V. Dolejší, M. Feistauer, V. Sobotıková

Analysis of the Discontinuous Galerkin Method for Nonlinear Convection–Diffusion Problems

Preprint No. MATH-knm-2004/4

30.8.2004

Submitted *Comput. Methods Appl. Mech. Eng.*

Department of Numerical Mathematics, Faculty of Mathematics and Physics,

Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic

http://www.karlin.mff.cuni.cz/ms-preprints/

Cover print by Reprostředisko, Malostranské nám. 25, Praha 1 - Malá Strana

Design by Mirko Rokyta, picture of snowflake by Ondřej Kalenda

Current series administrator: V. Dolejsi

Analysis of the Discontinuous Galerkin Method for Nonlinear Convection–Diffusion Problems ⋆

V. Dolejší [a], M. Feistauer [a,∗], V. Sobotíková [b]

[a] *Charles University Prague, Faculty of Mathematics and Physics, Sokolovská 83, 186 75 Praha, Czech Republic*

[b] *Department of Mathematics, Faculty of Electrical Engineering, Czech Technical University, Technická 2, 166 27 Praha 6, Czech Republic*

Abstract

The subject-matter is the analysis of the discontinuous Galerkin finite element method applied to a nonlinear convection-diffusion problem. In the contrary to the standard FEM the requirement of the conforming properties is omitted. This allows us to consider general polyhedral elements with mutually disjoint interiors. We do not require their convexity, but assume only that they are star-shaped. We present an error analysis for the case of a nonsymmetric discretization of diffusion terms. Theoretical results are accompanied by numerical experiments.

Key words: nonlinear convection-diffusion equation, discontinuous Galerkin finite element method, nonsymmetric stabilization of diffusive terms, interior and boundary penalty, asymptotic error estimates, numerical experiments
1991 MSC: 65M12, 65M15, 65M60

⋆ This research started during the visit of V. Dolejší and M. Feistauer at "Forschungsinstitut für Mathematik" at ETH in Zürich. It was supported under the Grants No. 201/02/0684 and No. 201/00/D116 of the Czech Grant Agency and the Grants No. MSM 113200007 and No. MSM 210000010 of the Ministry of Education of the Czech Republic. The authors acknowledge the support of these institutions and are very much obliged to Professor C. Schwab for fruitful discussions.
∗ Corresponding author.
Email addresses: `dolejsi@karlin.mff.cuni.cz` (V. Dolejší), `feist@karlin.mff.cuni.cz` (M. Feistauer), `sobotik@math.feld.cvut.cz` (V. Sobotíková).

1 Introduction

In the finite element method (FEM), an important question arises: should we prefer to use *conforming* finite elements or *nonconforming* finite elements? Conforming (i.e., continuous) finite element approximations are suitable for problems with sufficiently regular solutions. However, singularly perturbed problems or nonlinear conservation laws of fluid dynamics have solutions with steep gradients or discontinuities and their approximations by conforming finite elements suffer from the Gibbs phenomenon. One way how to avoid this drawback is to use a suitable stabilization as, e. g., the streamline diffusion method or Galerkin least squares method and shock capturing stabilization.

However, the approximation of discontinuous solutions to conservation laws by continuous functions does not appear quite natural. From this point of view, it seems that for conservation laws, the finite volume method using piecewise constant approximations could be more suitable, because the FV approximations are discontinuous on interelement interfaces, which allows better resolution of shock waves and contact discontinuities. On the other hand, the increase of accuracy in finite volume schemes seems to be problematic. A combination of ideas and techniques of the FV and FE methods yields the *discontinuous Galerkin finite element method* (DGFEM) using advantages of both approaches and allowing to obtain schemes with a higher order accuracy in a natural way. The DGFEM is based on the idea to approximate the solution of an initial-boundary value problem by piecewise polynomial functions over a finite element mesh without any requirement on interelement continuity. The DGFEM was applied to nonlinear conservation laws already in 1989 by Cockburn and Shu ([8]). It was used for the numerical simulation of compressible flow later by Bassi and Rebay in [3] and [2] and by Baumann and Oden in [4]. Problems with nonlinear diffusion are discussed in [22]. During several recent years the DGFE schemes have been extensively developed and become more and more popular. For a survey see, e. g., [9] and [7]. Some aspects of the DGFEM and applications to gas dynamics are also discussed in [1,10–13,16,17].

In this paper we are concerned with error analysis of the discontinuous Galerkin finite element method applied to a nonlinear convection-diffusion equation. In standard DGFE discretizations usually triangular (in 2D) or tetrahedral (in 3D) elements with standard properties mentioned, e. g., in [6] or [5] are used. However, it appears that in the DGFEM the requirement of conformity of finite element meshes is not necessary. Moreover, more general elements can be used. For example, in [11] we analyzed a special method using elements in the form of convex polygons (in 2D) and convex polyhedra (in 3D).

It is an interesting question, if it is possible to use even more general nonconvex elements, similarly as in the finite volume method, where, e.g., dual finite volumes are in general nonconvex. See, for example, [15]. This question has also another natural motivation. In the mesh generation (particularly in 3D) one often proceeds in such a way that the computational domain is first split in a finite number of polyhedral elements, in general nonconvex, which are subsequently divided in tetrahedra. Usually the number of tetrahedra is much larger than the number of the previous polyhedral elements. From the point of view of efficiency of numerical schemes with respect to adaptive procedures, it is desirable to be able to use a computational mesh formed by tetrahedra as well as larger polyhedral nonconvex elements. In this paper, we present an error analysis of the DGFEM applied on general polyhedral meshes and using the nonsymmetric discretization of diffusion terms. Theoretical results are accompanied by numerical experiments.

2 Continuous problem

Let $\Omega \subset I\!R^d$ ($d = 2$ or 3) be a bounded polyhedral domain and $T > 0$. We set $Q_T = \Omega \times (0,T)$. By $\overline{\Omega}$ and $\partial\Omega$ we denote the closure and boundary of Ω, respectively. Let us consider the following *initial-boundary value problem*: Find $u : Q_T \to I\!R$ such that

$$\frac{\partial u}{\partial t} + \sum_{s=1}^{d} \frac{\partial f_s(u)}{\partial x_s} = \varepsilon \, \Delta u + g \quad \text{in } Q_T, \tag{1}$$

$$u\Big|_{\partial\Omega\times(0,T)} = u_D, \tag{2}$$

$$u(x,0) = u^0(x), \quad x \in \Omega. \tag{3}$$

We assume that the data satisfy the following conditions:

$$\begin{aligned}
&\text{a) } f_s \in C^1(I\!R),\ f_s(0) = 0,\ s = 1,\dots,d, \\
&\text{b) } \varepsilon > 0, \\
&\text{c) } g \in C([0,T]; L^2(\Omega)), \\
&\text{d) } u_D \text{ is the trace of some } u^* \in C([0,T]; H^1(\Omega)) \cap L^\infty(Q_T) \\
&\quad \text{on } \partial\Omega \times (0,T), \\
&\text{e) } u^0 \in L^2(\Omega).
\end{aligned} \tag{4}$$

We use the standard notation for function spaces (see, e. g., [19]): $L^p(\Omega)$, $L^p(Q_T)$ denote the Lebesgue spaces, $W^{k,p}(\Omega)$, $H^k(\Omega) = W^{k,2}(\Omega)$ are the Sobolev spaces, $L^p(0,T;X)$ is the Bochner space of functions p-integrable over the interval $(0,T)$ with values in a Banach space X, $C([0,T];X)$ ($C^1([0,T];X)$) is the space of continuous (continuously differentiable) mappings of the interval $[0,T]$ into X. By $H_0^1(\Omega)$ we denote the subspace of all functions from $H^1(\Omega)$ with zero traces on $\partial\Omega$.

The assumption that $f_s(0) = 0$, $s = 1,\dots,d$, does not cause any loss of generality, as can be seen from equation (1). The functions f_s, called fluxes, represent convective terms, $\varepsilon > 0$ is the diffusion coefficient. The diffusion term can be, in general, more complicated (in some cases even nonlinear as in [22]). It is also possible to consider mixed Dirichlet–Neumann boundary conditions. For simplicity we prescribe the Dirichlet condition on the whole boundary.

A sufficiently regular function satisfying (1) – (3) pointwise is called a *classical solution*. It is suitable to introduce the concept of a weak solution. To this end we use the following *notation*:

$$(u,v) = \int_\Omega uv\,\mathrm{d}x, \quad u,\, v \in L^2(\Omega) \tag{5}$$

(scalar product in $L^2(\Omega)$),

$$a(u,v) = \varepsilon \int_\Omega \nabla u \cdot \nabla v\,\mathrm{d}x, \quad u,\, v \in H^1(\Omega), \tag{6}$$

$$b(u,v) = \int_\Omega \sum_{s=1}^{d} \frac{\partial f_s(u)}{\partial x_s} v\,\mathrm{d}x, \quad u \in H^1(\Omega) \cap L^\infty(\Omega),\ v \in L^2(\Omega), \tag{7}$$

$$\|u\|_{H^1(\Omega)} = \left(\int_\Omega \left(|u|^2 + |\nabla u|^2 \right) \mathrm{d}x \right)^{1/2}, \quad u \in H^1(\Omega), \tag{8}$$

(norm in $H^1(\Omega)$),

$$|u|_{H^1(\Omega)} = \left(\int_\Omega |\nabla u|^2\,\mathrm{d}x \right)^{1/2}, \quad u \in H^1(\Omega), \tag{9}$$

(seminorm in $H^1(\Omega)$). It is known that $|\cdot|_{H^1(\Omega)}$ is a norm on $H_0^1(\Omega)$ equivalent to $\|\cdot\|_{H^1(\Omega)}$.

The *weak solution* of (1) – (3) is defined as a function u satisfying the conditions

$$\text{a)} \quad u - u^* \in L^2(0,T;H_0^1(\Omega)), \quad u \in L^\infty(Q_T), \tag{10}$$

b) $\dfrac{\mathrm{d}}{\mathrm{d}t}(u(t), v) + b(u(t), v) + a(u(t), v) = (g(t), v)$

for all $v \in H_0^1(\Omega)$ in the sense of distributions on $(0, T)$,

c) $u(0) = u_0$ in Ω.

By $u(t)$ we denote the function on Ω such that $u(t)(x) = u(x, t)$, $x \in \Omega$.

With the aid of techniques from [20] and [21], it is possible to prove that there exists a unique weak solution. Moreover, it satisfies the condition $\partial u/\partial t \in L^2(Q_T)$. Then (10), b) can be rewritten as

$$\left(\frac{\partial u(t)}{\partial t}, v\right) + b(u(t), v) + a(u(t), v) = (g(t), v) \tag{11}$$

for all $v \in H_0^1(\Omega)$ and almost every $t \in (0, T)$.

We shall assume that the weak solution u is sufficiently regular, namely,

$$\frac{\partial u}{\partial t} \in L^2(0, T; H^{p+1}(\Omega)), \tag{12}$$

where an integer $p \geq 1$ will denote a given degree of approximation. It is possible to show that such a solution u satisfies equation (1) pointwise almost everywhere and

$$u \in C([0, T]; H^{p+1}(\Omega)). \tag{13}$$

3 Discretization of the problem

3.1 Triangulations

Let T_h $(h > 0)$ be a partition of the closure $\overline{\Omega}$ of the domain Ω into a finite number of closed d-dimensional (convex or nonconvex) polyhedra K with mutually disjoint interiors. We call T_h a triangulation of $\overline{\Omega}$ and *do not* require the conforming properties from the finite element method. In 2D problems we choose usually $K \in T_h$ as triangles or quadrilaterals. In 3D, $K \in T_h$ can be, e. g., tetrahedra, pyramids or hexahedra, but we can construct even more general elements K, as dual finite volumes from [15]. In the analysis carried out in Section 4 it is important to assume that all elements $K \in T_h$ are star-shaped.

