An assessment of discretizations for convection-dominated convection-diffusion equations

Matthias Augustin [1]

University of Kaiserslautern, Department of Mathematics Geomathematics Group, Building 49, P.O. Box 30 49, 67653 Kaiserslautern, Germany

Alfonso Caiazzo [2] André Fiebach [3] Jürgen Fuhrmann [4]

Weierstrass Institute for Applied Analysis and Stochastics, Leibniz Institute in Forschungsverbund Berlin e. V. (WIAS), Mohrenstr. 39, 10117 Berlin, Germany

Volker John [5]

Weierstrass Institute for Applied Analysis and Stochastics, Leibniz Institute in Forschungsverbund Berlin e. V. (WIAS), Mohrenstr. 39, 10117 Berlin, Germany, and Free University of Berlin, Department of Mathematics and Computer Science, Arnimallee 6, 14195 Berlin, Germany

Alexander Linke [6]

Weierstrass Institute for Applied Analysis and Stochastics, Leibniz Institute in Forschungsverbund Berlin e. V. (WIAS), Mohrenstr. 39, 10117 Berlin, Germany

Rudolf Umla [7]

Department of Earth Science and Engineering, South Kensington Campus, Imperial College London, SW7 2AZ, UK

Abstract

The performance of several numerical schemes for discretizing convection-dominated convection-diffusion equations will be investigated with respect to accuracy and efficiency. Accuracy is considered in measures which are of interest in applications. The study includes an exponentially fitted finite volume scheme, the Streamline-Upwind Petrov–Galerkin (SUPG) finite element method, a spurious oscillations at layers diminishing (SOLD) finite element method, a finite element method with continuous interior penalty (CIP) stabilization, a discontinuous Galerkin (DG) finite element method, and a total variation diminishing finite element method (FEMTVD). A detailed assessment of the schemes based on the Hemker example will be presented.

Preprint submitted to Elsevier Science *8 August 2011*

Key words: dominating convection, exponentially fitted finite volume scheme, stabilized finite element methods, Hemker problem

1 Introduction

Scalar convection-diffusion equations model the transport of species by diffusion and convection. In many applications, convection is larger by orders of magnitude than diffusion, which is challenging from the numerical point of view. Therefore, it is crucial to identify appropriate methods for the accurate and efficient numerical solution of convection-dominated convection-diffusion equations.

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a domain with boundary $\partial\Omega$. A steady-state linear scalar convection-diffusion equation has the form

$$-\varepsilon\Delta u + \mathbf{b} \cdot \nabla u = f \quad \text{in } \Omega, \qquad u = u_b \quad \text{on } \partial\Omega, \tag{1}$$

where $\varepsilon > 0$ is the diffusion coefficient and $\mathbf{b}(\mathbf{x})$ is the convection field. For simplicity of presentation, the equation is equipped with Dirichlet conditions on the whole boundary.

Note that reaction is not included in (1). It is well known that dominating reaction leads to different instabilities than dominating diffusion [12,16]. This paper will study only discretizations which were developed for the convection-dominated regime.

The development and analysis of numerical schemes for solving equations of form (1) in the case of dominant convection have already a long tradition. Overviews of the state of the art for many approaches can be found in [33,39]. However, there is still no method that has been proven to be a universal choice in applications. This unsatisfactory situation stimulated considerably the research during the last decade, and a number of methods have been proposed or studied in detail. Several methods possess additional computational overhead, e.g., due to the solution of non-linear equations or to the use of extended matrix stencils. Sometimes, methods combine concepts from different approaches

[1] email: augustin@mathematik.uni-kl.de
[2] email: caiazzo@wias-berlin.de
[3] email: fiebach@wias-berlin.de
[4] email: fuhrmann@wias-berlin.de
[5] corresponding author, email: john@wias-berlin.de
[6] email: linke@wias-berlin.de
[7] email: r.umla09@imperial.ac.uk

for discretizing partial differential equations, like finite volume methods and finite element methods.

Numerical analysis considers generally the accuracy of discretizations with respect to certain norms of vector or function spaces. However, in practical applications, such norms are often of minor interest, while other properties, like positivity preservation or mass conservation, become much more important. If the unknown quantity is a concentration or a density, then a method that does not guarantee positiveness, e.g. due to spurious oscillations (undershoots), is often of little usefulness in practice. If a given process conserves mass or if the equation fulfills a maximum principle, then the choice of a numerical method in applications is often motivated by the desire that the discrete equation should inherit these important properties, or at least, approximate them well. For linear discretizations, the preservation of qualitative properties usually restricts the freedom in the choice of the used mesh family, which can pose severe difficulties for mesh generators. This aspect is one reason to consider also non-linear schemes. Last but not least, the costs of a numerical scheme are of interest in applications. Altogether, the wishes on numerical schemes include high accuracy (with respect to appropriate measures), the presence of properties from physics in the discrete solution, and reasonable computational costs.

The current paper studies several discretizations which are based on the finite volume or the finite element methodology.

As finite volume scheme, an exponentially fitted scheme is considered [13], which is sometimes called Scharfetter–Gummel scheme. Based on a Delaunay triangulation, fluxes between control volumes (Voronoi boxes) are computed by an application of the one-dimensional Il'in–Allen–Southwell formula [33].

The most popular finite element method for solving (1) is certainly the Streamline-Upwind Petrov–Galerkin (SUPG) or Streamline-Diffusion finite element method (SDFEM), introduced in [4,18]. This method adds artificial diffusion in streamline direction by means of a residual-based stabilization term.

Finite element schemes that add to the SUPG scheme another stabilization term acting in crosswind direction are called shock capturing or, more precisely, Spurious Oscillations at Layers Diminishing (SOLD) methods. These methods are in general non-linear. The numerical investigations include one of the best SOLD methods from recent studies for the P_1 and Q_1 finite element [21,23], which was proposed in [29]. This method will be used here also with higher order finite elements.

An alternative to residual-based stabilizations are continuous interior penalty (CIP) methods. These methods achieve stability by penalizing the jumps of the first derivative of the computed solution across faces of the mesh cells.

They have been proposed in [9] and analyzed in [7].

Discontinuous Galerkin (DG) methods combine ideas from finite elements (variational formulation, piecewise polynomial solution) and from finite volumes (discontinuous solution). A stabilization is introduced in these methods by the localization of the ansatz and test functions. In the presented numerical studies, a method which was proposed in [28] will be included.

Last, a total variation diminishing finite element method (FEMTVD) proposed in [30] will be studied. Also this scheme combines ideas from finite element methods (variational formulation, piecewise polynomial solution) and from finite volume methods (consideration of fluxes). In contrast to the finite element schemes mentioned above, the FEMTVD scheme introduces the stabilization by manipulations at the algebraic level (matrices and vectors) and not by modifying the bilinear form or the finite element spaces.

There are even more proposals of stabilized methods, see [33] for an overview, which are not included in the studies presented in this paper. Among these methods are Galerkin least squares methods, residual-free bubble finite element methods, local projection stabilization schemes, and methods whose construction requires extensive a priori knowledge about the solution. Schemes from the last class, like the tailored finite point method [15] or layer-adapted schemes with Shishkin or Bakhvalov meshes, are from our point of view unsuitable to be used in applications. However, in our opinion, the considered schemes include most of the important finite volume and finite element approaches for solving steady-state convection-dominated scalar equations.

The considered schemes were studied at a number of examples and a large amount of data was collected. Of course, the assessment of a multitude of different discretizations with respect to the properties given above cannot be done comprehensively in a single paper of reasonable length. Some characteristic results had to be selected for presentation. Instead of discussing several examples shortly, we decided to present the assessment of the methods at one example in detail. This example, the so-called Hemker problem [17], allows the physical interpretation of a heat flux from a circular body in the direction of the convection. It possesses several features which are often present in applications, like non-straight boundaries, a boundary layer, and interior layers. During the evaluation of the results, it turned out that the advantages and drawbacks of the considered discretizations can be highlighted quite well with this example. References to numerical studies at other examples will be provided in the description of the methods.

The paper is organized as follows. Section 2 will introduce the considered discretizations. The numerical studies are presented in Section 3 and a summary is given in Section 4.

2 The Studied Stabilized Discretizations

2.1 An Exponentially Fitted Voronoi Box Finite Volume Method

Consider a boundary conforming Delaunay triangulation [38] of Ω into simplices. The vertices of this triangulation are denoted by $\{\mathbf{x}_i\}_{i=1}^N$. For the studied finite volume discretization, a secondary grid of control volumes $\{V_i\}_{i=1}^N$ is constructed. The control volume V_i around the vertex $\mathbf{x}_i$, also called the Voronoi box, is defined by

$$V_i = \{\mathbf{x} \in \overline{\Omega} \ : \ |\mathbf{x} - \mathbf{x}_i| < |\mathbf{x} - \mathbf{x}_j|, \ 1 \leq j \leq N, j \neq i\},$$

where $|\cdot|$ denotes the Euclidean norm of a vector. The boundary conforming Delaunay property of the triangulation allows the explicit construction of V_i. If $\mathbf{x}_i$ is situated in the interior of Ω, it suffices to connect the circumcenters of the triangles adjacent to $\mathbf{x}_i$, see Fig. 1. If $\mathbf{x}_i$ is situated at the boundary, the lines connecting $\mathbf{x}_i$ with the adjacent edge midpoints, and the lines connecting these edge midpoints with the circumcenters of the adjacent triangles are used to describe the part of ∂V_i which belongs to $\partial\Omega$, see Fig. 1. Note that the property of the triangulation of being boundary conforming Delaunay is weaker than the condition to be weakly acute.

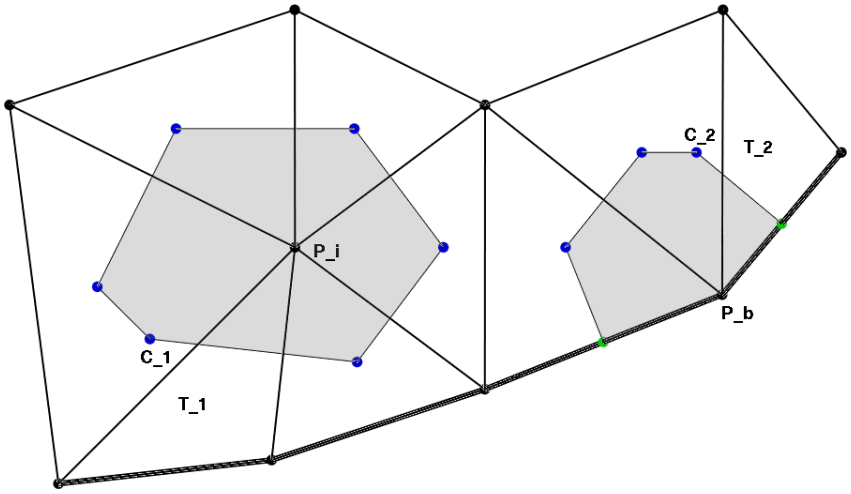

Fig. 1. Snapshot of a boundary conforming Delaunay grid. The boundary is marked by a bold line. Voronoi boxes are shown around an interior point `P_i` and a boundary point `P_b`. Note that the circumcenters `C_1` and `C_2` are situated outside their corresponding triangles `T_1` and `T_2`.

