Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

A numerical assessment of finite element discretizations for convection-diffusion-reaction equations satisfying discrete maximum principles

Volker John[1], Petr Knobloch[2], Ondřej Pártl[3]

submitted: July 15, 2022

[1] Weierstrass Institute
Mohrenstr. 39, 10117 Berlin, Germany
und Freie Universität Berlin
Arnimallee 6, 14195 Berlin, Germany
E-Mail: volker.john@wias-berlin.de

[2] Charles University
Faculty of Mathematics and Physics
Department of Numerical Mathematics
Sokolovská 83, 18675 Praha 8, Czech Republic
E-Mail: knobloch@karlin.mff.cuni.cz

[3] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: ondrej.partl@wias-berlin.de

No. 2946
Berlin 2022

2020 *Mathematics Subject Classification.* 65N30.

Key words and phrases. Convection-diffusion-reaction equations, P_1 finite element methods, algebraic stabilizations, upwind methods, discrete maximum principles.

The work of Petr Knobloch has been supported by the grant No. 22-01591S of the Czech Science Foundation.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

A numerical assessment of finite element discretizations for convection-diffusion-reaction equations satisfying discrete maximum principles

Volker John, Petr Knobloch, Ondřej Pártl

Abstract

Numerical studies are presented that investigate finite element methods satisfying discrete maximum principles for convection-diffusion-reaction equations. Two linear methods and several nonlinear schemes, some of them proposed only recently, are included in these studies, which consider a number of two-dimensional examples. The evaluation of the results examines the accuracy of the numerical solutions with respect to quantities of interest, like layer widths, and the efficiency of the simulations.

1 Introduction

Scalar convection-diffusion-reaction equations are part of many coupled systems that model processes in nature or industry. Such equations describe the conservation of concentrations, e.g., in combination with chemical reactions, or the conservation of energy. In practice, the physical consistency of a numerical method is of extreme importance. An essential aspect of physical consistency is the requirement that the numerical solution must not possess unphysical values, like negative concentrations. The computed solution of convection-diffusion-reaction equations in coupled systems is usually an input data for other equations of that system. Then, unphysical values of the solution might lead to unphysical data in the other equations. As a result, the simulations might become unstable and they might finally blow up, as it is our own experience in [26]. Altogether, the property that a method for solving numerically convection-diffusion-reaction equations provably leads to solutions without unphysical values, also called spurious oscillations, is a strong reason that this method is accepted by practitioners.

The (weak) solutions of scalar convection-diffusion-reaction equations obey, under certain conditions on their coefficients and data, maximum principles, e.g., see [16, Chapter 3.1]. The non-occurrence of spurious oscillations in numerical solutions is guaranteed by so-called discrete maximum principles (DMPs). It is well known that the satisfaction of DMPs is not straightforward, e.g., compare the recent survey [7] on DMP-respecting finite element methods. Often, DMPs have been proved only for continuous piecewise linear (P_1) finite elements on simplicial grids. For linear discretizations, one has to satisfy in general conditions

DOI 10.20347/WIAS.PREPRINT.2946 Berlin 2022

on the triangulation. An alternative are nonlinear discretizations, but even such methods might require conditions on the grid.

There are different (local) regimes that are distinguished for convection-diffusion-reaction equations: the diffusion-dominated, the reaction-dominated, and the convection-dominated regime. In many applications, the convection-dominated regime is present since the convective transport by a flow field is usually much stronger than the transport by molecular diffusion. It is well known that one has to apply a so-called stabilized discretization in this regime unless the grid is sufficiently fine, e.g., see [36]. The reason is that the solution usually exhibits so-called layers, which are very thin structures with a large norm of the gradient. In particular, the thickness of layers is typically much smaller than the affordable mesh width, so that they cannot be resolved. Applying a standard Galerkin finite element method leads to a numerical solution that is globally polluted with significant spurious oscillations. Numerical studies, e.g., in [22, 24], show, however, that the vast majority of stabilized discretizations proposed in the literature does not satisfy a global DMP and computes solutions with notable spurious oscillations in a vicinity of layers. Also in the reaction-dominated regime solutions might possess layers. The occurrence of layers motivates the proposal of nonlinear discretizations because an appropriate discretization has to work in a different way in a vicinity of layers and away from layers.

As documented in [7], in recent years there has been an enormous progress in the development of nonlinear discretizations for which DMPs can be proved. However, apart from two algebraically stabilized discretizations, there is no comprehensive numerical comparison of methods satisfying DMPs. This paper aims to fill this gap to some extent. The studied methods include the edge-averaged finite element method, which is a linear method, a linear and a nonlinear upwind finite element method, and algebraically stabilized schemes with various limiters.

There are various benchmark problems for convection-diffusion-reaction equations defined on two-dimensional domains whose solutions exhibit layers. In this situation, usually an analytic expression of the solution is not known and typically quantities of interest are defined that measure the accuracy of computed solutions. This approach is also pursued in the current paper. Since there are no spurious oscillations in the numerical solutions obtained with the studied schemes, a typical quantity of interest is the sharpness of layers. As already mentioned, most of these schemes are nonlinear. Thus, from the practical point of view, the cost for solving the nonlinear problems is of importance. Also this issue will be addressed. Finally, because it is proved for two methods that they satisfy DMPs for rather general elliptic problems, it is demonstrated exemplarily that they can be applied for computing oscillation-free solutions of anisotropic and heterogeneous diffusion problems on a standard grid.

Convection-diffusion-reaction equations will be introduced in Section 2. Then, Section 2.1 and Section 2.2 provide a description of the studied methods. For the sake of brevity, this presentation concentrates on main ideas for constructing the methods and main properties. It is referred to the literature for concrete formulas and technical details. In Section 2.3, a survey is given on available comparisons of methods that are included in our studies. Our numerical studies are presented in Section 3. Finally, Section 4 contains the main conclusions from these studies.

2 The Convection-Diffusion-Reaction Equation and the Studied Numerical Methods

All considered examples are defined in two dimensions. Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with polygonal boundary $\partial\Omega$, which is Lipschitz. The boundary is decomposed in the Dirichlet boundary $\partial\Omega_{\mathrm{D}}$ and the Neumann boundary $\partial\Omega_{\mathrm{N}}$, i.e., $\partial\Omega = \overline{\partial\Omega_{\mathrm{D}}} \cup \overline{\partial\Omega_{\mathrm{N}}}$, with $\partial\Omega_{\mathrm{D}} \cap \partial\Omega_{\mathrm{N}} = \emptyset$. Then, the considered stationary linear convection-diffusion-reaction boundary value problem is given by

$$\begin{array}{rcll} -\nabla\cdot(\mathbb{E}\nabla u) + \boldsymbol{b}\cdot\nabla u + cu &=& f & \text{in } \Omega, \\ u &=& u_{\mathrm{D}} & \text{on } \partial\Omega_{\mathrm{D}}, \\ (\mathbb{E}\nabla u)\cdot\boldsymbol{n} &=& 0 & \text{on } \partial\Omega_{\mathrm{N}}. \end{array} \tag{1}$$

In (1), $\mathbb{E}$ is the diffusion matrix, which is assumed to be symmetric and strictly positive definite on $\overline{\Omega}$. The convection vector is denoted by $\boldsymbol{b}$, the reaction field by c, the sources by f, the Dirichlet data by u_{D}, and $\boldsymbol{n}$ is the outward pointing unit normal vector on $\partial\Omega$. All the data of (1) may depend on the points $\boldsymbol{x} \in \Omega$. Concerning the regularity of this data, it will be assumed that

$$\mathbb{E} \in (L^\infty(\Omega))^{2\times 2},\ \boldsymbol{b} \in \left(W^{1,\infty}(\Omega)\right)^2,\ c \in L^\infty(\Omega),\ f \in L^2(\Omega),\ u_{\mathrm{D}} \in H^{1/2}(\partial\Omega_{\mathrm{D}}),$$

where standard notations for Lebesgue and Sobolev spaces are used. In addition, it will be assumed that

$$-\frac{1}{2}\nabla\cdot\boldsymbol{b} + c \geq 0 \quad \text{in } \Omega \tag{2}$$

and that the so-called inlet boundary $\overline{\{\boldsymbol{x}\in\partial\Omega \,:\, \boldsymbol{b}(\boldsymbol{x})\cdot\boldsymbol{n}(\boldsymbol{x}) < 0\}}$ is a subset of $\overline{\partial\Omega_{\mathrm{D}}}$. With all these assumptions, it can be proved by applying the Lax–Milgram theorem, that the standard weak problem corresponding to (1), which is obtained by applying integration by parts to the diffusive term, is well posed.

Now the finite element discretizations of the weak problem corresponding to (1) will be described. For the sake of brevity, the presentation of these discretizations will be descriptive. Formal mathematical presentations can be found in the literature and corresponding citations are provided.

Let $\mathcal{T}_h$ be a triangulation of the domain Ω consisting of triangles, which is assumed to be admissible in the sense of [13, p. 38, p. 51]. Only discretizations with conforming P_1 finite elements will be considered. This choice is motivated by the fact that there are severe restrictions to satisfy DMPs for higher order finite elements, e.g., see [37]. The degrees of freedom are the values of the finite element functions at the vertices of $\mathcal{T}_h$. Let N be the total number of vertices and let $M < N$ be the number of vertices that are not contained in $\overline{\partial\Omega_{\mathrm{D}}}$. Without loss of generality, the degrees of freedom will be numbered so that the Dirichlet values are at the end.