In our further considerations we shall use the following *notation*. We set $h_K = \mathrm{diam}(K)$, $h = \max_{K \in T_h} h_K$. By ρ_K we denote the radius of the largest d-dimensional ball inscribed into K and by $|K|$ we denote the d-dimensional

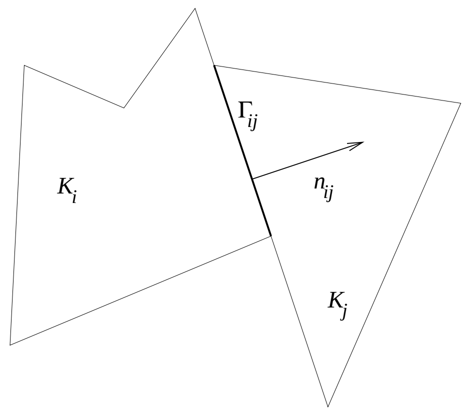

Fig. 1. Neighbouring elements $K_i,\ K_j$

Lebesgue measure of K. All elements of T_h will be numbered so that $T_h = \{K_i\}_{i\in I}$, where $I \subset Z^+ = \{0, 1, 2, \ldots\}$ is a suitable index set. If two elements K_i, $K_j \in T_h$, $i \neq j$, contain a nonempty common open face, we call them *neighbours*. We set in this case

$$\Gamma_{ij} = \partial K_i \cap \partial K_j. \tag{14}$$

(See Figure 1, showing a possible 2D situation.) For $i \in I$ we set

$$s(i) = \{j \in I; K_j \text{ is a neighbour of } K_i\}. \tag{15}$$

The boundary $\partial\Omega$ is formed by a finite number of faces of elements K_i adjacent to $\partial\Omega$. We denote all these boundary faces by S_j, where $j \in I_b \subset Z^- = \{-1, -2, \ldots\}$ and set

$$\begin{aligned} \gamma(i) &= \{j \in I_b; S_j \text{ is a face of } K_i\}, \\ \Gamma_{ij} &= S_j \text{ for } K_i \in T_h \text{ such that } S_j \subset \partial K_i,\ j \in I_b. \end{aligned} \tag{16}$$

For K_i not containing any boundary face S_j we set $\gamma(i) = \emptyset$. Obviously, $s(i) \cap \gamma(i) = \emptyset$ for all $i \in I$. Now, if we write $S(i) = s(i) \cup \gamma(i)$, we have

$$\partial K_i = \bigcup_{j\in S(i)} \Gamma_{ij}, \quad \partial K_i \cap \partial\Omega = \bigcup_{j\in\gamma(i)} \Gamma_{ij}. \tag{17}$$

Furthermore, we use the following notation: $\boldsymbol{n}_{ij} = ((n_{ij})_1, \ldots, (n_{ij})_d)$ – unit outer normal to ∂K_i on the face Γ_{ij} (see Figure 1), $|\Gamma_{ij}|$ – $(d-1)$-dimensional measure of Γ_{ij} (i. e., the length of Γ_{ij}, if $d = 2$, and the area of Γ_{ij}, if $d = 3$), $d(\Gamma_{ij}) = \mathrm{diam}(\Gamma_{ij})$. Obviously, $\boldsymbol{n}_{ij} = -\boldsymbol{n}_{ji}$ and, from the point of view of point sets, we have $\Gamma_{ij} = \Gamma_{ji}$, if $i \in I$, $j \in s(i)$ (and, hence, $j \in I$, $i \in s(j)$).

It is obvious that for $i \in I$ and $j \in S(i)$ we have

$$|K_i| \leq h_{K_i}^d \leq h^d, \tag{18}$$

$$d(\Gamma_{ij}) \leq h_{K_i} \leq h. \tag{19}$$

3.2 Broken Sobolev spaces

Over the triangulation T_h we define the so-called *broken Sobolev space*

$$H^k(\Omega, T_h) = \{v; v|_K \in H^k(K) \ \forall \, K \in T_h\} \tag{20}$$

and define there the norm

$$\|v\|_{H^k(\Omega,T_h)} = \left(\sum_{K \in T_h} \|v\|^2_{H^k(K)} \right)^{1/2} \tag{21}$$

and the seminorm

$$|v|_{H^k(\Omega,T_h)} = \left(\sum_{K \in T_h} |v|^2_{H^k(K)} \right)^{1/2}. \tag{22}$$

For $v \in H^1(\Omega, T_h)$, $i \in I$ and $j \in s(i)$ we introduce the following notation:

$$\begin{aligned} v|_{\Gamma_{ij}} &= \text{the trace of } v|_{K_i} \text{ on } \Gamma_{ij}, \\ v|_{\Gamma_{ji}} &= \text{the trace of } v|_{K_j} \text{ on } \Gamma_{ji}, \\ \langle v \rangle_{\Gamma_{ij}} &= \frac{1}{2} \left(v|_{\Gamma_{ij}} + v|_{\Gamma_{ji}} \right), \\ [v]_{\Gamma_{ij}} &= v|_{\Gamma_{ij}} - v|_{\Gamma_{ji}}. \end{aligned} \tag{23}$$

Obviously, $\langle v \rangle_{\Gamma_{ij}} = \langle v \rangle_{\Gamma_{ji}}$, but $[v]_{\Gamma_{ij}} = -[v]_{\Gamma_{ji}}$. On the other hand, $[v]_{\Gamma_{ij}} \boldsymbol{n}_{ij} = [v]_{\Gamma_{ji}} \boldsymbol{n}_{ji}$.

3.3 Variational formulation of the problem

In order to derive the discrete problem, we start from the weak solution u satisfying (12), multiply equation (1) by an arbitrary $v \in H^2(\Omega, T_h)$, integrate over each $K_i \in T_h$ and apply Green's theorem. We obtain the identity

$$\int\limits_{K_i} \frac{\partial u}{\partial t}(t)\, v \,\mathrm{d}x + \int\limits_{\partial K_i} \sum_{s=1}^{d} f_s(u(t))\, n_s\, v \,\mathrm{d}S - \int\limits_{K_i} \sum_{s=1}^{d} f_s(u(t))\, \frac{\partial v}{\partial x_s} \,\mathrm{d}x \tag{24}$$
$$+\varepsilon \int\limits_{K_i} \nabla u(t) \cdot \nabla v \,\mathrm{d}x - \varepsilon \int\limits_{\partial K_i} (\nabla u(t) \cdot \boldsymbol{n})\, v \,\mathrm{d}S = \int\limits_{K_i} g(t)\, v \,\mathrm{d}x.$$

Here $\boldsymbol{n} = (n_1, \ldots, n_d)$ denotes the unit outer normal to ∂K_i. Summing (24) over all $K_i \in T_h$, using the relations

$$[u]_{\Gamma_{ij}} = 0, \quad \langle \nabla u \rangle_{\Gamma_{ij}} = \nabla u|_{\Gamma_{ij}} = \nabla u|_{\Gamma_{ji}} \tag{25}$$

valid for the solution u satisfying (12), relations (17) and the boundary condition (2), we obtain the identity

$$\left(\frac{\partial u}{\partial t}(t), v\right) + \sum_{i \in I} \sum_{j \in S(i)} \int\limits_{\Gamma_{ij}} \sum_{s=1}^{d} f_s(u(t))\, (n_{ij})_s\, v|_{\Gamma_{ij}} \,\mathrm{d}S \tag{26}$$
$$- \sum_{i \in I} \int\limits_{K_i} \sum_{s=1}^{d} f_s(u(t))\, \frac{\partial v}{\partial x_s} \,\mathrm{d}x + \varepsilon \sum_{i \in I} \int\limits_{K_i} \nabla u(t) \cdot \nabla v \,\mathrm{d}x$$
$$- \varepsilon \sum_{i \in I} \sum_{\substack{j \in s(i) \\ j < i}} \int\limits_{\Gamma_{ij}} \Big(\langle \nabla u(t) \rangle \cdot \boldsymbol{n}_{ij} [v] - \langle \nabla v \rangle \cdot \boldsymbol{n}_{ij} [u(t)] \Big) \,\mathrm{d}S$$
$$- \varepsilon \sum_{i \in I} \sum_{j \in \gamma(i)} \int\limits_{\Gamma_{ij}} \Big((\nabla u(t) \cdot \boldsymbol{n}_{ij})\, v - (\nabla v \cdot \boldsymbol{n}_{ij})\, u(t) \Big) \,\mathrm{d}S$$
$$+ \varepsilon \left(\sum_{i \in I} \sum_{j \in s(i)} \int\limits_{\Gamma_{ij}} \sigma [u(t)]\, [v] \,\mathrm{d}S + \sum_{i \in I} \sum_{j \in \gamma(i)} \int\limits_{\Gamma_{ij}} \sigma\, u(t)\, v \,\mathrm{d}S \right)$$
$$= \int\limits_{\Omega} g(t)\, v \,\mathrm{d}x + \varepsilon \sum_{i \in I} \sum_{j \in \gamma(i)} \int\limits_{\Gamma_{ij}} (\nabla v \cdot \boldsymbol{n}_{ij})\, u_D(t) \,\mathrm{d}S$$
$$+ \varepsilon \sum_{i \in I} \sum_{j \in \gamma(i)} \int\limits_{\Gamma_{ij}} \sigma u_D(t)\, v \,\mathrm{d}S,$$

where the function $\sigma : \bigcup_{i \in I} \bigcup_{j \in S(i)} \Gamma_{ij} \to I\!R$ is defined by

$$\sigma|_{\Gamma_{ij}} = \frac{1}{d(\Gamma_{ij})}. \tag{27}$$

We see that we have added the terms

$$\varepsilon \int\limits_{\Gamma_{ij}} \left\{ (\nabla v \cdot \boldsymbol{n}_{ij})\, u(t) + \sigma\, u(t)\, v \right\} \,\mathrm{d}S$$

and

$$\varepsilon \int_{\Gamma_{ij}} \{(\nabla v \cdot \boldsymbol{n}_{ij})\, u_D(t) + \sigma\, u_D(t)\, v\}\, \mathrm{d}S,$$

$i \in I,\ j \in \gamma(i)$, to the left-hand side and the right-hand side of (24), respectively. Due to the boundary condition (2), these terms are equal. Above and in the sequel, in the integrals $\int_{\Gamma_{ij}}$, under $\langle\cdot\rangle$ and $[\cdot]$ we understand $\langle\cdot\rangle_{\Gamma_{ij}}$ and $[\cdot]_{\Gamma_{ij}}$, respectively.

On the basis of the above considerations, we introduce for $u, v \in H^2(\Omega, T_h)$ the forms

$$\begin{aligned}
a_h(u,v) = {} & \varepsilon \sum_{i\in I} \int_{K_i} \nabla u \cdot \nabla v \,\mathrm{d}x && (28)\\
& - \varepsilon \sum_{i\in I} \sum_{\substack{j\in s(i)\\ j<i}} \int_{\Gamma_{ij}} \Big(\langle \nabla u\rangle \cdot \boldsymbol{n}_{ij} [v] - \langle \nabla v\rangle \cdot \boldsymbol{n}_{ij} [u] \Big)\, \mathrm{d}S \\
& - \varepsilon \sum_{i\in I} \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} \Big((\nabla u \cdot \boldsymbol{n}_{ij})\, v - (\nabla v \cdot \boldsymbol{n}_{ij})\, u \Big)\, \mathrm{d}S, \\
\tilde{b}_h(u,v) = {} & \sum_{i\in I} \sum_{j\in S(i)} \int_{\Gamma_{ij}} \sum_{s=1}^{d} f_s(u)\, (n_{ij})_s\, v|_{\Gamma_{ij}}\, \mathrm{d}S && (29)\\
& - \sum_{i\in I} \int_{K_i} \sum_{s=1}^{d} f_s(u)\, \frac{\partial v}{\partial x_s}\, \mathrm{d}x, \\
\ell_h(v)\,(t) = {} & (g(t), v) + \varepsilon \sum_{i\in I} \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} u_D(t)\, (\nabla v \cdot \boldsymbol{n}_{ij})\, \mathrm{d}S && (30)\\
& + \varepsilon \sum_{i\in I} \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} \sigma u_D(t)\, v\, \mathrm{d}S, \\
J_h^\sigma(u,v) = {} & \sum_{i\in I} \sum_{j\in s(i)} \int_{\Gamma_{ij}} \sigma [u]\, [v]\, \mathrm{d}S + \sum_{i\in I} \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} \sigma u v\, \mathrm{d}S. && (31)
\end{aligned}$$

In the discrete problem the form J_h^σ represents a *stabilization interior and boundary penalty* replacing the requirement of the continuity of the approximate solution in conforming FE methods. Let us note that we use here the so-called *nonsymmetric formulation of the diffusion terms.*

3.4 Discretization

Now we shall introduce the discrete problem approximating identity (26) with the aid of the DGFE method. We define the space of discontinuous piecewise polynomial functions

$$S_h = S^{p,-1}(\Omega, T_h) = \{v; v|_K \in P_p(K)\ \forall K \in T_h\}, \tag{32}$$

where $P_p(K)$ denotes the space of all polynomials on K of degree $\le p$, where the integer $p \ge 1$ is a given degree of approximation.

