $\boldsymbol{b}$ points into the vertex zone or the edge zone of a_1 as depicted in Fig. 1. Then

$$\boldsymbol{b} \in \mathrm{VZ}_1 \Longleftrightarrow \boldsymbol{b}\cdot\nabla\varphi_1 > 0, \quad \boldsymbol{b}\cdot\nabla\varphi_2 \leqslant 0, \quad \boldsymbol{b}\cdot\nabla\varphi_3 \leqslant 0,$$
$$\boldsymbol{b} \in \mathrm{EZ}_1 \Longleftrightarrow \boldsymbol{b}\cdot\nabla\varphi_1 < 0, \quad \boldsymbol{b}\cdot\nabla\varphi_2 > 0, \quad \boldsymbol{b}\cdot\nabla\varphi_3 > 0,$$

where we write $\nabla\varphi_i$ instead of $\nabla\varphi_i|_K$ for simplicity.

If $\boldsymbol{b} \in \mathrm{VZ}_1$, then (3) holds and A^K is of nonnegative type for

$$C_1^K = \tfrac{2}{3}, \qquad C_2^K = C_3^K = -\tfrac{1}{3}.$$

If A^K has three rows, this is the only possibility how to choose these constants. On the other hand, if $\boldsymbol{b} \in \mathrm{EZ}_1$, then it is generally not possible to choose the constants C_1^K, C_2^K, C_3^K in such a way that (3) holds and A^K is of nonnegative type. However, Mizukami and Hughes made the important observation that u still solves Eq. (1) if we replace $\boldsymbol{b}$ by any function $\tilde{\boldsymbol{b}}$ such that $\tilde{\boldsymbol{b}} - \boldsymbol{b}$ is orthogonal to ∇u. This suggests to define the constants C_i^K in such a way that the matrix A^K is of nonnegative type for $\boldsymbol{b}$ replaced by a function $\tilde{\boldsymbol{b}}$ pointing into a vertex zone. Since ∇u is not known a priori, we obtain a nonlinear problem where the constants C_i^K depend on the discrete solution u_h which we want to compute.

Let us assume that $\boldsymbol{b}\cdot\nabla u_h|_K \neq 0$ and let $\boldsymbol{w} \neq \boldsymbol{0}$ be a vector orthogonal to $\nabla u_h|_K$. Then there exists $\alpha \in \mathbb{R}$ such that $\boldsymbol{b} + \alpha\boldsymbol{w} \in \mathrm{VZ}_2$ or $\boldsymbol{b} + \alpha\boldsymbol{w} \in \mathrm{VZ}_3$. The dashed and dotted arcs in Fig. 2 indicate to which part of the plane the vector $\boldsymbol{w}$ should point from the barycentre of K if the first or the second possibility should arrive. To simplify the presentation, let us introduce the sets

$$V_k = \{\alpha \in \mathbb{R}; \boldsymbol{b} + \alpha\boldsymbol{w} \in \mathrm{VZ}_k\}, \quad k = 2,3.$$

Mizukami and Hughes show that, depending on V_2 and V_3, the following values of the constants C_i^K should be used:

$$V_2 \neq \emptyset \quad \text{and} \quad V_3 = \emptyset$$
$$\Rightarrow \quad C_2^K = \tfrac{2}{3}, \quad C_1^K = C_3^K = -\tfrac{1}{3}, \tag{4}$$
$$V_2 = \emptyset \quad \text{and} \quad V_3 \neq \emptyset$$
$$\Rightarrow \quad C_3^K = \tfrac{2}{3}, \quad C_1^K = C_2^K = -\tfrac{1}{3}, \tag{5}$$
$$V_2 \neq \emptyset \quad \text{and} \quad V_3 \neq \emptyset$$
$$\Rightarrow \quad C_1^K = -\tfrac{1}{3}, \quad C_2^K + C_3^K = \tfrac{1}{3},$$
$$C_2^K \geqslant -\tfrac{1}{3}, \quad C_3^K \geqslant -\tfrac{1}{3}. \tag{6}$$

If, for some $k \in \{2,3\}$, the set V_k is nonempty, we choose $\alpha_k \in V_k$ and define the matrix $\tilde{A}^{K,k}$ with entries

$$\tilde{a}_{ij}^{K,k} = ((\boldsymbol{b} + \alpha_k\boldsymbol{w})\cdot\nabla\varphi_j, \tilde{\varphi}_i)_K, \quad i,j = 1,2,3 \ \ (a_i \in \Omega \cup \Gamma^{\mathrm{N}}),$$

where $\tilde{\varphi}_i$ are defined using C_i^K's from (4) if $k = 2$ and using C_i^K's from (5) if $k = 3$. As we have seen above, the matrix

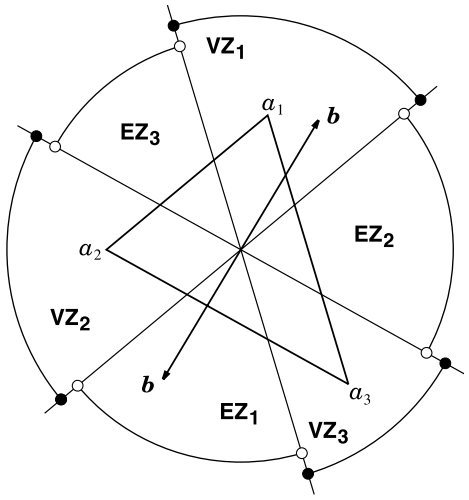

Fig. 1. Definition of edge zones and vertex zones.

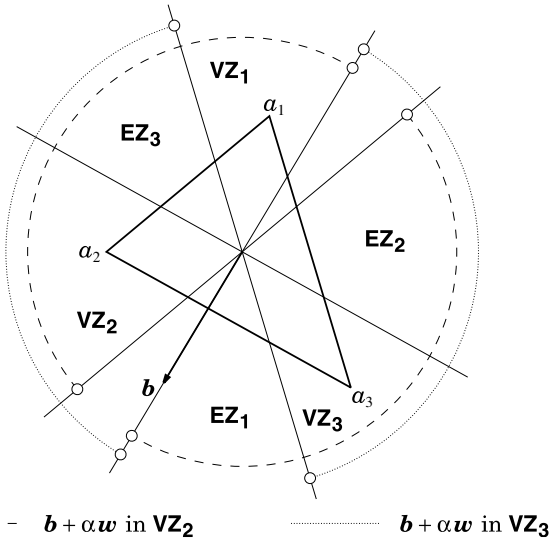

Fig. 2. Orientations of $\boldsymbol{w}$ for which $\boldsymbol{b} + \alpha\boldsymbol{w} \in \mathrm{VZ}_2$ or $\boldsymbol{b} + \alpha\boldsymbol{w} \in \mathrm{VZ}_3$.

$\widetilde{A}^{K,k}$ is of nonnegative type. Let us assume that V_2 or V_3 is empty and let V_k be the nonempty set. Since $u_h|_K = u_1\varphi_1 + u_2\varphi_2 + u_3\varphi_3$, the vector $U = (u_1, u_2, u_3)$ satisfies for $i = 1,2,3$ (with $a_i \in \Omega \cup \Gamma^{\mathrm{N}}$)

$$(A^K U)_i = (\boldsymbol{b}\cdot\nabla u_h, \tilde{\varphi}_i)_K = ((\boldsymbol{b} + \alpha_k \boldsymbol{w})\cdot\nabla u_h, \tilde{\varphi}_i)_K = (\widetilde{A}^{K,k}U)_i.$$

In case (6), we have

$$A^K = (C_2^K + \tfrac{1}{3})A^{K,2} + (C_3^K + \tfrac{1}{3})A^{K,3},$$

where $A^{K,2}$ and $A^{K,3}$ are matrices defined like A^K but using C_i^K's from (4) and (5), respectively. Consequently, for $i = 1,2,3$ (with $a_i \in \Omega \cup \Gamma^{\mathrm{N}}$), we obtain

$$(A^K U)_i = (C_2^K + \tfrac{1}{3})(\widetilde{A}^{K,2}U)_i + (C_3^K + \tfrac{1}{3})(\widetilde{A}^{K,3}U)_i.$$

Thus, in all three cases (4)–(6), the discrete solution satisfies

$$(A^K U)_i = (\widetilde{A}^K U)_i, \quad i = 1,2,3 \ \ (a_i \in \Omega \cup \Gamma^{\mathrm{N}}), \tag{7}$$

where $\widetilde{A}^K$ is a matrix of nonnegative type. In case (6), Mizukami and Hughes suggest to set

$$C_i^K = \frac{\boldsymbol{b}\cdot\nabla\varphi_i}{3|\boldsymbol{b}\cdot\nabla\varphi_1|}, \quad i = 1,2,3. \tag{8}$$

This choice is also considered if $\boldsymbol{b} \in \mathrm{EZ}_1$ satisfies $\boldsymbol{b}\cdot\nabla u_h|_K = 0$. If $\boldsymbol{b} = \boldsymbol{0}$, Mizukami and Hughes set $C_i^K = 0$ for $i = 1,2,3$.

The above choice of the constants C_i^K assures that the discrete solution always satisfies (7) with a matrix $\widetilde{A}^K$ of nonnegative type. Denoting by D the matrix having the entries $d_{ij} = (\nabla\varphi_j, \nabla\varphi_i)$, $i = 1,\ldots,M_h$, $j = 1,\ldots,N_h$, and by $\widetilde{A}$ the $M_h \times N_h$ matrix made up of the local matrices $\widetilde{A}^K$, we see that the vector of coefficients of the discrete solution u_h with respect to the basis $\{\varphi_i\}_{i=1}^{N_h}$ of the space V_h is the solution of a linear system with the matrix $C \equiv \varepsilon D + \widetilde{A}$. Since the triangulation $\mathscr{T}_h$ is of weakly acute type, it is easily seen that the matrix $\{(\nabla\varphi_j, \nabla\varphi_i)_K\}_{i,j=1}^3$ is of nonnegative type. Consequently, the matrices D and C also are of nonnegative type. Moreover, since the matrix $\{d_{ij}\}_{i,j=1}^{M_h}$ is nonsingular, the matrix $\{c_{ij}\}_{i,j=1}^{M_h}$ also is nonsingular. This implies that u_h satisfies the discrete maximum principle (see e.g. [8]). Thus, for any $G \subset \overline{\Omega}$ being a union of closures of elements of $\mathscr{T}_h$, we have

$$(f, \tilde{\varphi}_i) \leqslant 0 \quad \forall a_i \in \operatorname{int} G \Rightarrow \quad \max_G u_h = \max_{\partial G} u_h, \tag{9}$$

$$(f, \tilde{\varphi}_i) \geqslant 0 \quad \forall a_i \in \operatorname{int} G \Rightarrow \quad \min_G u_h = \min_{\partial G} u_h, \tag{10}$$

which shows that the discrete solution does not contain any spurious oscillations.

3. Improvement of the Mizukami–Hughes method in boundary layer regions

The Mizukami–Hughes method often provides accurate and oscillation-free discrete solutions, see the examples in [14,9]. However, in some cases, we observed that the discrete solution was not correct. We shall demonstrate this on a simple example which was also considered in [14].

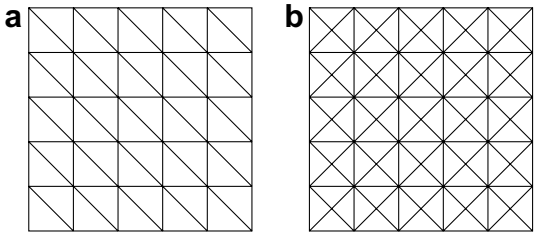

Fig. 3. Considered types of triangulations.