2.1 Linear Discretizations

A linear discretization of a scalar convection-diffusion-reaction problem leads to a linear system of equations of the form

$$\mathbb{A}u = \begin{pmatrix} \mathbb{A}_{\mathrm{I}} & \mathbb{A}_{\mathrm{B}} \\ \mathbb{O} & \mathbb{I} \end{pmatrix} \begin{pmatrix} u_{\mathrm{I}} \\ u_{\mathrm{B}} \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{3}$$

where $u_{\mathrm{I}} \in \mathbb{R}^M$ are the values of the non-Dirichlet degrees of freedom and $u_{\mathrm{B}} \in \mathbb{R}^{N-M}$ are values of the degrees of freedom at $\overline{\partial\Omega_{\mathrm{D}}}$. The matrix $\mathbb{A}_{\mathrm{I}} \in \mathbb{R}^{M\times M}$ contains the discretization of the non-Dirichlet degrees of freedom, $\mathbb{A}_{\mathrm{B}} \in \mathbb{R}^{M\times(N-M)}$ contains the connection of them to the Dirichlet data, and $\mathbb{I} \in \mathbb{R}^{(N-M)\times(N-M)}$ is the identity matrix. The vector $f \in \mathbb{R}^M$ represents the discretization of the right-hand side of the boundary value problem and the vector $g \in \mathbb{R}^{N-M}$ is a suitable approximation of the Dirichlet data.

The following definition introduces very important concepts for studying DMPs for linear discretizations.

Definition 2.1 (Matrix of non-negative type, monotone matrix, M-matrix) *A matrix* $(a_{ij})_{j=1,\ldots,n}^{i=1,\ldots,m} \in \mathbb{R}^{m\times n}$ *(*$m, n \in \mathbb{N}$*) will be said to be of non-negative type if*

$$a_{ij} \le 0 \quad \forall\, i \ne j, 1 \le i \le m, 1 \le j \le n, \tag{4}$$

$$\sum_{j=1}^{n} a_{ij} \ge 0 \quad \forall\, 1 \le i \le m. \tag{5}$$

A matrix $\mathbb{A} = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n}$ *is called to be monotone, if* $\mathbb{A}$ *is non-singular and its inverse possesses only non-negative entries, in symbol* $\mathbb{A}^{-1} \ge 0$*.*

A monotone matrix that satisfies (4) *is called an M-matrix.*

It can be shown that the set of invertible matrices of non-negative type is a proper subset of the set of M-matrices.

Let $a \in \mathbb{R}$, then its positive and negative parts are defined by

$$a^+ := \max\{a, 0\} \ge 0 \quad \text{and} \quad a^- := \min\{a, 0\} \le 0.$$

The same notation will be used for the positive and negative parts of a real-valued function.

Definition 2.2 (Local DMPs.) *A solution of* (3) *is said to satisfy the local DMP if*

$$f_i \le 0 \implies u_i \le \max_{j\ne i, a_{ij}\ne 0} u_j^+, \quad f_i \ge 0 \implies u_i \ge \min_{j\ne i, a_{ij}\ne 0} u_j^- \quad \forall\, i = 1, \ldots, M.$$

It turns out that the local DMPs are satisfied if and only if the system matrix $\mathbb{A}$ from (3) is of non-negative type and has positive diagonal entries, e.g., see [5, 7]. If, in addition, all row sums of $(a_{ij})_{j=1,\ldots,N}^{i=1,\ldots,M}$ vanish, i.e., condition (5) is satisfied with the equal sign, then a stronger version of the local DMP can be proved

$$f_i \le 0 \implies u_i \le \max_{j\neq i, a_{ij}\neq 0} u_j, \quad f_i \ge 0 \implies u_i \ge \max_{j\neq i, a_{ij}\neq 0} u_j \quad \forall\, i = 1,\ldots,M.$$

Definition 2.3 (Global DMPs.) *A solution of* (3) *is said to satisfy the global DMP if*

$$\begin{aligned} f_i \le 0 \quad \forall\, i = 1,\ldots,M \quad &\implies \quad \max_{i=1,\ldots,N} u_i \le \max_{j=M+1,\ldots,N} u_j^+, \\ f_i \ge 0 \quad \forall\, i = 1,\ldots,M \quad &\implies \quad \min_{i=1,\ldots,N} u_i \ge \min_{j=M+1,\ldots,N} u_j^-. \end{aligned}$$

A sufficient and necessary condition for the satisfaction of the global DMP was already proved in [12]. This condition states that $\mathbb{A}$ has to be monotone and that the row sums of $-\left(\mathbb{A}_{\mathrm{I}}\right)^{-1}\mathbb{A}_{\mathrm{B}}$ are smaller than 1. In practice, often a sufficient condition, also proved in [12], is used: $\mathbb{A}$ is monotone and (5) is satisfied. If (5) holds with the equal sign for all rows, one can prove a stronger version of the global DMP

$$\begin{aligned} f_i \le 0 \quad \forall\, i = 1,\ldots,M \quad &\implies \quad \max_{i=1,\ldots,N} u_i = \max_{j=M+1,\ldots,N} u_j, \\ f_i \ge 0 \quad \forall\, i = 1,\ldots,M \quad &\implies \quad \min_{i=1,\ldots,N} u_i = \min_{j=M+1,\ldots,N} u_j. \end{aligned}$$

In the following description of the numerical methods, all comments concerning DMPs, matrices of non-negative type and so on refer to P_1 finite elements.

The standard finite element approach, which just replaces the infinite-dimensional spaces of the weak formulation with finite element subspaces, is called *Galerkin finite element method*. Consider a pure diffusion problem with $\mathbb{E} = \varepsilon\mathbb{I}$, $\varepsilon \in \mathbb{R}, \varepsilon > 0$, $\boldsymbol{b} = \boldsymbol{0}$, $c = 0$. Then, for $\Omega \subset \mathbb{R}^2$, the Galerkin finite element method leads to a matrix of non-negative type if the grid is of Delaunay type, see [39]. If any inner edge has at least one vertex in Ω, then this condition is also necessary. This result was extended to tensor-valued diffusion coefficients in [18]. For a reaction-diffusion problem with $\mathbb{E} = \varepsilon\mathbb{I}$, $\varepsilon \in \mathbb{R}, \varepsilon > 0$, $\boldsymbol{b} = \boldsymbol{0}$, $c(\boldsymbol{x}) \ge 0$, one can derive a sufficient condition for the matrix to be of non-negative type. This condition is satisfied in the reaction-dominated regime only if the mesh is sufficiently fine, see [7, Lemma 4.4] for the case that c is a positive constant. The reason is that the finite element mass matrix possesses only non-negative entries so that condition (4) is not satisfied as long as the mesh is too coarse. A remedy consists in utilizing mass lumping, which introduces a first order consistency error. For mass lumping, the system matrix is of non-negative type under the same conditions as for the diffusion problem. It was already mentioned in the introduction that the Galerkin finite element method should not be applied for convection-diffusion-reaction equations, in particular if convection dominates.

One of the oldest finite element methods suited for the numerical solution of problems with dominating convection is the *linear upwind method* proposed in [2]. This method uses the

Galerkin method for discretizing the diffusive term. There is no reactive term in the model problem studied in [2]. But based on the discussion of discretizing this term for the Galerkin discretization, it follows that the mass lumping approach should be applied if the satisfaction of DMPs on affordable grids is the goal. Finally, the convective term is discretized with a finite volume approach. To this end, the convective term is considered in divergence form $\nabla \cdot (\boldsymbol{b}u)$ and a dual domain D_i is defined for each vertex $\boldsymbol{x}_i$ of the triangulation. Then, integrating the convective term over D_i and applying integration by parts leads to integrals on the boundary of D_i, containing the convective fluxes across this boundary. In [2], a simple upwind approach is used to approximate these fluxes, for the concrete definition it is referred to [2]. It is proved in that paper that the proposed choice of the numerical fluxes leads to a convection matrix with zero row sums and non-positive off-diagonal entries if the mesh is non-obtuse. Consequently, for such meshes, the complete method, with mass lumping for the reactive term, satisfies the sufficient condition for the global DMP from [12].

The *edge-averaged finite element (EAFE) scheme*, proposed and analyzed in [39], is based on the divergence form of the convection-diffusion equation

$$-\nabla \cdot (\varepsilon \nabla u - \boldsymbol{b}u) = f \quad \text{in } \Omega, \tag{6}$$

i.e., the diffusion is isotropic and homogeneous. For simplicity, let us explain the idea of the method for $\partial\Omega_{\mathrm{D}} = \partial\Omega$. Multiplying (6) with a test function and applying integration by parts leads to the bilinear form

$$(\varepsilon \nabla u - \boldsymbol{b}u, \nabla v) =: (\boldsymbol{J}(u), \nabla v)\,, \tag{7}$$

where $\boldsymbol{J}(u)$ is called total flux. The derivation of the EAFE scheme is based on the observation that, for P_1 finite elements, it is possible to write the term $(\boldsymbol{a}, \nabla v_h)_K$ for any mesh cell K, $\boldsymbol{a} \in \mathbb{R}^2$, and v_h in the finite element space as a sum over the edges of the triangulation involving tangent components of $\boldsymbol{a}$. To employ this observation for a discretization of (7), the total flux $\boldsymbol{J}(u)$ is replaced by a piecewise constant approximation. To approximate its tangent components, a function χ_E is defined on each edge E, whose derivative is the negative of the convection in tangential direction divided by the diffusion. If the sign of the tangent component of the convection does not change, the function χ_E is increasing in the direction from the downwind node to the upwind node. Then a formula that provides a connection between χ_E, u, and the tangent component of $\boldsymbol{J}(u)$ on E can be derived in a straightforward way. This formula contains the weighting function e^{χ_E} that creates the upwind property of the EAFE scheme. For Delaunay meshes and under irreducibility assumptions, it was shown that the system matrix of the EAFE scheme is an M-matrix which implies positivity preservation. The satisfaction of DMPs is proved if the convection field $\boldsymbol{b}$ is constant. It is shown in [3] that the EAFE scheme is equivalent to the Scharfetter–Gummel finite volume scheme.

If (6) is equipped with the boundary conditions of (1) (with $\mathbb{E} = \varepsilon\mathbb{I}$), then the integration by parts leads to the additional term $(\boldsymbol{b}\cdot\boldsymbol{n}\, u, v)_{\partial\Omega_{\mathrm{N}}}$ in (7). When deriving the discrete problem, we applied a lumping to this term so that it contributes to diagonal entries of the system matrix only.

Note that both upwinding strategies discussed above reduce (locally) to the Galerkin discretization if the convection field (locally) vanishes.