We derive the discrete problem in such a way that in (26) $u(t)$ is approximated by $u_h(t) \in S_h$ and v is replaced by $v_h \in S_h$. Similarly as in the finite volume method, for $i \in I$, $j \in S(i)$, the fluxes $\int_{\Gamma_{ij}} \sum_{s=1}^d f_s(u)\, n_s\, v \,\mathrm{d}S$ are approximated with the aid of the so-called *numerical flux* $H(u, u', \boldsymbol{n})$ by the expression $\int_{\Gamma_{ij}} H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ji}}, \boldsymbol{n}_{ij})\, v|_{\Gamma_{ij}} \,\mathrm{d}S$. Of course, if $j \in \gamma(i)$, then $\Gamma_{ij} \subset \partial\Omega$ and it is necessary to specify the meaning of $u|_{\Gamma_{ji}}$. Here we use the extrapolation, i.e. we set $u|_{\Gamma_{ji}} := u|_{\Gamma_{ij}}$.

In this way we obtain the approximation $b_h(u, v)$ of the convection form $\tilde{b}_h(u, v)$:

$$\begin{aligned} b_h(u,v) = \sum_{i\in I} \Bigg(& \sum_{j\in s(i)} \int_{\Gamma_{ij}} H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ji}}, \boldsymbol{n}_{ij})\, v|_{\Gamma_{ij}} \,\mathrm{d}S \\ & + \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ij}}, \boldsymbol{n}_{ij})\, v|_{\Gamma_{ij}} \,\mathrm{d}S \Bigg) \\ & - \sum_{i\in I} \int_{K_i} \sum_{s=1}^d f_s(u) \frac{\partial v}{\partial x_s} \,\mathrm{d}x, \quad u,\, v \in H^1(\Omega, T_h), \quad u \in L^\infty(\Omega). \end{aligned} \tag{33}$$

We shall assume that the numerical flux has the following properties:

Assumptions (H):

(1) $H(u, v, \boldsymbol{n})$ is defined in $I\!R^d \times B_1$, where $B_1 = \{\boldsymbol{n} \in I\!R^d; |\boldsymbol{n}| = 1\}$, and *Lipschitz-continuous* with respect to u, v:

$$\begin{gathered} |H(u, v, \boldsymbol{n}) - H(u^*, v^*, \boldsymbol{n})| \le C_2(|u - u^*| + |v - v^*|), \\ u,\, v,\, u^*,\, v^* \in I\!R,\ \boldsymbol{n} \in B_1. \end{gathered} \tag{34}$$

(2) $H(u, v, \boldsymbol{n})$ is *consistent*:

$$H(u, u, \boldsymbol{n}) = \sum_{s=1}^d f_s(u)\, n_s, \quad u \in I\!R,\ \boldsymbol{n} = (n_1, \dots, n_d) \in B_1. \tag{35}$$

(3) $H(u, v, \boldsymbol{n})$ is *conservative*:

$$H(u,v,\boldsymbol{n}) = -H(v,u,-\boldsymbol{n}), \quad u,\, v \in I\!R,\ \boldsymbol{n} \in B_1. \tag{36}$$

In virtue of (34) and (35), the functions $f_s,\ s = 1,\dots,d$, are Lipschitz-continuous with constant $L_f = 2C_2$.

From (4), a) and (35) we see that

$$H(0,0,\boldsymbol{n}) = 0 \quad \forall\, \boldsymbol{n} \in B_1. \tag{37}$$

By (29), (33) and (35),

$$b_h(u,v) = \tilde{b}_h(u,v) \quad \forall u \in H^2(\Omega),\ \forall v \in H^2(\Omega, T_h). \tag{38}$$

Since $S_h \subset H^2(\Omega, T_h)$, the forms (28) – (31) make sense for $u := u_h,\ v := v_h \in S_h$. We define the *semidiscrete approximate solution* as a function $u_h : Q_T \to I\!R$ satisfying the conditions

$$\begin{aligned}
&\text{a)} \quad u_h \in C^1([0,T]; S_h), \\
&\text{b)} \quad \left(\frac{\partial u_h}{\partial t}(t), v_h\right) + a_h(u_h(t), v_h) + b_h(u_h(t), v_h) \\
&\qquad + \varepsilon J_h^\sigma(u_h(t), v_h) = \ell_h(v_h)\,(t) \quad \forall\, v_h \in S_h,\ \forall\, t \in [0,T], \\
&\text{c)} \quad (u_h(0), v_h) = (u^0, v_h) \quad \forall\, v_h \in S_h.
\end{aligned} \tag{39}$$

The discrete problem (39), a) – c) is equivalent to a system of ordinary differential equations (ODE). (This approach to the numerical solution of initial-boundary value problems via the space semidiscretization is called the method of lines.) If we apply some ODE solver to problem (39), a) – c), we obtain the fully discrete problem. In what follows we shall be concerned with the analysis of the semidiscrete problem (39), a) – c).

Taking into account that the exact solution with property (12) satisfies $[u]_{\Gamma_{ij}} = 0$, $u|_{\partial\Omega\times(0,T)} = u_D$ and using (26), (29) and (38), we find that u satisfies the identity

$$\left(\frac{\partial u}{\partial t}(t), v_h\right) + a_h(u(t), v_h) + b_h(u(t), v_h) + \varepsilon J_h^\sigma(u(t), v_h) = \ell_h(v_h)\,(t) \tag{40}$$

for all $v_h \in S_h$ and almost all $t \in (0,T)$, which implies the Galerkin orthogonality property of the error.

4 Some auxiliary results

In this section we shall be concerned with some important approximation results and properties of the form b_h.

4.1 Geometry of the mesh

Let us consider a system $\{T_h\}_{h\in(0,h_0)}$, $h_0 > 0$, of partitions of the domain Ω, i.e. $T_h = \{K_i\}_{i\in I_h}$, $I_h \subset Z^+$. For the sake of simplicity, we shall write I instead of I_h $(h \in (0, h_0))$ and the dependence of index sets $I, I_b, s(i), \gamma(i)$ and $S(i)$ on h will not be emphasized by the notation.

Let us assume that the system $\{T_h\}_{h\in(0,h_0)}$ has the following properties:

(A1) Each element $K \in T_h$, $h \in (0, h_0)$, is a *star-shaped* domain with respect to at least one point $x_K = (x_{K1}, \dots, x_{Kd}) \in K^\circ$, where K° is the interior of K. We assume that

i) there exists a constant $\kappa > 0$ independent of K and h such that

$$\frac{\max_{x\in\partial K} |x - x_K|}{\min_{x\in\partial K} |x - x_K|} \le \kappa \quad \forall K \in T_h,\ h \in (0, h_0), \tag{41}$$

ii) element K can be divided into a finite number of closed simplexes:

$$K = \bigcup_{S\in\omega(K)} S; \tag{42}$$

there exists a positive constant C_3 independent of K, S and h such that

$$\frac{h_S}{\rho_S} \le C_3 \quad \forall S \in \omega(K) \qquad \text{(shape regularity)}, \tag{43}$$

where h_S is the diameter of S, ρ_S is the radius of the largest d-dimensional ball inscribed into S and, moreover,

$$1 \le \frac{h_K}{h_S} \le \tilde{\kappa} < \infty \quad \forall S \in \omega(K), \tag{44}$$

where $\tilde{\kappa}$ is a constant independent of K, S and h.

(A2) There exists a constant $C_4 > 0$ such that

$$h_{K_i} \le C_4\, d(\Gamma_{ij}), \quad i \in I, j \in S(i), h \in (0, h_0). \tag{45}$$

Let us note that these properties can be verified, e.g., in the case of dual finite volumes constructed over a regular simplicial mesh.

In what follows we shall cite some technical results formulated as Lemmas 1–4. Because of completeness of the exposition, we shall prove them in Appendix.

Lemma 1 *Let assumption* (A1) *be valid. Then for each* $K \in T_h,\ h \in (0, h_0)$, *there exist axiparallel boxes* $B_0,\ B_1$ *such that*

i) $B_0 \subset K \subset B_1$,
ii) centres of B_0 *and* B_1 *are identical and equal to* x_K,
iii) we have

$$1 \le \frac{h_{B_1}}{h_{B_0}} \le \bar{\kappa} < \infty, \tag{46}$$

where $\bar{\kappa}$ *is a constant independent of* K *and* h *and the numbers* h_{B_0} *and* h_{B_1} *are the half of sizes of* B_0 *and* B_1, *respectively, i.e.* $B_l = \prod_{s=1}^{d}[x_{Ks} - h_{B_l}, x_{Ks} + h_{B_l}]$, $l = 0, 1,$ *and*

$$h_K \le 2h_{B_1}. \tag{47}$$

4.2 Some important inequalities and estimates

Under the above assumptions, the following results can be established.

Lemma 2 (Multiplicative trace inequality) *There exists a constant* $C_5 > 0$ *independent of* v, h *and* K *such that*

$$\|v\|^2_{L^2(\partial K)} \le C_5 \left(\|v\|_{L^2(K)} \, |v|_{H^1(K)} + h_K^{-1} \|v\|^2_{L^2(K)} \right), \tag{48}$$
$$K \in T_h,\ v \in H^1(K),\ h \in (0, h_0).$$

Lemma 3 (Inverse inequality) *There exists a constant* $C_6 > 0$ *independent of* v, h *and* K *such that*

$$|v|_{H^1(K)} \le C_6 h_K^{-1} \|v\|_{L^2(K)}, \quad v \in P_p(K),\ K \in T_h,\ h \in (0, h_0). \tag{49}$$

An important question represent approximation properties of the space S_h constructed over the mesh T_h.

Lemma 4 *There exist a constant* $C_7 > 0$ *independent of* v *and* h *and a mapping* $\Pi : H^1(K) \to P_p(K)$ *such that*

$$\|\Pi v - v\|_{L^2(K)} \le C_7 h_K^{p+1} |v|_{H^{p+1}(K)}, \tag{50}$$

$$
\begin{aligned}
|\Pi v - v|_{H^1(K)} &\le C_7 h_K^p |v|_{H^{p+1}(K)},\\
|\Pi v - v|_{H^2(K)} &\le C_7 h_K^{p-1} |v|_{H^{p+1}(K)}
\end{aligned}
$$

for all $v \in H^{p+1}(K),\ K \in T_h$ *and* $h \in (0, h_0)$.

4.3 Properties of the form b_h

Let assumptions (4), a), Assumptions (H) in Section 3.4 and Assumptions (A1), (A2) in Section 4.1 be satisfied. We shall be concerned with the Lipschitz-continuity of the form b_h.