As the circumcenter of a triangle is the intersection point of its mid-perpendiculars, one obtains that the edge connecting two neighboring points $\mathbf{x}_i$ and $\mathbf{x}_j$ of the triangulation is orthogonal to the corresponding part of boundary of the Voronoi box. This fact allows to identify the normal direction of this boundary with the direction of the line connecting $\mathbf{x}_i$ and $\mathbf{x}_j$.

To define the discrete formulation, equation (1) is rewritten in divergence form

$$\nabla \cdot (-\varepsilon \nabla u + \mathbf{b}u) + \tilde{c}u = f \quad \text{in } \Omega, \qquad u = u_b \quad \text{on } \partial\Omega. \tag{2}$$

Clearly, (2) and (1) are identical for $\tilde{c} = -\nabla \cdot \mathbf{b}$.

The considered finite volume method (FVM) approximates the solution of (2) by a piecewise constant function, whose degrees of freedom are assigned to the vertices $\{\mathbf{x}_i\}_{i=1}^N$. Integrating (2) on Ω gives

$$\begin{aligned}
\int_\Omega (\nabla \cdot (-\varepsilon \nabla u + \mathbf{b}u) + \tilde{c}u)\, d\mathbf{x} &= \int_\Omega f \; d\mathbf{x}, \\
\sum_{i=1}^N \int_{V_i} (\nabla \cdot (-\varepsilon \nabla u + \mathbf{b}u) + \tilde{c}u)\, d\mathbf{x} &= \sum_{i=1}^N \int_{V_i} f \; d\mathbf{x}, \\
\sum_{i=1}^N \int_{\partial V_i} (-\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n}_i \; d\gamma + \sum_{i=1}^N \int_{V_i} \tilde{c}u \; d\mathbf{x} &= \sum_{i=1}^N \int_{V_i} f \; d\mathbf{x}.
\end{aligned} \tag{3}$$

The volume integrals in (3) are approximated by a simple quadrature rule

$$\int_{V_i} \tilde{c}u \; d\mathbf{x} \approx \tilde{c}(\mathbf{x}_i) u(\mathbf{x}_i) |V_i|, \quad \int_{V_i} f \; d\mathbf{x} \approx f(\mathbf{x}_i)|V_i|,$$

with $|V_i|$ being the measure of V_i. Denote by γ_{ij} the planar part of the boundary of the control volume between $\mathbf{x}_i$ and its neighbor $\mathbf{x}_j$. By construction, γ_{ij} is perpendicular to $\mathbf{h}_{ij} = \mathbf{x}_j - \mathbf{x}_i$. Hence, a unit normal on γ_{ij} is given by $\mathbf{n}_{ij} = \mathbf{h}_{ij}/|\mathbf{h}_{ij}|$. The intersection of $\mathbf{h}_{ij}$ and γ_{ij} is called $\mathbf{x}_{ij}$. In a first step, the integrals on the boundaries of the control volumes in (3) are approximated by a simple quadrature rule

$$\begin{aligned}
\int_{\partial V_i} (-\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n}_i \; d\gamma &= \sum_{j=1}^{N_i} \int_{\gamma_{ij}} (-\varepsilon \nabla u + \mathbf{b}u) \cdot \mathbf{n}_{ij} \; d\gamma \\
&\approx -\varepsilon \frac{|\gamma_{ij}|}{|\mathbf{h}_{ij}|} (u(\mathbf{x}_j) - u(\mathbf{x}_i)) + \frac{|\gamma_{ij}|}{|\mathbf{h}_{ij}|} (\mathbf{b}u)(\mathbf{x}_{ij}) \cdot \mathbf{h}_{ij}.
\end{aligned}$$

In a second step, $(\mathbf{b}u)(\mathbf{x}_{ij}) \cdot \mathbf{h}_{ij}$ has to be approximated by a numerical flux. This approximation is essentially a one-dimensional problem defined on $\mathbf{h}_{ij}$. The considered exponentially fitted scheme treats convection and diffusion together. It is obtained with the help of the Bernoulli function $B(\xi) = \xi/(\exp(\xi) - 1)$. The approximation reads as follows

$$\varepsilon(u(\mathbf{x}_j) - u(\mathbf{x}_i)) + (\mathbf{b}u)(\mathbf{x}_{ij}) \cdot \mathbf{h}_{ij} \approx \varepsilon \Big(B(-2Pe(\mathbf{x}_{ij}))u(\mathbf{x}_i) - B(2Pe(\mathbf{x}_{ij}))u(\mathbf{x}_j) \Big),$$

where

$$Pe(\mathbf{x}) = \frac{b_{ij}|\mathbf{h}_{ij}|}{2\varepsilon}$$

is the signed local Péclet number and

$$b_{ij} = \frac{1}{|\gamma_{ij}|} \int_{\gamma_{ij}} \mathbf{b} \cdot \mathbf{n}_{ij} \, d\gamma$$

is the average normal flux of $\mathbf{b}$ through γ_{ij}.

This scheme can be derived from the solution of a two point boundary value problem involving convection and diffusion terms projected on the grid edge joining $\mathbf{x}_i$ and $\mathbf{x}_j$, see [10]. It corresponds in one dimension to the Il'in–Allen–Southwell finite difference scheme [1,20,33], also called, within the community of semiconductor device simulations, Scharfetter–Gummel scheme [34].

Dirichlet boundary values are set directly in the vertices at the boundary. In the case of homogeneous Neumann boundary conditions, this approach yields a linear system of equations with a matrix A which has column sum zero, non-positive off-diagonal entries, and non-negative main diagonal entries, and whose graph is connected. Setting at least one Dirichlet boundary condition makes this matrix diagonally dominant, finally resulting in the M-matrix property. Furthermore, if $\mathbf{b}$ is divergence-free in all interior nodes, all rows of the matrix have sum zero such that a local maximum principle holds [13].

The foundations of a general convergence theory for finite volume methods for elliptic problems, including the method considered here, were laid in [11]. A short survey of literature, information on orders of convergence, and numerical experiments can be found in [14]. Here, it should be only mentioned that in general, for H^2 regular problems, the exponentially fitted scheme is second order convergent on square meshes in the discrete L^2 norm. Experimental evidence [14] shows that for a number of cases on triangular meshes, the exponentially fitted scheme is second order convergent as well.

2.2 The SUPG Finite Element Method

Finite element methods are based on the variational form of the underlying equation. Consider a convection-diffusion equation of form (1) and let $\tilde{u}_b \in H^1(\Omega)$ be an extension of the boundary condition u_b. Multiplying (1) with a test function $v \in V = H_0^1(\Omega)$, integrating on Ω, and applying integration by parts to the diffusive term lead to the variational problem: Find $u \in H^1(\Omega)$ such that $u - \tilde{u}_b \in V$ and

$$a(u, v) = (f, v) \quad \forall v \in V \tag{4}$$

with

$$a(u, v) = \varepsilon(\nabla u, \nabla v) + (\mathbf{b} \cdot \nabla u, v).$$

The Galerkin finite element discretization is obtained from (4) by replacing the space V by a finite element subspace V_h and by approximating $\tilde{u}_b$ by a finite element interpolant $\tilde{u}_{bh}$. Consider for simplicity a conforming finite element method, i.e. $V_h \subset V$. Then, the Galerkin finite element method reads: Find $u_h \in H^1(\Omega)$ such that $u_h - \tilde{u}_{bh} \in V_h$ and

$$a(u_h, v_h) = (f, v_h) \quad \forall v_h \in V_h.$$

It is well known that in the convection-dominated regime, the Galerkin finite element method often leads to solutions that are globally polluted with spurious oscillations. A stabilization of this method is necessary.

The SUPG method [4,18] is one of the most popular stabilized finite element methods. Basically, this method adds diffusion in the direction of the streamlines to the Galerkin finite element method. It reads as follows: Find $u_h \in H^1(\Omega)$ such that $u_h - \tilde{u}_{bh} \in V_h$ and

$$a(u_h, v_h) + \sum_{K \in \mathcal{T}_h} (R_h(u_h), \delta_K \mathbf{b} \cdot \nabla v_h)_K = (f, v_h) \quad \forall v_h \in V_h, \tag{5}$$

where $\mathcal{T}_h$ denotes the triangulation of Ω, $\{K\}$ are the mesh cells,

$$R_h(u_h) = -\varepsilon \Delta_h u_h + \mathbf{b} \cdot \nabla_h u_h - f$$

is the residual of the strong form of the equation, the index h at the differential operators denotes their restriction to a mesh cell K, δ_K are the stabilization parameters, and $(\cdot, \cdot)_K$ denotes the $L^2(K)$ inner product. Obviously, the SUPG method is a consistent, residual-based stabilization.

Results concerning the numerical analysis of the SUPG method are summarized in [33]. The analysis gives guidelines for the choice of the stabilization parameters δ_K. Several possible choices that can be used in practice were discussed in detail in [21]. In the numerical studies presented in Section 3, the same type of stabilization parameter is used as in [21]

$$\delta_K(\mathbf{x}) = \frac{\bar{h}_K}{2p|\mathbf{b}(\mathbf{x})|} \xi(Pe_K(\mathbf{x})), \quad Pe_K(\mathbf{x}) = \frac{|\mathbf{b}(\mathbf{x})|\bar{h}_K}{2p\varepsilon}, \quad \xi(\alpha) = \coth \alpha - \frac{1}{\alpha},$$

where $\bar{h}_K$ is an approximation of the length of the mesh cell K in the direction of the convection, see [21] for details, Pe_K is the local mesh cell Péclet number, and p is the degree of the local finite element space.

The properties of solutions computed with the SUPG method are well known: the computed layers are quite sharp, but non-negligible spurious oscillations (under- and overshoots) appear in a vicinity of the layers. In particular, positivity is not preserved which is sometimes a severe drawback of this method in the simulation of problems arising in physics or chemistry, see, e.g., [27].

Because of the observation that solutions computed with the SUPG method often possess spurious oscillations in a vicinity of layers, a number of Spurious Oscillations at Layers Diminishing (SOLD) methods have been developed, starting with [19]. One can distinguish between isotropic and anisotropic SOLD methods. The main idea of these methods consists in adding a stabilization term to the SUPG method (5) that introduces also diffusion orthogonal to the streamlines, so-called crosswind diffusion. Generally, this term is non-linear. A critical survey of these methods and a number of numerical studies for the P_1 and Q_1 finite element can be found in [21,22,23]. These studies showed that many SOLD methods, on the one hand, improve the accuracy of the solutions, compared with the SUPG solution. But, on the other hand, none of the proposed SOLD methods could universally compute solutions without the undesirable features of SUPG solutions. The numerical studies presented here will include one of the anisotropic SOLD methods that has been proven to be among the best SOLD methods in [21,22,23].