Let $\Omega = (0,1)^2$ and, like in [14], let us consider uniform triangulations $\mathscr{T}_h$ of Ω of the type depicted in Fig. 3(a), which consist of $2(N\times N)$ equal right-angled isosceles triangles ($N = 5$ in Fig. 3(a)). Let $N = 10$ and let us consider the problem (1) and (2) with

$$\varepsilon = 10^{-7}, \quad \boldsymbol{b} = (1,0), \quad f = 1, \quad \Gamma^{\mathrm{D}} = \partial\Omega, \quad u_{\mathrm{b}} = 0. \tag{11}$$

The discrete solution obtained using the Mizukami–Hughes method is indistinguishable from the discrete solution corresponding to $\varepsilon \to 0$. For $\varepsilon \to 0$, we easily find that the discrete solution is nodally exact, i.e.,

$$u_h(x,y) = x \quad \text{for } (x,y) \in [0,0.9]\times[0.1,0.9]. \tag{12}$$

Changing $\boldsymbol{b}$ to $\boldsymbol{b} = (1,\alpha)$ with $|\alpha| \ll 1$, we expect that the discrete solution basically remains the same. However, Fig. 4(a) corresponding to $\alpha = -0.0001$ shows that the discrete solution changes dramatically. The reason is that the small change of $\boldsymbol{b}$ causes a significant change of the constants C_i^K for elements $K \in \mathscr{T}_h$ having an edge at the upper part of the boundary of Ω, see Fig. 5(a) and (b). Note that we can set $\boldsymbol{w} = (1,0)$ for these elements K since $u_h = 0$ on $\partial\Omega$. Let us mention that Fig. 4(a) does not show a violation of the discrete maximum principle since $(f, \tilde{\varphi}_i) > 0$ for all $i \in \{1,\ldots,M_h\}$ and the right-hand side of (10) is satisfied for any admissible set G.

It is obvious that a small change of $\boldsymbol{b}$ should only lead to a small change of the constants C_i^K and hence a first idea to improve the behaviour of the method might be to use the vertex-zone definition of C_i^K's also for $\boldsymbol{b}$ which is not contained in a vertex zone but is very near to it. However, the problems also appear for vectors $\boldsymbol{b}$ which cannot be considered to lie near a vertex zone, e.g. for $\alpha \in [-0.5,-0.1]$. For such α, a nodally exact solution (again for $\varepsilon \to 0$) should satisfy

$$u_h(x,y) = x \quad \text{for } (x,y) \in [0,0.9]\times[0.1,0.2]. \tag{13}$$

Let us assume that

(A1) the constants C_i^K are defined as described in Section 2 if $\boldsymbol{b}$ lies in a vertex zone;
(A2) $C_j^K = -\frac{1}{3}$ if $\boldsymbol{b} \in \mathrm{EZ}_j^K$ for some index j.

Then, for $\varepsilon = 0$, it is easy to show that the necessary condition for the validity of (13) is that, for any element K having the vertices $(x,0)$, $(x,0.1)$, $(x-0.1,0.1)$ with $x \in$

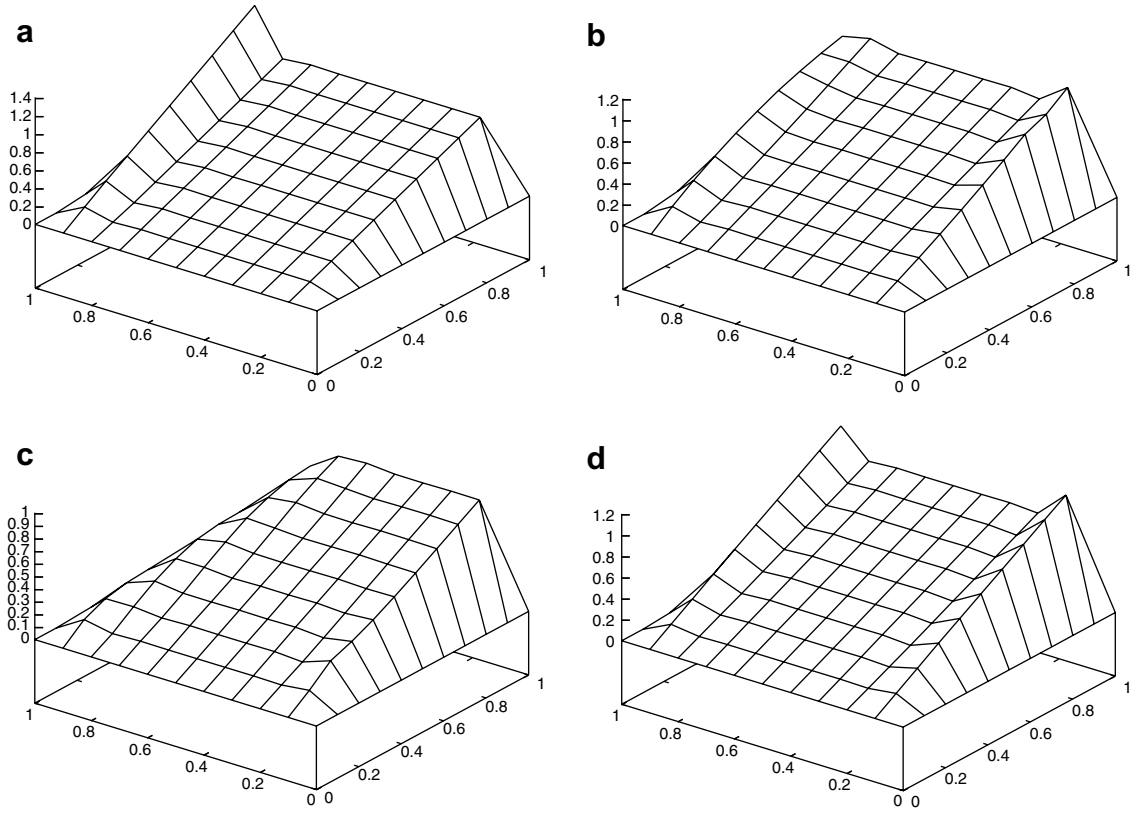

Fig. 4. Mizukami–Hughes discrete solution for data (11) with $\boldsymbol{b}$ replaced by the indicated vectors: (a) $\boldsymbol{b}=(1,-0.0001)$, (b) $\boldsymbol{b}=(1,-0.1)$, (c) $\boldsymbol{b}=(1,-0.4)$ and (d) $\boldsymbol{b}=(1,0)$, $\mathscr{T}_h$ from Fig. 3(b).

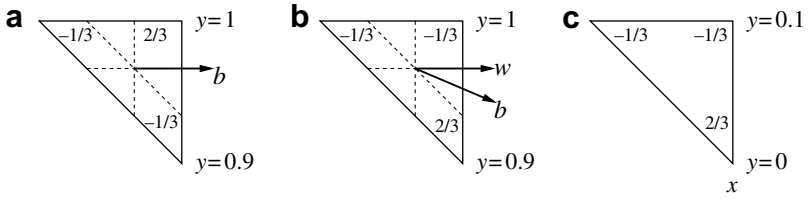

Fig. 5. Values of the constants C_i^K: (a) $\boldsymbol{b}=(1,0)$, (b) $\boldsymbol{b}=(1,\alpha)$, $\alpha\in(-1,0)$ and (c) optimal values.

$\{0.1,0.2,\ldots,0.9\}$, the constants C_i^K are equal to the values depicted in Fig. 5(c). Since $u_h(x,0)=0$, we have $\nabla u_h=(1,10x)$ and we can set $\boldsymbol{w}=(1,-0.1/x)$. Hence the Mizukami–Hughes method gives the optimal values of C_i^K's only if $x>0.1/|\alpha|$. Thus, for $\alpha=-0.1$, the discrete solution u_h is wrong along the whole lower part of the boundary of Ω (cf. Fig. 4(b)) whereas, for $\alpha=-0.4$, the values of C_i^K's are correct for elements with $x>0.25$ and hence u_h is better although still wrong (cf. Fig. 4(c)).

The problems observed above also appear for the data (11) if we consider a triangulation of Ω of the type depicted in Fig. 3(b) which consists of $4(N\times N)$ equal right-angled isosceles triangles ($N=5$ in Fig. 3(b)). For $N=10$, the discrete solution corresponding to the Mizukami–Hughes method is shown in Fig. 4(d) and, as wee see, it is wrong (the solution is visualized using its values at the same points as in Fig. 4(a)–(c)). For $\varepsilon=0$ and under the assumptions (A1) and (A2), the discrete solution satisfies (12) only if, on elements K with vertices $(x,0)$, $(x,0.1)$, $(x-0.05,0.05)$ or $(x,0.9)$, $(x,1)$, $(x-0.05,0.95)$ where $x\in\{0.1,0.2,\ldots,0.9\}$, we set $C_i^K=-\frac{1}{3}$ for i corresponding to $(x,0.1)$ or $(x,0.9)$. Whereas, for the examples mentioned above, we could think of redefining C_i^K's employing the relation between $\boldsymbol{b}$ and $\boldsymbol{w}$ in some more sophisticated way, now this is not possible since $\boldsymbol{w}=\boldsymbol{b}$. Moreover, the direction of ∇u_h on K also cannot be employed since it changes if $f=-1$ is used instead of $f=1$ whereas the values of C_i^K's have to remain the same.

In view of the above discussed and many other numerical experiments, we conclude that the definition of C_i^K's for $\boldsymbol{b}$ lying in an edge zone is not appropriate if K lies in the numerical boundary layer. The only remedy we have found is to set $C_i^K=-\frac{1}{3}$ for all i corresponding to inner vertices. This leads us to the following requirement:

(A3) $C_i^K=-\frac{1}{3}$ for all $i=1,2,3$ if $\overline{K}\cap\overline{\Gamma^{\mathrm{D}}}\neq\emptyset$ and if $\boldsymbol{b}\in\mathrm{EZ}_j^K$ for some $j\in\{1,2,3\}$.

Note that the constants C_i^K corresponding to vertices $a_i \in \overline{\Gamma^D}$ do not influence the discrete solution so that we could also define them in such a way that (3) is formally satisfied.

The requirement (A3) is not sufficient to avoid wrong discrete solutions on a triangulation of the type from Fig. 3(b) if $\boldsymbol{b} = (1, \alpha)$ with $\alpha \neq 0$. In this case we require that

(A3*) $C_i^K = -\frac{1}{3}$ for all $i = 1, 2, 3$ if all vertices of K are connected by edges to vertices on $\overline{\Gamma^D}$ and if $\boldsymbol{b} \in \mathrm{EZ}_j^K$ for some $j \in \{1, 2, 3\}$.

For $\boldsymbol{b} = (1, 0)$, this stronger requirement is not needed on a triangulation of the type from Fig. 3(b): for $N = 10$, there exists a unique $u_h \in V_h$ satisfying (12) and such that $\boldsymbol{b} \cdot \nabla u_h = f$ on any element of $\mathcal{T}_h$ with vertices of the type (x, y), $(x, y + 0.1)$, $(x + 0.05, y + 0.05)$ and on any element having an edge on the boundary of $(0, 0.9) \times (0.1, 0.9)$. Assuming (A3), it is easy to verify that this u_h solves the discrete problem with $\varepsilon = 0$. However, generally, (A3*) is a necessary condition for obtaining a nodally exact solution.

4. Continuous dependence of C_i^K's on the orientation of the convection $\boldsymbol{b}$

Let us consider the situation depicted in Fig. 5(a). Since $\boldsymbol{b}$ lies in a vertex zone, the values of the constants C_i^K are independent of the discrete solution u_h. Now, like in the preceding section, let us change $\boldsymbol{b}$ to $\boldsymbol{b} = (1, \alpha)$ with $\alpha < 0$, $|\alpha| \ll 1$. Then $\boldsymbol{b}$ lies in an edge zone which we denote EZ_1 and the constants C_i^K are determined according to (4)–(6). Assuming that both V_2 and V_3 are nonempty, the formula (8) replaces the value $\frac{2}{3}$ in Fig. 5(a) by $\frac{1+\alpha}{3}$ and the value $-\frac{1}{3}$ at the vertex with $y = 0.9$ by $-\frac{\alpha}{3}$. Thus, the definition of the constants C_i^K is discontinuous with respect to the orientation of $\boldsymbol{b}$. This does not seem to be reasonable and our numerical experiments show that it may deteriorate the quality of the discrete solution. Therefore, in this section, we propose another way how to compute the constants C_i^K in case (6).