2.2 Nonlinear Discretizations

Nonlinear discretizations take into account that the solutions of convection-diffusion-reaction equations generally behave differently in different subregions of the domain. There are layer regions, on the one hand, whose numerical simulation requires a substantial stabilization. And on the other hand, there are regions where the solution changes gently and only little or even no stabilization is necessary. In nonlinear stabilized discretizations, the stabilization term depends on the unknown numerical solution.

The criteria for proving DMPs formulated for linear discretizations do not apply for nonlinear ones. Instead, one considers only the entries of rows that correspond to a degree of freedom where an extremum of the finite element solution is encountered. Thus, instead of studying every equation of the system, like for linear discretizations, one needs to investigate only those equations where the discrete solution attains a local extremum. For precise formulations of conditions for which DMPs are satisfied for nonlinear stabilized discretizations, it is referred to [10, 31, 8, 7]. Let us mention that the local DMPs usually hold with a larger index sets than considered in Definition 2.2. Typically, the maxima and minima are computed over all nodes connected by an edge with x_i.

The first nonlinear DMP-satisfying approach proposed for the numerical solution of convection–diffusion equations is the Mizukami–Hughes method [35], which was improved and further developed in [28, 29, 30]. It is a nonlinear Petrov–Galerkin method, herein called *nonlinear upwind method*, where an upwind effect is created by solution-dependent weighting functions which guarantee that the approximate solution satisfies a linear system with a matrix of non-negative type. The weighting functions are obtained by modifying standard P_1 basis functions by adding suitable constants on mesh cells making up their supports. The choice of these constants is motivated by the requirement that the local convection matrices are of non-negative type, and it depends on the direction of the convection vector with respect to the edges of the mesh cells. In general, the orientation of the gradient of the unknown numerical solution has to be also taken into account, which makes the method nonlinear. We refer to [28] for the definitions of these constants used for the computations presented in this paper. For special triangulations and convection fields, the method can become linear, which will be the case in Examples 3.1 and 3.4 below. The DMPs for the nonlinear upwind method were proved for non-obtuse meshes.

Algebraically stabilized discretizations start with the linear algebraic system of equations (3) obtained with the Galerkin finite element method. Defining the so-called artificial diffusion matrix

$$\mathbb{D} = (d_{ij})_{i,j=1}^{N}, \quad d_{ij} = -\max\left\{0, a_{ij}, a_{ji}\right\} \quad \text{for } i \neq j, \quad d_{ii} = -\sum_{j=1, j\neq i}^{N} d_{ij},$$

where an extension of the matrix $(a_{ij})_{j=1,\ldots,N}^{i=1,\ldots,M}$ from (3) to all nodes is considered, a first version of an algebraically stabilized scheme is of the form

$$\begin{aligned} \sum_{j=1}^N a_{ij}u_j + \sum_{j=1}^N \big(1-\alpha_{ij}\left(u_{\mathrm{IB}}\right)\big) d_{ij}\left(u_j - u_i\right) &= f_i \qquad \text{for } i = 1,\ldots,M, \\ u_i &= g_{i-M} \quad \text{for } i = M+1,\ldots,N, \end{aligned} \tag{8}$$

with $u_{\mathrm{IB}} = (u_{\mathrm{I}}, u_{\mathrm{B}})^T$. The correction factors $\alpha_{ij}(u_{\mathrm{IB}})$, the so-called limiters, are chosen to be in $[0,1]$ and it is required that $\alpha_{ij}(u_{\mathrm{IB}}) = \alpha_{ji}(u_{\mathrm{IB}})$ for $i \neq j$. If all of them vanish, then one obtains a linear system of equations where, by construction, the system matrix is of non-negative type. The corresponding discretization can be interpreted as globally adding numerical diffusion and it leads to very inaccurate solutions. In practice, numerical diffusion is only necessary in a vicinity of layers, hence the limiters should be small only in layer regions. Away from layers, the limiters should be close to 1 to achieve a high accuracy in these regions.

The first finite element method of type (8) was proposed in [32] and it is called *AFC scheme with Kuzmin limiter*, see also [5, 8, 7] for presentations of the concrete form of the limiter. This limiter possesses an upwind character, because in the step that ensures the symmetry condition of the limiter, the information from the upwind node (degree of freedom) is taken. This method was studied analytically in [5], in particular with respect to the DMP. If mass lumping is applied to the reactive term, then it was found that a sufficient condition for the satisfaction of the DMP is that the triangulation is non-obtuse. In two dimensions, the Delaunay property suffices. In [31], the condition for the satisfaction of the local DMP was weakened to the requirement that

$$\min\{a_{ij}, a_{ji}\} \le 0 \quad \forall\, i = 1,\ldots,M\,,\ j = 1,\ldots,N\,,\ i \neq j. \tag{9}$$

It is also shown in [31] that the local DMP is generally not satisfied if this condition is violated.

In the stabilization term of (8), one can write for $i \neq j$

$$\big(1-\alpha_{ij}\left(u_{\mathrm{IB}}\right)\big) d_{ij} = -\max\left\{0, \left(1-\alpha_{ij}\left(u_{\mathrm{IB}}\right)\right) a_{ij}, \left(1-\alpha_{ji}\left(u_{\mathrm{IB}}\right)\right) a_{ji}\right\}\ .$$

These rewriting makes it possible to get rid of the symmetry requirement on the limiters α_{ij}. In this way, after a reformulation and modification of the Kuzmin limiter in [25], the *Monotone Upwind-type Algebraically Stabilized (MUAS) method* was derived, which satisfies the DMPs on arbitrary meshes (without a lumping of the reaction term) and preserves the favorable features of the AFC scheme with the Kuzmin limiter. In particular, if the condition (9) is satisfied (which is the case for many computations presented in this paper), the MUAS method coincides with the AFC scheme with the Kuzmin limiter, up to a small modification of a multiplicative factor in the definition of the limiter functions.

The so-called *AFC scheme with BJK limiter* was proposed in [6]. The construction of this limiter resembles the classical Zalesak limiter [40] for time-dependent problems. Of course, the length of the time step that enters the definition of the Zalesak limiter is not available for steady-state problems. Instead, a quantity that depends on the local geometry of the grid is utilized for computing the BJK limiter. This type of limiter is constructed so far only for P_1 finite elements. In [6], it is proved that the AFC scheme with BJK limiter satisfies the DMPs on arbitrary simplicial grids. In addition, it is shown that this method is linearity preserving, i.e., the stabilization term vanishes if the solution of the convection-diffusion-reaction equation is a linear function. It is conjectured in [6] that this property is of advantage for the accuracy of the results, but so far a mathematical proof of this conjecture is an open problem.

The *AFC scheme with BBK limiter*, proposed in [4], checks the local smoothness of the numerical solution u_i at a node $\boldsymbol{x}_i$, i.e., at a vertex of a simplex. This check considers the ratio of the absolute value of the sum of differences of u_i to the values of the solution in all neighbor nodes and the sum of the absolute values of these differences. Then, the stabilization is defined based on this ratio. This definition contains two user-chosen parameters, one of them, γ_0, depending on the data of the problem and the other one, p, determining the decay of the numerical diffusion with the distance to critical points, see [4] for a detailed discussion.

An algebraic stabilization for hyperbolic conservation laws was proposed in [33]. It is shown in this paper that the steady-state limit of the nonlinear discrete problem is well defined, independently of the time step. Thus, this scheme can be used for discretizing steady-state equations. Although diffusion is not considered in [33], the so-called *AFC scheme with monolithic convex (MC) limiter* was applied in [21] for the solution of a steady-state convection-diffusion equation with promising results, which is the reason why we included this method in our numerical studies. Concretely, only the convective term was limited with the diffusive and reactive term transferred to the right-hand side and after limiting, unlimited discretizations of both terms are added. The limiter is defined by a convex combination of fluxes from the previous and the current iteration, whose definition is based on intermediate solutions, so-called bar states. A local DMP for the algebraic system corresponding to the discretization of the steady-state (transport) problem considered in [33] is proved in this paper.

2.3 Available Numerical Studies

In [22, 24], comprehensive numerical studies of so-called spurious oscillations at layers diminishing (SOLD) methods are reported. DMPs can be proved for some of the considered methods in these studies: the improved version of the nonlinear upwind method from [28], which is described above, and several so-called edge stabilization methods from [9, 10, 11]. The solutions obtained with the nonlinear upwind method were indeed free of spurious oscillations, but there were cases in which the used iterative scheme for solving the nonlinear problem failed to converge. The numerical solutions computed with all edge stabilization methods possessed spurious oscillations, often small, but sometimes not. In fact, the DMPs for these schemes are proved if a stabilization parameter is sufficiently large. However, it was not possible with the approach utilized in [22, 24] to solve the nonlinear problems for sufficiently large parameters. The same experience is reported in [7]. Already in the original papers [9, 10, 11] small spurious oscillations of the numerical solutions are reported. In summary, there is currently no approach for solving the nonlinear problems arising in the nonlinear edge stabilizations from [9, 10, 11] for stabilization parameters that ensure the satisfaction of DMPs. For this reason, these methods are not included in the study presented in the current paper.

In [1], numerical studies at the popular Hemker problem, defined in [17], see also Example 3.2, are presented. From the methods considered in the current paper, only the AFC scheme with Kuzmin limiter and the finite volume Scharfetter–Gummel scheme, which is equivalent to the EAFE method according to [3], were included. They were the only stabilized methods that computed solutions without spurious oscillations. However, it was observed in [5] that the AFC scheme with Kuzmin limiter reduces errors in certain norms in an optimal way only on special meshes.

A numerical comparison of AFC schemes can be found in [8]. For the Hemker problem and another two-dimensional example that describes the transport of an impulse in a long channel, see Example 3.1, the AFC schemes with Kuzmin and BJK limiters were considered. The results with BJK limiter were often somewhat more accurate, but it took generally more iterations to solve the corresponding systems of nonlinear equations. In addition, a three-dimensional problem, which is a Hemker-type problem with space-dependent convection field, is briefly studied in [8]. Also the AFC scheme with BBK limiter was included. Results for just one grid are presented, which are in agreement with those for the two-dimensional problems solved using the Kuzmin and BJK limiters. The solution computed with the BBK limiter showed even somewhat more smearing than the solution obtained with the Kuzmin limiter.