Lemma 5 *There exist constants* $C_8 > 0,\ C_9 > 0$ *such that*

$$
\begin{aligned}
&|b_h(u,v) - b_h(\bar{u},v)| \le C_8 \left(J_h^\sigma(v,v)^{1/2} + |v|_{H^1(\Omega,T_h)} \right) \qquad (51)\\
&\times \left(\|u-\bar{u}\|_{L^2(\Omega)} + \left(\sum_{i\in I} h_{K_i} \|u-\bar{u}\|^2_{L(\partial K_i)} \right)^{1/2} \right),\\
&\qquad\qquad u,\ \bar{u},\ v \in H^1(\Omega, T_h),\ h \in (0,h_0),
\end{aligned}
$$

and

$$
\begin{aligned}
&|b_h(u_h,v_h) - b_h(\bar{u}_h,v_h)| \qquad (52)\\
&\le C_9 \left(J_h^\sigma(v_h,v_h)^{1/2} + |v_h|_{H^1(\Omega,T_h)} \right) \|u_h - \bar{u}_h\|_{L^2(\Omega)},\\
&\qquad\qquad u_h,\ \bar{u}_h,\ v_h \in S_h,\ h \in (0,h_0).
\end{aligned}
$$

PROOF. By (33), for $u,\ \bar{u},\ v \in H^1(\Omega, T_h)$,

$$
\begin{aligned}
b_h(u,v) - b_h(\bar{u},v) = &\underbrace{-\sum_{i\in I}\int_{K_i} \sum_{s=1}^{d} \left(f_s(u) - f_s(\bar{u})\right) \frac{\partial v}{\partial x_s}\,\mathrm{d}x}_{:=\sigma_1} \qquad (53)\\
&+\underbrace{\sum_{i\in I}\sum_{j\in s(i)}\int_{\Gamma_{ij}} \Big(H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) - H(\bar{u}|_{\Gamma_{ij}}, \bar{u}|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \Big) v|_{\Gamma_{ij}}\,\mathrm{d}S}_{:=\sigma_2}\\
&+\underbrace{\sum_{i\in I}\sum_{j\in \gamma(i)}\int_{\Gamma_{ij}} \Big(H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ij}}, \boldsymbol{n}_{ij}) - H(\bar{u}|_{\Gamma_{ij}}, \bar{u}|_{\Gamma_{ij}}, \boldsymbol{n}_{ij}) \Big) v|_{\Gamma_{ij}}\,\mathrm{d}S}_{:=\sigma_3}.
\end{aligned}
$$

From the Lipschitz-continuity of the functions $f_s,\ s=1,\dots,d$, we have

$$|\sigma_1| \le L_f \sum_{i\in I} \int_{K_i} \sum_{s=1}^{d} |u-\bar{u}| \left|\frac{\partial v}{\partial x_s}\right| \,\mathrm{d}x \le \sqrt{d} L_f \|u-\bar{u}\|_{L^2(\Omega)} |v|_{H^1(\Omega,T_h)}. \quad (54)$$

Using the conservativity (36) of H we find that

$$\begin{aligned}
\sigma_2+\sigma_3 =& \sum_{i\in I} \left\{ \sum_{j\in s(i)} \int_{\Gamma_{ij}} v|_{\Gamma_{ij}} \Big(H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) - H(\bar{u}|_{\Gamma_{ij}}, \bar{u}|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \Big) \,\mathrm{d}S \right.\\
&\left. + \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} v|_{\Gamma_{ij}} \Big(H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ij}}, \boldsymbol{n}_{ij}) - H(\bar{u}|_{\Gamma_{ij}}, \bar{u}|_{\Gamma_{ij}}, \boldsymbol{n}_{ij}) \Big) \,\mathrm{d}S \right\}\\
=& \sum_{i\in I} \left\{ \frac{1}{2} \sum_{j\in s(i)} \int_{\Gamma_{ij}} \Big(v|_{\Gamma_{ij}} - v|_{\Gamma_{ji}} \Big) \right.\\
&\times \Big(H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) - H(\bar{u}|_{\Gamma_{ij}}, \bar{u}|_{\Gamma_{ji}}, \boldsymbol{n}_{ij}) \Big) \,\mathrm{d}S\\
&\left. + \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} v|_{\Gamma_{ij}} \Big(H(u|_{\Gamma_{ij}}, u|_{\Gamma_{ij}}, \boldsymbol{n}_{ij}) - H(\bar{u}|_{\Gamma_{ij}}, \bar{u}|_{\Gamma_{ij}}, \boldsymbol{n}_{ij}) \Big) \,\mathrm{d}S \right\}.
\end{aligned}$$

This, the Lipschitz-continuity (34) of H, the Cauchy inequality and (19) imply that

$$\begin{aligned}
|\sigma_2+\sigma_3| \le& C_2 \sum_{i\in I} \left\{ \frac{1}{2} \sum_{j\in s(i)} \int_{\Gamma_{ij}} \left|[v]_{\Gamma_{ij}}\right| \Big(|u-\bar{u}|_{\Gamma_{ij}} + |u-\bar{u}|_{\Gamma_{ji}} \Big) \,\mathrm{d}S \right.\\
&\left. + 2 \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} |v|_{\Gamma_{ij}}|\, |u-\bar{u}|_{\Gamma_{ij}} \,\mathrm{d}S \right\} \qquad (55)\\
\le& C_2 \sum_{i\in I} \left\{ \sum_{j\in s(i)} \int_{\Gamma_{ij}} \left|[v]_{\Gamma_{ij}}\right| |u-\bar{u}|_{\Gamma_{ij}} \,\mathrm{d}S \right.\\
&\left. + 2 \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} |v|_{\Gamma_{ij}}|\, |u-\bar{u}|_{\Gamma_{ij}} \,\mathrm{d}S \right\}\\
\le& 2C_2 \sum_{i\in I} \left\{ \left(\sum_{j\in s(i)} \int_{\Gamma_{ij}} \frac{[v]^2}{d(\Gamma_{ij})} \,\mathrm{d}S + \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} \frac{v^2}{d(\Gamma_{ij})} \,\mathrm{d}S \right)^{1/2} \right.
\end{aligned}$$

$$\times \left(\sum_{j \in s(i)} d(\Gamma_{ij}) \int_{\Gamma_{ij}} (u - \bar{u})^2 \, \mathrm{d}S \right.$$

$$\left. \left. + \sum_{j \in \gamma(i)} d(\Gamma_{ij}) \int_{\Gamma_{ij}} (u - \bar{u})^2 \, \mathrm{d}S \right)^{1/2} \right\}$$

$$\leq 2C_2 \, J_h^\sigma(v,v)^{1/2} \left(\sum_{i \in I} h_{K_i} \|u - \bar{u}\|^2_{L^2(\partial K_i)} \right)^{1/2} .$$

Now from (54) and (55) we get (51) with $C_8 = \max(\sqrt{d} L_f, 2C_2)$.

Further, let $u_h, \ \bar{u}_h, \ v_h \in S_h$. We want to prove estimate (52). In the above we set $u := u_h, \ \bar{u} := \bar{u}_h$ and $v := v_h$. Using the multiplicative trace inequality (48), the discrete Cauchy inequality and the inverse inequality (49), we obtain for $\varphi \in S_h$

$$\begin{aligned} \sum_{i \in I} h_{K_i} \|\varphi\|^2_{L^2(\partial K_i)} &\leq C_5 \sum_{i \in I} \left(\|\varphi\|^2_{L^2(K_i)} + h_{K_i} \|\varphi\|_{L^2(K_i)} \, |\varphi|_{H^1(K_i)} \right) \\ &\leq C_5 \left(\|\varphi\|^2_{L^2(\Omega)} + \|\varphi\|_{L^2(\Omega)} \left(\sum_{i \in I} h^2_{K_i} |\varphi|^2_{H^1(K_i)} \right)^{1/2} \right) \\ &\leq C_5 \left(\|\varphi\|^2_{L^2(\Omega)} + \|\varphi\|_{L^2(\Omega)} C_6 \left(\sum_{i \in I} \|\varphi\|^2_{L^2(K_i)} \right)^{1/2} \right) \\ &= C_5 (1 + C_6) \|\varphi\|^2_{L^2(\Omega)}. \end{aligned} \tag{56}$$

Now relations (55) and (56), where we put $\varphi := u_h - \bar{u}_h$, yield

$$|\sigma_2 + \sigma_3| \leq 2C_2 (C_5(1 + C_6))^{1/2} \, J_h^\sigma(v_h, v_h)^{1/2} \|u_h - \bar{u}_h\|_{L^2(\Omega)}. \tag{57}$$

From (53), (54) and (57) we get the final result

$$\begin{aligned} &|b_h(u_h, v_h) - b_h(\bar{u}_h, v_h)| \\ &\leq C_9 \left(J_h^\sigma(v_h, v_h)^{1/2} + |v_h|_{H^1(\Omega, T_h)} \right) \|u_h - \bar{u}_h\|_{L^2(\Omega)} \end{aligned} \tag{58}$$

with $C_9 = \max(\sqrt{d} L_f, 2C_2 (C_5(1 + C_6))^{1/2})$. □

Lemma 6 *There exists a constant* $C_{10} > 0$ *such that*

$$\begin{aligned} &|b_h(u, v_h) - b_h(\Pi u, v_h)| \leq \\ &\leq C_{10} h^{p+1} \left(J_h^\sigma(v_h, v_h)^{1/2} + |v_h|_{H^1(\Omega, T_h)} \right) |u|_{H^{p+1}(\Omega)}, \\ &\qquad u \in H^{p+1}(\Omega), \ v_h \in S_h, \ h \in (0, h_0), \end{aligned} \tag{59}$$

where Πu *is the* S_h*-interpolant of* u *from Lemma 4.*

PROOF. We start from (51) with $u \in H^{p+1}(\Omega)$, $\bar{u} := \Pi\, u$ and $v := v_h \in S_h$. Using the multiplicative trace inequality (48), the discrete Cauchy inequality and (50), we obtain

$$\begin{aligned}\sum_{i\in I} h_{K_i} \|u - \Pi u\|^2_{L^2(\partial K_i)} \leq & \sum_{i\in I} C_5 \Big(\|u - \Pi u\|^2_{L^2(K_i)} \\ & + h_{K_i} \|u - \Pi u\|_{L^2(K_i)} |u - \Pi u|_{H^1(K_i)} \Big) \quad (60) \\ \leq & \sum_{i\in I} C_5 C_7^2 \left(h_{K_i}^{2p+2} + h_{K_i} h_{K_i}^{p+1} h_{K_i}^{p} \right) |u|^2_{H^{p+1}(K_i)} \\ \leq & \, 2C_5 C_7^2 h^{2p+2} |u|^2_{H^{p+1}(\Omega)}.\end{aligned}$$

This estimate, (50) and (51) yield (59) with $C_{10} = C_7 C_8 (\sqrt{2C_5} + 1)$. $\square$

5 Error estimates for the method of lines

Our goal is now to analyze the error estimates of the approximate solution u_h obtained by the method of lines (39) under the assumption that the exact solution u satisfies (12).

Let $\Pi u(t)$ be the S_h-interpolation of $u(t)$ ($t \in [0, T]$) from Lemma 4. We set

$$\xi = u_h - \Pi u \in S_h, \quad \eta = \Pi u - u \in H^{p+1}(\Omega, T_h). \tag{61}$$

Then the error e_h can be expressed as

$$e_h = u_h - u = \xi + \eta. \tag{62}$$

From (38) – (40), where we substitute $v_h := \xi$, it follows that

$$\begin{aligned}&\left(\frac{\partial \xi}{\partial t}, \xi \right) + a_h(\xi, \xi) + \varepsilon J_h^\sigma(\xi, \xi) \qquad (63) \\ &= b_h(u, \xi) - b_h(u_h, \xi) - \left(\frac{\partial \eta}{\partial t}, \xi \right) - a_h(\eta, \xi) - \varepsilon J_h^\sigma(\eta, \xi).\end{aligned}$$

(Of course, $\xi = \xi(t),\ \eta = \eta(t)$ for $t \in [0,T]$, but we do not emphasize the dependence on t by our notation, if it is not necessary.) In what follows, we estimate the individual terms on the right-hand side of (63).