Let us again consider an element K with vertices a_1, a_2 and a_3. If $\boldsymbol{b} \in \mathrm{VZ}_2$, then $C_2^K = \frac{2}{3}$, $C_3^K = -\frac{1}{3}$ and, if $\boldsymbol{b} \in \mathrm{VZ}_3$, then $C_2^K = -\frac{1}{3}$, $C_3^K = \frac{2}{3}$. Thus, if $\boldsymbol{b} \in \mathrm{EZ}_1$, it is sensible to set

$$C_2^K = F\left(\frac{\alpha_3}{\alpha_2 + \alpha_3}\right), \quad C_3^K = F\left(\frac{\alpha_2}{\alpha_2 + \alpha_3}\right),$$

where α_2 and α_3 are the angles depicted in Fig. 6 and $F : [0, 1] \to \left[-\frac{1}{3}, \frac{2}{3}\right]$ is a continuous monotone function satisfying $F(0) = -\frac{1}{3}$ and $F(1) = \frac{2}{3}$. It is convenient to replace F by the function

$$G(x) = 2F\left(\frac{x+1}{2}\right) - \frac{1}{3}.$$

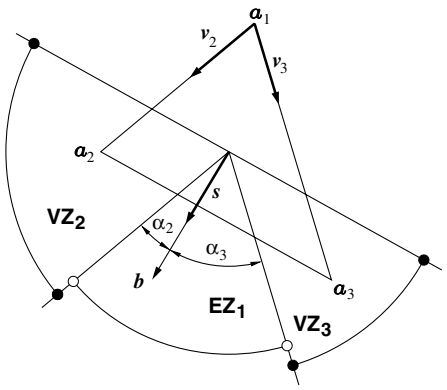

Fig. 6. Definition of angles α_2 and α_3 and of vectors $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ and $\boldsymbol{s}$.

Then

$$C_2^K = \frac{1}{6} + \frac{1}{2} G\left(\frac{\alpha_3 - \alpha_2}{\alpha_2 + \alpha_3}\right), \quad C_3^K = \frac{1}{6} + \frac{1}{2} G\left(\frac{\alpha_2 - \alpha_3}{\alpha_2 + \alpha_3}\right)$$

and G is a continuous monotone function satisfying

$$G : [-1, 1] \to [-1, 1], \quad G(-1) = -1, \quad G(1) = 1. \tag{14}$$

Moreover, (6) implies that G is odd.

To make the computation of the constants C_i^K cheaper, we use the approximation

$$\frac{\alpha_3 - \alpha_2}{\alpha_2 + \alpha_3} \approx \frac{\sin\left[\frac{1}{2}(\alpha_3 - \alpha_2)\right]}{\sin\left[\frac{1}{2}(\alpha_2 + \alpha_3)\right]} = \frac{\cos\alpha_2 - \cos\alpha_3}{1 - \cos(\alpha_2 + \alpha_3)},$$

which is certainly acceptable for $\alpha_2 + \alpha_3 \leqslant \frac{\pi}{2}$. Note that, denoting by $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ unit vectors pointing from a_1 to a_2 and a_3, respectively, and by $\boldsymbol{s}$ the unit vector in the direction of $\boldsymbol{b}$ (cf. Fig. 6), we have

$$\frac{\cos\alpha_2 - \cos\alpha_3}{1 - \cos(\alpha_2 + \alpha_3)} = \frac{(\boldsymbol{v}_2 - \boldsymbol{v}_3) \cdot \boldsymbol{s}}{1 - \boldsymbol{v}_2 \cdot \boldsymbol{v}_3}.$$

Thus, we arrive at the formulas

$$C_2^K = \frac{1}{6} + \frac{1}{2} G\left(\frac{(\boldsymbol{v}_2 - \boldsymbol{v}_3) \cdot \boldsymbol{s}}{1 - \boldsymbol{v}_2 \cdot \boldsymbol{v}_3}\right), \quad C_3^K = \frac{1}{3} - C_2^K, \tag{15}$$

where G is a continuous monotone odd function satisfying (14). We performed a lot of numerical experiments which revealed that a good choice for the function G is to simply set

$$G(x) = x.$$

5. Continuous dependence of C_i^K's on the orientation of ∇u_h

Let us consider the situation depicted in Fig. 2, i.e., $\boldsymbol{b} \in \mathrm{EZ}_1$. If the vector $\boldsymbol{w}$ points from the barycentre of K into the part of EZ_1 marked by the dashed arc, then the constants C_i^K are determined by (4) and hence $C_2^K = \frac{2}{3}$ and $C_3^K = -\frac{1}{3}$. However, as soon as $\boldsymbol{w}$ comes into the interior of VZ_3, the values of these constants change to values given

by (15). Consequently, the constants C_i^K depend on the orientation of $\boldsymbol{w}$ (and hence of ∇u_h) in a discontinuous way. Our numerical experiences show that, in some cases, this prevents the nonlinear iterative process from converging. Therefore, in the following, we describe a modification of the formula (15) taking into account the orientation of $\boldsymbol{w}$. We assume that $\boldsymbol{b}\cdot\nabla u_h|_K \neq 0$.

We shall need some additional notation which is introduced in Fig. 7. Here, the straight dashed lines are axes of the angles between the two lines which cross at the barycentre of K and are parallel to the edges of K containing the vertex a_1. One of these angles is the same as the angle ω_1 of K at a_1 and we introduce a unit vector $\boldsymbol{v}$ in the direction of the axis of this angle pointing as in Fig. 7. Without loss of generality, we may assume that $|\boldsymbol{w}| = 1$ and that $\boldsymbol{w}\cdot\boldsymbol{v} \geqslant 0$. Therefore, the dashed and dotted arcs in Fig. 7, which have the same meaning as in Fig. 2, are restricted to the corresponding half plane. We denote by δ the angle between $\boldsymbol{w}$ and the part of the boundary of EZ_1 which is 'nearer' to $\boldsymbol{w}$ (cf. Fig. 7). Like in Fig. 6, we introduce the angles α_2 and α_3 and the unit vectors $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ and $\boldsymbol{s}$.

If $\boldsymbol{w} \in \overline{\mathrm{EZ}}_1$, the constants C_i^K are uniquely determined by (4) and (5). Thus, let us consider the case (6) and let $j,k \in \{2,3\}$, $j \neq k$, be such that $\boldsymbol{w} \in \mathrm{VZ}_j \cup \mathrm{EZ}_k$ ($j = 3$ in Fig. 7). It suffices to discuss the choice of C_j^K since $C_1^K = -\frac{1}{3}$ and $C_k^K = \frac{1}{3} - C_j^K$. Obviously, $\alpha_j \in (0,\omega_1)$ and $\delta \in (0,\kappa]$ with $\kappa = \frac{\pi}{2} - \frac{\omega_1}{2}$. We shall require the following values of C_j^K in the limit cases:

$\delta = \kappa \Rightarrow C_j^K$ is determined by (15),

$\alpha_j \to 0,\ \delta \nrightarrow 0 \Rightarrow C_j^K$ is determined by (15) ($\Rightarrow C_j^K \to \frac{2}{3}$),

$\delta \to 0,\ \alpha_j \nrightarrow 0 \Rightarrow C_j^K \to -\frac{1}{3}$.

If $\alpha_j \to 0$, $\delta \to 0$, then $\boldsymbol{b}\cdot\nabla u_h|_K \approx 0$ and hence the choice of C_i^K's is not important since $A^K U \approx 0$ in (7). Denoting by $\overline{C}_j^K$ the value of C_j^K determined by (15), we set

$$C_j^K = \overline{C}_j^K \Phi(\alpha_j,\delta) - \tfrac{1}{3}[1 - \Phi(\alpha_j,\delta)],$$

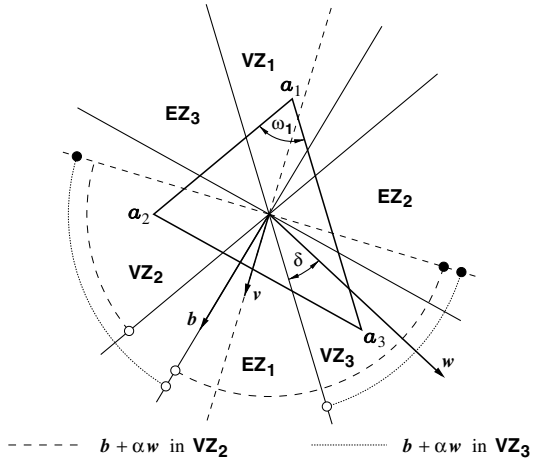

Fig. 7. Definition of angles ω_1 and δ and of the vector $\boldsymbol{v}$.

where $\Phi:([0,\omega_1]\times[0,\kappa])\backslash(0,0) \to [0,1]$ is a continuous function. The above requirements imply that

$$\Phi(\alpha_j,\kappa) = \Phi(0,\delta) = 1, \quad \Phi(\alpha_j,0) = 0 \quad \forall \alpha_j \in (0,\omega_1],\ \delta \in (0,\kappa].$$

Since the direction of $\boldsymbol{w}$ may strongly vary during the nonlinear iterative process, the constant C_j^K should be mainly determined by (15) and the orientation of $\boldsymbol{w}$ should influence C_j^K only if δ/κ is smaller than α_j/ω_1. Therefore, we set

$$\Phi(\alpha_j,\delta) = \min\left\{1, \frac{2\sin\delta}{r_j \sin\kappa}\right\}, \tag{16}$$

where

$$r_j = \begin{cases} \dfrac{\sin\alpha_j}{\sin\frac{\omega_1}{2}} & \text{if } \alpha_j < \dfrac{\omega_1}{2}, \\ 1 & \text{if } \alpha_j \geqslant \dfrac{\omega_1}{2}. \end{cases}$$

Of course, many other formulas for $\Phi(\alpha_j,\delta)$ can also be used. Let us mention that the computation of (16) is inexpensive since, denoting by $\boldsymbol{v}_j^\perp$ a unit vector orthogonal to $\boldsymbol{v}_j$, we have

$$\sin\kappa = \boldsymbol{v}\cdot\boldsymbol{v}_j, \quad \sin\frac{\omega_1}{2} = |\boldsymbol{v}\cdot\boldsymbol{v}_j^\perp|, \quad \sin\delta = |\boldsymbol{w}\cdot\boldsymbol{v}_j^\perp|,$$
$$\sin\alpha_j = |\boldsymbol{s}\cdot\boldsymbol{v}_j^\perp|.$$

Remark 1. The dependence of the constants C_i^K on $\boldsymbol{w}$ is also discontinuous if the orientation of $\boldsymbol{w}$ passes the direction of $\boldsymbol{s}$ (i.e., of $\boldsymbol{b}$). However, this does not seem to be important since, if $\boldsymbol{w} \approx \boldsymbol{s}$, we have $\boldsymbol{b}\cdot\nabla u_h|_K \approx 0$ and hence $A^K U \approx 0$ in (7).