The goal of [19] was to investigate solvers for the nonlinear problems arising from AFC schemes with Kuzmin and BJK limiters. For the scheme with Kuzmin limiter, P_1 and Q_1 finite elements were considered. Hemker problems in two and three dimensions and the three-dimensional Hemker-type problem from [8] were studied for mildly and strongly convection-dominated situations. It is concluded that it was usually much easier to solve the problems for the Kuzmin limiter than for the BJK limiter. In addition, it was often not possible to compute the solution of the scheme with BJK limiter in the strongly convection-dominated regime and on fine grids within the prescribed number of iterations. A method was identified that is recommended as solver for the nonlinear problems, both in two and three dimensions. This method, called *fixed point rhs*, was used in the numerical simulations for the current paper.

Numerical comparisons of the AFC scheme with Kuzmin limiter and the MUAS method are presented in [25], of course only for situations where these methods do not coincide. Such situations are grids in two dimensions that are not Delaunay and the reaction-dominated regime, in which the proof of the DMPs for the AFC scheme with Kuzmin limiter requires the use of mass lumping. It was demonstrated in [25] that the MUAS method outperformed in all these situations the AFC scheme with Kuzmin limiter.

There are two main topics of [20]. The first one consists in studying the behavior of the AFC methods with Kuzmin limiter, with BJK limiter, and the MUAS method on adaptively refined grids. And the second one is the correct incorporation of grids with hanging nodes in the methodology of algebraically stabilized schemes. Three standard two-dimensional examples from the literature, including the Hemker problem, were considered. It was shown that the AFC scheme with Kuzmin limiter does not converge on grids with conforming closure which are not Delaunay, if the numerical problem becomes locally diffusion-dominated, whereas the other two methods converged optimally. Again, it was observed that generally the most iterations for solving the nonlinear problem are required by the AFC scheme with BJK limiter.

A very brief study, at a two-dimensional example with layers and a bump, is presented in [7], which includes the linear and nonlinear upwind methods and the AFC methods with Kuzmin, BJK, and BBK limiters, respectively. The main conclusions of this study consist in the statements that all nonlinear methods compute much more accurate results than the linear upwind method and that there are only small differences in the results obtained with the nonlinear methods.

In [21], among others, a steady-state convection-diffusion problem in three dimensions was studied for the AFC schemes with Kuzmin, BJK, and MC limiters. The emphasis of the comparison was on solvers for the nonlinear problem, where the MPI parallelized version of the code ParMooN, which was used also for most of the numerical studies of the present paper, was utilized. It could be observed that the computing times for the method with BJK limiter were about one order of magnitude larger than for the other methods. However, the solution obtained with the Kuzmin limiter contained some overshoots.

In summary, there are already a number of numerical studies available in the literature. On the one hand, the AFC schemes with Kuzmin and BJK limiters have been compared already comprehensively and their principal behavior is well understood: the results obtained with the BJK limiter possess usually sharper layers if the nonlinear systems can be solved, and solving the nonlinear systems is generally easier with the Kuzmin limiter. On the other hand, comparisons of these two methods with the other methods described in Sections 2.1 and 2.2 are rare, and this paper tries to fill this gap to some extent.

3 Numerical Studies

The numerical studies consider several examples defined on two-dimensional domains. Most of the used methods were implemented in two codes, one of them ParMooN, see [38], in order to verify the computed solutions. The simulations were performed on HP Compute Servers *HPE Synergy 660 Gen10 4xXeon, Eighteen-Core 3100 MHz*.

A fixed point iteration was utilized for solving the nonlinear problems, concretely the method *fixed point rhs*, which turned out to be the most efficient method in the numerical studies in [19]. For the AFC scheme with BBK limiter, however, this method did not converge, and the method *fixed point matrix* from [19] was used instead. The fixed point iteration was stopped if the Euclidean norm of the residual vector times the square root of the number of degrees of freedom (including Dirichlet nodes), i.e. $\sqrt{N}$, was below 10^{-8}. If we could observe spurious

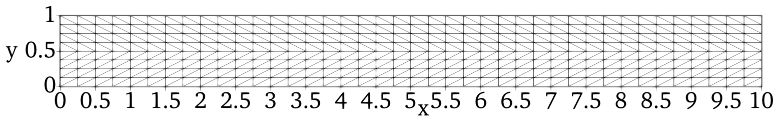

Figure 1: Example 3.1. Initial grid (level 0).

oscillations in the numerical solutions, then they were always only of the order of the stopping tolerance. Alternatively, the iteration was stopped after $\max_{\text{it}} = 10\,000$ iterations.

In neither of the two used codes all methods are implemented, and the concrete adaption of the damping parameter in the fixed point iteration differs. For these reasons, we decided to present computing times (instead of iteration numbers) to provide an impression on the efficiency of the methods. To be concrete, the computing times for the AFC schemes were those obtained with ParMooN and the times for the nonlinear upwind scheme with the other code. As solver for the linear systems of equations, the sparse direct solver UMFPACK, see [14], was applied. In addition, we explored as alternative an inexact solve using the restarted GMRES with at most 10 000 iterations and a Jacobi preconditioner. The presented computing times do not include the times for computing quantities of interest.

The only method that contains user-chosen parameters is the AFC scheme with BBK limiter. In the numerical studies, we tried to find for each example a good choice of these parameters.

Example 3.1 (Transport of an impulse) The reason for investigating briefly this example is the experience that algebraic stabilizations do not work satisfactory for this example, see [27] for the time-dependent equation and [8] for the steady-state problem. In [8], only visual presentations of numerical solutions are provided, and here a quantitative evaluation of the results will be given.

Let $\Omega = (0,10) \times (0,1)$, $\varepsilon = 10^{-10}$, $\boldsymbol{b} = (1,0)^T$, and $c = f = 0$. The inlet condition at $x = 0$ is an impulse defined by

$$u(0,y) = \begin{cases} 1 & \text{if } y \in [0.375, 0.625], \\ 0 & \text{else}, \end{cases}$$

at $y = 0$ and $y = 1$ the homogeneous Dirichlet boundary condition $u = 0$ was considered, and at the outlet, a homogeneous Neumann boundary condition was prescribed. With this setup, the impulse is transported from left to right and because of the very small diffusion coefficient, its value at the outlet point $(10, 0.5)$ should be very close to 1.

Figure 1 presents the initial grid, level 0. This grid is aligned with the convection field and it should allow an almost perfect transport of the impulse to the outlet.

A typical solution, for which a notable smearing of the impulse can be observed, is depicted in Figure 2. A quantitative comparison of the method using the value $|1 - u_h(10, 0.5)|$ is shown in Figure 3. It can be seen that there are considerable differences between the methods. The most accurate results were obtained for the nonlinear upwind method, the

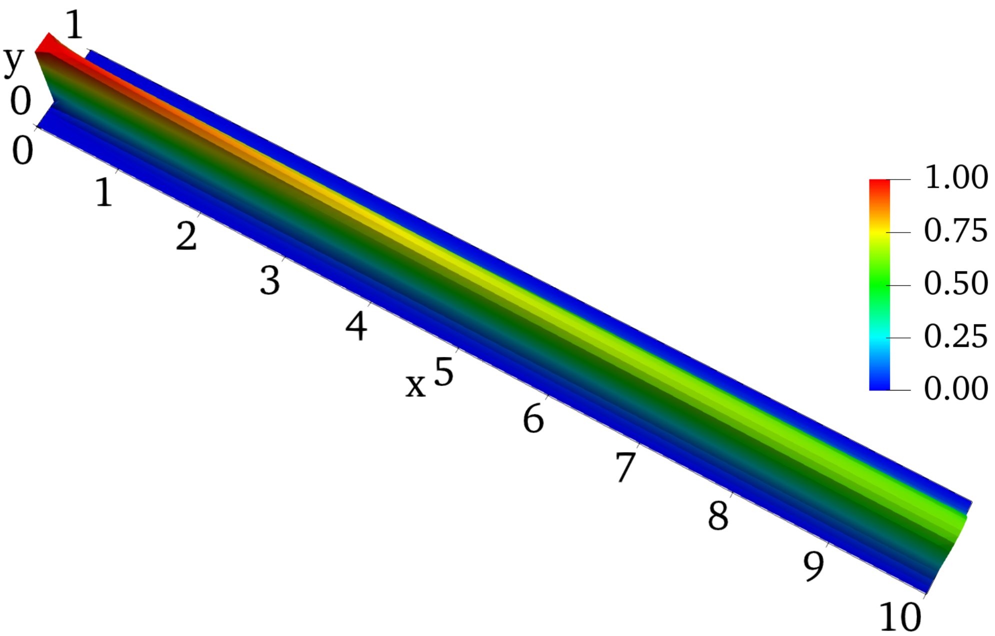

Figure 2: Example 3.1. Solution computed using the MUAS method, level 1.

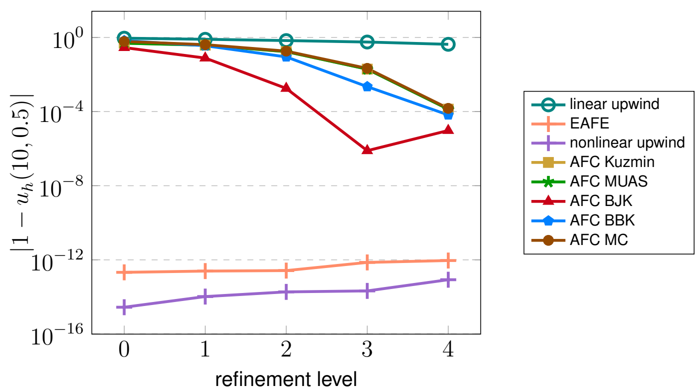

Figure 3: Example 3.1. Values $|1 - u_h(10, 0.5)|$ for the studied methods. The curves for the AFC schemes with Kuzmin and MC limiters and for the MUAS method are on top of each other.