This method, proposed in [29] and modified in [21] (method KLR02_3 in [21], C93 in [22]), adds a non-linear crosswind diffusion to the SUPG method (5). It has the form: Find $u_h \in H^1(\Omega)$ such that $u_h - \tilde{u}_{bh} \in V_h$ and

$$a(u_h, v_h) + \sum_{K \in \mathcal{T}_h} (R_h(u_h), \delta_K \mathbf{b} \cdot \nabla v_h)_K + (\tilde{\varepsilon} D \nabla u_h, \nabla v_h) = (f, v_h) \quad \forall v_h \in V_h,$$

with

$$D = \begin{cases} I - \dfrac{\mathbf{b} \otimes \mathbf{b}}{|\mathbf{b}|^2} & \text{if } \mathbf{b} \neq \mathbf{0}, \\ 0 & \text{if } \mathbf{b} = \mathbf{0}, \end{cases}$$

I being the identity tensor, and

$$\tilde{\varepsilon} = \max \left\{ 0, \sigma_{\text{sold}} \frac{\text{diam}(K)|R_h(u_h)|}{2|\nabla u_h|} - \varepsilon \right\}.$$

The parameter σ_{sold} has to be chosen by the user. Large values of σ_{sold} introduce a rather large amount of crosswind diffusion and increase the non-linearity of the discrete equation. In this case, numerical studies in [23] showed that iterative schemes for solving the non-linear equation need more iterations (or even fail to converge) and the computational overhead increases. Moreover, it was shown in [23] that a constant choice of σ_{sold} is in general not optimal. Approaches for choosing the parameter appropriately in a non-constant way are just under development [24]. To our best knowledge, comprehensive numerical studies with this SOLD method and higher order finite elements cannot be found so far in the literature.

Continuous Interior Penalty (CIP) methods have been studied in detail during the last decade for several types of equations, e.g., see [5,7]. The basic idea of these methods consists in the penalization of discontinuities across faces of the first derivative of the computed solution.

The formulation of the considered CIP method reads as follows: Find $u_h \in H^1(\Omega)$ such that $u_h - \tilde{u}_{bh} \in V_h$ and

$$a(u_h, v_h) + \sum_{E \in \mathcal{E}_h} \sigma_{\text{cip}} h_E^2 \left(\mathbf{b} \cdot [\nabla u_h]_E, \mathbf{b} \cdot [\nabla v_h]_E\right)_E = (f, v_h) \quad \forall v_h \in V_h, \qquad (6)$$

where $\mathcal{E}_h$ is the set of all interior faces, h_E is a measure of face E (length of edge E in 2D), σ_{cip} is a user-chosen parameter, $[\cdot]_E$ denotes the jump of a function across E in the direction of the unit normal $\mathbf{n}_E$

$$[w]_E(\mathbf{x}) := \lim_{s \to 0} \left(w(\mathbf{x} + s\mathbf{n}_E) - w(\mathbf{x} - s\mathbf{n}_E)\right), \quad \mathbf{x} \in E,$$

and $(\cdot, \cdot)_E$ denotes the $L^2(E)$ inner product. Note that the definition of the jump is not unique as there are two unit normals that differ in their sign. However, it can be seen in (6) that the concrete choice of $\mathbf{n}_E$ does not play any role in the CIP method. The considered CIP method (6) is the method abbreviated by ES in the numerical studies of [7].

Compared with the SUPG method, the formulation (6) possesses two advantages: it does not introduce new non-symmetric terms and it does not require the computation of second order derivatives. However, the stabilization term establishes connections between all degrees of freedom on neighboring mesh cells. Hence, the matrix stencil is denser compared with the SUPG method. Extensive studies on the impact of the parameter σ_{cip} and comparisons with the SUPG method on a number of examples have been carried out in [40].

Similarly to the SUPG method, a non-linear term of SOLD-type can be added to the CIP method, as proposed in [6]. A main difficulty in the application of this approach is that it contains two user-chosen parameters σ_{cip} and σ_{sold}. We performed numerical studies with the combination of the CIP method (6) and the SOLD term $(\tilde{\varepsilon} D \nabla u_h, \nabla v_h)$ introduced in Section 2.3. This SOLD term was chosen because the term proposed in [6] was not among the best approaches in [21]. It turned out that the results of the CIP-SOLD method were generally worse than the results of the SOLD method from Section 2.3. Thus, for the sake of brevity, the CIP-SOLD approach is not included in the numerical studies presented below.

2.5 Discontinuous Galerkin Finite Element Methods, Interior Penalty Methods

Discontinuous Galerkin (DG) finite element methods approximate the solution with piecewise polynomial but discontinuous functions. The coupling of the discontinuous functions occurs in the bilinear form by means of integrals across the faces. To some extent, these methods can be considered as a combination of ideas from finite volume methods (discontinuous approximations) and finite element methods (the basis is a variational formulation).

The presented numerical studies consider a DG finite element method from [28]. For a discontinuous finite element space V_h, it reads as follows: Find $u_h \in V_h$ such that for all $v_h \in V_h$

$$
\begin{aligned}
&\sum_{K\in\mathcal{T}_h} \varepsilon(\nabla u_h, \nabla v_h)_K + (\mathbf{b}\cdot\nabla u_h, v_h)_K \\
&\quad -\varepsilon \sum_{E\in\mathcal{E}_h} \left[\gamma \left([u_h]_E, \langle \nabla v_h \cdot \mathbf{n}_E \rangle_E \right)_E + \left(\langle \nabla u_h \cdot \mathbf{n}_E \rangle_E, [v_h]_E \right)_E \right] \\
&\quad - \sum_{K\in\mathcal{T}_h} \left(\mathbf{b}\cdot\mathbf{n}_{\partial K} \lfloor u_h \rfloor_K, v_h^+ \right)_{\partial^- K \setminus \partial\Omega} + \sigma_{\mathrm{DG}} \sum_{E\in\mathcal{E}_h} \left([u_h]_E, [v_h]_E \right)_E \\
&\quad -\varepsilon \sum_{E\in\partial\Omega} \left[\gamma (u_h, \nabla v_h \cdot \mathbf{n}_E)_E + (\nabla u_h \cdot \mathbf{n}_E, v_h)_E \right] \\
&\quad - \sum_{K\in\mathcal{T}_h} \left(\mathbf{b}\cdot\mathbf{n}_{\partial K} u_h^+, v_h^+ \right)_{\partial^- K \cap \partial\Omega} + 2\sigma_{\mathrm{DG}} \sum_{E\in\partial\Omega} (u_h, v_h)_E \\
&= \sum_{K\in\mathcal{T}_h} (f, v_h)_K - \varepsilon \sum_{E\in\partial\Omega} (u_{bh}, \nabla v_h \cdot \mathbf{n}_E)_E \\
&\quad - \sum_{K\in\mathcal{T}_h} \left(\mathbf{b}\cdot\mathbf{n}_{\partial K} u_{bh}, v_h^+ \right)_{\partial^- K \cap \partial\Omega} + 2\sigma_{\mathrm{DG}} \sum_{E\in\partial\Omega} (u_{bh}, v_h)_E .
\end{aligned}
\tag{7}
$$

Here, $\langle\cdot\rangle_E$ denotes the arithmetic mean of a function at the face E

$$
\langle w \rangle_E(\mathbf{x}) = \frac{1}{2} \left(w|_{\partial K \cap E}(\mathbf{x}) + w|_{\partial K' \cap E}(\mathbf{x}) \right), \quad \mathbf{x} \in E,
$$

where E is the face between the mesh cells K and K'. The inflow boundary of a mesh cell K is denoted by $\partial^- K$

$$
\partial^- K = \{ \mathbf{x} \in \partial K \ : \ \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}_{\partial K} < 0 \} .
$$

The jump of a function in the direction of the convection is defined by

$$
\lfloor w \rfloor_K(\mathbf{x}) = w^+ - w^- = \lim_{s\to 0, s>0} w(\mathbf{x} + s\mathbf{b}) - \lim_{s\to 0, s>0} w(\mathbf{x} - s\mathbf{b}), \quad \mathbf{x} \in \partial K.
$$

Note, the sign of this jump on an edge E might change on this edge, depending on the sign of $\mathbf{b}\cdot\mathbf{n}_{\partial K}$. For $\gamma = 1$, the method is called symmetric interior

penalty method (SIP), for $\gamma = -1$ non-symmetric interior penalty method (NIP) and for $\gamma = 0$ incomplete interior penalty method. For the sake of brevity, only the SIP method is considered in the numerical studies presented below. Note that the parameter γ is always scaled with ε such that in the convection-dominated regime the choice of γ possesses only little impact on the obtained results.

The nodal functionals of the degrees of freedom in DG finite element methods are defined as integrals on the mesh cells. Hence, Dirichlet boundary conditions have to be imposed weakly, as it is done in (7), because there is no degree of freedom whose nodal functional is a point value on the faces at the boundary. Analogously to the CIP method, the integrals on the faces of the mesh cells couple all degrees of freedom on neighboring mesh cells. In addition, a DG method possesses more degrees of freedom on the same grid than a continuous finite element method, like the SUPG method, of the same order.

Comprehensive numerical studies with respect to the parameter σ_{DG} and comparisons with the SUPG method can be found in [3].

2.6 A Total Variation Diminishing Finite Element Method

The Total Variation Diminishing Finite Element Method (FEMTVD), which also combines ideas from finite element and finite volume schemes, works on the algebraic level, see [30]. It starts by discretizing (1) with a high order discretization, like the Galerkin finite element method. Then, the resulting matrices and vectors are modified at the algebraic level in two steps. Firstly, the matrices are changed in order to obtain a positivity preserving, but still too diffusive, scheme. Secondly, the diffusion is locally removed where it is not needed. This is done by an appropriate anti-diffusive contribution on the right hand side.

The Galerkin finite element method applied to (1) leads to an algebraic equation of the form

$$A\underline{u} = \underline{f}, \quad A \in \mathbb{R}^{N\times N},\ \underline{u}, \underline{f} \in \mathbb{R}^N. \tag{8}$$

Define the matrix $D = (d_{ij}) \in \mathbb{R}^{N\times N}$ by

$$d_{ij} = \begin{cases} -\max\{0, a_{ij}, a_{ji}\} & \text{for } i \neq j, \\ -\displaystyle\sum_{j=1, j\neq i}^{N} d_{ij} & \text{otherwise}. \end{cases}$$

This symmetric matrix is a discrete diffusion operator. Its row and column sums are zero. By construction, the matrix $\tilde{A} = A + D$ does not possess positive off-diagonals and it holds $\tilde{a}_{ii} \geq a_{ii} > 0$, $i = 1, \ldots, N$. These are two

properties that are necessary for $\tilde{A}$ being an M-matrix. Moreover, since the amount of mass obtained by node i is subtracted from node j and vice versa, adding D to A does not change the original mass conservation properties, [31].

Now, observe that the i-th row of D can be decomposed as follows

$$(Du)_i = \sum_{j=1}^{N} d_{ij}u_j = \sum_{j=1,j\neq i}^{N} d_{ij}u_j - \sum_{j=1,j\neq i}^{N} d_{ij}u_i = \sum_{j=1,j\neq i}^{N} d_{ij}(u_j - u_i) = \sum_{j=1}^{N} \phi_{ij},$$

where $\phi_{ij} = d_{ij}(u_j - u_i) = -\phi_{ji}$ are the so-called internodal fluxes. This representation leads to an equivalent formulation of the Galerkin scheme (8)

$$A\underline{u} = \underline{f} \quad \Longleftrightarrow \quad \tilde{A}\underline{u} = \underline{f} + D\underline{u} = \underline{f} + \left(\sum_{j=1}^{N} \phi_{ij}\right)_{i=1}^{N}. \tag{9}$$

The FEMTVD scheme considers, instead of the Galerkin scheme (9), the following discrete system

$$\tilde{A}\underline{u} = \underline{f} + \left(\sum_{j=1}^{N} \alpha_{ij}\phi_{ij}\right)_{i=1}^{N}, \quad 0 \leq \alpha_{ij} \leq 1.$$

Clearly, the Galerkin finite element method is recovered by $\alpha_{ij} = 1$, $i, j = 1, \ldots, N$. The goal of the FEMTVD method consists in defining the corrections α_{ij} such that, on the one hand, positivity is preserved and, on the other hand, the amount of artificial diffusion is reduced significantly where artificial diffusion is not needed.