6. Summary of the improved Mizukami–Hughes method

In this section we summarize the definitions of the constants C_i^K introduced in the previous sections. Let us consider any element $K \in \mathscr{T}_h$ and let a_1, a_2 and a_3 be its vertices. If $\boldsymbol{b} \neq \boldsymbol{0}$, we assume that $\boldsymbol{b}$ points into the vertex zone or the edge zone of a_1 (cf. Fig. 1) and we denote

$$\boldsymbol{s} = \frac{\boldsymbol{b}}{|\boldsymbol{b}|}, \quad \boldsymbol{v}_2 = \frac{a_2 - a_1}{|a_2 - a_1|}, \quad \boldsymbol{v}_3 = \frac{a_3 - a_1}{|a_3 - a_1|}, \quad \boldsymbol{v} = \frac{\boldsymbol{v}_2 + \boldsymbol{v}_3}{|\boldsymbol{v}_2 + \boldsymbol{v}_3|}.$$

Further, we introduce unit vectors $\boldsymbol{w}$, $\boldsymbol{v}^\perp$, $\boldsymbol{v}_2^\perp$ and $\boldsymbol{v}_3^\perp$ such that

$$\boldsymbol{w}\cdot\nabla u_h|_K = 0, \quad \boldsymbol{v}^\perp\cdot\boldsymbol{v} = 0, \quad \boldsymbol{v}_2^\perp\cdot\boldsymbol{v}_2 = 0, \quad \boldsymbol{v}_3^\perp\cdot\boldsymbol{v}_3 = 0,$$
$$\boldsymbol{w}\cdot\boldsymbol{v} \geqslant 0, \quad \boldsymbol{v}^\perp\cdot\boldsymbol{v}_3 \geqslant 0.$$

Finally, we recall the spaces V_2 and V_3 introduced in Section 2. Then the constants C_1^K, C_2^K and C_3^K are determined according to the algorithm in Fig. 8. It is obvious that the improved method preserves the general properties of the original Mizukami–Hughes method, particularly, it satisfies the discrete maximum principle discussed at the end of Section 2.

IF $\boldsymbol{b} = \boldsymbol{0}$ THEN

$$C_1^K = C_2^K = C_3^K = 0$$

ELSE IF $\boldsymbol{b} \in \mathrm{VZ}_1$ THEN

$$C_1^K = \tfrac{2}{3}, \qquad C_2^K = C_3^K = -\tfrac{1}{3}$$

ELSE IF $\overline{K} \cap \overline{\Gamma^D} \neq \emptyset$ THEN

$$C_1^K = C_2^K = C_3^K = -\tfrac{1}{3}$$

ELSE IF $\mathscr{T}_h$ is not of the type from Fig. 3(a) and all vertices of K are connected by edges to vertices on $\overline{\Gamma^D}$ THEN

$$C_1^K = C_2^K = C_3^K = -\tfrac{1}{3}$$

ELSE IF $\boldsymbol{b} \cdot \nabla u_h|_K = 0$ THEN

$$C_1^K = -\tfrac{1}{3}, \quad C_2^K = C_3^K = \tfrac{1}{6}$$

ELSE IF $V_2 \neq \emptyset$ & $V_3 = \emptyset$ THEN

$$C_2^K = \tfrac{2}{3}, \quad C_1^K = C_3^K = -\tfrac{1}{3}$$

ELSE IF $V_2 = \emptyset$ & $V_3 \neq \emptyset$ THEN

$$C_3^K = \tfrac{2}{3}, \quad C_1^K = C_2^K = -\tfrac{1}{3}$$

ELSE IF $\boldsymbol{w} \cdot \boldsymbol{v}^\perp < 0$ THEN

$$r_2 = \min\left\{1, \frac{|\boldsymbol{s} \cdot \boldsymbol{v}_2^\perp|}{|\boldsymbol{v} \cdot \boldsymbol{v}_2^\perp|} + 1 - \operatorname{sgn}(\boldsymbol{b} \cdot \boldsymbol{v}_2)\right\},$$

$$\Phi = \min\left\{1, \frac{2\,|\boldsymbol{w} \cdot \boldsymbol{v}_2^\perp|}{r_2\, \boldsymbol{v} \cdot \boldsymbol{v}_2}\right\},$$

$$C_2^K = -\frac{1}{3} + \frac{1}{2}\,\Phi\left[1 + \frac{(\boldsymbol{v}_2 - \boldsymbol{v}_3) \cdot \boldsymbol{s}}{1 - \boldsymbol{v}_2 \cdot \boldsymbol{v}_3}\right],$$

$$C_3^K = \frac{1}{3} - C_2^K, \qquad C_1^K = -\tfrac{1}{3}$$

ELSE

$$r_3 = \min\left\{1, \frac{|\boldsymbol{s} \cdot \boldsymbol{v}_3^\perp|}{|\boldsymbol{v} \cdot \boldsymbol{v}_3^\perp|} + 1 - \operatorname{sgn}(\boldsymbol{b} \cdot \boldsymbol{v}_3)\right\},$$

$$\Phi = \min\left\{1, \frac{2\,|\boldsymbol{w} \cdot \boldsymbol{v}_3^\perp|}{r_3\, \boldsymbol{v} \cdot \boldsymbol{v}_3}\right\},$$

$$C_3^K = -\frac{1}{3} + \frac{1}{2}\,\Phi\left[1 + \frac{(\boldsymbol{v}_3 - \boldsymbol{v}_2) \cdot \boldsymbol{s}}{1 - \boldsymbol{v}_2 \cdot \boldsymbol{v}_3}\right],$$

$$C_2^K = \frac{1}{3} - C_3^K, \qquad C_1^K = -\tfrac{1}{3}.$$

Fig. 8. Definition of the constants C_i^K in the improved Mizukami–Hughes method.

7. Numerical results

In this section we demonstrate the properties of the improved Mizukami–Hughes method by means of several standard test problems formulated in Examples 1–7 below and taken (in a slightly modified form) from [9,12,14]. In all these examples we consider $\varepsilon = 10^{-7}$ and, except for Example 6, $\Omega = (0,1)^2$. Unless otherwise specified, we use a triangulation of the type depicted in Fig. 3(a). The number of elements will be determined by the parameter N introduced at the beginning of Section 3. In Examples 1–3, the convection vector $\boldsymbol{b}$ is defined using an angle θ which is assumed to satisfy $\theta \in (0, \pi/2)$. To simplify the definitions of various parts of $\partial\Omega$, we introduce the sets

$$\Gamma_1 = (\{0\} \times (0,1]) \cup ([0,1) \times \{1\}),$$
$$\Gamma_2 = (\{0\} \times (0.7,1]) \cup ([0,1) \times \{1\}).$$

In the captions of figures we denote by MH the original Mizukami–Hughes method [14] and by IMH the improved Mizukami–Hughes method introduced in this paper. Let us mention that the discrete solutions obtained using the SUPG method [2] contain spurious oscillations for all the examples except for Example 6.

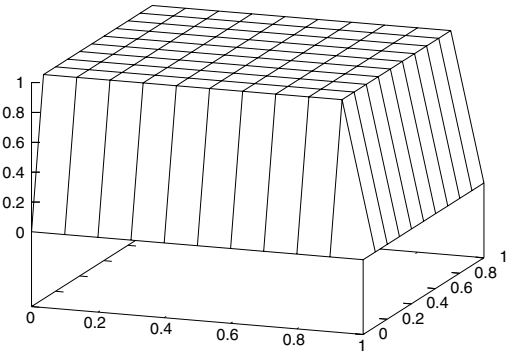

Fig. 9. Example 1, IMH, $N = 10$.

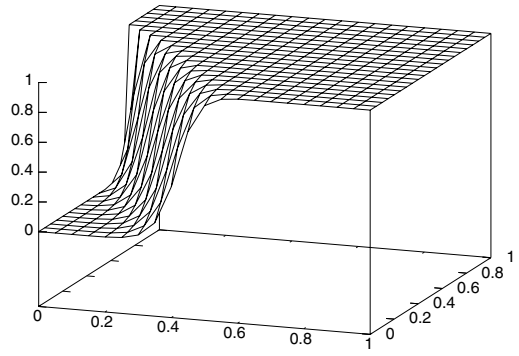

Fig. 10. Example 2, $\theta = \frac{\pi}{3}$, IMH, $N = 20$.

Example 1 (*Convection skew to the mesh with boundary layers*)

$$\boldsymbol{b} = (\cos\theta, -\sin\theta), \quad f = 0, \quad \Gamma^{\mathrm{D}} = \partial\Omega,$$
$$u_{\mathrm{b}} = 1 \quad \text{on } \Gamma_1, \quad u_{\mathrm{b}} = 0 \quad \text{on } \Gamma^{\mathrm{D}} \setminus \Gamma_1.$$

Both the original and the improved Mizukami–Hughes method give the same discrete solution which is nodally exact (cf. Fig. 9). This easily follows from the definition of the constants C_i^K. However, for $\theta \neq \pi/4$, it is rather difficult to compute the discrete solution of the original Mizukami–Hughes method due to the discontinuous dependence of C_i^K's on the orientation of ∇u_h. On the other hand, the computation of the discrete solution of the improved Mizukami–Hughes method needs only a few nonlinear iterations.

Example 2 (*Convection skew to the mesh with an inner layer*)

$$\boldsymbol{b} = (\cos\theta, -\sin\theta), \quad f = 0, \quad g = 0, \quad \Gamma^{\mathrm{D}} = \Gamma_1,$$
$$u_{\mathrm{b}} = 1 \quad \text{on } \Gamma_2, \quad u_{\mathrm{b}} = 0 \quad \text{on } \Gamma^{\mathrm{D}} \setminus \Gamma_2.$$

For $\theta = \pi/4$, the vector $\boldsymbol{b}$ points into vertex zones in all elements of the triangulation and it is easy to see that, for both methods, the discrete solution is constant along the diagonals in Fig. 3(a) if $\varepsilon \to 0$. Consequently, both the original and the improved Mizukami–Hughes method give the same nodally exact discrete solution. If $\theta \neq \pi/4$, the discrete solutions are not nodally exact but they are similar for both methods. Fig. 10 shows the discrete solution for $\theta = \pi/3$ and $N = 20$ obtained using the improved Mizukami–Hughes method. Fig. 11 compares the outflow profiles along the x-axis for the two methods and the exact solution of the hyperbolic limit of (1). The solution of the improved method seems to be slightly better. Like for the previous example, the discrete solution is much more difficult to compute for the original Mizukami–Hughes method.

Example 3 (*Convection skew to the mesh with inner and boundary layers*)

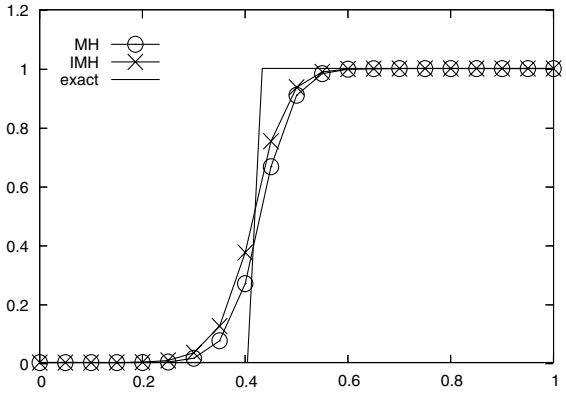

Fig. 11. Example 2, $\theta = \frac{\pi}{3}$, $N = 20$, MH, IMH and exact solution on $y = 0$.

$$\boldsymbol{b} = (\cos\theta, -\sin\theta), \quad f = 0, \quad \Gamma^{\mathrm{D}} = \partial\Omega,$$
$$u_{\mathrm{b}} = 1 \quad \text{on } \Gamma_2, \quad u_{\mathrm{b}} = 0 \quad \text{on } \Gamma^{\mathrm{D}} \setminus \Gamma_2.$$

This test problem is more complicated than the previous one since, in addition to the inner layer, it also involves one or two boundary layers. The relation between the original and the improved Mizukami–Hughes method is similar as in the previous example. Fig. 12 shows the discrete solution obtained using the improved Mizukami–Hughes method for $\theta = \pi/3$ and $N = 20$.

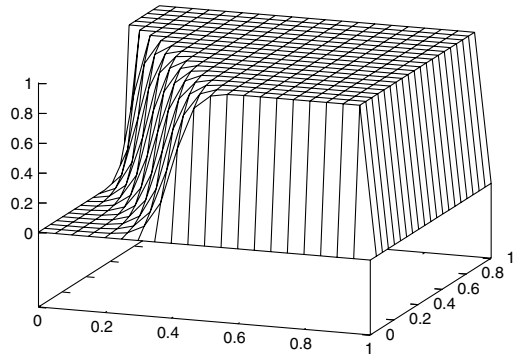

Fig. 12. Example 3, $\theta = \frac{\pi}{3}$, IMH, $N = 20$.