EAFE scheme and the AFC scheme with BJK limiter. The AFC method with BBK limiter (with the parameter values $p = 8$ and $\gamma_0 = 0.2$) is slightly more accurate than the remaining AFC schemes. On the given grid, the AFC scheme with the Kuzmin limiter and the MUAS method coincide. The limiters of the AFC schemes are by construction for multi-dimensional situations and they do not identify the one-dimensional character of this example properly, leading to the introduction of unnecessary diffusion. The least accurate method, exhibiting the most smearing, is the linear upwind method.

Information on computing times and the behavior of the nonlinear solver are presented in Figure 4. In this case, it was more efficient to apply the sparse direct solver for the

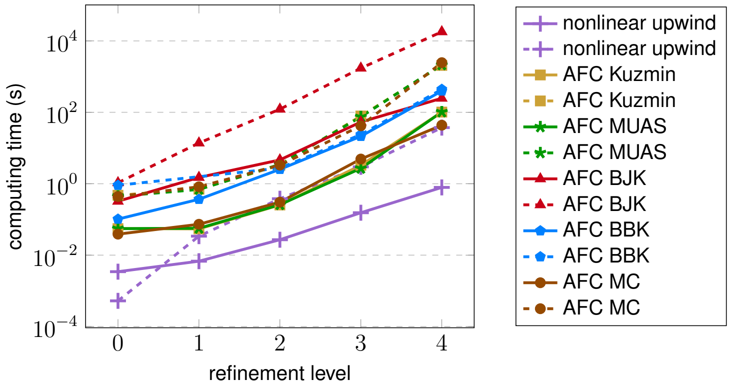

Figure 4: Example 3.1. Computing times with the direct (solid line) and iterative (dashed line) linear solver, respectively. The AFC scheme with BBK limiter and the iterative linear solver did not converge for the level 1. The AFC scheme with BJK limiter was stopped after reaching $\max_{\text{it}}$ for levels 3 and 4. The nonlinear upwind scheme reduces to a linear scheme for this example.

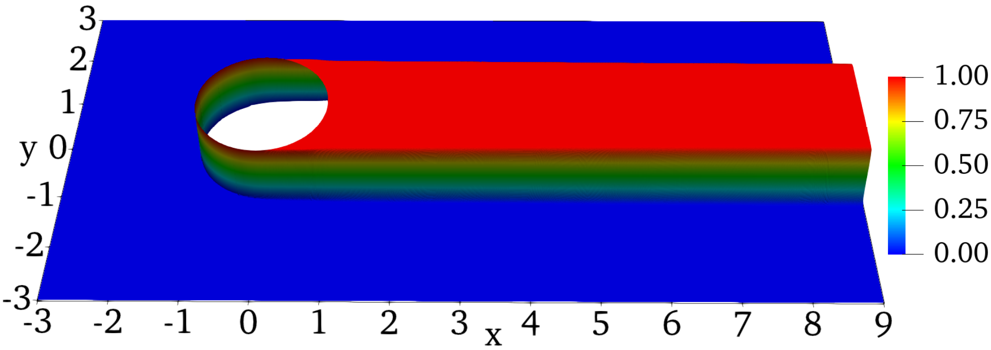

Figure 5: Example 3.2. Solution computed using the AFC scheme with MC limiter on level 5.

linear system of equations. Then, the problems for the AFC schemes with Kuzmin and MC limiters and for the MUAS method could be solved considerably faster than the systems for the AFC schemes with BBK and BJK limiters.

Example 3.2 (The Hemker problem) The Hemker problem is a standard benchmark problem for steady-state convection-diffusion equations whose setup was defined in [17]. The domain is given by $\Omega = \{(-3,9)\times(-3,3)\}\setminus\{(x,y)\ :\ x^2+y^2 \le 1\}$. In our simulations, the coefficients $\varepsilon = 10^{-4}$, $\boldsymbol{b} = (1,0)^T$, and $c = f = 0$ were used. Thus, the convection is from left to right. The boundary conditions were of Dirichlet type at $x = -3$, with $u = 0$, and at the circle, with $u = 1$. On all other boundaries, homogeneous Neumann boundary conditions were prescribed. Hence, the solution exhibits a boundary layer at the left-hand side of the circle, which continues to the top and bottom of the circle. From there, two inner layers start in the direction of the flow field, see Figure 5.

The initial grid for the simulations (level 0) is depicted in Figure 6. It was generated with

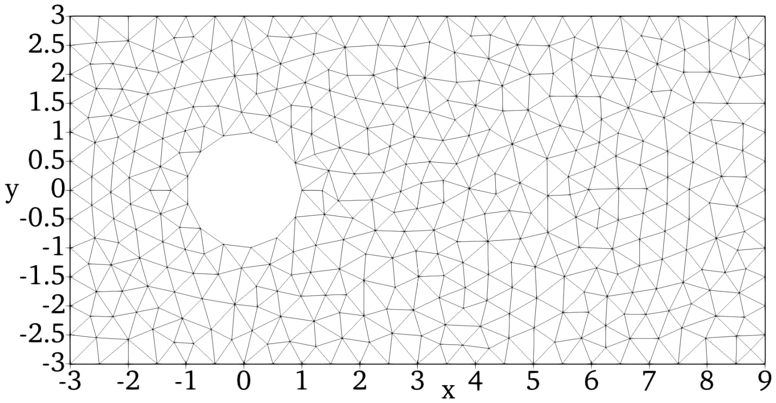

Figure 6: Example 3.2. Initial grid (level 0).

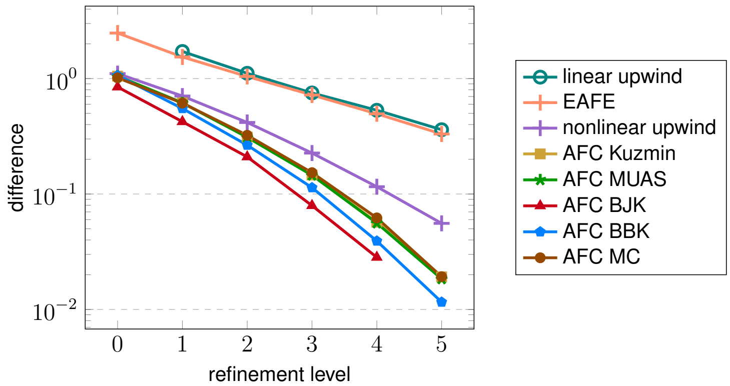

Figure 7: Example 3.2. Difference between the layer width of the computed solutions and the reference layer width. The solution computed via the linear upwind on level 0 is so smeared that the layer is not defined. The nonlinear problem for the AFC method with BJK limiter on level 5 could not be solved.

GMSH, [15], and it has the Delaunay property. For the chosen diffusion coefficient, there is a reference solution reported in [1]. In particular, a measure for the layer widths at $x = 4$, i.e., downwind of the circle, is defined and the reference value 0.0723 is provided. We utilized this feature to assess the accuracy of the studied methods, see Figure 7.

The numerical results presented in this figure show a clear difference between linear and nonlinear methods. The most accurate method is the AFC scheme with BJK limiter, followed on most levels by the AFC scheme with BBK limiter (with the parameter values $p = 8$ and $\gamma_0 = 0.6$). For this example, one can observe differences between the results computed with the AFC scheme with Kuzmin limiter and with the MUAS method on finer grids. Even if the initial grid is Delaunay, where both methods almost coincide in the diffusion-dominated case, the red refinement of a Delaunay grid is Delaunay only if all angles of the initial grid are non-obtuse. This situation is not given for the grid presented in Figure 6. The differences between both methods are very small. The least accurate methods are the linear upwind method and the EAFE scheme.

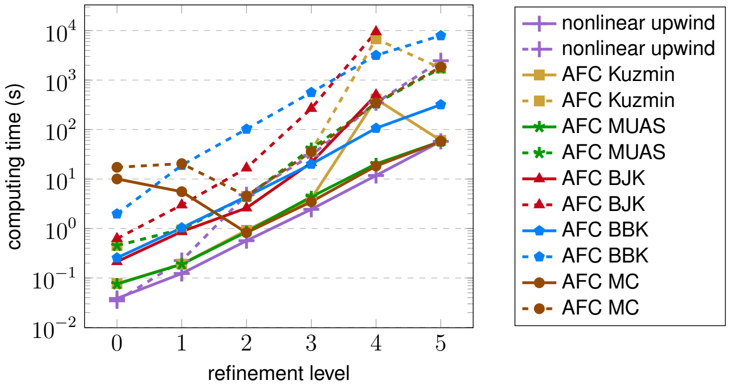

Figure 8: Example 3.2. Computing times with the direct (solid line) and iterative (dashed line) linear solver, respectively. The AFC scheme with BJK limiter did not converge for the level 5. The AFC scheme with MC limiter and the iterative linear solver was stopped after reaching $\max_{it}$ for level 0.

Figure 8 presents the computing times. Again, it can be seen that it was more efficient to use the sparse direct solver for solving the arising linear systems of equations. The simulations took similar times for the nonlinear upwind method, the AFC schemes with Kuzmin and MC limiters, and the MUAS method. Considerably longer computing times were necessary for the AFC methods with the other two limiters, even the failure of convergence for the AFC method with BJK limiter on the finest grid was observed.