The numerical studies presented in Section 3 use the following algorithm, proposed in [30], for constructing the correction factors. For each pair of neighboring nodes i and j with $\tilde{a}_{ji} \leq \tilde{a}_{ij} \leq 0$:

(1) compute the sum of the anti-diffusive fluxes

$$P_i^+ := P_i^+ + \max\{0, \phi_{ij}\}, \quad P_i^- := P_i^- + \min\{0, \phi_{ij}\},$$

(2) compute upper and lower bounds for the anti-diffusive fluxes

$$\begin{aligned} Q_i^+ &:= Q_i^+ + \max\{0, -\phi_{ij}\}, & Q_j^+ &:= Q_j^+ + \max\{0, \phi_{ij}\}, \\ Q_i^- &:= Q_i^- + \min\{0, -\phi_{ij}\}, & Q_j^- &:= Q_j^- + \min\{0, \phi_{ij}\}, \end{aligned}$$

(3) compute the correction factors

$$R_i^{\pm} = \min\{1, Q_i^{\pm}/P_i^{\pm}\}, \quad \alpha_{ij} = \begin{cases} R_i^+ \text{ if } \phi_{ij} > 0, \\ R_i^- \text{ else.} \end{cases}$$

Since the internodal fluxes ϕ_{ij} are based on a current solution, the FEMTVD scheme is non-linear. To our best knowledge, error estimates are not known for this method.

3 Numerical Studies

For the assessment of the methods presented in Section 2, appropriate test examples are necessary. It turned out that the definition of such examples is rather difficult for convection-dominated convection-diffusion equations. The use of prescribed smooth solutions is not helpful since these solutions do not possess layers, which is the characteristic feature of solutions of convection-dominated equations. In our experience, even the Galerkin finite element method gives reasonably good results if the solution is smooth, independently of the size of the diffusion. There are proposals of analytically known solutions with layers, e.g., in [26]. However, the diffusion coefficient enters in the definition of the right hand side $f(\mathbf{x})$ of these examples in such a way that $f(\mathbf{x})$ possesses layers. For small diffusion, the quadrature error of the right hand side dominates the results even for high order quadrature rules. For this reason, the examples from [26] are appropriate only for moderately small diffusion. Last, there are those examples whose solution possesses layers but for which an analytical expression of the solution is not known.

Numerical studies were performed at different examples and at all kinds of problems mentioned above. Of course, the presentation of the whole set of results is infeasible. Instead of showing short studies for several examples, it is, in our opinion, more interesting to present a comprehensive study of one example which, on the one hand, possesses characteristics of problems arising in applications and, on the other hand, reveals typical features of the numerical methods. We selected the so-called Hemker problem, which was proposed in [17]. References to further numerical studies of other examples were given already in the presentation of the individual methods.

All simulations with the finite volume method were performed with the code PDELIB 2 [32] and the simulations with the finite element methods with the code MOONMD [25]. Delaunay triangulations with triangular mesh cells were generated with TRIANGLE [37]. The linear systems were solved with the sparse direct solvers PARDISO [35,36] (in PDELIB 2) and UMFPACK [8] (in MOONMD). Non-linear problems for the SOLD scheme and the FEMTVD scheme were solved by a fixed point iteration with Anderson acceleration [2,41]. The Anderson acceleration keeps vectors from previous iterations and computes with their help second order information. There is a close relation of this technique to quasi-Newton (secant updating) methods. To be precise, a standard fixed point iteration was applied in the first m steps. Then, the An-

derson acceleration was used with the vectors from the previous m iterations. In the numerical studies presented below, $m = 5$ was used. This approach was often twice or even more faster than the fixed point iteration with automatic damping as described in [23]. The starting iterate for the non-linear schemes was the solution of the SUPG method and the iterations were stopped if the Euclidean norm of the residual vector was below 10^{-12} times the square root of the number of degrees of freedom.

The Hemker problem is defined in $\Omega = \{[-3, 9] \times [-3, 3]\} \setminus \{(x, y) \ : \ x^2 + y^2 < 1\}$, the coefficients are $\mathbf{b} = (1, 0)^T$ and $f = 0$, and the boundary conditions are given by

$$u(x, y) = \begin{cases} 0, & \text{for } x = -3, \\ 1, & \text{for } x^2 + y^2 = 1, \\ \varepsilon \nabla u \cdot \mathbf{n} = 0, & \text{else.} \end{cases}$$

The presented numerical studies consider the diffusion $\varepsilon = 10^{-4}$. For this value, we were able to solve problem (1) on a very fine grid with the Galerkin finite element method (Q_1, 48 252 416 d.o.f.) and we could obtain in this way reference curves for cuts of the solution. This numerical solution is presented in Fig. 2. Its values are contained in $[0, 1]$.

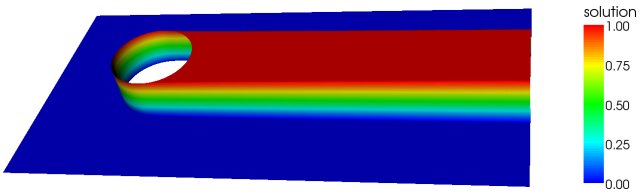

Fig. 2. Solution of the Hemker problem for $\varepsilon = 10^{-4}$.

The Hemker problem can be thought of as a model of a hot column (circle) with normalized temperature $T = 1$, where the heat is transported in the direction of the convection. In this setting, a boundary layer appears in the upwind direction at the circle, reaching from the bottom $(0, -1)$ to the top $(0, 1)$ of the circle. On the bottom and the top of the circle, interior layers start which spread in the direction of the convection.

Computations on triangular and quadrilateral grids will be presented. For the quadrilateral grids, meshes of higher refinement level were obtained by red refinement of an initial grid, thereby increasing the quality of the approximation of the circle. The triangular grids were generated all with TRIANGLE such that the number of mesh cells increased by a factor of about four from level to level.

The alignment of grids is in our opinion an admissible and reasonable approach if the convection is known. However, in applications it might occur that the convection is the computed solution of another equation and it is a priori un-

known. Then, aligned grids cannot be constructed and for this reason, results for an unstructured triangular grid will be presented, too.

3.1 Aligned Triangular and Quadrilateral Grids

Fig. 3 shows the initial grids (level 0). The coarsest quadrilateral grid consists of 184 mesh cells and the coarsest aligned triangular grid of 259 mesh cells. Numbers of degrees of freedom for different refinement levels, including Dirichlet nodes, are given in Table 1. It can be seen that the considered grids were not very coarse, but also not too fine, such that, on the one hand, reasonable results can be expected and, on the other hand, the impact of the stabilizations is essential. For the higher order discretizations, isoparametric finite elements were used at the circle. For shortness of presentation, only results up to second order elements will be given in detail. The results obtained for third order finite elements (used in the SUPG, SOLD, CIP, and DG method) were very similar to those of the second order finite elements.

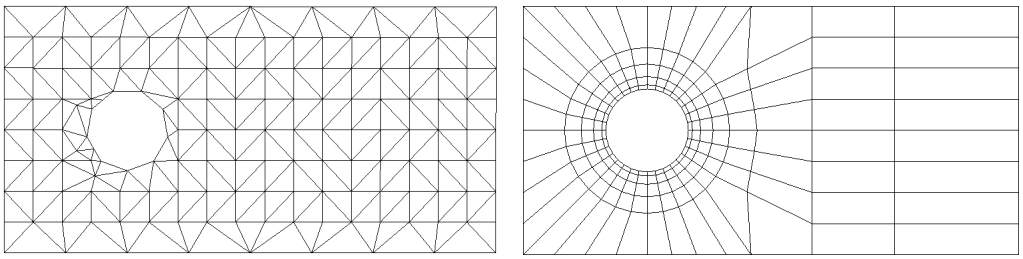

Fig. 3. Hemker problem, initial aligned triangular and quadrilateral grids (level 0).

Table 1
Hemker problem, degrees of freedom for the grids corresponding to the coarsest grids from Fig. 3.

	P_1		P_2		Q_1		Q_2	
level	others	DG	others	DG	others	DG	others	DG
0	151	777	561	1554	219	736	806	1656
3	9271	54 132	36 586	108 264	12 056	47 104	47 664	105 984
4	36 148	214 011	—	—	47 664	188 416	189 536	423 936
5	—	—	—	—	189 536	753 664	—	—

As already discussed in the introduction, the accuracy of the discretizations will be assessed with measures which are of importance in applications.

Under- and overshoots of the computed solutions are studied in some detail in Figs. 4 and 5. The values for the undershoots are defined by the minimal value of the discrete solution and for the overshoots by the maximal value

subtracted by one. For the finite volume method, the values of the projection into the space P_1 were considered. It was observed in [24] that the critical points with respect to the undershoots are the transition points from the boundary layer to the interior layers on bottom and top of the circle. The finite volume scheme and the FEMTVD scheme led always to solutions with values in $[0, 1]$. They are clearly the best schemes with respect to the criterion of under- and overshoots. The SOLD scheme reduced the under- and overshoots of the SUPG method. If the SOLD parameter was sufficiently large, then the under- and overshoots were often almost suppressed. However, note that for parameters larger than the ones presented in Figs. 4 and 5, the iteration for solving the non-linear problem did not converge. The solutions obtained with the CIP scheme showed often smaller undershoots for first order elements than for second order elements. But in all cases, these undershoots were not negligible. For the DG finite element method, always large undershoots could be observed. They were often outside the range of the diagrams.

Cuts of the computed solutions at $x = 4$ were used to study the smearing of the interior layers and the accuracy of the solutions away from the circle, see Fig. 6. For the computation of the cut line at $x = 4$, 10 001 equally distributed points in $y \in [-3, 3]$, and for the cut line at $y = 1$, 20 001 equally distributed points in $x \in [-2, 8]$ were used. With respect to all further evaluations of the simulations, one of the parameters of the parameter-dependent methods (SOLD, CIP, DG) was chosen for each finite element which led to solutions with comparable small under- and overshoots, since under- and overshoots of numerical solutions are often considered to be particularly undesired features in applications. For the results obtained with the DG finite element method, the average values of the solutions were taken on the edges.