Example 4 (*Convection with a constant source term*)

$$\boldsymbol{b} = (1, \alpha), \quad \alpha \in (-0.5, 0.5), \quad f = 1, \quad \Gamma^{\mathrm{D}} = \partial\Omega, \quad u_{\mathrm{b}} = 0.$$

This problem was already considered in Section 3 where we have seen that the original Mizukami–Hughes method gives wrong discrete solutions (cf. Fig. 4). As we see from Fig. 13, the discrete solutions of the improved Mizukami–Hughes method seem to be correct in all cases considered in Section 3. Moreover, the improved Mizukami–Hughes method gives the discrete solutions shown in Fig. 13(a)–(c) also if we use a triangulation of the type depicted in Fig. 3(b).

Example 5 (*Convection with a nonconstant source term*)

$$\boldsymbol{b} = (1, \alpha), \quad \alpha \in (-0.5, 0.5), \quad \Gamma^{\mathrm{D}} = \partial\Omega, \quad u_{\mathrm{b}} = 0,$$
$$f = 1 \quad \text{in } (0, \tfrac{1}{2}) \times (0, 1), \quad f = -1 \quad \text{in } (\tfrac{1}{2}, 1) \times (0, 1).$$

Like in the previous example, both methods coincide and give a nodally exact solution for $\alpha = 0$ and a triangulation of the type depicted in Fig. 3(a). This is no longer true if we use $\alpha \neq 0$ or a triangulation of the type depicted in Fig. 3(b). Figs. 14 and 15 demonstrate that the original Mizukami–Hughes method generally gives wrong discrete solutions whereas the solutions of the improved Mizukami–Hughes method seem to be correct.

Example 6 (*Donut problem*). We consider $\Omega = (0,1)^2 \backslash \Gamma$ with $\Gamma = \{\frac{1}{2}\} \times (0, \frac{1}{2})$. The convection field $\boldsymbol{b}$ is defined by

$$\boldsymbol{b}(x, y) = (-y + \tfrac{1}{2}, x - \tfrac{1}{2})$$

so that it represents a vortex around the midpoint of the unit square in the counter-clockwise direction. Therefore, Γ represents an inflow boundary denoted by Γ^{in} if we approach Γ from the right but it also represents an outflow boundary denoted by Γ^{out} if we approach it from the left. We set

$$f = 0, \quad g = 0, \quad \Gamma^{\mathrm{D}} = \Gamma^{\mathrm{in}} \cup \partial[(0,1)^2], \quad \Gamma^{\mathrm{N}} = \Gamma^{\mathrm{out}},$$
$$u_{\mathrm{b}} = 0 \quad \text{on } \Gamma^{\mathrm{D}} \setminus \Gamma^{\mathrm{in}}, \quad u_{\mathrm{b}}(\tfrac{1}{2}, y) = \sin(\pi(1 - 2y)) \quad \text{for } y \in (0, \tfrac{1}{2}).$$

For this problem, an almost nodally exact discrete solution can be obtained using the SUPG method and it is interesting to see to what extent the discrete solution deteriorates if other stabilized methods are used. The solution of the improved Mizukami–Hughes discretization is shown in Fig. 16 and is similar to the solution obtained using the original Mizukami–Hughes method. Fig. 17 shows a comparison of the discrete solutions of the two Mizukami–Hughes methods and the exact solution of the hyperbolic limit of (1) by means of cuts through graphs of the solutions along the line $x = 1/2$. It seems that the improved

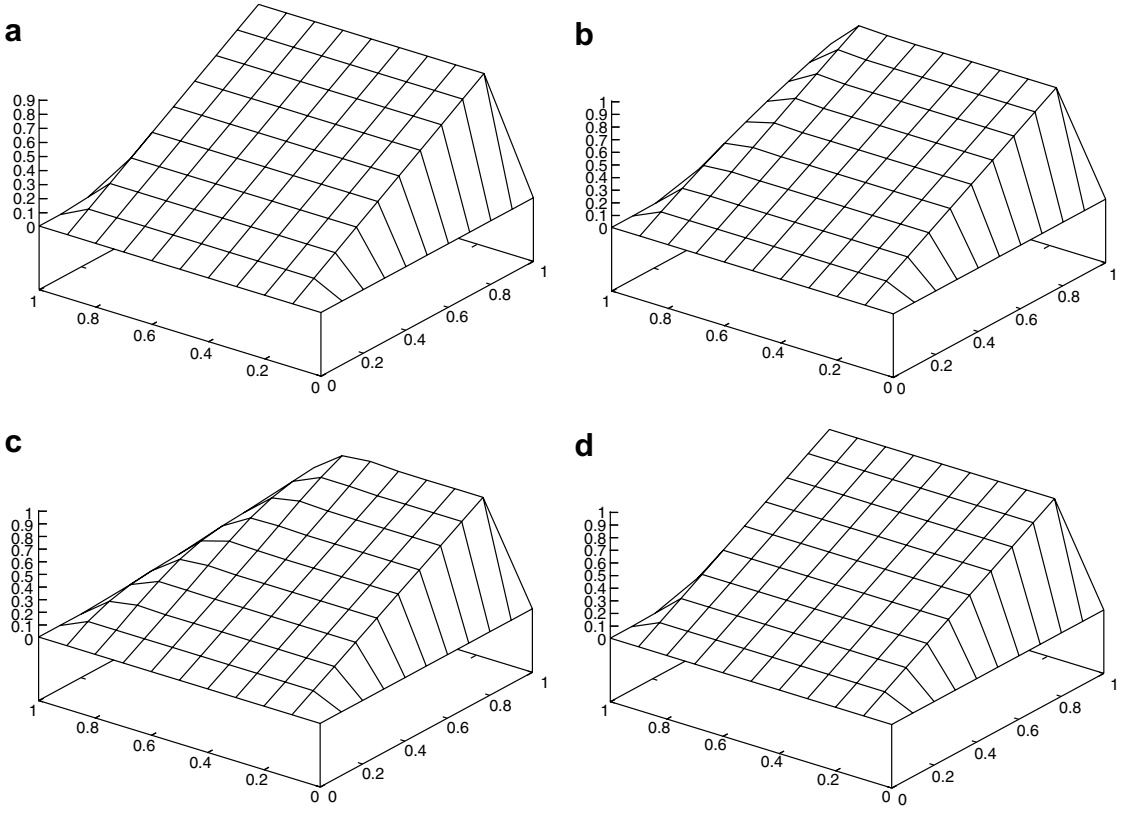

Fig. 13. Example 4, IMH, $N = 10$: (a) $\alpha = -0.0001$, (b) $\alpha = -0.1$, (c) $\alpha = -0.4$ and (d) $\alpha = 0$, $\mathscr{T}_h$ from Fig. 3(b).

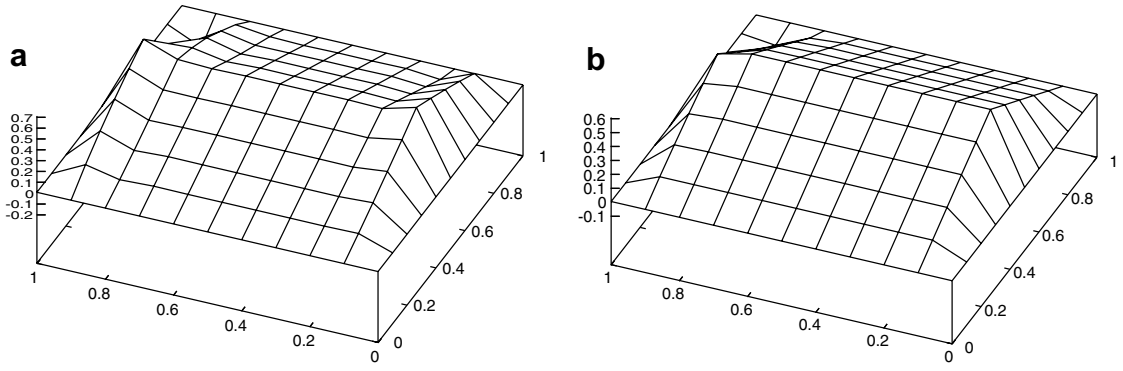

Fig. 14. Example 5, $\alpha = -0.1$, $N = 10$: (a) MH and (b) IMH.

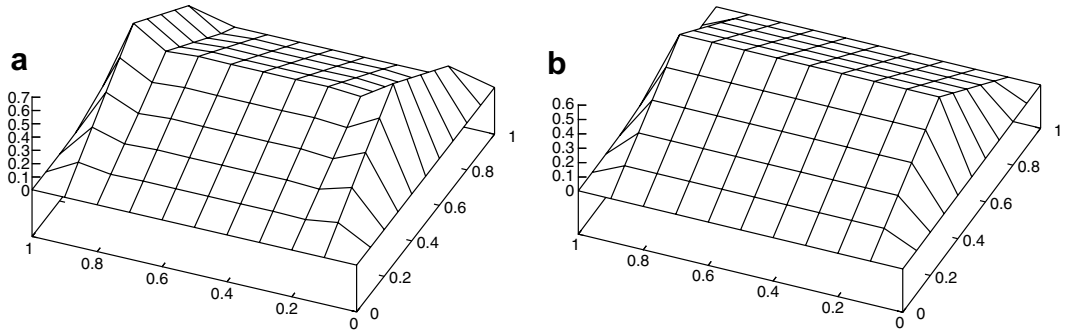

Fig. 15. Example 5, $\alpha = 0$, $\mathscr{T}_h$ from Fig. 3(b), $N = 10$: (a) MH and (b) IMH.

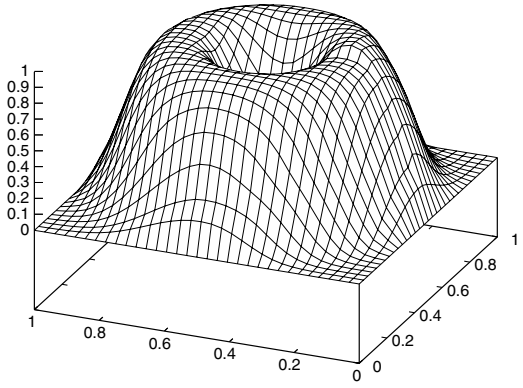

Fig. 16. Example 6, IMH, $N = 32$.

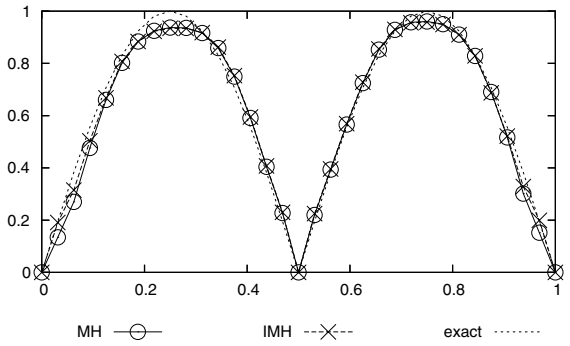

Fig. 17. Example 6, $N = 32$, MH, IMH and exact solution on $x = 1/2$.

Mizukami–Hughes method gives a slightly better solution. The two discrete solutions are comparable with best discrete solutions obtained using discontinuity-capturing methods mentioned in the introduction.

Example 7 (*Problem with known exact solution*)

$$\boldsymbol{b} = (2,3), \quad \Gamma^{\mathrm{D}} = \partial\Omega.$$

The functions f and u_{b} are chosen in such a way that

$$u(x,y) = xy^2 - y^2 \exp\left(\frac{2(x-1)}{\varepsilon}\right) - x\exp\left(\frac{3(y-1)}{\varepsilon}\right) + \exp\left(\frac{2(x-1)+3(y-1)}{\varepsilon}\right)$$

is the exact solution of (1) and (2).