Example 3.3 (Non-constant convection) This problem is defined in $\Omega = (0,1)^2$ with the convection field $\boldsymbol{b} = (-y, x)^T$ and with vanishing reaction field c and right-hand side f. There are Dirichlet boundary conditions on $[0,1] \times \{0\}$, $\{1\} \times [0,1]$, and $[0,1] \times \{1\}$:

$$u(x,y) = \begin{cases} 1 - \frac{1}{4}\left(1 - \cos\left(\frac{1/3+\xi-x}{2\xi}\pi\right)\right)^2 & x \in \left[\frac{1}{3} - \xi, \frac{1}{3} + \xi\right], y = 0, \\ 1 & x \in \left(\frac{1}{3} + \xi, \frac{2}{3} - \xi\right), y = 0, \\ 1 - \frac{1}{4}\left(1 - \cos\left(\frac{x-2/3+\xi}{2\xi}\pi\right)\right)^2 & x \in \left[\frac{2}{3} - \xi, \frac{2}{3} + \xi\right], y = 0, \\ 0 & \text{remaining Dirichlet boundary}, \end{cases}$$

with $\xi = 10^{-3}$. Thus, the inlet boundary condition possesses the form of a regularized step at the lower boundary. This inlet profile is transported to the outlet boundary $\{0\} \times (0,1)$, where a homogeneous Neumann boundary condition is prescribed. A sketch of the solution is presented in Figure 9.

For $\varepsilon = 10^{-5}$, we computed a reference solution with the Q_2 Galerkin finite element method on a uniform grid consisting of squares with edge length $1/2048$ (more than 67 million degrees of freedom). As in [23], the width of each layer at the outlet is defined to be the length of the interval with $u(0,y) \in [0.1, 0.9]$. To calculate the widths, the outlet

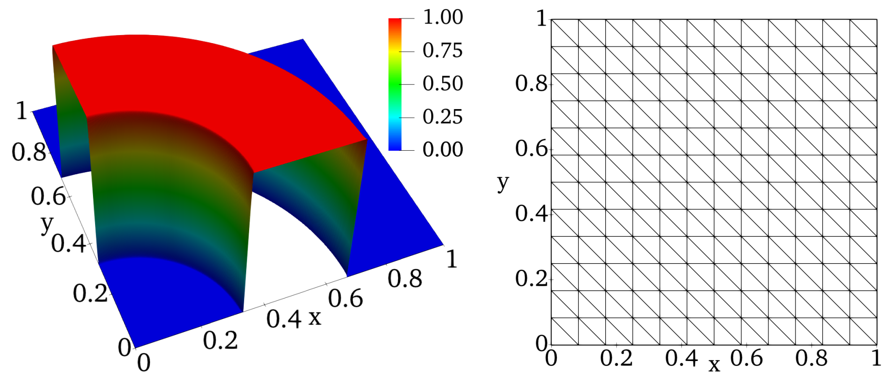

Figure 9: Example 3.3. Solution computed using the AFC scheme with BBK limiter on level 6 (left), initial grid (level 0) (right).

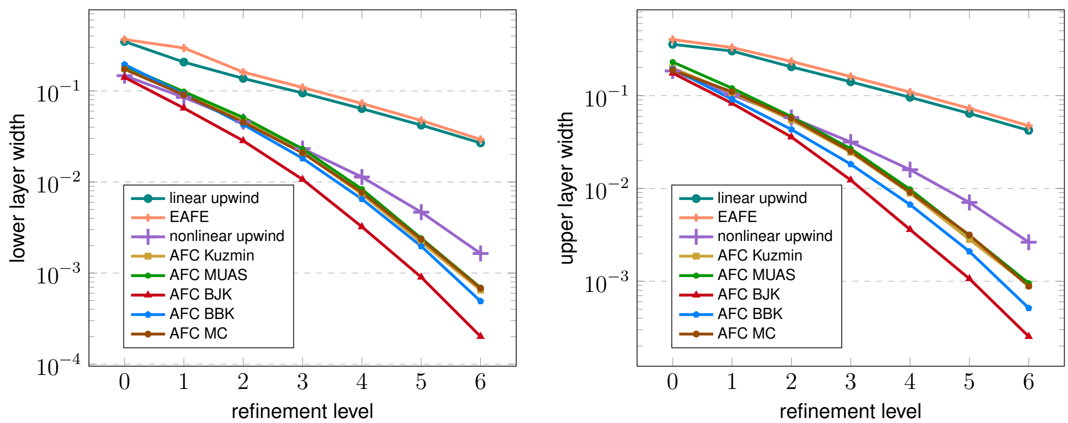

Figure 10: Example 3.3. Differences between the computed lower and upper layer widths and their reference values.

boundary was decomposed into 100000 equidistant intervals, the values of the solution at the nodes of this decomposition were computed and then linearly interpolated. Using this procedure, we also calculated the width of the outflow profile, which we defined as the length of the interval with $u(0, y) \geq 0.1$. Based on these results, the following reference values were obtained:

- width of the lower layer (left layer in Figure 9): 0.01439869,
- width of the upper layer (right layer in Figure 9): 0.01439637,
- outflow profile width: 0.3482541.

Comparing these values with the results obtained on a grid with $h = 1/1024$, we realized that at least the first three non-zero digits of these values are reliable.

Figure 9 presents the initial grid, level 0. The results with respect to the layer widths are presented in Figures 10 and 11. Their evaluation is similar as for the previous example. The results from the AFC method with BJK limiter are the most accurate ones, followed by the results obtained with the AFC method with the BBK limiter (for the parameter values $p = 8$ and $\gamma_0 = 0.6$). But the results with respect to the layer widths are usually only slightly better

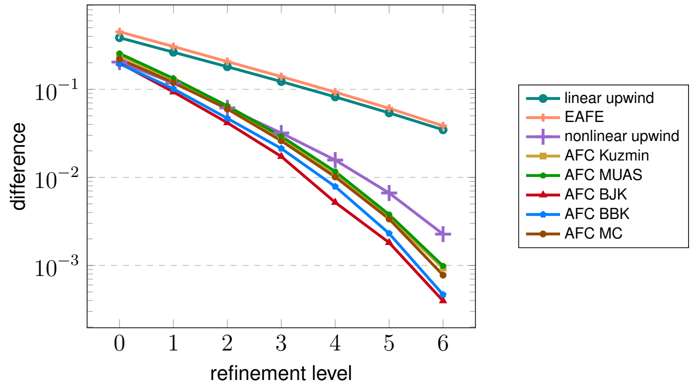

Figure 11: Example 3.3. Difference between the computed and the reference outflow profile widths.

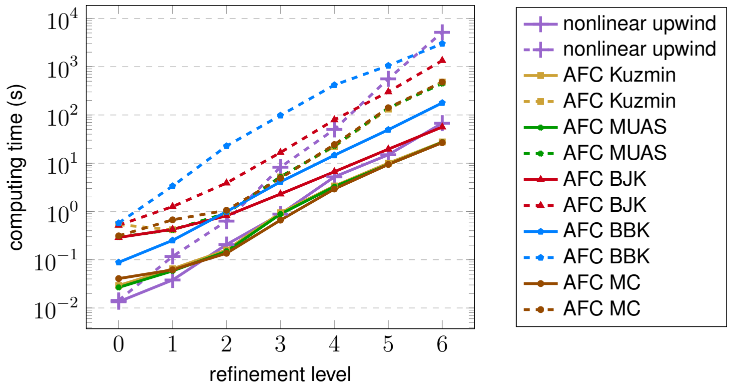

Figure 12: Example 3.3. Computing times, with the direct (solid line) and iterative (dashed line) linear solver, respectively.

than those for the other nonlinear schemes. Again, there is a great gap between the accuracy of the linear and nonlinear discretizations.

Also for this example, it turned out to be more efficient to use the direct solver for the linear systems of equations, compare Figure 12. On finer grids, the fastest methods were the AFC schemes with Kuzmin and MC limiters and the MUAS method. The longest computing times were needed for the AFC method with BBK limiter.

Example 3.4 (Solution with two interior layers caused by an inhomogeneous right-hand side) This example was considered in [24], where it proved to be quite hard for many stabilized methods. It is defined by $\Omega = (0,1)^2$, $\varepsilon = 10^{-8}$, $\boldsymbol{b} = (1,0)^T$, $c = 0$, and the piecewise polynomial right-hand side

$$f = \begin{cases} 16(1-2x) & \text{for } (x,y) \in [0.25, 0.75]^2, \\ 0 & \text{else.} \end{cases}$$

Homogeneous Dirichlet boundary conditions are prescribed on the whole boundary of Ω. The solution, see Figure 13 (left) for its numerical approximations, possesses two interior

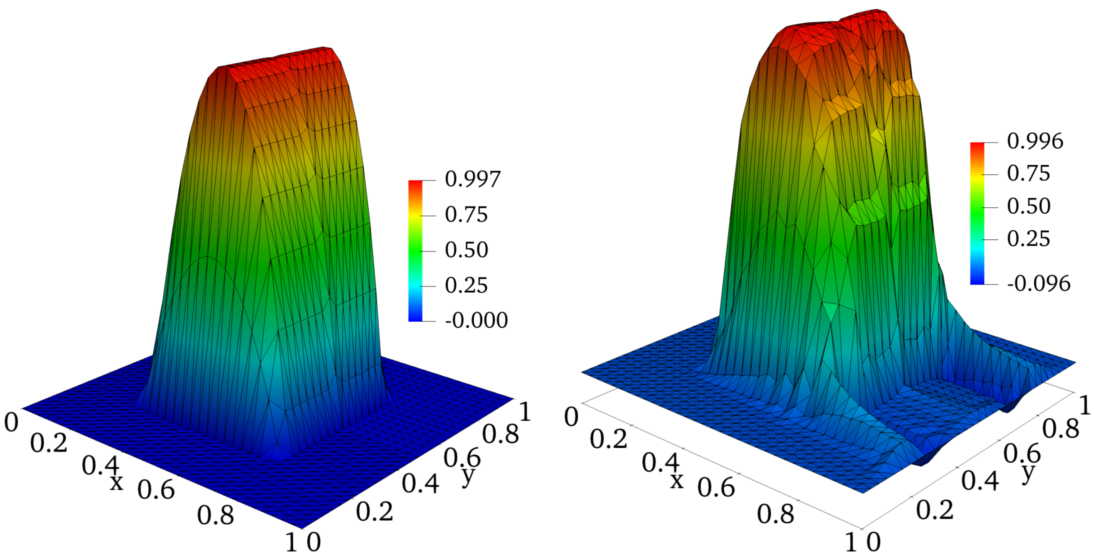

Figure 13: Example 3.4. Solutions computed using the EAFE scheme (left) and the AFC scheme with BJK limiter (right) on level 2.