The Cauchy inequality implies that

$$\left|\left(\frac{\partial \eta}{\partial t}, \xi\right)\right| \le \left\|\frac{\partial \eta}{\partial t}\right\|_{L^2(\Omega)} \|\xi\|_{L^2(\Omega)}, \tag{64}$$

$$J_h^\sigma(\eta,\xi) \le \left(J_h^\sigma(\eta,\eta)\right)^{1/2} \left(J_h^\sigma(\xi,\xi)\right)^{1/2}. \tag{65}$$

Lemma 7 *For all* $h \in (0, h_0)$ *we have*

$$\begin{aligned} \|\eta\|_{L^2(\Omega)} &\le C_7 h^{p+1} |u|_{H^{p+1}(\Omega)}, \\ |\eta|_{H^1(\Omega,T_h)} &\le C_7 h^p |u|_{H^{p+1}(\Omega)}, \\ |\eta|_{H^2(\Omega,T_h)} &\le C_7 h^{p-1} |u|_{H^{p+1}(\Omega)}, \\ \left\|\frac{\partial \eta}{\partial t}\right\|_{L^2(\Omega)} &\le C_7 h^{p+1} \left|\frac{\partial u}{\partial t}\right|_{H^{p+1}(\Omega)}, \end{aligned} \tag{66}$$

where $C_7 > 0$ *is the constant from Lemma 4.*

PROOF. The proof is a consequence of (12), (13), (50) and the relation

$$\Pi(\partial u/\partial t) = \partial(\Pi u)/\partial t, \tag{67}$$

following from the construction of the interpolation operator Π in [23]. $\square$

Lemma 8 *For* $h \in (0, h_0)$ *we have*

$$\begin{aligned} &|b_h(u,\xi) - b_h(u_h,\xi)| \\ &\le C_{11}\left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)}\right)\left(h^{p+1}|u|_{H^{p+1}(\Omega)} + \|\xi\|_{L^2(\Omega)}\right), \end{aligned} \tag{68}$$

where $C_{11} > 0$ *is a constant independent of* h *and* ξ*.*

PROOF. We can write

$$b_h(u,\xi) - b_h(u_h,\xi) = (b_h(u,\xi) - b_h(\Pi u,\xi)) + (b_h(\Pi u,\xi) - b_h(u_h,\xi)). \tag{69}$$

From Lemmas 5 and 6 we find that

$$|b_h(\Pi u,\xi) - b_h(u_h,\xi)| \le C_9 \left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)}\right) \|\Pi u - u_h\|_{L^2(\Omega)}, \tag{70}$$
$$|b_h(u,\xi) - b_h(\Pi u,\xi)| \le C_{10} h^{p+1} \left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)}\right) |u|_{H^{p+1}(\Omega)}.$$

By (61), (69) and (70),

$$\begin{aligned}|b_h(u,\xi) - b_h(u_h,\xi)| \le \max(C_9, C_{10}) & \left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)}\right) \\ & \times \left(h^{p+1}|u|_{H^{p+1}(\Omega)} + \|\xi\|_{L^2(\Omega)}\right),\end{aligned}$$

which is (68) with $C_{11} = \max(C_9, C_{10})$. $\square$

Lemma 9 *There exists a constant $C_{12} > 0$ independent of u, h, ξ, ε such that*

$$|a_h(\eta,\xi)| \le C_{12} \varepsilon h^p |u|_{H^{p+1}(\Omega)} \left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)}\right), \quad h \in (0,h_0). \tag{71}$$

PROOF. In view of (28), we can write

$$\begin{aligned} a_h(\eta,\xi) &= \sigma_1 + \sigma_2 + \sigma_3, \\ \sigma_1 &:= \varepsilon \sum_{i\in I} \int_{K_i} \nabla\eta \cdot \nabla\xi \, \mathrm{d}x, \\ \sigma_2 &:= -\varepsilon \sum_{i\in I} \sum_{\substack{j\in s(i)\\ j<i}} \int_{\Gamma_{ij}} \left(\langle \nabla\eta \rangle \cdot \boldsymbol{n}_{ij} [\xi] - \langle \nabla\xi \rangle \cdot \boldsymbol{n}_{ij} [\eta] \right) \mathrm{d}S, \\ \sigma_3 &:= -\varepsilon \sum_{i\in I} \sum_{j\in\gamma(i)} \int_{\Gamma_{ij}} \left((\nabla\eta \cdot \boldsymbol{n}_{ij})\, \xi - (\nabla\xi \cdot \boldsymbol{n}_{ij})\, \eta \right) \mathrm{d}S. \end{aligned} \tag{72}$$

We estimate the individual terms in (72) similarly as in [18].

The Cauchy inequality and estimates (66) imply that

$$|\sigma_1| \le \varepsilon C_7 h^p |u|_{H^{p+1}(\Omega)} |\xi|_{H^1(\Omega,T_h)}. \tag{73}$$

Now we shall estimate σ_2. The Cauchy inequality, (31), (27) and (18) yield the estimates

$$\begin{aligned}|\sigma_2| \le \varepsilon & \left(\sum_{i\in I} \sum_{j\in s(i)} d(\Gamma_{ij}) \|\nabla\eta\|^2_{L^2(\Gamma_{ij})} \right)^{1/2} \left(\sum_{i\in I} \sum_{j\in s(i)} \frac{1}{d(\Gamma_{ij})} \|[\xi]\|^2_{L^2(\Gamma_{ij})} \right)^{1/2} \\ +\varepsilon & \left(\sum_{i\in I} \sum_{j\in s(i)} d(\Gamma_{ij}) \|\nabla\xi\|^2_{L^2(\Gamma_{ij})} \right)^{1/2} \left(\sum_{i\in I} \sum_{j\in s(i)} \frac{1}{d(\Gamma_{ij})} \|[\eta]\|^2_{L^2(\Gamma_{ij})} \right)^{1/2}\end{aligned} \tag{74}$$

$$
\begin{aligned}
&\le \varepsilon \left(\sum_{i\in I} h_{K_i} \|\nabla \eta\|^2_{L^2(\partial K_i)} \right)^{1/2} J_h^\sigma(\xi,\xi)^{1/2} \\
&\quad + \varepsilon \left(\sum_{i\in I} h_{K_i} \|\nabla \xi\|^2_{L^2(\partial K_i)} \right)^{1/2} J_h^\sigma(\eta,\eta)^{1/2}.
\end{aligned}
$$

Taking into account that $\eta = \Pi u - u \in H^2(\Omega, T_h)$, we have

$$
\|\nabla \eta\|_{L^2(K_i)} = |\eta|_{H^1(K_i)}, \quad |\nabla \eta|_{H^1(K_i)} = |\eta|_{H^2(K_i)}, \quad i \in I. \tag{75}
$$

Similarly as in (60), from (48) (where we set $v := |\nabla \eta|$) and (66) we find that

$$
\|\nabla \eta\|^2_{L^2(\partial K_i)} \le 2C_5 C_7^2 |u|^2_{H^{p+1}(K_i)} h_{K_i}^{2p-1}. \tag{76}
$$

Hence,

$$
\left(\sum_{i\in I} h_{K_i} \|\nabla \eta\|^2_{L^2(\partial K_i)} \right)^{1/2} \le \sqrt{2C_5}\, C_7\, h^p |u|_{H^{p+1}(\Omega)}. \tag{77}
$$

From (48) and (49), where we put $v := |\nabla \xi|$, we get

$$
\begin{aligned}
\sum_{i\in I} h_{K_i} \|\nabla \xi\|^2_{L^2(\partial K_i)} &\le \sum_{i\in I} C_5 \left(\|\nabla \xi\|^2_{L^2(K_i)} + h_{K_i} \|\nabla \xi\|_{L^2(K_i)} |\nabla \xi|_{H^1(K_i)} \right) \\
&\le \sum_{i\in I} C_5 \left(\|\nabla \xi\|^2_{L^2(K_i)} + C_6 \|\nabla \xi\|^2_{L^2(K_i)} \right) \\
&= C_5 (1 + C_6) |\xi|^2_{H^1(\Omega, T_h)}.
\end{aligned} \tag{78}
$$

Now it remains to estimate the term $J_h^\sigma(\eta,\eta)$. In virtue of (31), (45), (48) and (50), we have

$$
\begin{aligned}
J_h^\sigma(\eta,\eta) &\le 4 \sum_{i\in I} \sum_{j\in S(i)} \int_{\Gamma_{ij}} \frac{1}{d(\Gamma_{ij})} |\eta|_{\Gamma_{ij}}|^2 \,\mathrm{d}S \qquad (79) \\
&\le 4C_4 \sum_{i\in I} \frac{1}{h_{K_i}} \int_{\partial K_i} |\eta|^2 \,\mathrm{d}S \\
&\le 4C_4\, C_5 \sum_{i\in I} \frac{1}{h_{K_i}} \left(\|\eta\|_{L^2(K_i)} |\eta|_{H^1(K_i)} + h_{K_i}^{-1} \|\eta\|^2_{L^2(K_i)} \right) \\
&\le 8C_4\, C_5\, C_7^2\, h^{2p} |u|^2_{H^{p+1}(\Omega)}.
\end{aligned}
$$

Finally, from (74) – (79) we conclude that

$$
|\sigma_2| \le \sigma(\varepsilon, h, u, \xi) := \varepsilon\, C\, h^p |u|_{H^{p+1}(\Omega)} \left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)} \right), \tag{80}
$$

where $C > 0$ is a constant independent of u, ξ, h, ε.

In a similar way we obtain the estimate of σ_3:

$$|\sigma_3| \leq \sigma(\varepsilon, h, u, \xi). \tag{81}$$

Estimates (73), (80) and (81) immediately imply (71). $\square$

In the convergence proof the following version of Gronwall's lemma will be applied:

Lemma 10 *Let* y, q, z, $r \in C([0,T])$, $r \geq 0$ *and let*

$$y(t) + q(t) \leq z(t) + \int_0^t r(s)\, y(s)\, \mathrm{d}s, \quad t \in [0,T]. \tag{82}$$

Then

$$\begin{aligned} & y(t) + q(t) + \int_0^t r(\vartheta)\, q(\vartheta) \exp\left(\int_\vartheta^t r(s)\, \mathrm{d}s\right) \mathrm{d}\vartheta \\ \leq\, & z(t) + \int_0^t r(\vartheta)\, z(\vartheta) \exp\left(\int_\vartheta^t r(s)\, \mathrm{d}s\right) \mathrm{d}\vartheta, \quad t \in [0,T]. \end{aligned} \tag{83}$$

PROOF. The proof can be carried out in a similar way as in [14], Par. 8.2.29. $\square$

Now we shall formulate the *main result*.

Theorem 11 *Let assumptions (4), a) – e),* (H) *from Section 3.4 and* (A1) – (A2) *from Section 4.1 be satisfied. Let u be the exact strong solution of problem (10) satisfying (12) and let u_h be the approximate solution defined by (39). Then the error $e_h = u_h - u$ satisfies the estimate*

$$\max_{t\in[0,T]} \|e_h(t)\|^2_{L^2(\Omega)} + \varepsilon \int_0^T \left(|e_h(\vartheta)|^2_{H^1(\Omega,T_h)} + J_h^\sigma(e_h(\vartheta), e_h(\vartheta)) \right) \mathrm{d}\vartheta \tag{84}$$

$$
\begin{aligned}
&+C_{13}(1+\varepsilon)\int_0^T \exp\left(C_{13}\frac{1+\varepsilon}{\varepsilon}(T-\vartheta)\right) \\
&\times\left\{\int_0^\vartheta \left(|e_h(\alpha)|^2_{H^1(\Omega,T_h)} + J_h^\sigma(e_h(\alpha),e_h(\alpha))\right)\mathrm{d}\alpha\right\}\mathrm{d}\vartheta \\
\le\, & 4h^{2p}\left\{\left(\frac{c_1(\varepsilon,h)}{\varepsilon}+c_2+\varepsilon C_{15}\right)\exp\left(C_{13}\frac{1+\varepsilon}{\varepsilon}T\right)+C_{14}\right\},
\end{aligned}
$$

with constants C_{13}, C_{14}, C_{15}, $c_2 > 0$ independent of h, ε and with a function $c_1(\varepsilon, h)$, specified in the proof.