The function u contains two typical exponential boundary layers and hence this example represents a suitable tool for gauging the accuracy of numerical methods for the solution of convection–diffusion problems. The discrete solution obtained using the improved Mizukami–Hughes method for $N = 20$ can be seen in Fig. 18. Fig. 19 shows the discrete solution computed using the SUPG method [2] with the so-called optimal definition of the stabilization parameter and element size defined as the element diameter in the direction of the flow. We consider the SUPG method here since it is known to approximate solutions with layers on non-layer-adapted meshes at least outside the layers

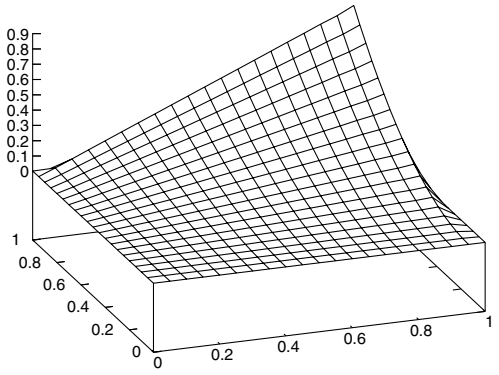

Fig. 18. Example 7, IMH, $N = 20$.

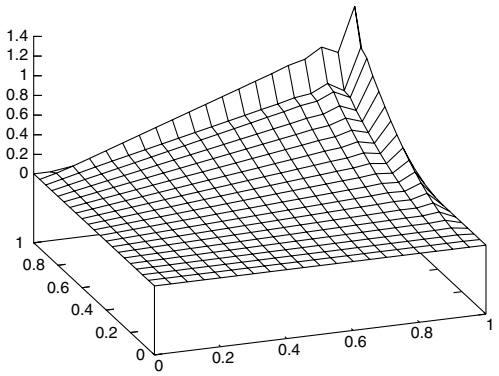

Fig. 19. Example 7, SUPG, $N = 20$.

very precisely. Therefore, it is interesting to compare the accuracy of the SUPG method with the accuracy of the improved Mizukami–Hughes method. We measure the errors of the discrete solutions by means of the norm in $L^2(\Omega)$ denoted by $\|\cdot\|_{0,\Omega}$ and using the discrete L^∞ norm on Ω denoted by $\|\cdot\|_{0,\infty,h}$ and defined as the maximum of the absolute values of errors at vertices of the triangulation. In addition, we consider this type of norms on the domain $\Omega^* \equiv (0, 0.8)^2$ which does not contain a neighborhood of the layers. The respective norms are denoted by $\|\cdot\|^*_{0,\Omega}$ and $|\cdot|^*_{0,\infty,h}$. Finally, we evaluate the $H^1(\Omega^*)$ seminorm denoted by $|\cdot|^*_{1,\Omega}$. Because of the boundary layers it makes no sense to show the $H^1(\Omega)$ seminorm. Like in the previous examples, the bilinear forms of the discrete problems were computed exactly whereas the right-hand sides were evaluated using quadrature formulas which are exact for piecewise cubic f. The evaluation of the L^2 norms (respectively, the H^1 seminorm) was exact for piecewise quadratic (respectively, cubic) functions. The obtained results are shown in Tables 1 and 2 and we see that, outside the layers, both methods converge with optimal convergence orders. On fine meshes, the SUPG method is more precise in Ω^* than the modified Mizukami–Hughes method, particularly, with respect to the discrete L^∞ norm. However, on the whole domain Ω, the SUPG solution does not converge in the discrete L^∞ norm since the magnitude

Table 1
Example 7, errors of the improved Mizukami–Hughes method

N	$\|\cdot\|_{0,\Omega}$	$\|\cdot\|_{0,\infty,h}$	$\|\cdot\|^*_{0,\Omega}$	$\|\cdot\|^*_{1,\Omega}$	$\|\cdot\|^*_{0,\infty,h}$
20	$5.91-2$	$7.02-3$	$3.68-4$	$2.05-2$	$2.15-3$
40	$4.20-2$	$3.93-3$	$1.13-4$	$1.02-2$	$6.71-4$
80	$2.98-2$	$2.07-3$	$3.14-5$	$5.06-3$	$1.87-4$
160	$2.11-2$	$1.05-3$	$8.30-6$	$2.52-3$	$4.94-5$
Order	0.50	0.98	1.92	1.01	1.92

Table 2
Example 7, errors of the SUPG method

N	$\|\cdot\|_{0,\Omega}$	$\|\cdot\|_{0,\infty,h}$	$\|\cdot\|^*_{0,\Omega}$	$\|\cdot\|^*_{1,\Omega}$	$\|\cdot\|^*_{0,\infty,h}$
20	$4.91-2$	$5.08-1$	$3.33-4$	$2.49-2$	$9.37-3$
40	$3.51-2$	$5.70-1$	$3.95-5$	$1.00-2$	$2.32-4$
80	$2.50-2$	$6.02-1$	$9.80-6$	$4.99-3$	$7.06-6$
160	$1.78-2$	$6.18-1$	$2.45-6$	$2.49-3$	$1.74-6$
Order	0.49	−0.04	2.00	1.00	2.02

of the spurious oscillations visible in Fig. 19 does not decrease for decreasing h as long as h is significantly larger than the width of the boundary layers. On the other hand, the solution of the modified Mizukami–Hughes method converges on the whole domain Ω with first order of accuracy in the discrete L^∞ norm and does not contain any spurious oscillations as we can also see from Fig. 18.

8. Application of the Mizukami–Hughes method to convection–diffusion–reaction equations

In this section we extend the Mizukami–Hughes method described in the preceding sections to convection–diffusion–reaction equations

$$-\varepsilon\Delta u + \boldsymbol{b}\cdot\nabla u + cu = f \quad \text{in } \Omega, \tag{17}$$

where c is a given function. Our aim again is to derive a numerical method satisfying the discrete maximum principle and hence we shall assume that $c \geqslant 0$ since otherwise no maximum principle generally holds for Eq. (17). Again, we consider the singularly perturbed case, i.e., $\varepsilon \ll |\boldsymbol{b}| + c$.

The discrete solution u_h of (17), (2) is defined by

$$\begin{aligned}
&u_h \in V_h,\\
&\varepsilon(\nabla u_h, \nabla\varphi_i) + (\boldsymbol{b}\cdot\nabla u_h + cu_h, \tilde{\varphi}_i)\\
&\quad = (f, \tilde{\varphi}_i) + (g, \varphi_i)_{\Gamma^N}, \quad i = 1, \ldots, M_h,\\
&u_h(a_i) = u_D(a_i), \quad i = M_h + 1, \ldots, N_h.
\end{aligned}$$

Like for $\boldsymbol{b}$, we assume that c is piecewise constant.

For any $K \in \mathcal{T}_h$, the local reaction matrix R^K has entries

$$r^K_{ij} = (c\varphi_j, \tilde{\varphi}_i)_K = \tfrac{1}{3}c|_K \mathrm{meas}_2(K)\left(\tfrac{1}{4} + C^K_i + \tfrac{1}{4}\delta_{ij}\right)$$

(with $i = 1, \ldots, M_h, j = 1, \ldots, N_h, a_i, a_j \in \overline{K}$), where δ_{ij} is the Kronecker symbol. We define the matrix $S^K \equiv A^K + R^K$ (with entries s^K_{ij}), where A^K is the local convection matrix

introduced in Section 2. Like before, we want to define the constants C_i^K in such a way that the matrix S^K is of nonnegative type or at least satisfies an analogue of (7), i.e.,

$$S^K U = \widetilde{S}^K U, \tag{18}$$

where $\widetilde{S}^K$ is a matrix of nonnegative type. Note that

$$\sum_{\substack{j=1\\ a_j\in\overline{K}}}^{N_h} s_{ij}^K = \sum_{\substack{j=1\\ a_j\in\overline{K}}}^{N_h} r_{ij}^K = c|_K \mathrm{meas}_2(K)\left(\tfrac{1}{3}+C_i^K\right)$$
$$\forall i\in\{1,\dots,M_h\},\quad a_i\in\overline{K},$$

and hence the first condition in (3) is necessary for S^K to be of nonnegative type.

On the other hand, the second condition in (3) cannot be fulfilled in general. To see this, let us denote the vertices of K by a_1, a_2, a_3 and let us assume that $\boldsymbol{b}\in \mathrm{VZ}_1$ (from now on we shall write $\boldsymbol{b}$, c, $\nabla\varphi_i$ instead of $\boldsymbol{b}|_K$, $c|_K$, $\nabla\varphi_i|_K$, respectively). Then s_{21}^K and s_{31}^K may be nonpositive only if $C_2^K < -\frac{1}{4}$ and $C_3^K < -\frac{1}{4}$. A necessary condition for s_{12}^K and s_{13}^K to be nonpositive is

$$\tfrac{2}{3}c\left(\tfrac{1}{4}+C_1^K\right) \leqslant \boldsymbol{b}\cdot\nabla\varphi_1\left(\tfrac{1}{3}+C_1^K\right).$$

If $C_1^K \in (\frac{1}{2},\frac{2}{3}]$, which is necessary for the validity of (3), this inequality will not be satisfied for $c \geqslant 2\boldsymbol{b}\cdot\nabla\varphi_1$. Hence the validity of the second condition in (3) cannot be generally required.

Fortunately, the second condition in (3) is not needed to assure that (7) holds with a matrix $\widetilde{A}^K$ of nonnegative type. It is easy to check that (7) still holds if those constants in the definition of A^K, for which larger values than $-\frac{1}{3}$ are prescribed, are replaced by any values from the interval $[-\frac{1}{3},\infty)$. Thus, our idea is first to compute the constants C_i^K according to the algorithm in Fig. 8 and then possibly to decrease some of the constants in such a way that (18) holds with a matrix $\widetilde{S}^K$ of nonnegative type. Since, for $c>0$, the matrix R^K is of nonnegative type if and only if all the constants C_i^K are from the interval $\left[-\frac{1}{3},-\frac{1}{4}\right]$, a constant C_i^K provided by the algorithm in Fig. 8 will not be decreased if $C_i^K \leqslant -\frac{1}{4}$. If $C_i^K > -\frac{1}{4}$, it is never necessary to decrease this constant below the value $-\frac{1}{4}$.

Now let us describe the new definition of the constants C_i^K in detail. We again denote the vertices of K by a_1, a_2, a_3 and assume that $\boldsymbol{b}\in\mathrm{VZ}_1$. Then, according to Fig. 8, $C_2^K = C_3^K = -\frac{1}{3}$ and hence we only have to assure that s_{12}^K and s_{13}^K are nonpositive, which is the case if and only if

$$36\left(\boldsymbol{b}\cdot\nabla\varphi_j+\tfrac{1}{3}c\right)\left(\tfrac{1}{3}+C_1^K\right) \leqslant c,\quad j=2,3. \tag{19}$$

Of course, the constant $C_1^K = \frac{2}{3}$ provided by the algorithm in Fig. 8 generally does not satisfy this inequality. Therefore, denoting

$$\xi = 36\max\{0, \boldsymbol{b}\cdot\nabla\varphi_2+\tfrac{1}{3}c,\ \boldsymbol{b}\cdot\nabla\varphi_3+\tfrac{1}{3}c\},$$

we set

$$C_1^K := \min\left\{\frac{2}{3}, -\frac{1}{3}+\frac{c}{\xi}\right\}$$

(if $c=\xi=0$, we define $c/\xi=\infty$). Since $\boldsymbol{b}\cdot\nabla\varphi_j \leqslant 0$ for $j=2,3$, we really have $C_1^K \geqslant -\frac{1}{4}$.