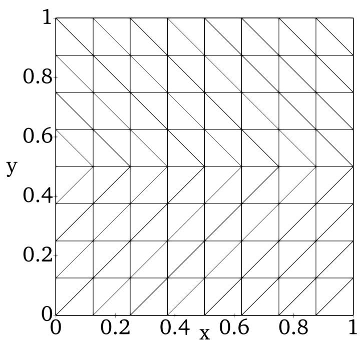

Figure 14: Example 3.4. Initial grid (level 0).

layers, at $(0.25, 0.75) \times \{0.25\}$ and at $(0.25, 0.75) \times \{0.75\}$. From the reduced equation, i.e., the equation for $\varepsilon = 0$, one finds that the solution should be very close to $(4x - 1)(3-4x)$ in $(0.25, 0.75)^2$. Since the right-hand side possesses negative values, a maximum principle cannot be applied to deduce that the solution is non-negative.

Because one can guess from the data of the problem that the solution might be of complicated shape at the lines where the right-hand side is discontinuous, grids should be used that take this situation into account, i.e., which have edges at these places. The initial grid utilized in our simulations is depicted in Figure 14.

As in [24], we used the values

$$\mathrm{osc} := - \min_{x\in[0.4,0.6]} u_h(x,y), \quad \mathrm{diff} := \max_{x\geq 0.8} u_h(x,y) - \min_{x\geq 0.8} u_h(x,y), \quad y \in [0,1],$$

to measure oscillations of the discrete solution u_h. These values are provided in Figure 15 for the different methods. One can observe that, for $x \in [0.4, 0.6]$, some of the methods

DOI 10.20347/WIAS.PREPRINT.2946

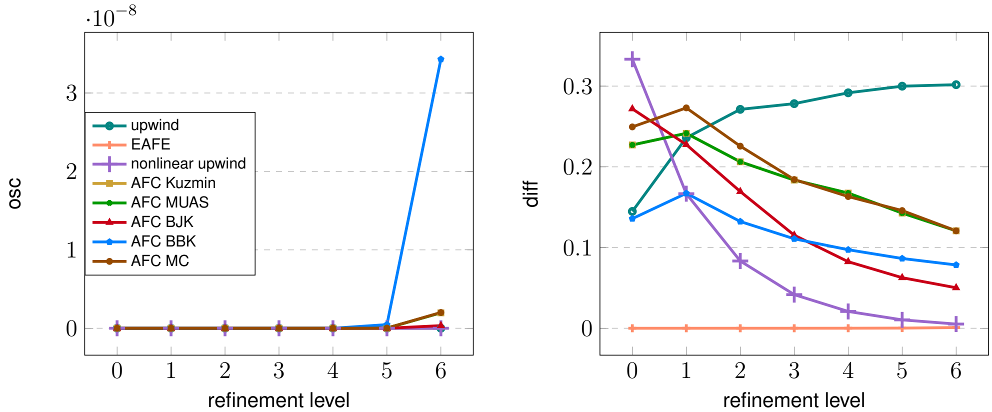

Figure 15: Example 3.4. Quantities of interest for the different methods.

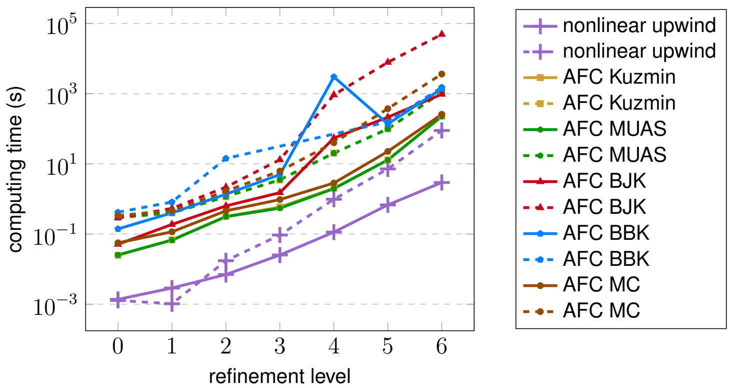

Figure 16: Example 3.4. Computing times with the direct (solid line) and iterative (dashed line) linear solver, respectively. The AFC scheme with BBK limiter and the iterative solver did not converge for the levels 3 and 4. The AFC scheme with BJK limiter was stopped after reaching $\max_{\mathrm{it}}$ for levels 4–6. The same applies for the AFC scheme with BBK limiter and the direct solver for level 4. The nonlinear upwind method reduces to a linear method for this example.

slightly violate the positivity on fine meshes; the largest violation is observed for the AFC scheme with BBK limiter (for the parameter values $p = 8$ and $\gamma_0 = 0.6$). Since the violations are of the order of the stopping criterion for solving the nonlinear problem, the positivity preservation for $x \in [0.4, 0.6]$ can be considered to be satisfied. On the contrary, for $x \geq 0.8$, almost all methods have severe problems to provide an accurate approximation of the solution, see Figure 13 (right) where a typical behavior for $x \geq 0.8$ can be seen. In this case, the most successful method is the EAFE scheme for which the numerical solution is close to zero for $x \geq 0.8$ on all meshes. The nonlinear upwind method, for which the value of diff linearly converges to zero, is the second best method.

With respect to the computing times, see Figure 16, similar conclusions as in the previous examples can be drawn. It should be briefly noted that one can see for the AFC method

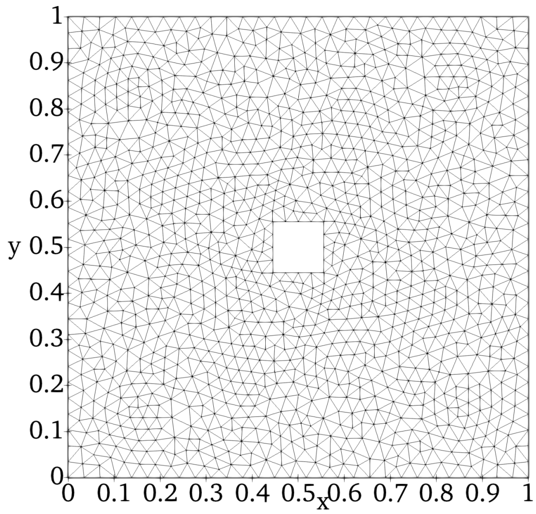

Figure 17: Example 3.5. Delaunay triangulation consisting of 2428 triangles, corresponding to 1280 degrees of freedoms.

with BBK limiter, level 4, that it can happen that the nonlinear scheme converges with the inexact solver for the linear systems of equations but does not with the direct solver.

Example 3.5 (Anisotropic and heterogeneous diffusion) Some of the studied methods are proved to satisfy the DMPs for rather general elliptic equations and on arbitrary simplicial grids. This example, taken from [34], demonstrates that these methods can indeed be used even for solving a diffusion problem where the diffusion tensor is anisotropic (not a multiple of the unit tensor) and heterogeneous (values depend on the position).

Let $\Omega = (0,1)^2 \setminus [4/9, 5/9]^2$. The diffusion tensor is given by

$$\mathbb{E} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1000 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad \theta = \pi \sin(x)\cos(y).$$

The other coefficients of problem (1) are given by $\boldsymbol{b} = \mathbf{0}$, $c = f = 0$. The problem is equipped with Dirichlet boundary conditions $u = 0$ at the outer boundary and $u = 2$ at the inner boundary. Hence, the solution satisfies a maximum principle.

It is shown in [34] that a standard Galerkin discretization applied on an unstructured grid, which seems to be of Delaunay type, leads to a solution that violates the global DMP. In particular, negative values are present in the numerical solution. Our numerical simulations utilized the grid depicted in Figure 17. The schemes that should work on arbitrary regular simplicial triangulations are the AFC schemes with BJK limiter and the MUAS method. Figure 18 presents the corresponding results. It can be seen that they are without spurious oscillations. The number of iterations for solving the nonlinear problem was 230 if the BJK limiter was used and 106 if the MUAS method was utilized.

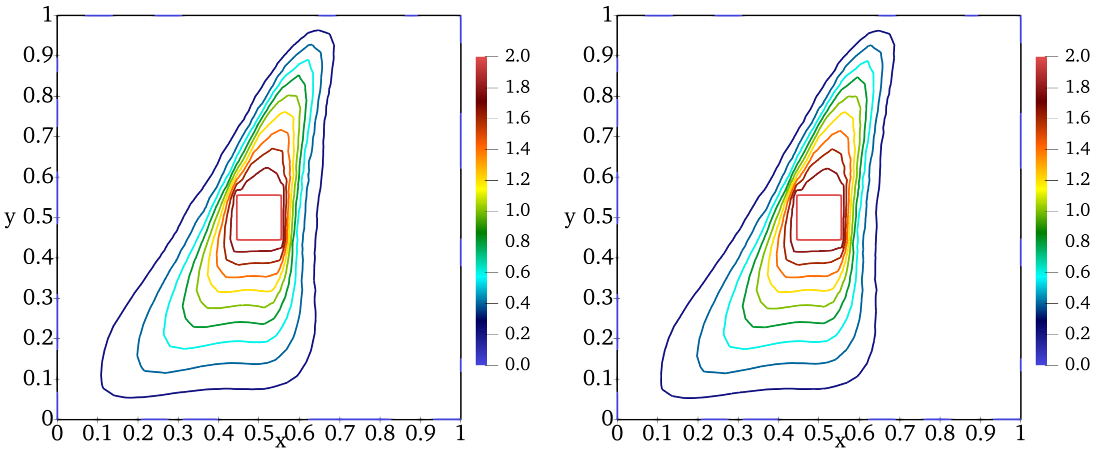

Figure 18: Example 3.5. Contour lines of solutions computed using the MUAS method (left) and using the AFC scheme with BJK limiter (right).