Fig. 7 presents the width of the computed interior layers, where for symmetry reasons only the interior layer at $y = 1$ was considered. For these pictures, the width of the interior layer is defined to be the length of the interval $y_{\text{layer,num}} = y_1 - y_0$ in which the solution falls from $u(4, y_0) = 0.9$ to $u(4, y_1) = 0.1$. For the reference solution, there is $y^{\text{ref}}_{\text{layer,num}} = 0.0723$. The SUPG method led to solutions with comparatively sharp layers, which is also well known from other examples. The same can be observed for the DG finite element method. A strong smearing of the layers can be observed for the FVM on coarse grids. The situation becomes better on the finer grids. The CIP method computed solutions with strongly smeared layers for parameters that led to comparatively small under- and overshoots. Also the contrary could be observed in the studies. There are parameters for the CIP method where the solution had much sharper layers, but then the under- and overshoots were considerably larger then for the parameters from Fig. 7. Strongly smeared layers can be seen in the solutions obtained with the SOLD method. This supports the observations from [21,23]. The results of the FEMTVD method showed a very large smearing on the triangular grid. Thus, one can say that

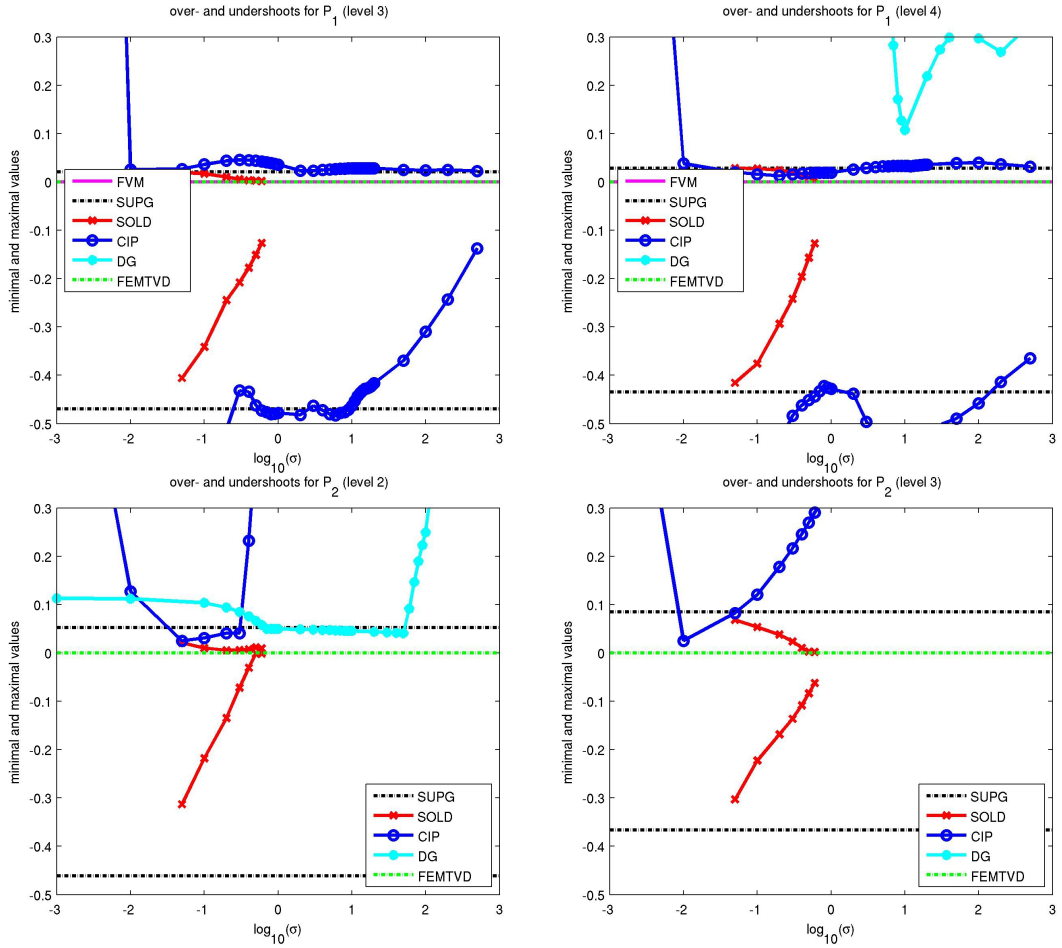

Fig. 4. Hemker problem, over- and undershoots on the triangular grids; P_1 (top, including the finite volume scheme), P_2 (bottom), refinement level 3 (4) on left (right) side. The lines of FVM and FEMTVD (upper plots) are on top of each other. Variations of the stabilization parameter σ are relevant only for SOLD, CIP, and DG.

the suppression of the over- and undershoots by the FVM (on coarser grids), the SOLD method, and the FEMTVD scheme is paid by a notable increase of the smearing of the interior layers.

Differences to the cut lines at $y = 1$ and $x = 4$ were used to assess the quality of the solutions at the layers, see Figs. 8 – 11. The differences to the reference cut lines were computed in each point, giving a vector of errors $\mathbf{e}$. In Figs. 8 – 11, the errors are given in the maximum norm $\|\mathbf{e}\|_\infty$ and the averaged Euclidean norm $\|\mathbf{e}\|_2 := (n^{-1}\sum_{i=1}^{n} e_i^2)^{1/2}$.

The results in Fig. 8 show that the solution obtained with the FVM matched the cut lines worse than the solutions computed with the other methods. This is because the interior layers were not computed at the correct positions. They were situated somewhat too close to the walls. For the FEMTVD scheme, it can be well observed that the produced solutions possess small wiggles

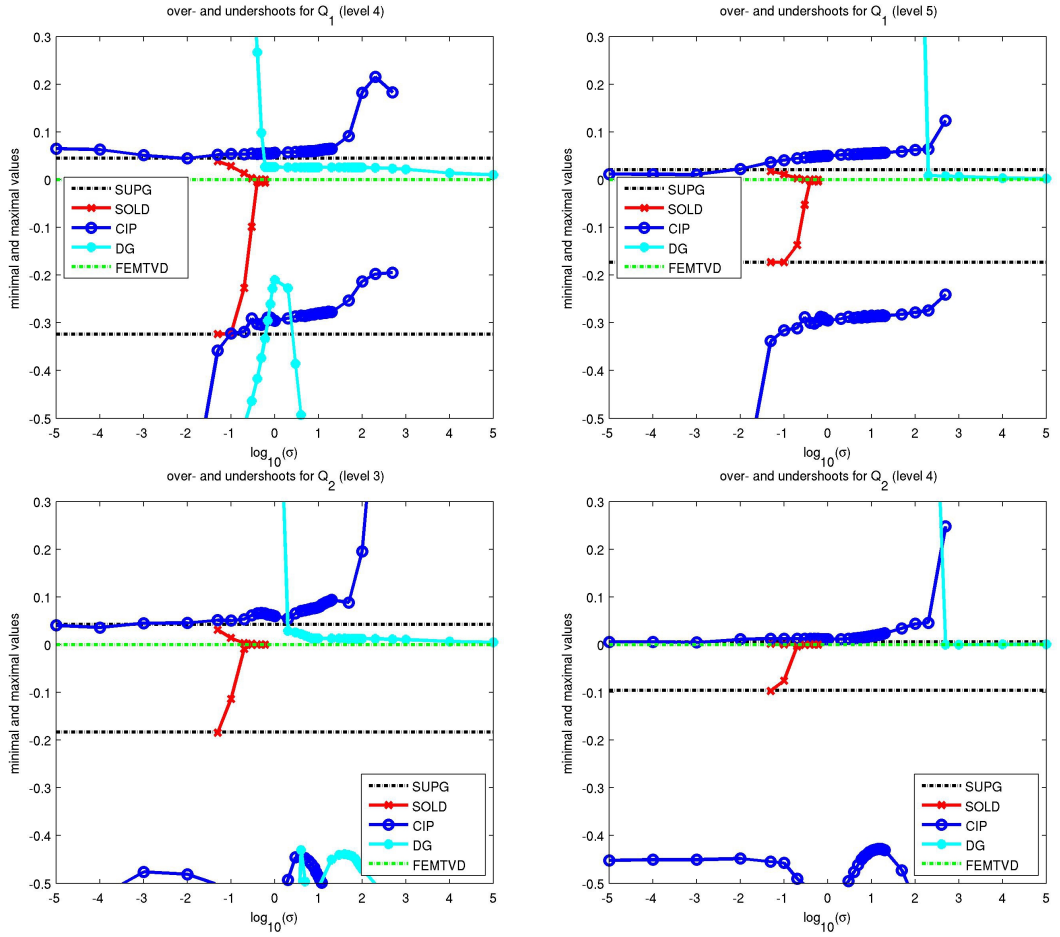

Fig. 5. Hemker problem, over- and undershoots on the quadrilateral grids; Q_1 (top), Q_2 (bottom), refinement level 4 (5) on left (right) side. Variations of the stabilization parameter σ are relevant only for SOLD, CIP, and DG.

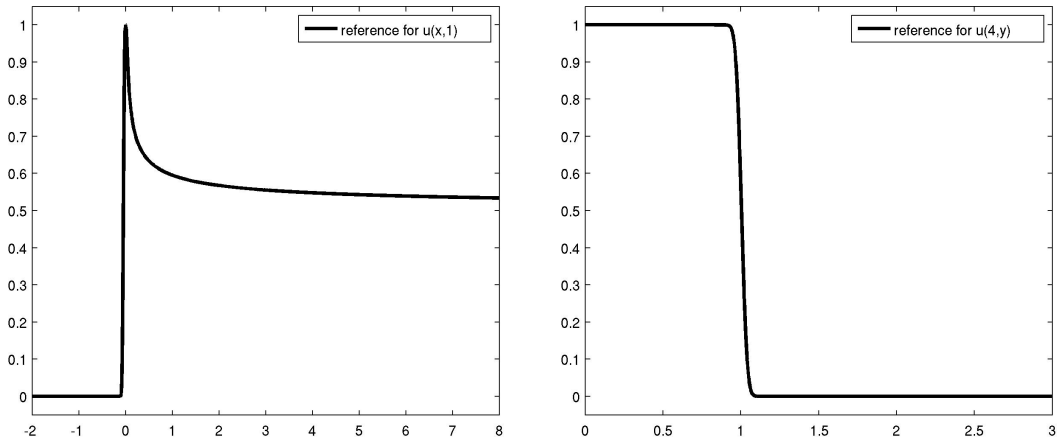

Fig. 6. Hemker problem, cut lines of the reference solution.

along $y = 1$. Thus, although this method led to solutions without under- and overshoots, it is not oscillation-free and a local maximum principle does not hold. The errors to the cut line at $y = 1$ were comparatively small on the

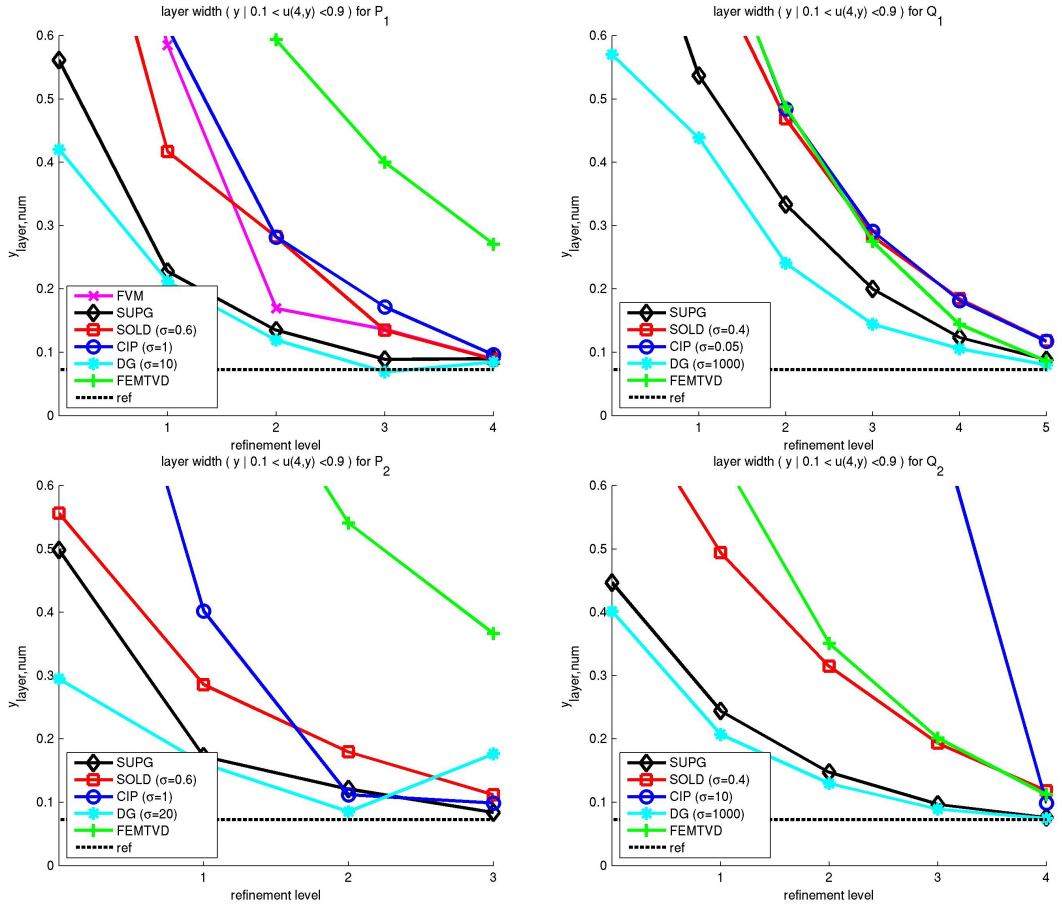

Fig. 7. Hemker problem, layer width at $x = 4$ and $y = 1$; P_1 (including the finite volume scheme), Q_1, P_2, Q_2 (left to right, top to bottom).