PROOF. As in (61), we set $\xi = u_h - \Pi u \in S_h$, $\eta = \Pi u - u$. Then (62) holds: $e_h = u_h - u = \xi + \eta$. From (63), (86) and the relations

$$
a_h(\xi,\xi) = \varepsilon|\xi|^2_{H^1(\Omega,T_h)} \tag{85}
$$

and

$$
\left(\frac{\partial \xi}{\partial t},\xi\right) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\xi\|^2_{L^2(\Omega)} \tag{86}
$$

we get

$$
\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\xi\|^2_{L^2(\Omega)} + \varepsilon|\xi|^2_{H^1(\Omega,T_h)} + \varepsilon J_h^\sigma(\xi,\xi) \\
=\, & b_h(u,\xi) - b_h(u_h,\xi) - \left(\frac{\partial \eta}{\partial t},\xi\right) - a_h(\eta,\xi) - \varepsilon J_h^\sigma(\eta,\xi) \quad \text{in } (0,T).
\end{aligned} \tag{87}
$$

Now from (64) – (68), (71) and (79), using Young's inequality, we derive the estimates

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\|\xi\|^2_{L^2(\Omega)} + 2\varepsilon|\xi|^2_{H^1(\Omega,T_h)} + 2\varepsilon J_h^\sigma(\xi,\xi) \\
\le\, & 2C_{11}\left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)}\right)\left(\|\xi\|_{L^2(\Omega)} + h^{p+1}|u|_{H^{p+1}(\Omega)}\right) \\
&+2C_7 h^{p+1}\left|\frac{\partial u}{\partial t}\right|_{H^{p+1}(\Omega)}\|\xi\|_{L^2(\Omega)} \\
&+2C_{12}\,\varepsilon\, h^p|u|_{H^{p+1}(\Omega)}\left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)}\right) \\
&+2\varepsilon\sqrt{8C_4C_5}C_7 h^p|u|_{H^{p+1}(\Omega)}J_h^\sigma(\xi,\xi)^{1/2} \\
\le\, & 2C_{13}^*\left(J_h^\sigma(\xi,\xi)^{1/2} + |\xi|_{H^1(\Omega,T_h)}\right)\left(\|\xi\|_{L^2(\Omega)} + (\varepsilon h^p + h^{p+1})|u|_{H^{p+1}(\Omega)}\right)
\end{aligned} \tag{88}
$$

$$
\begin{aligned}
&+2C_7h^{p+1}\|\xi\|_{L^2(\Omega)}\left|\frac{\partial u}{\partial t}\right|_{H^{p+1}(\Omega)}\\
\le\, &\varepsilon J_h^\sigma(\xi,\xi)+\varepsilon|\xi|^2_{H^1(\Omega,T_h)}+C_{13}\left(1+\frac{1}{\varepsilon}\right)\|\xi\|^2_{L^2(\Omega)}\\
&+\frac{C_{13}}{\varepsilon}\left(\left(\varepsilon^2h^{2p}+h^{2p+2}\right)|u|^2_{H^{p+1}(\Omega)}\right)+C_7^2h^{2p+2}\left|\frac{\partial u}{\partial t}\right|^2_{H^{p+1}(\Omega)}\quad \text{in } (0,T),
\end{aligned}
$$

where $C_{13}^*=\max\{C_{11},C_{12}+\sqrt{8C_4C_5}C_7\}$, $C_{13}=\max\{1,8(C_{13}^*)^2\}$. Hence,

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\|\xi(t)\|^2_{L^2(\Omega)}+\varepsilon|\xi(t)|^2_{H^1(\Omega,T_h)}+\varepsilon J_h^\sigma(\xi(t),\xi(t)) \\
\le\, &C_{13}\left(1+\frac{1}{\varepsilon}\right)\|\xi(t)\|^2_{L^2(\Omega)}+\frac{C_{13}}{\varepsilon}\left(\left(\varepsilon^2h^{2p}+h^{2p+2}\right)|u(t)|^2_{H^{p+1}(\Omega)}\right)\\
&+C_7^2h^{2p+2}\left|\frac{\partial u(t)}{\partial t}\right|^2_{H^{p+1}(\Omega)},\quad t\in(0,T).
\end{aligned}
\tag{89}
$$

From assumption (12) it follows that the right-hand side of (89) is integrable over $(0,T)$. In virtue of (39), c) and estimate (66) we find that $\|\xi(0)\|_{L^2(\Omega)}\le\|\eta(0)\|_{L^2(\Omega)}$ and, thus,

$$
\|\xi(0)\|^2_{L^2(\Omega)}\le C_7^2h^{2p+2}|u^0|^2_{H^{p+1}(\Omega)}. \tag{90}
$$

The integration of (89) from 0 to $t\in[0,T]$ yields

$$
\begin{aligned}
&\|\xi(t)\|^2_{L^2(\Omega)}+\varepsilon\int_0^t\left(|\xi(\vartheta)|^2_{H^1(\Omega,T_h)}+J_h^\sigma(\xi(\vartheta),\xi(\vartheta))\right)\mathrm{d}\vartheta \\
\le\, &C_{13}\left(1+\frac{1}{\varepsilon}\right)\int_0^t\|\xi(\vartheta)\|^2_{L^2(\Omega)}\,\mathrm{d}\vartheta+\frac{C_{13}}{\varepsilon}h^{2p}\int_0^t\left(\left(\varepsilon^2+h^2\right)|u(\vartheta)|^2_{H^{p+1}(\Omega)}\right)\mathrm{d}\vartheta\\
&+C_7^2h^{2p+2}\int_0^t\left|\frac{\partial u(\vartheta)}{\partial t}\right|^2_{H^{p+1}(\Omega)}\mathrm{d}\vartheta+C_7^2h^{2p+2}|u^0|^2_{H^{p+1}(\Omega)},\quad t\in[0,T].
\end{aligned}
\tag{91}
$$

Now the application of Gronwall's Lemma 10, where we set

$$
\begin{aligned}
y(t)&=\|\xi(t)\|_{L^2(\Omega)},\\
q(t)&=\varepsilon\int_0^t\left(|\xi(\vartheta)|^2_{H^1(\Omega,T_h)}+J_h^\sigma(\xi(\vartheta),\xi(\vartheta))\right)\mathrm{d}\vartheta,\\
r(t)&=C_{13}\frac{1+\varepsilon}{\varepsilon},
\end{aligned}
\tag{92}
$$

$$
\begin{aligned}
z(t) &= h^{2p}\left(\frac{c_1}{\varepsilon}+c_2h^2\right),\\
c_1 &= c_1(\varepsilon,h) = C_{13}\left(\left(\varepsilon^2+h^2\right)\|u\|^2_{L^2(0,T;H^{p+1}(\Omega))}\right),\\
c_2 &= C_7^2\left(\|\partial u/\partial t\|^2_{L^2(0,T;H^{p+1}(\Omega))}+|u^0|^2_{H^{p+1}(\Omega)}\right),
\end{aligned}
$$

and the relations

$$
\begin{aligned}
&\int_\vartheta^t r(s)\mathrm{d}s = C_{13}\frac{1+\varepsilon}{\varepsilon}(t-\vartheta), \qquad (93)\\
&\int_0^t r(\vartheta)\,z(\vartheta)\exp\left(\int_\vartheta^t r(s)\,\mathrm{d}s\right)\mathrm{d}\vartheta = \left(\frac{c_1}{\varepsilon}+c_2h^2\right)h^{2p}\left(\exp\left(C_{13}\frac{1+\varepsilon}{\varepsilon}t\right)-1\right)
\end{aligned}
$$

imply that

$$
\begin{aligned}
&\|\xi(t)\|^2_{L^2(\Omega)}+\varepsilon\int_0^t\left(|\xi(\vartheta)|^2_{H^1(\Omega,T_h)}+J_h^\sigma(\xi(\vartheta),\xi(\vartheta))\right)\mathrm{d}\vartheta \qquad (94)\\
&+C_{13}(1+\varepsilon)\int_0^t\Bigg(\exp\left(C_{13}\frac{1+\varepsilon}{\varepsilon}(t-\vartheta)\right)\\
&\qquad\times\left\{\int_0^\vartheta\left(|\xi(\alpha)|^2_{H^1(\Omega,T_h)}+J_h^\sigma(\xi(\alpha),\xi(\alpha))\right)\mathrm{d}\alpha\right\}\Bigg)\mathrm{d}\vartheta\\
&\le\left(\frac{c_1}{\varepsilon}+c_2h^2\right)h^{2p}\exp\left(C_{13}\frac{1+\varepsilon}{\varepsilon}t\right),\quad t\in[0,T].
\end{aligned}
$$

Now, since $e_h=\xi+\eta$, we can write

$$
\begin{aligned}
\|e_h\|^2_{L^2(\Omega)} &\le 2\left(\|\xi\|^2_{L^2(\Omega)}+\|\eta\|^2_{L^2}\right), \qquad (95)\\
|e_h|^2_{H^1(\Omega,T_h)} &\le 2\left(|\xi|^2_{H^1(\Omega,T_h)}+|\eta|^2_{H^1(\Omega,T_h)}\right),\\
J_h^\sigma(e_h,e_h) &\le 2\left(J_h^\sigma(\xi,\xi)+J_h^\sigma(\eta,\eta)\right).
\end{aligned}
$$

Using assumption (12), from (79), (94), (95) and (66) we deduce that

$$
\begin{aligned}
&\|e_h(t)\|^2_{L^2(\Omega)}+\varepsilon\int_0^t\left(|e_h(\vartheta)|^2_{H^1(\Omega,T_h)}+J_h^\sigma(e_h(\vartheta),e_h(\vartheta))\right)\mathrm{d}\vartheta \qquad (96)\\
&+C_{13}(1+\varepsilon)\int_0^t\Bigg(\exp\left(C_{13}\frac{1+\varepsilon}{\varepsilon}(t-\vartheta)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\qquad\qquad \times \left\{ \int_0^{\vartheta} \left(|e_h(\alpha)|^2_{H^1(\Omega,T_h)} + J_h^\sigma(e_h(\alpha), e_h(\alpha)) \right) \mathrm{d}\alpha \right\} \Bigg) \mathrm{d}\vartheta \\
&\le 2h^{2p} \Bigg\{ \left(\frac{c_1}{\varepsilon} + c_2 h^2 \right) \exp\left(C_{13} \frac{1+\varepsilon}{\varepsilon} t \right) \\
&\quad + \varepsilon (1 + 8C_4C_5)\, C_7^2 \int_0^t |u(\vartheta)|^2_{H^{p+1}(\Omega)}\, \mathrm{d}\vartheta \\
&\quad + C_{13}(1+\varepsilon)\,(1 + 8C_4C_5)\, C_7^2 \\
&\qquad \times \int_0^t \exp\left(C_{13} \frac{1+\varepsilon}{\varepsilon} (t - \vartheta) \right) \int_0^{\vartheta} |u(\alpha)|^2_{H^{p+1}(\Omega)} \mathrm{d}\alpha\, \mathrm{d}\vartheta \\
&\quad + C_7^2 h^2 |u(t)|^2_{H^{p+1}(\Omega)} \Bigg\} \\
&\le 2h^{2p} \left\{ \left(\frac{1}{\varepsilon} c_1(\varepsilon, h) + c_2 h^2 + \varepsilon C_{15} \right) \exp\left(C_{13} \frac{1+\varepsilon}{\varepsilon} T \right) + h^2 C_{14} \right\},
\end{aligned}
$$

where

$$
\begin{aligned}
C_{14} &= C_7^2 \|u\|^2_{L^\infty(0,T;H^{p+1}(\Omega))}, \\
C_{15} &= (1 + 8C_4C_5) C_7^2 |u|^2_{L^2(0,T;H^{p+1}(\Omega))}.
\end{aligned} \tag{97}
$$

This already implies (84). $\square$

Remark 12 *Estimate (84) implies that*

$$
\|u - u_h\|_{L^\infty(0,T;L^2(\Omega))} = O(h^p) \quad \textit{for } h \to 0+\,. \tag{98}
$$

Comparing this result with the approximation properties (50) and (66) implying that

$$
\|u - \Pi u\|_{L^\infty(0,T;L^2(\Omega))} = O(h^{p+1}), \tag{99}
$$

we see that the error estimate (84) is suboptimal. This is caused by the technique used in the theoretical analysis (application of Young's inequality and Gronwall's lemma) in order to overcome the nonlinearity in the convective terms. There is a question, whether this estimate can be improved. To this end, numerical experiments are introduced in the next section. They are performed for piecewise linear elements ($p = 1$) on triangular meshes and applied to the initial-boundary value problem for the viscous Burgers equation. The results from Table 1 indicate that the experimental order of convergence in the $L^\infty(L^2)$-norm behaves as the expected optimal order $O(h^2)$. It is a challenging problem to prove optimal error estimates for nonlinear convection-diffusion problems.

It would be interesting to perform numerical experiments on general nonconvex elements used in the theoretical part. However, it is quite difficult to construct such numerical experiments and, therefore, we leave it, together with the use of higher degree elements ($p > 1$), to our further research (in progress).

It is clear that estimate (84) cannot be used for $\varepsilon \to 0+$, because it blows up exponentially with respect to $1/\varepsilon$. In the case that equation (1) contains a dominating reaction term $c\,u$ on the left hand side with $c \geq C_{13}(1 + 1/\varepsilon)$, the blow up is of order $O(1/\varepsilon)$. The nonlinearity of the convective terms represents a serious obstacle for obtaining a uniform error estimate with respect to $\varepsilon \to 0+$.