Table 1: Evaluation of the numerical results with respect to accuracy.

scheme	Ex 3.1	Ex. 3.2	Ex. 3.3	Ex. 3.4
linear upwind	$--$	$--$	$--$	$--$
EAFE	++	$--$	$--$	++
nonlinear upwind	++	○	○	+/++
AFC Kuzmin	○	+	+	○
AFC MUAS	○	+	+	○
AFC BJK	+	++	++	+
AFC BBK	○	++	+	+
AFC MC	○	+	+	○

4 Conclusions

A summary of the results with respect to accuracy is given in Table 1 and with respect to efficiency in Table 2. In these tables, '++' means very good, '+' good, '○' average, '$-$' sufficient, and '$--$' bad. The grades were given on the basis of our own conclusions from the presented results. With respect to efficiency, the computing times with the direct solver for the linear systems of equations were taken into account because this approach was almost always more efficient than the use of an iterative solver.

One can see that among the linear schemes, the EAFE method usually outperformed the linear upwind scheme. There is no example where the EAFE method was notably worse, and there are special examples for which the EAFE scheme fits particularly well. But in more general examples, the solutions computed with the EAFE scheme were quite inaccurate compared with the solutions from the nonlinear schemes. Among the latter schemes, the nonlinear upwind method was sometimes the most accurate, at the more special Ex-

Table 2: Evaluation of the numerical results with respect to computing time.

scheme	Ex 3.1	Ex. 3.2	Ex. 3.3	Ex. 3.4
linear upwind	++	++	++	++
EAFE	++	++	++	++
nonlinear upwind	++	+	○	++
AFC Kuzmin	○	+	+	○
AFC MUAS	○	+	+	○
AFC BJK	—	—	○	—
AFC BBK	—	○	—	—
AFC MC	○	+	+	○

amples 3.1 and 3.4. The most accurate solutions for more general problems were usually obtained with the AFC scheme with BJK limiter, followed by the AFC scheme with BBK limiter. However, the solution of the arising nonlinear systems was difficult for both methods, sometimes even not possible. In addition, the BBK limiter requires two user-chosen parameters, which we consider as a drawback. The other schemes showed a good balance of accuracy and efficiency. There are only minor differences in the results between them. Because of being able to satisfy DMPs on general simplicial grids, we recommend the use of the MUAS method instead of the AFC scheme with Kuzmin limiter. We prefer the MUAS method also in comparison with AFC scheme with MC limiter, since the theoretical background for convection-diffusion-reaction equations is currently further developed for the MUAS method.

In future, it is certainly worthwhile to extend the numerical studies to three-dimensional problems. Because, in our opinion, there is a lack of appropriate three-dimensional benchmark problems, this extension has to go along with the construction of such problems.

References

[1] M. Augustin, A. Caiazzo, A. Fiebach, J. Fuhrmann, V. John, A. Linke, and R. Umla. An assessment of discretizations for convection-dominated convection-diffusion equations. *Comput. Methods Appl. Mech. Engrg.*, 200(47-48):3395–3409, 2011.

[2] K. Baba and M. Tabata. On a conservative upwind finite element scheme for convective diffusion equations. *RAIRO Anal. Numér.*, 15(1):3–25, 1981.

[3] R. E. Bank, W. M. Coughran jr., and L. C. Cowsar. The finite volume Scharfetter-Gummel method for steady convection diffusion equations. *Comput. Vis. Sci.*, 1(3):123–136, 1998.

[4] G. R. Barrenechea, E. Burman, and F. Karakatsani. Edge-based nonlinear diffusion for finite element approximations of convection-diffusion equations and its relation to algebraic flux-correction schemes. *Numer. Math.*, 135(2):521–545, 2017.

[5] G. R. Barrenechea, V. John, and P. Knobloch. Analysis of algebraic flux correction schemes. *SIAM J. Numer. Anal.*, 54(4):2427–2451, 2016.

[6] G. R. Barrenechea, V. John, and P. Knobloch. An algebraic flux correction scheme satisfying the discrete maximum principle and linearity preservation on general meshes. *Math. Models Methods Appl. Sci.*, 27(3):525–548, 2017.

[7] G. R. Barrenechea, V. John, and P. Knobloch. Finite element methods respecting the discrete maximum principle for convection-diffusion equations. Technical Report 10.48550/ARXIV.2204.07480, arXiv, 2022.

[8] G. R. Barrenechea, V. John, P. Knobloch, and R. Rankin. A unified analysis of algebraic flux correction schemes for convection-diffusion equations. *SeMA J.*, 75(4):655–685, 2018.

[9] E. Burman and A. Ern. Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection–diffusion-reaction equation. *Comput. Methods Appl. Mech. Engrg.*, 191(35):3833–3855, 2002.

[10] E. Burman and A. Ern. Stabilized Galerkin approximation of convection-diffusion-reaction equations: discrete maximum principle and convergence. *Math. Comp.*, 74(252):1637–1652, 2005.

[11] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193(15-16):1437–1453, 2004.

[12] P. G. Ciarlet. Discrete maximum principle for finite-difference operators. *Aequationes Math.*, 4:338–352, 1970.

[13] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland, Amsterdam, 1978.

[14] T. A. Davis. Algorithm 832: UMFPACK V4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Software*, 30(2):196–199, 2004.

[15] C. Geuzaine and J.-F. Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. *Internat. J. Numer. Methods Engrg.*, 79(11):1309–1331, 2009.

[16] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Springer-Verlag, Berlin, 1983.

[17] P. W. Hemker. A singularly perturbed model problem for numerical computation. *J. Comput. Appl. Math.*, 76(1-2):277–285, 1996.

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

[19] A. Jha and V. John. A study of solvers for nonlinear AFC discretizations of convection-diffusion equations. *Comput. Math. Appl.*, 78(9):3117–3138, 2019.

[20] A. Jha, V. John, and P. Knobloch. Adaptive grids in the context of algebraic stabilizations for convection-diffusion-reaction equations. Technical Report 10.48550/ARXIV.2007.08405v4, arXiv, 2022.

[21] A. Jha, O. rej Pártl, N. Ahmed, and D. Kuzmin. An assessment of solvers for algebraically stabilized discretizations of convection-diffusion-reaction equations. *Journal of Numerical Mathematics*, in press, 2022.

[22] V. John and P. Knobloch. On spurious oscillations at layers diminishing (SOLD) methods for convection-diffusion equations. I. A review. *Comput. Methods Appl. Mech. Engrg.*, 196(17-20):2197–2215, 2007.

[23] V. John and P. Knobloch. On the performance of SOLD methods for convection-diffusion problems with interior layers. *Int. J. Comput. Sci. Math.*, 1(2-4):245–258, 2007.

[24] V. John and P. Knobloch. On spurious oscillations at layers diminishing (SOLD) methods for convection-diffusion equations. II. Analysis for P_1 and Q_1 finite elements. *Comput. Methods Appl. Mech. Engrg.*, 197(21-24):1997–2014, 2008.

[25] V. John and P. Knobloch. On algebraically stabilized schemes for convection-diffusion-reaction problems. Technical Report 10.48550/ARXIV.2111.08697, arXiv, 2021.

[26] V. John, T. Mitkova, M. Roland, K. Sundmacher, L. Tobiska, and A. Voigt. Simulations of population balance systems with one internal coordinate using finite element methods. *Chem. Eng. Sci.*, 64(4):733 – 741, 2009.

[27] V. John and J. Novo. On (essentially) non-oscillatory discretizations of evolutionary convection-diffusion equations. *J. Comput. Phys.*, 231(4):1570–1586, 2012.

[28] P. Knobloch. Improvements of the Mizukami-Hughes method for convection-diffusion equations. *Comput. Methods Appl. Mech. Engrg.*, 196(1-3):579–594, 2006.

[29] P. Knobloch. Application of the Mizukami–Hughes method to bilinear finite elements. In M. Beneš, M. Kimura, and T. Nakaki, editors, *Proceedings of Czech–Japanese Seminar in Applied Mathematics 2006*, volume 6 of *COE Lecture Note*, pages 137–147. Faculty of Mathematics, Kyushu University, 2007.

[30] P. Knobloch. Numerical solution of convection-diffusion equations using a nonlinear method of upwind type. *J. Sci. Comput.*, 43(3):454–470, 2010.

[31] P. Knobloch. On the discrete maximum principle for algebraic flux correction schemes with limiters of upwind type. In Z. Huang, M. Stynes, and Z. Zhang, editors, *Boundary and Interior Layers, Computational and Asymptotic Methods BAIL 2016*, volume 120 of *Lect. Notes Comput. Sci. Eng.*, pages 129–139. Springer-Verlag, Cham, 2017.

[32] D. Kuzmin. Algebraic flux correction for finite element discretizations of coupled systems. In M. Papadrakakis, E. Oñate, and B. Schrefler, editors, *Proceedings of the Int. Conf. on Computational Methods for Coupled Problems in Science and Engineering*, pages 1–5. CIMNE, Barcelona, 2007.

[33] D. Kuzmin. Monolithic convex limiting for continuous finite element discretizations of hyperbolic conservation laws. *Comput. Methods Appl. Mech. Engrg.*, 361:112804, 28 pages, 2020.

[34] X. Li and W. Huang. An anisotropic mesh adaptation method for the finite element solution of heterogeneous anisotropic diffusion problems. *J. Comput. Phys.*, 229(21):8072–8094, 2010.

[35] A. Mizukami and 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(2):181–193, 1985.

[36] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust numerical methods for singularly perturbed differential equations*. Springer-Verlag, Berlin, 2008.

[37] T. Vejchodský. *Discrete maximum principles*. Charles University in Prague, 2011. Habilitation Thesis.

[38] U. Wilbrandt, C. Bartsch, N. Ahmed, N. Alia, F. Anker, L. Blank, A. Caiazzo, S. Ganesan, S. Giere, G. Matthies, R. Meesala, A. Shamim, J. Venkatesan, and V. John. ParMooN—A modernized program package based on mapped finite elements. *Comput. Math. Appl.*, 74(1):74–88, 2017.

[39] J. Xu and L. Zikatanov. A monotone finite element scheme for convection-diffusion equations. *Math. Comp.*, 68(228):1429–1446, 1999.

[40] S. T. Zalesak. Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.*, 31(3):335–362, 1979.