Now let us assume that $\boldsymbol{b}\in\mathrm{EZ}_1$ and that K does not have the properties formulated in (A3) and (A3*) at the end of Section 3. If $\boldsymbol{b}\cdot\nabla u_h|_K = 0$, then $A^K U = 0$ and we set $C_1^K = -\frac{1}{3}$ and $C_2^K = C_3^K = -\frac{1}{4}$, which guarantees that the matrix R^K is of nonnegative type. Let $\boldsymbol{b}\cdot\nabla u_h|_K \neq 0$ and let the vector $\boldsymbol{w}$ be defined like in Section 6. It is convenient to denote for $\alpha\in\mathbb{R}$ and $j,k\in\{1,2,3\}$, $j\neq k$,

$$\xi_j(\alpha) = 36(\boldsymbol{b}\cdot\nabla\varphi_j + \alpha\boldsymbol{w}\cdot\nabla\varphi_j + \tfrac{1}{3}c),$$
$$\xi_{jk}(\alpha) = \max\{0,\xi_j(\alpha),\xi_k(\alpha)\}.$$

Let us first assume that $V_2\neq\emptyset$ and $V_3=\emptyset$. Then $C_1^K = C_3^K = -\frac{1}{3}$ and, like in Section 2, we deduce that (18) holds with $\widetilde{S}^K = \left(\frac{1}{3}+C_2^K\right)\widetilde{A}^{K,2} + R^K$. The matrix $\widetilde{A}^{K,2}$ was defined in Section 2 using an arbitrarily chosen $\alpha_2\in V_2$ and its first and third row consist of zeros. Therefore, we only have to assure that the entries $\widetilde{s}_{21}^K$ and $\widetilde{s}_{23}^K$ of the matrix $\widetilde{S}^K$ are nonpositive *for some* $\alpha_2\in V_2$. Like in (19), we get the condition that, for some $\alpha\in V_2$,

$$\xi_j(\alpha)\left(\tfrac{1}{3}+C_2^K\right) \leqslant c,\quad j=1,3.$$

The set V_2 is a closed interval and hence it is easy to compute

$$\xi = \min_{\alpha\in V_2}\xi_{13}(\alpha).$$

Thus, it suffices to set

$$C_2^K := \min\left\{\frac{2}{3}, -\frac{1}{3}+\frac{c}{\xi}\right\}.$$

Since $\xi_{13}(\alpha)\leqslant 12c$ for any $\alpha\in V_2$, we again have $C_2^K \geqslant -\frac{1}{4}$. The case $V_2=\emptyset$, $V_3\neq\emptyset$ is treated analogously.

If both V_2 and V_3 are nonempty, then $C_1^K = -\frac{1}{3}$ but the constants C_2^K and C_3^K provided by the algorithm in Fig. 8 may be so large that (18) does not hold for any matrix $\widetilde{S}^K$ of nonnegative type. Therefore, like above, we set

$$C_2^K := \min\left\{C_2^K, -\frac{1}{3}+\frac{c}{\xi}\right\},\quad C_3^K := \min\left\{C_3^K, -\frac{1}{3}+\frac{c}{\xi'}\right\},$$

where

$$\xi = \min_{\alpha\in V_2}\xi_{13}(\alpha),\quad \xi' = \min_{\alpha\in V_3}\xi_{12}(\alpha).$$

Up to now, we have not mentioned the case when $\boldsymbol{b}=\boldsymbol{0}$ and hence $A^K=0$. We set $C_1^K = C_2^K = C_3^K = -\frac{1}{4}$, which leads to a matrix S^K with positive diagonal entries and zero off-diagonal entries.

The above modifications of the constants C_i^K assure that the discrete solution of (17), (2) always satisfies (18) with a matrix $\widetilde{S}^K$ of nonnegative type. Therefore (see the end of Section 2), the discrete solution satisfies the discrete maximum principle and hence it does not contain any spurious oscillations.

Let us illustrate the properties of the improved Mizukami–Hughes method with the above described definition of the constants C_i^K by means of two simple test problems

taken from [10,16]. Like in Section 7, we consider $\varepsilon = 10^{-7}$, $\Omega = (0,1)^2$ and triangulations of the type depicted in Fig. 3(a).

Example 8 (*Reaction without convection*)

$$\boldsymbol{b} = \boldsymbol{0}, \quad c = 1, \quad f = 1, \quad \Gamma^{\mathrm{D}} = \partial\Omega, \quad u_{\mathrm{b}} = 0.$$

Fig. 20 shows a discrete solution computed using the Galerkin discretization (corresponding to the Mizukami–Hughes method with all C_i^K's equal to zero) and we observe significant spurious oscillations along the whole boundary of Ω. On the other hand, the improved Mizukami–Hughes method gives a nodally exact discrete solution, see Fig. 21.

Example 9 (*Reaction with convection*)

$$\boldsymbol{b}(x,y) = (1-y^2, 0), \quad c = 25, \quad f = 0, \quad \Gamma^{\mathrm{D}} = \{0\} \times (0,1), \quad g = 0, \quad u_{\mathrm{b}} = 1.$$

The Galerkin solution (cf. Fig. 22) again exhibits spurious oscillations which become even larger if the SUPG method described in the previous section is applied. The discrete solution obtained using the improved Mizukami–Hughes method with C_i^K's defined by the algorithm in Fig. 8 is comparable with the SUPG solution. However, using the constants C_i^K introduced in this section, we obtain the discrete solution depicted in Fig. 23 where no spurious oscillations are present.

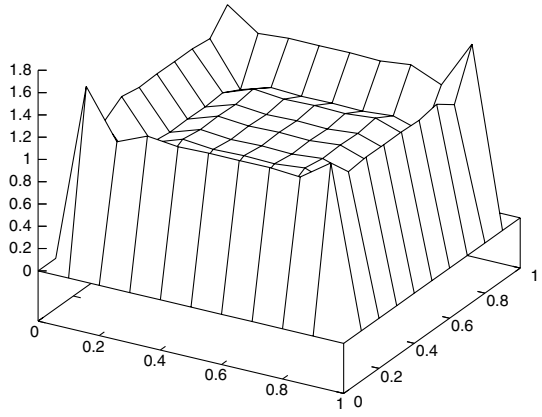

Fig. 20. Example 8, Galerkin, $N = 10$.

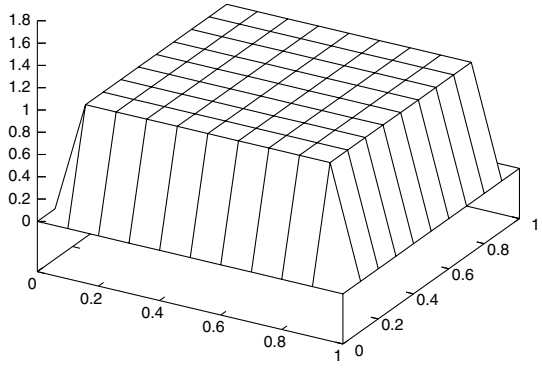

Fig. 21. Example 8, IMH, $N = 10$.

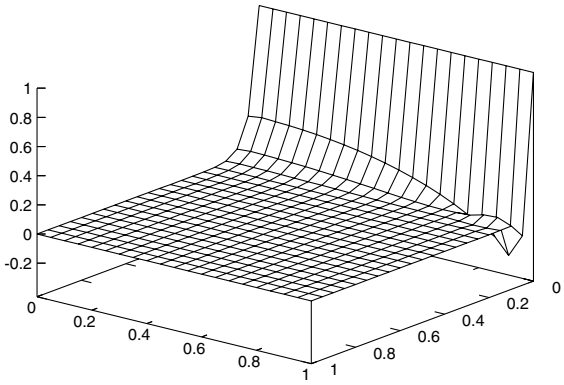

Fig. 22. Example 9, Galerkin, $N = 20$.

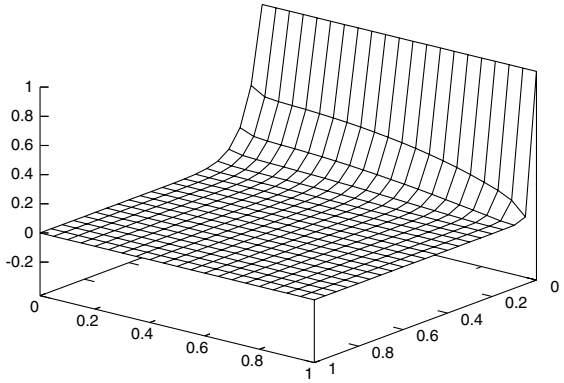

Fig. 23. Example 9, IMH, $N = 20$.

Remark 2. If the assumption that $\varepsilon \ll |\boldsymbol{b}| + c$ is not satisfied and the reaction term dominates the convection term (in particular, if $\boldsymbol{b} = \boldsymbol{0}$), the invalidity of the second condition in (3) may lead to a large error of the discrete solution. Therefore, in this case, instead of requiring that S^K or $\widetilde{S}^K$ be of nonnegative type, one should require this property of the matrix $D^K + S^K$ or $D^K + \widetilde{S}^K$, respectively, where D^K is the local diffusion matrix with entries $d_{ij}^K = \varepsilon(\nabla\varphi_j, \nabla\varphi_i)_K$. Assuming that D^K has three rows (since otherwise the second condition in (3) can always be fulfilled), there is at least one row of D^K whose all entries are different from zero. Therefore, adding D^K to S^K or $\widetilde{S}^K$ always enables to increase at least one of the constants C_i^K. In this way, the second condition in (3) can often be (nearly) satisfied since $d_{ij}^K \approx \varepsilon$ whereas $r_{ij}^K \approx c\ \mathrm{meas}_2(K)$.

9. The Mizukami–Hughes method in three dimensions

In this section, we briefly show how the ideas presented in Section 2 can be applied to the three-dimensional case.

We assume that Ω is a bounded three-dimensional domain with a polyhedral boundary $\partial\Omega$ and that we are given a triangulation $\mathscr{T}_h$ of Ω consisting of a finite number

of open tetrahedral elements K. The notation, assumptions and concepts introduced in Section 1 and at the beginning of Section 2 (by the end of the definition of the discrete solution) can be extended in a natural way to the three-dimensional case and hence we shall not mention them again.

Analogously as in Section 2, the local convection matrices A^K have entries

$$a_{ij}^K = (\boldsymbol{b}\cdot\nabla\varphi_j, \tilde{\varphi}_i)_K = (\boldsymbol{b}\cdot\nabla\varphi_j)|_K \mathrm{meas}_3(K)(\tfrac{1}{4} + C_i^K),$$

$i = 1,\dots,M_h$, $j = 1,\dots,N_h$, $a_i, a_j \in \overline{K}$. Therefore, we shall require that the constants C_i^K satisfy

$$C_i^K \geqslant -\tfrac{1}{4}\ \forall i \in \{1,\dots,N_h\},\ a_i \in \overline{K}, \quad \sum_{\substack{i=1\\ a_i\in\overline{K}}}^{N_h} C_i^K = 0. \tag{20}$$

Let K be any element of the triangulation $\mathscr{T}_h$ and let the vertices of K be a_1, a_2, a_3 and a_4. We divide the space $\mathbb{R}^3$ into 14 sets whose boundaries are formed by the four planes containing the barycentre a_c of K which are parallel to the faces of K. We denote these sets as vertex zones VZ_i, face zones FZ_i and edge zones EZ_{ij}, $i,j \in I$, $i<j$, where we used the index set $I = \{1,2,3,4\}$ for brevity. Precisely, the sets are defined in the following way:

$$\begin{aligned}
\mathrm{VZ}_i = \{&x\in\mathbb{R}^3; (x-a_c)\cdot\nabla\varphi_i > 0, (x-a_c)\cdot\nabla\varphi_k \leqslant 0\\
&\forall k \in I\setminus\{i\}\},\\
\mathrm{FZ}_i = \{&x\in\mathbb{R}^3; (x-a_c)\cdot\nabla\varphi_i < 0, (x-a_c)\cdot\nabla\varphi_k \geqslant 0\\
&\forall k \in I\setminus\{i\},\\
&\exists l \in I\setminus\{i\} : (x-a_c)\cdot\nabla\varphi_k > 0\ \forall k \in I\setminus\{i,l\}\},\\
\mathrm{EZ}_{ij} = \{&x\in\mathbb{R}^3; (x-a_c)\cdot\nabla\varphi_i > 0, (x-a_c)\cdot\nabla\varphi_j > 0,\\
&(x-a_c)\cdot\nabla\varphi_k < 0\ \forall k \in I\setminus\{i,j\}\}.
\end{aligned}$$

Again, we write $\nabla\varphi_i$ instead of $\nabla\varphi_i|_K$ for simplicity. Note that

$$\left(\bigcup_{i\in I}\mathrm{VZ}_i\right)\cup\left(\bigcup_{i\in I}\mathrm{FZ}_i\right)\cup\left(\bigcup_{i,j\in I, i<j}\mathrm{EZ}_{ij}\right) = \mathbb{R}^3\setminus\{a_c\}$$

and that all the 14 sets are mutually disjoint.