6 Numerical examples

In this section we present some numerical experiments carried out for the 2D viscous Burgers equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x_1} + u\frac{\partial u}{\partial x_2} = \varepsilon \Delta u + g \quad \text{in } \ \Omega \times (0, T), \tag{100}$$

equipped with the boundary condition (2) and the initial condition (3).

We set $\Omega = (-1, 1)^2$, $T = \infty$, $\varepsilon = 0.002$ and define the function g and the initial condition in such a way that the exact solution has the form

$$u(x_1, x_2, t) = (1 - x_1^2)^2(1 - x_2^2)^2 \left(1 - \frac{e^{-t}}{2}\right). \tag{101}$$

Figure 2 shows the exact solution (101) at time $t = 1$ and for $t \to \infty$ when we achieve the steady state solution

$$u(x_1, x_2) = (1 - x_1^2)^2(1 - x_2^2)^2. \tag{102}$$

In the latter case we write $t = \infty$.

We solve the initial-boundary value problem by the presented numerical method with piecewise linear ($p = 1$) triangular elements. In the definition of the form b_h the following numerical flux is used:

$$H(u_1, u_2, \boldsymbol{n}) = \begin{cases} \sum_{s=1}^2 f_s(u_1)n_s, & \text{if } A > 0 \\ \sum_{s=1}^2 f_s(u_2)n_s, & \text{if } A \leq 0 \end{cases}, \tag{103}$$

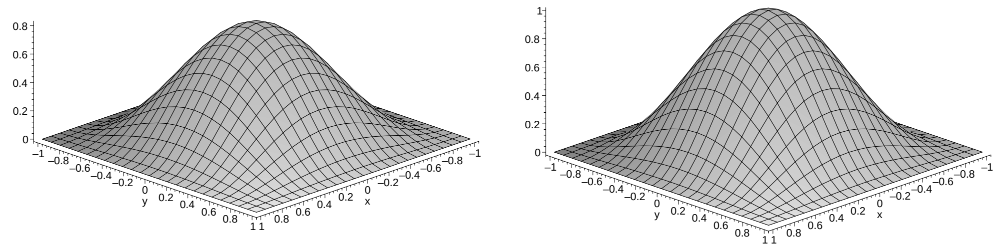

Fig. 2. The exact solution at $t = 1$ (left) and the steady state solution for $t \to \infty$ (right)

where

$$A = \sum_{s=1}^{2} f_s'(\bar{u}) n_s, \quad \bar{u} = \frac{1}{2}(u_1 + u_2). \tag{104}$$

The resulting system of ODE's is solved by the forward Euler method with a very small time step $\tau = 10^{-4}$, which guarantees the stability and sufficiently accurate resolution with respect to time.

The computational error of the solution is evaluated at time t over the domain Ω in the $L^2(\Omega)$-norm:

$$\|e_h(t)\|_{L^2(\Omega)} \equiv \|u_h(\cdot,t) - u(\cdot,t)\|_{L^2(\Omega)}. \tag{105}$$

We set $\tilde{e}_h = \|e_h(t)\|_{L^2(\Omega)}$ for $t := 1$ or $t := \infty$. We suppose that the error behaves according to the formula

$$\tilde{e}_h \approx C h^{\alpha}, \tag{106}$$

where $C > 0$ is a constant independent of h and α is the *order of accuracy* of the method. The numerical solution was computed on 6 unstructured, non-nested meshes T_{h_l}, $l = 1, \ldots, 6$, with different h_l, see Figure 3 and Table 1. We define the *local experimental order of convergence* by

$$\alpha_l = \frac{\log\left(\tilde{e}_{h_l}/\tilde{e}_{h_{l-1}}\right)}{\log\left(h_l/h_{l-1}\right)}, \quad l = 2, \ldots, 6. \tag{107}$$

The *global experimental order of convergence* $\bar{\alpha}$ is obtained by the least square method. Table 1 shows the L^2-error, the values of α_l, for $l = 2, \ldots, 6$ and $\bar{\alpha}$. In Figure 4 we see the convergence history for $h \to 0$. Figure 5 shows the computed approximate solutions.

The global experimental order of convergence $\bar{\alpha}$ is approximately equal to 2. It indicates that the order of the numerical scheme is in the L^2-norm optimal in contrast to our theoretical results, according to which $\tilde{e}_h = O(h)$ for $p = 1$.

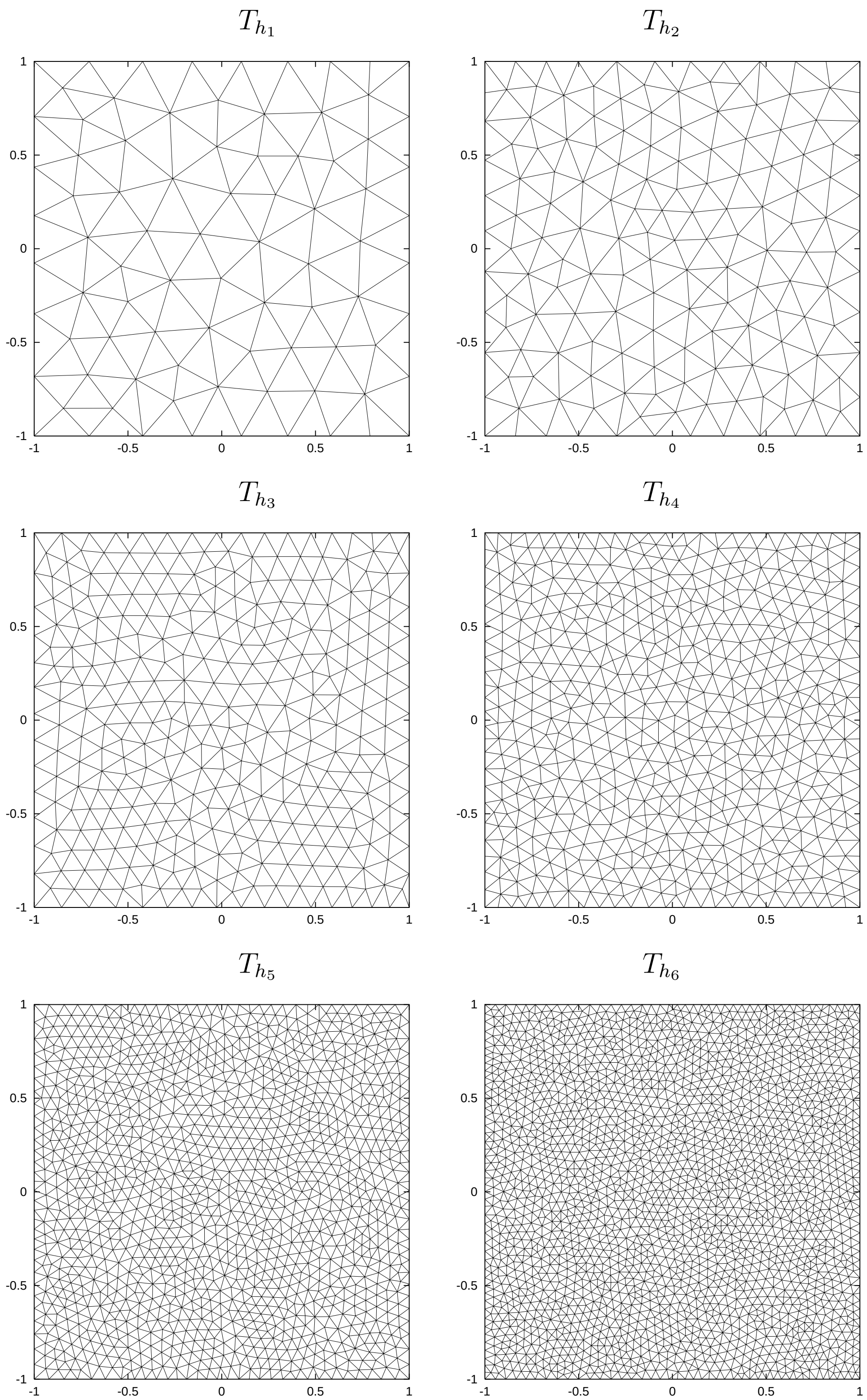

Fig. 3. Triangulations $T_{h_l},\ l = 1, \ldots, 6$, used for the computation

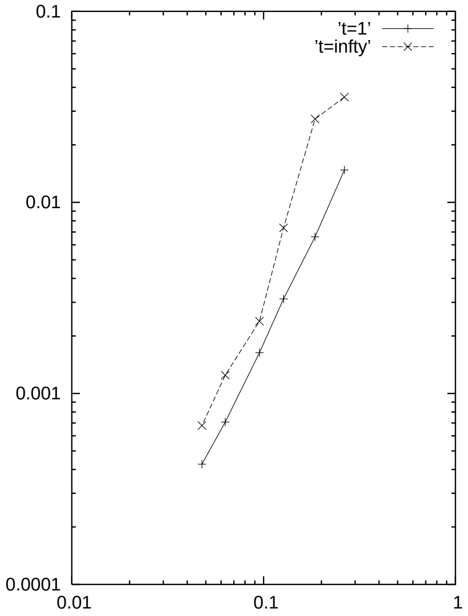

Fig. 4. Dependence of $e_h(t)$ on h in L^2-norm for $t := 1$ (full line) and $t := \infty$ (dashed line)

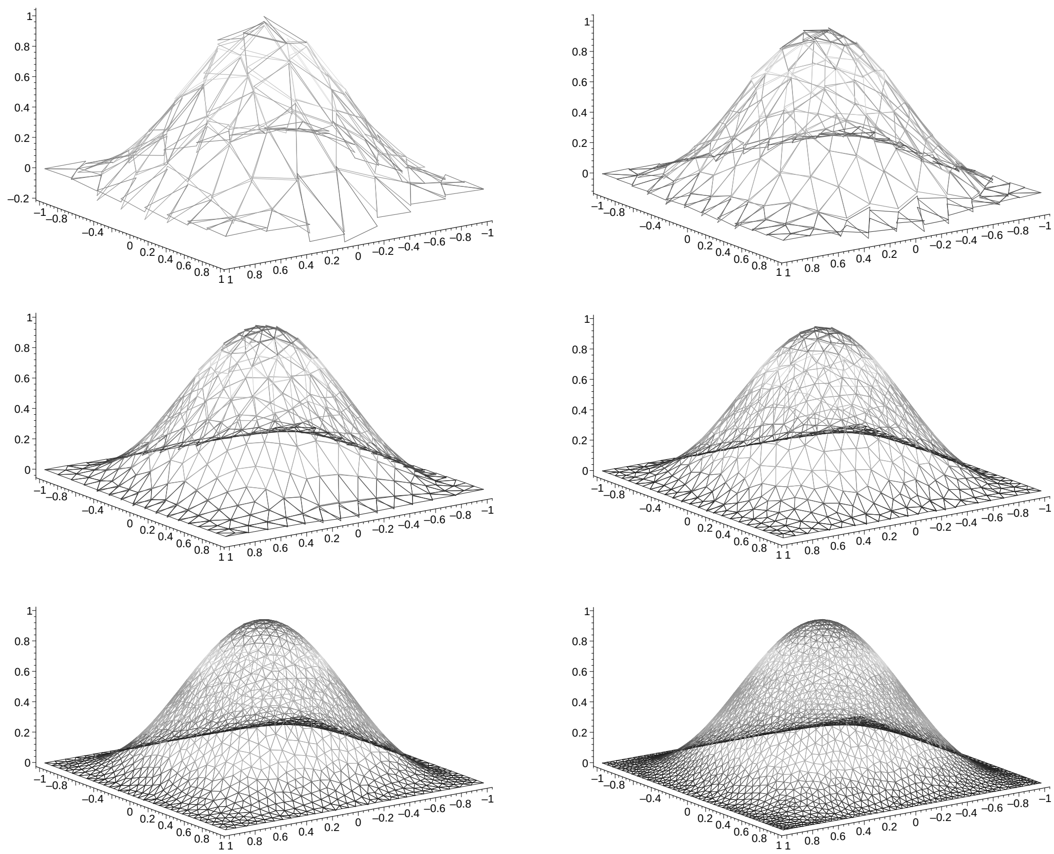

Fig. 5. Approximate solutions obtained on the triangulations T_{h_l}, $l = 1, \dots, 6$, for $t = \infty$