triangular grid but the errors to the cut line at $x = 4$ are very large. On this grid, the smallest errors with respect to the cut line at $y = 1$ were obtained with the DG method. Note that this method has more degrees of freedom on the grids than the other methods. The solution computed with the SUPG method possessed rather large errors at $y = 1$. But with respect to the cut line at $x = 4$, it was among the best solutions on the triangular grid.

On the quadrilateral grid, Fig. 9, the cut lines were matched often better by most of the methods than on the triangular grid. The most accurate results were obtained by the SUPG method and the DG method.

Similarly, concerning the second order finite elements, Figs. 10 and 11, the best agreement to the reference cut lines can be observed in general by the solutions computed with the SUPG method and with the DG method.

The solutions computed with the SOLD method and with the CIP method were generally not among the best results with respect to the profiles of the cuts.

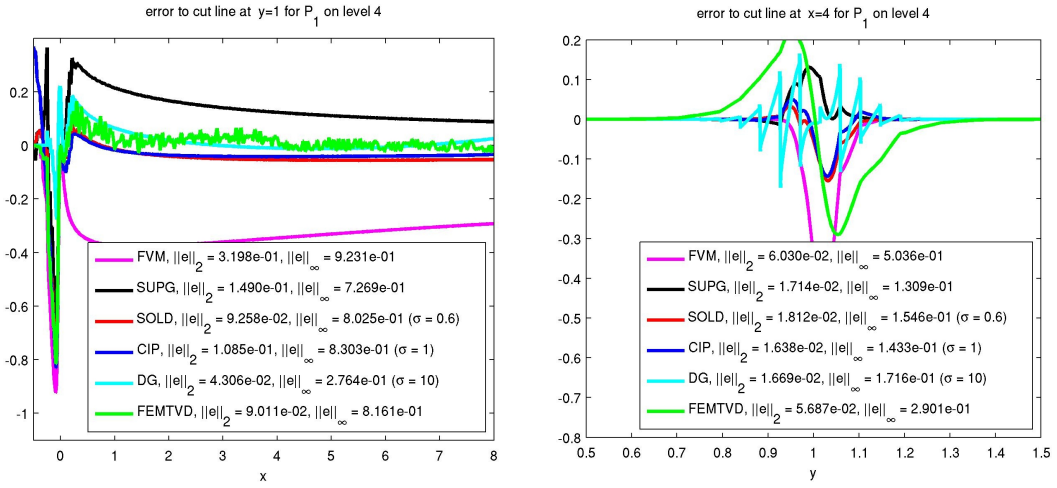

Fig. 8. Hemker problem, errors to the reference cut lines at $y = 1$ and $x = 4$, P_1 (including the finite volume scheme) on level 4.

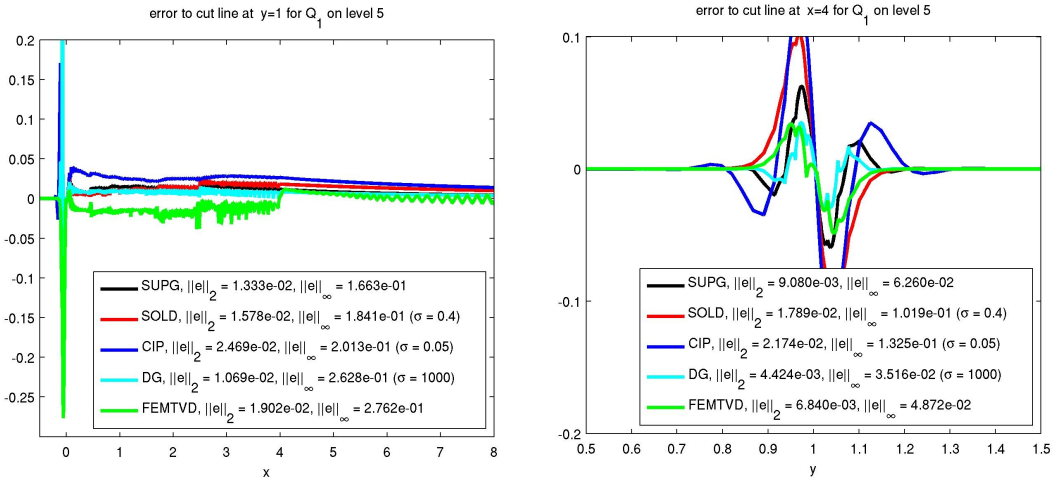

Fig. 9. Hemker problem, errors to the reference cut lines at $y = 1$ and $x = 4$, Q_1 on level 5.

Altogether, the solutions obtained with the DG method possessed in general small errors to the reference curves. Apart from the cut line at $y = 1$ on the triangular grid, the results computed with the SUPG method were likewise good. An incorrect position of the interior layer is the reason for the large errors of the solutions computed by FVM. Despite the suppression of global under- and overshoots, solutions obtained with the FEMTVD scheme possessed wiggles at the layers. Large differences to the reference curves could be observed often for the solutions obtained with the SOLD method and the CIP method.

From the point of view of applications, an important aspect is the efficiency of the different methods. The most relevant measure of efficiency is computing (CPU) time. This is a particularly fair measure if all simulations were performed with the same code, because the actual computing times depend on

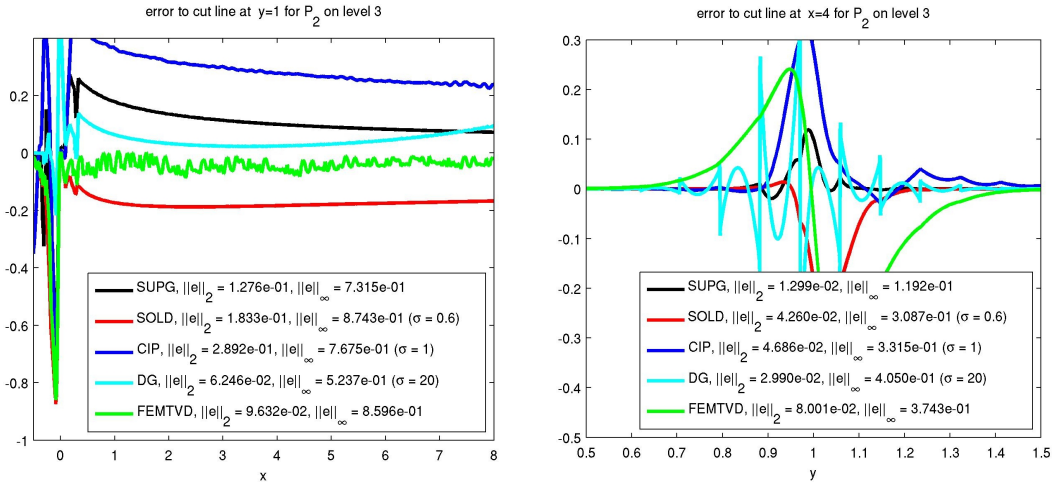

Fig. 10. Hemker problem, errors to the reference cut lines at $y = 1$ and $x = 4$, P_2 on level 3.

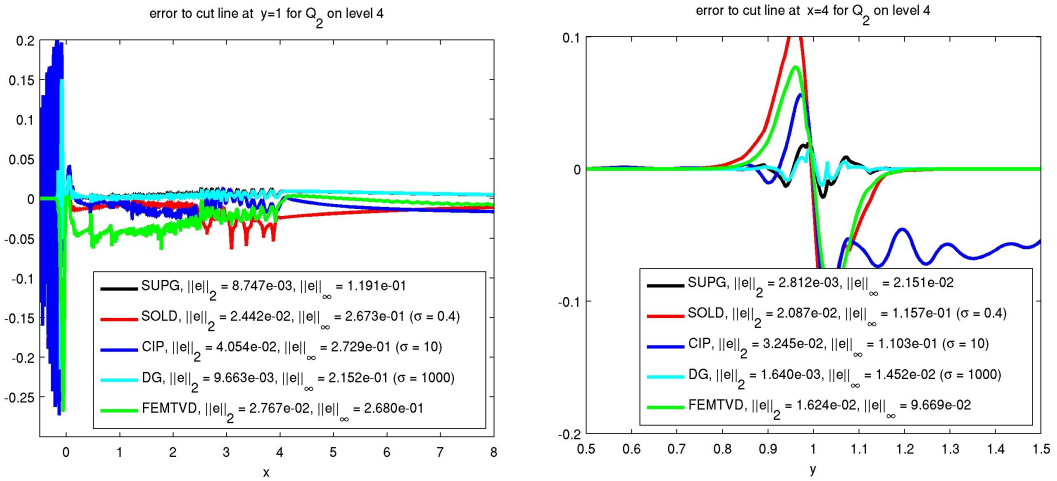

Fig. 11. Hemker problem, errors to the reference cut lines at $y = 1$ and $x = 4$, Q_2 on level 4.

details of the implementation. Since the FVM results were obtained with a different code than the other results, the computing times for this discretization could not be compared directly. In view of the similar overhead of the FVM discretization and the SUPG method (linear method, comparatively sparse matrix), the computing times of the SUPG method were used as reference for FVM, too. The other linear methods, CIP and DG, possess denser matrices than SUPG. DG has even considerably more degrees of freedom than the other methods on the same grid. The SOLD method and FEMTVD are non-linear schemes.