To get a better impression of the form of these sets, we introduce the points

$$u_{ij} = \frac{3a_i + a_j}{4}, \quad v_{ij} = \frac{a_i + \sum_{k\in I\setminus\{j\}} a_k}{4}, \quad i,j\in I,\ i\neq j.$$

Obviously, a point u_{ij} lies on the edge of K with end points a_i, a_j and a point v_{ij} lies on the face of K opposite the vertex a_j. It is easy to verify that the closure of $K\cap\mathrm{VZ}_i$ is a parallelepiped whose eight vertices are a_c, a_i, u_{ik}, v_{ik}, $k\in I\setminus\{i\}$, the closure of $K\cap\mathrm{FZ}_i$ is a tetrahedron whose four vertices are a_c, v_{ki}, $k\in I\setminus\{i\}$, and the closure of $K\cap\mathrm{EZ}_{ij}$ is a polyhedron with seven vertices a_c, u_{ij}, u_{ji}, v_{ik}, v_{jk}, $k\in I\setminus\{i,j\}$. Examples of an edge zone, a face zone and a vertex zone can be seen in Fig. 24. Note that, for any $k\in I$, all the nine points u_{ik} and v_{ij} with $i,j\in I\setminus\{k\}$, $i\neq j$, are contained in the plane containing a_c and being parallel to the face of K opposite a_k.

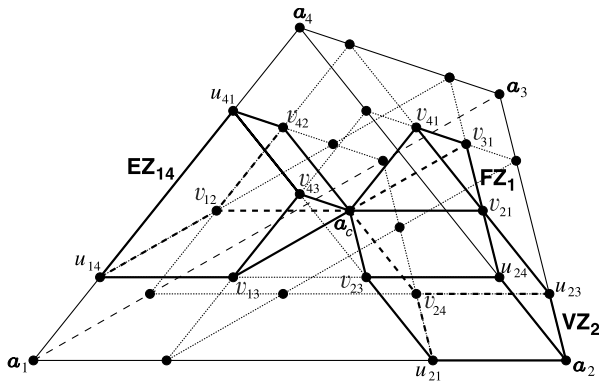

Fig. 24. Definition of edge zones, face zones and vertex zones.

Now let us discuss the definition of the constants $C_1^K,\dots,C_4^K$. If $\boldsymbol{b}$ points into a vertex zone, say VZ_j, $j\in I$, then (20) holds and A^K is of nonnegative type for

$$C_j^K = \tfrac{3}{4}, \quad C_k^K = -\tfrac{1}{4} \quad \forall k \in I\setminus\{j\}.$$

If A^K has four rows, this is the only possibility how to choose these constants.

Now let us assume that $\boldsymbol{b}$ does not point into any of the vertex zones. Then the constants C_i^K cannot be generally defined in such a way that (20) holds and the matrix A^K is of nonnegative type. Therefore, like in Section 2, we shall try to find such constants C_i^K, that the coefficient vector $U\in\mathbb{R}^4$ of $u_h|_K$ with respect to the basis $\{\varphi_i|_K\}_{i\in I}$ satisfies

$$A^K U = \widetilde{A}^K U, \tag{21}$$

where $\widetilde{A}^K$ is a matrix of nonnegative type. This is trivially satisfied if $\boldsymbol{b}\cdot\nabla u_h|_K = 0$ and hence we shall assume that $\boldsymbol{b}\cdot\nabla u_h|_K \neq 0$ in the following. Similarly as in Section 2, we introduce the sets

$$V_k = \{\tilde{\boldsymbol{b}}\in\mathbb{R}^3; (\tilde{\boldsymbol{b}}-\boldsymbol{b})\cdot\nabla u_h|_K = 0,\ a_c + \tilde{\boldsymbol{b}}\in\mathrm{VZ}_k\}, \quad k\in I.$$

If $\boldsymbol{b}$ points into the face zone FZ_j, $j\in I$, there exists $k\in I\setminus\{j\}$ such that $V_k\neq\emptyset$ and we may consider any constants C_i^K satisfying (20) and the following requirements:

$$V_k \neq \emptyset \quad \forall k\in I\setminus\{j\} \quad\Rightarrow\quad C_j^K = -\tfrac{1}{4}, \tag{22}$$

$$\exists k\in I\setminus\{j\} : V_k = \emptyset \quad\text{and}\quad V_l\neq\emptyset \quad \forall l\in I\setminus\{j,k\} \quad\Rightarrow\quad C_j^K = C_k^K = -\tfrac{1}{4}, \tag{23}$$

$$\exists k\in I\setminus\{j\} : V_k \neq \emptyset \quad\text{and}\quad V_l = \emptyset \quad \forall l\in I\setminus\{j,k\} \quad\Rightarrow\quad C_l^K = -\tfrac{1}{4} \quad \forall l\in I\setminus\{k\}. \tag{24}$$

If $\boldsymbol{b}$ points into the edge zone EZ_{jk}, $j,k\in I$, $j<k$, then $V_j\cup V_k\neq\emptyset$ and we consider any constants C_i^K satisfying (20) and the following requirements:

$$V_j\neq\emptyset \ \text{ and } \ V_k\neq\emptyset \quad\Rightarrow\quad C_l^K = -\tfrac{1}{4} \quad \forall l\in I\setminus\{j,k\}, \tag{25}$$

$$V_j\neq\emptyset \ \text{ and } \ V_k=\emptyset \quad\Rightarrow\quad C_l^K = -\tfrac{1}{4} \quad \forall l\in I\setminus\{j\}, \tag{26}$$

$$V_j=\emptyset \ \text{ and } \ V_k\neq\emptyset \quad\Rightarrow\quad C_l^K = -\tfrac{1}{4} \quad \forall l\in I\setminus\{k\}. \tag{27}$$

Let us assume that the constants C_i^K are defined according to (20) and (22)–(27) and let us introduce vectors $\tilde{\boldsymbol{b}}_1, \ldots, \tilde{\boldsymbol{b}}_4$ such that, for any $i \in I$,

$$\tilde{\boldsymbol{b}}_i = \boldsymbol{b} \quad \text{if } C_i^K = -\tfrac{1}{4}, \quad \tilde{\boldsymbol{b}}_i \in V_i \quad \text{if } C_i^K > -\tfrac{1}{4}.$$

We define a matrix $\widetilde{A}^K$ with entries

$$\tilde{a}_{ij}^K = (\tilde{\boldsymbol{b}}_i \cdot \nabla \varphi_j)|_K \mathrm{meas}_3(K)(\tfrac{1}{4} + C_i^K), \quad i, j \in I, \ a_i \in \Omega \cup \Gamma^{\mathrm{N}}.$$

Then $\widetilde{A}^K$ is of nonnegative type and (21) holds.

There are many possibilities how to satisfy the requirements (20) and (22)–(27) and, since (21) always holds with a matrix $\widetilde{A}^K$ of nonnegative type, the discrete solution u_h always satisfies the discrete maximum principle. However, not every choice of the constants C_i^K satisfying (20) and (22)–(27) is appropriate and we may encounter similar difficulties like those ones discussed in Sections 3–5. The derivation of suitable formulas for the constants C_i^K will be a subject of our further research.

10. Conclusions

In this paper we introduced several improvements of the Mizukami–Hughes method for the numerical solution of two-dimensional steady convection–diffusion equations. We have shown that the improved method satisfies the discrete maximum principle and we demonstrated by means of various numerical results that it gives very accurate discrete solutions with no spurious oscillations. Moreover, our extensive numerical tests (which will be published in a separate paper) revealed that none of the discontinuity-capturing methods mentioned in the introduction can be regarded as superior to the improved Mizukami–Hughes method. Therefore, the improved Mizukami–Hughes method seems to be one of the best choices for solving the problem (1) and (2) using conforming piecewise linear triangular finite elements if convection strongly dominates diffusion. We have also shown that the Mizukami–Hughes method can be extended to convection–diffusion–reaction equations and to the three-dimensional case but here further research is necessary.

Acknowledgements

The work is a part of the research project MSM 0021620839 financed by MSMT and it was partly supported by the Grant Agency of the Charles University in Prague under the Grant No. 344/2005/B-MAT/MFF.

References

[1] R.C. Almeida, R.S. Silva, A stable Petrov–Galerkin method for convection-dominated problems, Comput. Methods Appl. Mech. Engrg. 140 (1997) 291–304.
[2] A.N. Brooks, T.J.R. Hughes, Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations, Comput. Methods Appl. Mech. Engrg. 32 (1982) 199–259.
[3] E. Burman, A. Ern, Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection–diffusion–reaction equation, Comput. Methods Appl. Mech. Engrg. 191 (2002) 3833–3855.
[4] E. Burman, A. Ern, Stabilized Galerkin approximation of convection–diffusion–reaction equations: discrete maximum principle and convergence, Math. Comput. 74 (2005) 1637–1652.
[5] E. Burman, P. Hansbo, Edge stabilization for Galerkin approximations of convection–diffusion–reaction problems, Comput. Methods Appl. Mech. Engrg. 193 (2004) 1437–1453.
[6] E.G.D. do Carmo, G.B. Alvarez, A new upwind function in stabilized finite element formulations, using linear and quadratic elements for scalar convection–diffusion problems, Comput. Methods Appl. Mech. Engrg. 193 (2004) 2383–2402.
[7] P.G. Ciarlet, P.-A. Raviart, Maximum principle and uniform convergence for the finite element method, Comput. Methods Appl. Mech. Engrg. 2 (1973) 17–31.
[8] R. Codina, A discontinuity-capturing crosswind-dissipation for the finite element solution of the convection–diffusion equation, Comput. Methods Appl. Mech. Engrg. 110 (1993) 325–342.
[9] T.J.R. Hughes, M. Mallet, A. Mizukami, A new finite element formulation for computational fluid dynamics: II. Beyond SUPG, Comput. Methods Appl. Mech. Engrg. 54 (1986) 341–355.
[10] S. Idelsohn, N. Nigro, M. Storti, G. Buscaglia, A Petrov–Galerkin formulation for advection–reaction–diffusion problems, Comput. Methods Appl. Mech. Engrg. 136 (1996) 27–46.
[11] T. Ikeda, Maximum Principle in Finite Element Models for Convection–Diffusion Phenomena, Lecture Notes in Numerical and Applied Analysis, vol. 4, North-Holland, Amsterdam, 1983.
[12] V. John, J.M. Maubach, L. Tobiska, Nonconforming streamline-diffusion-finite-element-methods for convection–diffusion problems, Numer. Math. 78 (1997) 165–188.
[13] T. Knopp, G. Lube, G. Rapin, Stabilized finite element methods with shock capturing for advection–diffusion problems, Comput. Methods Appl. Mech. Engrg. 191 (2002) 2997–3013.
[14] A. Mizukami, T.J.R. Hughes, A Petrov–Galerkin finite element method for convection-dominated flows: an accurate upwinding technique for satisfying the maximum principle, Comput. Methods Appl. Mech. Engrg. 50 (1985) 181–193.
[15] Y.-T. Shih, H.C. Elman, Modified streamline diffusion schemes for convection–diffusion problems, Comput. Methods Appl. Mech. Engrg. 174 (1999) 137–151.
[16] T.E. Tezduyar, Y.J. Park, Discontinuity-capturing finite element formulations for nonlinear convection–diffusion–reaction equations, Comput. Methods Appl. Mech. Engrg. 59 (1986) 307–325.