Figs. 12 and 13 present plots of the different quality criteria of the numerical solutions versus CPU times. For the sake of brevity, only the results for P_1 and Q_1 are shown. These finite elements are certainly the most important ones in applications. The best results in the diagrams in Figs. 12 and 13 are those

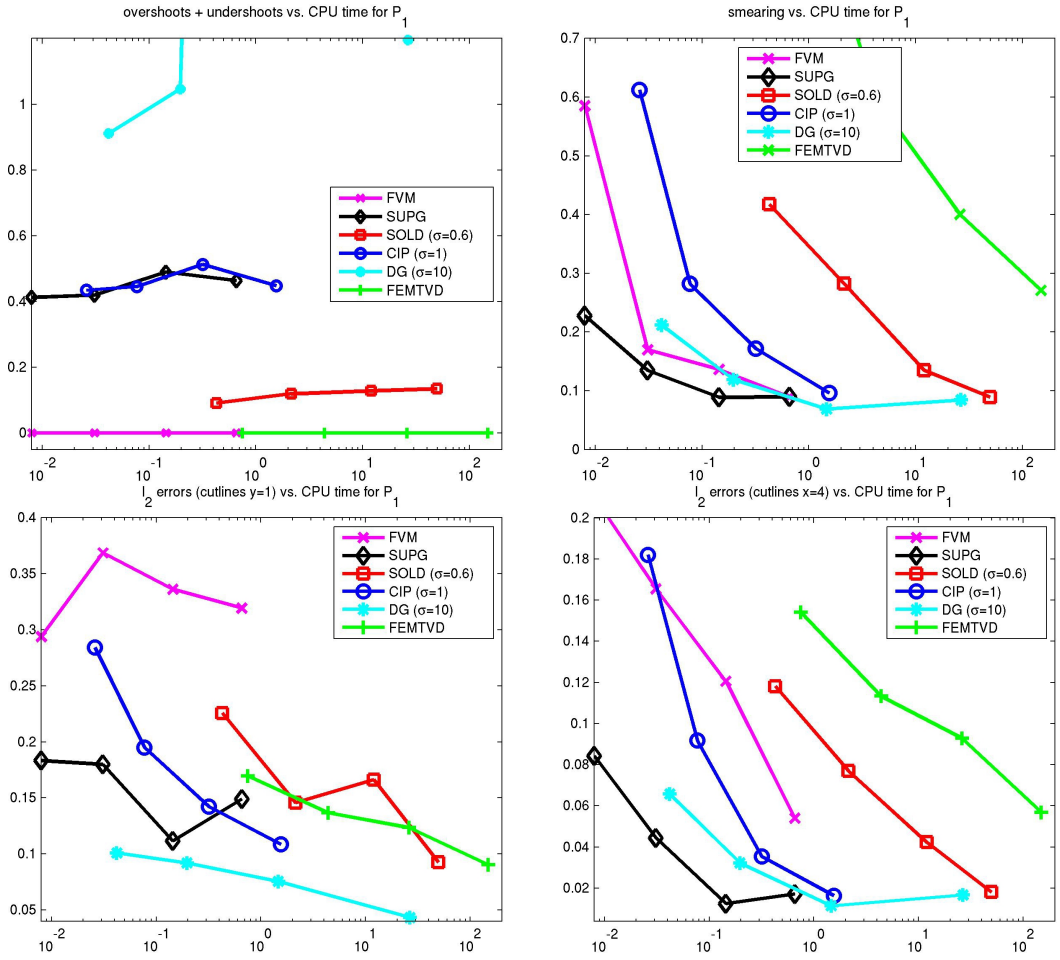

Fig. 12. Hemker problem, CPU times vs. quality measures, P_1, levels $1-4$.

in the lower left corner, since they are accurate and they were obtained in a short computing time.

The large computational costs of the non-linear schemes are clearly visible in these diagrams, since the corresponding curves are always on the right hand side of the pictures. Apart from the sum of under- and overshoots, the results obtained with these schemes were not particularly good. Excellent results with respect to the sum of under- and overshoots were obtained with the FVM discretization in a much shorter time. Altogether, the SOLD scheme and FEMTVD can be considered to be too inefficient. Although a considerable gain in efficiency could be obtained by applying the Anderson acceleration in addition to the fixed point iteration, a further substantial improvement of the non-linear iteration scheme is necessary to make these methods competitive.

Considering the diagrams with respect to smearing and with respect to the differences to the cut lines, the curves of the SUPG method are in general closest to the lower left corner, save for the error to the cut line at $y = 1$ on the triangular grid. These results indicate that this method possesses the

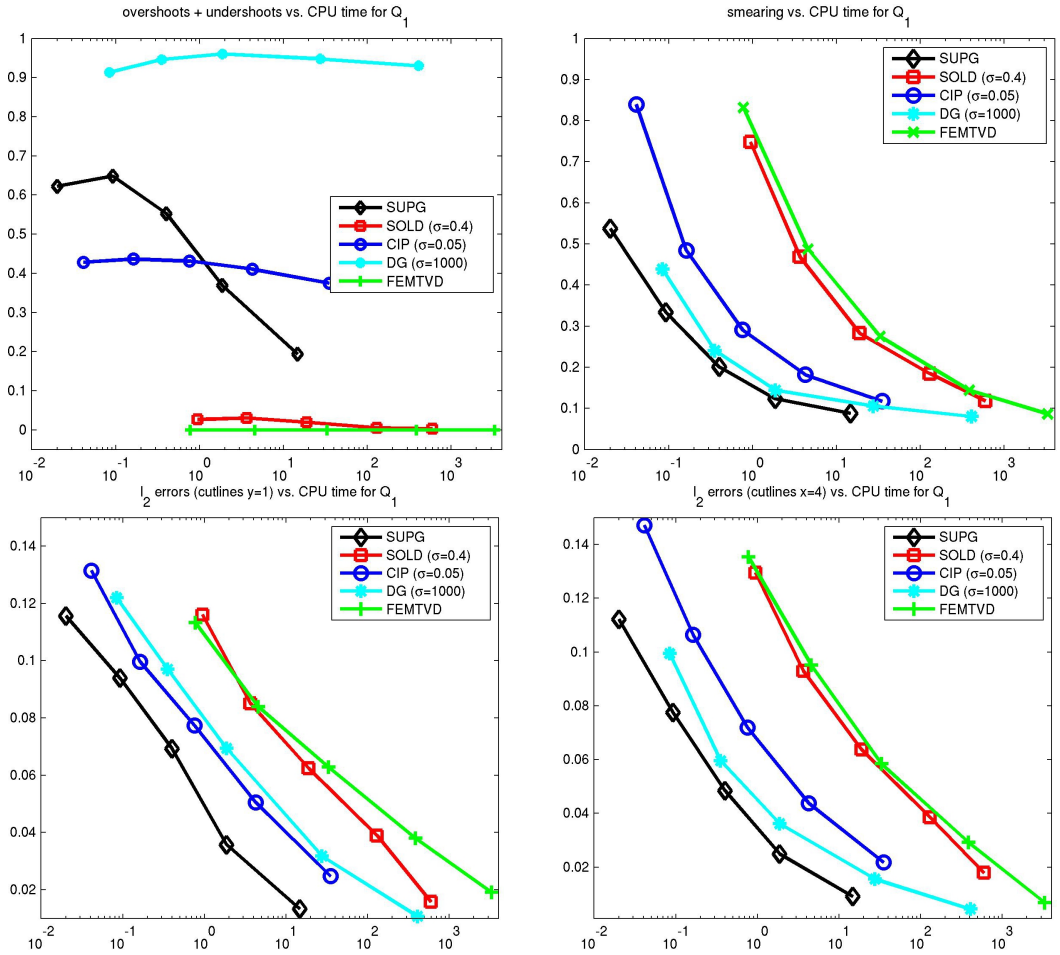

Fig. 13. Hemker problem, CPU times vs. quality measures, Q_1, levels $1-5$.

best ratio of quality (with respect to these criteria) and CPU times among the studied methods. For these criteria, also the curves of DG are rather close to the lower left corner of the diagrams. However, with respect to the under- and overshoots, the DG method was by far the most inaccurate scheme.

For the second order finite elements, essentially the same behavior of the studied methods with respect to errors versus CPU times could be observed.

3.2 Unstructured Triangular Grids

Some results will be presented which were obtained on unstructured triangular grids of the type presented in Fig. 14. The coarsest grid possesses 293 mesh cells. The grid of level 4 has 74 475 mesh cells which leads to 37 693 degrees of freedom for the continuous P_1 finite element space and to 223 425 degrees of freedom for the P_1 discontinuous finite element space.

For the sake of brevity, only results for linear finite elements and the finite

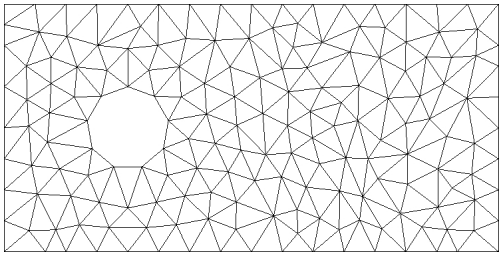

Fig. 14. Hemker problem, unstructured triangular grid (level 0).

volume method are presented in Figs. 15 and 16 since these methods are the most important ones in applications. The under- and overshoots on level 4 are shown in the left picture of Fig. 15. This picture is rather similar to the corresponding results in Fig. 4. The undershoots of the SUPG method are somewhat larger on the unstructured grid and the overshoots of DG are smaller. Again, FVM and FEMTVD do not possess under- and overshoots and the reduction of the under- and overshoots of SUPG by adding the SOLD term is clearly visible.

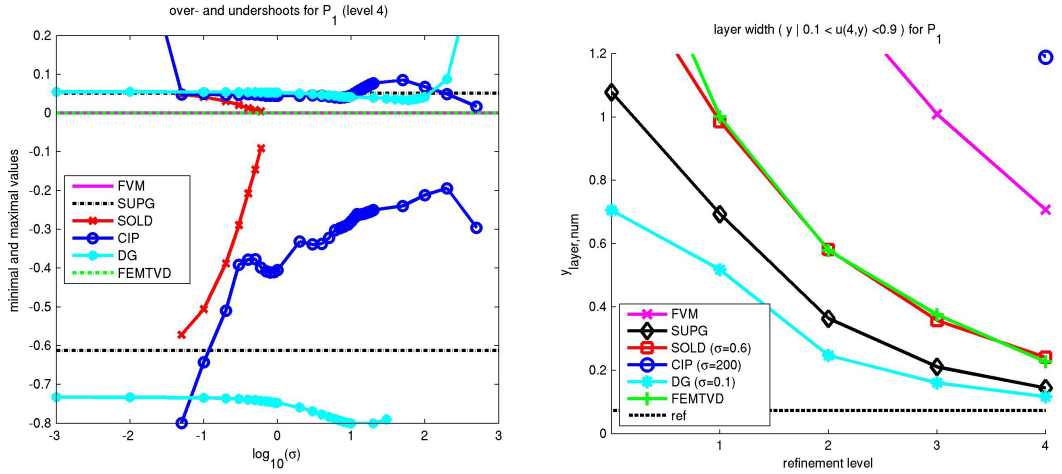

Fig. 15. Hemker problem, over- and undershoots on the unstructured triangular grid (left) and layer width at $x = 4$ and $y = 1$; P_1 (including the finite volume scheme). Variations of the stabilization parameter σ are relevant only for SOLD, CIP, and DG.

For the methods with user-chosen parameter, again such parameters were used in the further assessment of the results which led to solutions with comparatively small under- and overshoots. The choice of such a parameter is not so clear for CIP and DG since various parameters of different size gave similar results. For this reason, we decided to show results for parameters of different magnitude than for the align grid, that means for a large parameter used in the CIP method and for a small parameter used in the DG method. One can observe that there is only little influence of the parameter in the DG method on the results. In contrast, an extremely smeared solution was obtained with the large parameter in the CIP method.

The layer width of the interior layer at $x = 4$ and $y = 1$ is presented in the right picture of Fig. 15. For all methods, the layer is smeared more than in the case of the aligned grid, Fig. 7. The best results were obtained again with DG, followed by SUPG. Extremely smeared are the layers for the solutions computed with FVM and with the CIP method with the large user-chosen parameter.

Evaluations of the cut lines are presented in Fig. 16. It can be seen that the cut lines at $y = 1$ for FVM and SUPG were computed better than on the aligned grid. A possible reason is that there is no grid line of the aligned grid which matches exactly the position of the interior layer. This fact might lead to an adjustment of the computed layer to the grid line and consequently to a somewhat wrong position. For the cut line at $y = 1$, the results obtained with SUPG is best, followed by SOLD and DG (in the averaged Euclidean norm). The errors of the cut lines at $x = 4$ show clearly the extremely smeared solutions computed with FVM and CIP. With respect to this cut line, DG gave the best results, closely followed by SUPG.

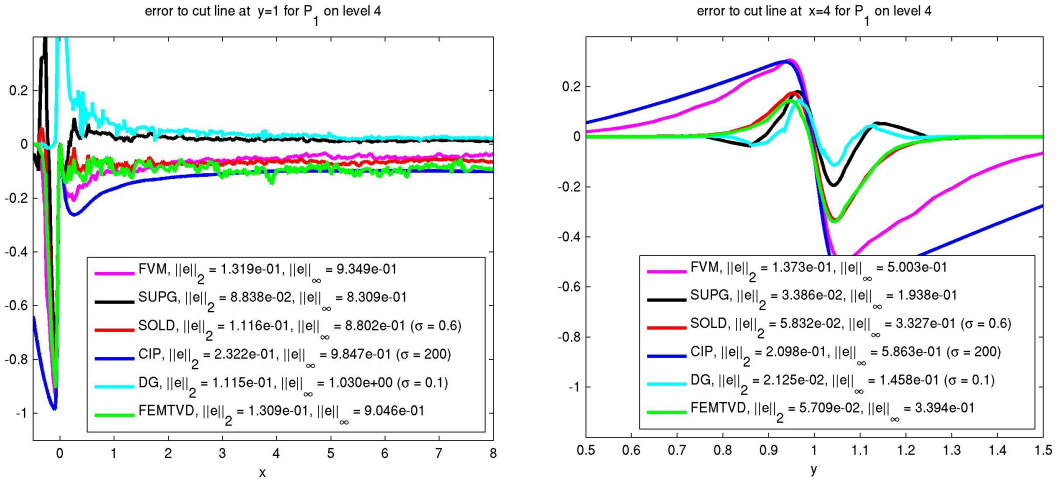

Fig. 16. Hemker problem, errors to the reference cut lines on the unstructured triangular grid at $y = 1$ and $x = 4$, P_1 (including the finite volume scheme) on level 4.

In summary, the numerical studies on the unstructured grid led to a very similar assessment of the methods like the studies on the aligned grid.

4 Summary

Discretizations for convection-dominated convection-diffusion equations which are based on finite volume and finite element ideas were studied numerically. The study of a particular example, the Hemker problem, highlighted advantages and drawbacks of the different approaches. Many of the consid-

ered schemes show non-negligible spurious oscillations. Those schemes, which lead to (nearly) oscillation-free solutions, showed deficits with respect to other aspects, like large smearing of layers, incorrect position of layers, or computing time. A favored method could not be identified. Advice can be given only for some special situations:

- if it is necessary to compute solutions without spurious oscillations: use FVM, taking care on the construction of an appropriate grid might be essential for reducing the smearing of the layers,
- if sharpness and position of layers are important and spurious oscillations can be tolerated: often the SUPG method is a good choice.

Both of these classical schemes are also very efficient. From the more modern approaches which were included in this study, FEMTVD stands out somewhat by suppressing under- and overshoots, but it is quite inefficient. The DG method showed small errors with respect to reference cut lines. However, the solutions possessed very large under- and overshoots. In summary, the use of the modern approaches (SOLD, CIP, DG, FEMTVD) was, considering all aspects, seldom beneficially compared with the classical FVM and SUPG method. Consequently, there is still the urgent need to construct better methods for discretizing convection-dominated equations than those which are currently available.

References

[1] D.N. Allen and R.V. Southwell. Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder. *Quart. J. Mech. and Appl. Math.*, 8:129–145, 1955.

[2] D.G. Anderson. Iterative procedures for nonlinear integral equations. *J. Assoc. Comput. Machinery*, 12:547 – 560, 1965.

[3] M. Augustin. Numerische Untersuchungen eines unstetigen Galerkin–Verfahrens zur Lösung der Konvektions–Diffusions–Gleichung. Diploma thesis, Universität des Saarlandes, FR 6.1 – Mathematik, 2009.

[4] A.N. Brooks and 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.*, 32:199 – 259, 1982.

[5] E. Burman. A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.*, 43:2012–2033, 2005.

[6] E. Burman and A. Ern. Stabilized Galerkin approximation of convection-diffusion-reaction equations: Discrete maximum principle and convergence. *Math. Comput.*, 74:1637 – 1652, 2005.

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

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

[9] J. Douglas and T. Dupont. Interior penalty procedures for elliptic and parabolic Galerkin methods. In R. Glowinski and J.L. Lions, editors, *Computing Methods in Applied Sciences (Second Internat. Sympos., Versailles, 1975)*, volume 58 of *Lecture Notes in Phys.*, pages 207 – 216. Springer-Verlag Berlin, 1976.

[10] R. Eymard, J. Fuhrmann, and K. Gärtner. A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local Dirichlet problems. *Numerische Mathematik*, 102(3):463 – 495, 2006.

[11] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, 2000.

[12] L.P. Franca and F. Valentin. On an improved unusual stabilized finite element method for the advective-reactive-diffusive equation. *Comput. Methods Appl. Mech. Engrg.*, 190:1785 – 1800, 2000.

[13] J. Fuhrmann and H. Langmach. Stability and existence of solutions of time-implicit finite volume schemes for viscous nonlinear conservation laws. *Appl. Numer. Math.*, 37:201 – 230, 2001.

[14] J. Fuhrmann, A. Linke, and H. Langmach. A numerical method for mass conservative coupling between fluid flow and solute transport. *Applied Numerical Mathematics*, 61(4):530 – 553, 2011.

[15] H. Han, Z. Huang, and R.B. Kellogg. A tailored finite point method for a singular perturbation problem on an unbounded domain. *J. Sci. Comput.*, 36:243 – 261, 2008.

[16] G. Hauke. A simple subgrid scale stabilized method for the advection-diffusion-reaction equation. *Comput. Methods Appl. Mech. Engrg.*, 191:2925 – 2947, 2002.

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

[18] T.J.R. Hughes and A.N. Brooks. A multidimensional upwind scheme with no crosswind diffusion. In T.J.R. Hughes, editor, *Finite Element Methods for Convection Dominated Flows, AMD vol.34*, pages 19 – 35. ASME, New York, 1979.

[19] T.J.R. Hughes, M. Mallet, and A. Mizukami. A new finite element formulation for computational fluid dynamics: II. beyond SUPG. *Comput. Methods Appl. Mech. Engrg.*, 54:341 – 355, 1986.

[20] A.M. Il'in. A difference scheme for a differential equation with a small parameter multiplying the second derivative. *Mat. zametki*, 6:237–248, 1969.

[21] V. John and P. Knobloch. A comparison of spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: Part I – a review. *Comput. Methods Appl. Mech. Engrg.*, 196:2197 – 2215, 2007.

[22] V. John and P. Knobloch. On the performance of SOLD methods for convection–diffusion problems with interior layers. *International Journal of Computing Science and Mathematics*, 1:245 – 258, 2007.

[23] V. John and P. Knobloch. A comparison of spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: Part II – analysis for P_1 and Q_1 finite elements. *Comput. Methods Appl. Mech. Engrg.*, 197:1997 – 2014, 2008.

[24] V. John, P. Knobloch, and S.B. Savescu. A posteriori optimization of parameters in stabilized methods for convection-diffusion problems – Part I. *Comput. Methods Appl. Mech. Engrg.*, 200:2916 – 2929, 2011.

[25] V. John and G. Matthies. MooNMD - a program package based on mapped finite element methods. *Comput. Visual. Sci.*, 6:163 – 170, 2004.

[26] V. John, J.M. Maubach, and L. Tobiska. Nonconforming streamline-diffusion-finite-element-methods for convection-diffusion problems. *Numer. Math.*, 78:165 – 188, 1997.

[27] 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. Engrg. Sci.*, 64:733 – 741, 2009.

[28] G. Kanschat. *Discontinuous Galerkin Methods for Viscous Incompressible Flow.* Advances in Numerical Mathematics. Teubner Research, 2007.

[29] T. Knopp, G. Lube, and G. Rapin. Stabilized finite element methods with shock capturing for advection–diffusion problems. *Comput. Methods Appl. Mech. Engrg.*, 191:2997 – 3013, 2002.

[30] D. Kuzmin. Algebraic flux corrections for finite element discretizations of coupled systems. In *Proceedings of the ECCOMAS Conference Computational Methods for Coupled Problems in Science and Engineering*, 2007.

[31] D. Kuzmin and M. Möller. Algebraic flux correction I. Scalar conservation laws. In R. Löhner D. Kuzmin and S. Turek, editors, *Flux-Corrected Transport: Principles, Algorithms and Applications*, pages 155 – 206. Springer, 2005.

[32] pdelib 2. URL: http://www.wias-berlin.de/software/pdelib/.

[33] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust Numerical Methods for Singularly Perturbed Differential Equations*, volume 24 of *Springer Series in Computational Mathematics*. Springer, 2nd edition, 2008.

[34] D.L. Scharfetter and H.K. Gummel. Large signal analysis of a silicon Read diode. *IEEE Trans. Elec. Dev.*, 16:64–77, 1969.

[35] O. Schenk and K. Gärtner. Solving unsymmetric sparse systems of linear equations with PARDISO. *Future Gen. Comp. Sys*, 20(3):475–487, 2004.

[36] O. Schenk, K. Gärtner, G. Karypis, S. Röllin, and M. Hagemann. PARDISO - sparse direct solver, version 3.0. URL: http://www.pardiso-project.org, 2007. Retrieved 2011-04-07.

[37] J.R. Shewchuk. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In M. C. Lin and D. Manocha, editors, *Applied Computational Geometry: Towards Geometric Engineering*, volume 1148 of *Lecture Notes in Computer Science*, pages 203–222. Springer, 1996.

[38] H. Si, K. Gärtner, and J. Fuhrmann. Boundary conforming Delaunay mesh generation. *Comput. Math. Math. Phys.*, 50:38–53, 2010.

[39] M. Stynes. Steady-state convection-diffusion problems. In A. Iserles, editor, *Acta Numerica*, pages 445 – 508. Cambridge University Press, 2005.

[40] R. Umla. Stabilisierte Finite-Elemente Verfahren für die Konvektions-Diffusions-Gleichung und die Oseen-Gleichung. Diploma thesis, Universität des Saarlandes Saarbrücken, 2009.

[41] H.F. Walker and P. Ni. Anderson acceleration for fixed-point iterations. *SIAM J. Numer. Anal.*, 2011. accepted for publication.