Table 1
Computational errors in L^2-norm, the values of $\alpha_l,\ l = 2, \ldots, 6$, and $\bar{\alpha}$

			$t = 1$		$t \to \infty$	
l	$\#T_{h_l}$	h_l	e_h	α_l	e_h	α_l
1	136	2.638E-01	1.4778E-02	–	3.5589E-02	–
2	276	1.854E-01	6.6063E-03	2.282	2.7337E-02	0.748
3	586	1.269E-01	3.1208E-03	1.977	7.3604E-03	3.459
4	1046	9.516E-02	1.6331E-03	2.251	2.3878E-03	3.913
5	2374	6.308E-02	7.0742E-04	2.035	1.2455E-03	1.583
6	4150	4.772E-02	4.2615E-04	1.817	6.7826E-04	2.179
global order of convergence $\bar{\alpha}$				2.078		2.497

7 Appendix

Proof of Lemma 1. Condition (41) implies that there exists a d-dimensional ball A_0 with centre x_K inscribed into K and there exists a d-dimensional ball A_1 with the centre x_K containing K. The ratio of diameters of A_0 and A_1 satisfies

$$1 \leq \frac{r_{A_1}}{r_{A_0}} = \frac{\max_{x \in \partial K} |x - x_K|}{\min_{x \in \partial K} |x - x_K|} \leq \kappa. \tag{108}$$

This implies the existence of axiparallel boxes B_0, B_1 satisfying conditions i) – ii) from Lemma 1. Obviously,

$$h_{B_1} = r_{A_1}, \quad h_{B_0}\sqrt{d} = r_{A_0}, \tag{109}$$

see Figure 6. Then (108) and (109) yield

$$\frac{h_{B_1}}{\sqrt{d} h_{B_0}} = \frac{r_{A_1}}{r_{A_0}} \leq \kappa \tag{110}$$

and the condition iii) is valid with $\bar{\kappa} = \sqrt{d}\kappa$. $\square$

Proof of Lemma 2. In [11], Lemma 3.1, we proved the multiplicative trace inequality for a convex element K. Here we consider the nonconvex case. Let $K \in T_h$ and $S \in \omega(K)$. Then (43) and Lemma 3.1 from [11] yield

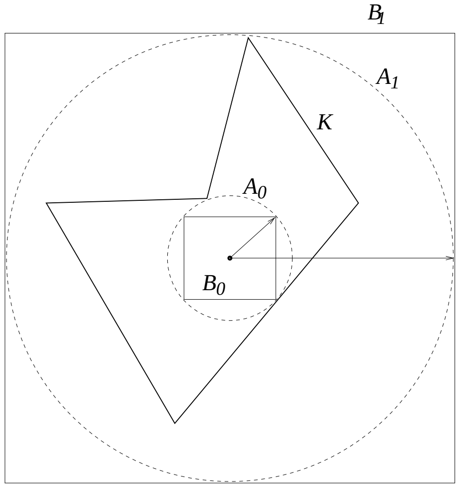

Fig. 6. The balls A_0, A_1 and axiparallel boxes B_0, B_1 used in the proof of Lemma 1

$$\|v\|^2_{L^2(\partial S)} \le C\left(\|v\|_{L^2(S)}\,|v|_{H^1(S)} + h_S^{-1}\|v\|^2_{L^2(S)}\right), \qquad S\in\omega(K),\ v\in H^1(K), \tag{111}$$

where $C = d\,C_3$ is independent of S, K, v and h. Using the inclusion $\partial K\subset \bigcup_{S\in\omega(K)}\partial S$, relations (111) and (44) and the Cauchy inequality, we find that

$$\begin{aligned}
\|v\|^2_{L^2(\partial K)} &\le \sum_{S\in\omega(K)} \|v\|^2_{L^2(\partial S)}\\
&\le C\left[\sum_{S\in\omega(K)}\left(\|v\|_{L^2(S)}\,|v|_{H^1(S)} + h_S^{-1}\|v\|^2_{L^2(S)}\right)\right]\\
&\le C\left[\left(\sum_{S\in\omega(K)}\|v\|^2_{L^2(S)}\right)^{1/2}\left(\sum_{S\in\omega(K)}|v|^2_{H^1(S)}\right)^{1/2} + \tilde\kappa h_K^{-1}\|v\|^2_{L^2(K)}\right]\\
&= C_5\left(\|v\|_{L^2(K)}\,|v|_{H^1(K)} + h_K^{-1}\|v\|^2_{L^2(K)}\right),
\end{aligned} \tag{112}$$

where $C_5 = C\tilde\kappa$. $\square$

Proof of Lemma 3. Let $K\in T_h$ and let B_0 and B_1 be the boxes from Lemma 1 satisfying (46). Without loss of generality we suppose that the centre of B_0 and B_1 is the origin of the coordinate system. Then obviously

$$[-1,1]^d \subset h_{B_0}^{-1}K \subset h_{B_0}^{-1}B_1, \tag{113}$$

where

$$h_{B_0}^{-1}K = \Big\{y = (y_1,\dots,y_d)\in I\!R^d;\ y_l = h_{B_0}^{-1}x_l, \tag{114}$$

$$l = 1, \dots, d,\ x = (x_1, \dots, x_d) \in K\}\,.$$

We define the reference boxes

$$\hat{B}_0 = [-1, 1]^d, \quad \hat{B}_1 = [-\bar{\kappa}, \bar{\kappa}]^d, \tag{115}$$

where $\bar{\kappa}$ is the constant from (46). It can be verified that $\|.\|_{L^2(\hat{B}_0)}$ is a norm on the space $P_p(\hat{B}_1)$. The equivalence of norms on the finite dimensional space $P_p(\hat{B}_1)$ implies that there exists a constant $C(p, \bar{\kappa})$ such that

$$\|v\|_{L^2(\hat{B}_1)} \le C(p, \bar{\kappa}) \|v\|_{L^2(\hat{B}_0)} \quad \forall v \in P_p(\hat{B}_1). \tag{116}$$

In what follows, by $\hat{v}$ we denote the function obtained by the scaling of $v \in P_p$ defined by (114). Furthermore, the transformation to the reference element and the equivalence of norms $\|\cdot\|_{L^2(\hat{B}_0)}$ and $\|\cdot\|_{H^1(\hat{B}_0)}$ on the finite dimensional space $P_p(\hat{B}_0)$ imply that

$$\begin{aligned}
\|\nabla v\|_{L^2(B_0)} &= h_{B_0}^{d/2-1} \|\hat{\nabla} \hat{v}\|_{L^2(\hat{B}_0)} \\
&\le h_{B_0}^{d/2-1} \|\hat{v}\|_{H^1(\hat{B}_0)} \\
&\le \tilde{C}(p) h_{B_0}^{d/2-1} \|\hat{v}\|_{L^2(\hat{B}_0)} \\
&= \tilde{C}(p) h_{B_0}^{-1} \|v\|_{L^2(B_0)} \quad \forall v \in P_p(B_0),
\end{aligned} \tag{117}$$

where $\tilde{C} = \tilde{C}(p)$ is a constant independent of h and v.

Let $v \in P_p(B_1)$. Then the transformation to the reference box $\hat{B}_0$ and relations (113) – (117), (46) and (47) yield

$$\begin{aligned}
\|\nabla v\|_{L^2(K)} &\le \|\nabla v\|_{L^2(B_1)} \\
&= (h_{B_1}/\bar{\kappa})^{d/2-1} \|\hat{\nabla} \hat{v}\|_{L^2(\hat{B}_1)} \\
&\le (h_{B_1}/\bar{\kappa})^{d/2-1} C(p, \bar{\kappa}) \|\hat{\nabla} \hat{v}\|_{L^2(\hat{B}_0)} \\
&= (h_{B_1}/\bar{\kappa})^{d/2-1} h_{B_0}^{1-d/2} C(p, \bar{\kappa}) \|\nabla v\|_{L^2(B_0)} \\
&\le C(p, \bar{\kappa}) \tilde{C}(p) h_{B_0}^{-1} \|v\|_{L^2(B_0)} \\
&\le C(p, \bar{\kappa}) \tilde{C}(p) h_{B_0}^{-1} \|v\|_{L^2(K)} \\
&\le \bar{\kappa} C(p, \bar{\kappa}) \tilde{C}(p) h_{B_1}^{-1} \|v\|_{L^2(K)} \\
&\le \bar{\kappa} C(p, \bar{\kappa}) \tilde{C}(p) 2 h_K^{-1} \|v\|_{L^2(K)} =: C_6 h_K^{-1} \|v\|_{L^2(K)},
\end{aligned} \tag{118}$$

which proves the lemma. $\square$

Proof of Lemma 4. If T_h is a simplicial mesh, then standard results from the finite element method can be employed (see, e. g., [6] or [5]). In our case, when general nonconvex elements are used, the approximation properties (50) are derived under assumptions (A1), (A2) in [23]. $\square$

References

[1] S. Adjerid, D. Devine, J. E. Flaherty, and L. Krivodonova, A posteriori error estimation for discontinuous Galerkin solutions of hyperbolic problems, Comput. Methods Appl. Mech. Engrg., 191 (2002) 1097–1112.

[2] F. Bassi and S. Rebay, A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier–Stokes equations, J. Comput. Phys., 131 (1997) 267–279.

[3] F. Bassi and S. Rebay, High-order accurate discontinuous finite element solution of the 2D Euler equations, J. Comput. Phys., 138 (1997) 251–285.

[4] C. E. Baumann and J. T. Oden, A discontinuous *hp* finite element method for the Euler and Navier-Stokes equations. Int. J. Numer. Methods Fluids, 31 (1999) 79–95.

[5] S. Brenner and R. L. Scott, The Mathematical Theory of Finite Element Methods, (Springer, New York, 1994).

[6] P. G. Ciarlet, The Finite Elements Method for Elliptic Problems, (North-Holland, Amsterdam, New York, Oxford, 1979).

[7] B. Cockburn, Discontinuous Galerkin methods for convection dominated problems, In T. J. Barth and H. Deconinck, editors, High–Order Methods for Computational Physics, Lecture Notes in Computational Science and Engineering 9, 69–224, (Springer, Berlin, 1999).

[8] B. Cockburn and C. W. Shu, TVB Runge–Kutta local projection discontinuous Galerkin finite element method for scalar conservation laws II: General framework, Math. Comp., 52 (1989) 411–435.

[9] B. Cockburn, G. E. Karniadakis, and C.-W. Shu (eds.), Discontinuous Galerkin methods. Lecture Notes in Computational Science and Engineering 11, (Springer, Berlin, 2000).

[10] V. Dolejší and M. Feistauer, On the discontinuous Galerkin method for the numerical solution of high-speed flow. In F. Brezzi, A. Buffa, S. Corsaro, and A. Murli, editors, Numerical Mathematics and Advanced Applications, ENUMATH 2001, pages 65–84, (Springer-Verlag Italia, Milano, 2003).

[11] V. Dolejší, M. Feistauer, and C. Schwab, A finite volume discontinuous Galerkin scheme for nonlinear convection–diffusion problems, Calcolo, 39 (2002) 1–40.

[12] V. Dolejší, M. Feistauer, and C. Schwab, On discontinuous Galerkin methods for nonlinear convection–diffusion problems and compressible flow, Mathematica Bohemica, 127 (2002) 163–179.

[13] V. Dolejší, M. Feistauer, and C. Schwab, On some aspects of the discontinuous Galerkin finite element method for conservation laws, Math. Comput. Simul., 61 (2003) 333–346.

[14] M. Feistauer, Mathematical Methods in Fluid Dynamics, (Longman Scientific & Technical, Harlow, 1993).

[15] M. Feistauer, J. Felcman, M. Lukáčová-Medvidová, and G. Warnecke, Error estimates of a combined finite volume – finite element method for nonlinear convection – diffusion problems. SIAM J. Numer. Anal., 36 (1999) 1528–1548.

[16] M. Feistauer, J.Felcman and I. Straškraba, Mathematical and Computational Methods for Compressible Flow, (Clarendon Press, Oxford, 2003).

[17] J. E. Flaherty, R. M. Loy, M. S. Shephart, and J. D. Teresco, Software for the parallel adaptive solution of conservation laws by discontinuous Galerkin methods, In B. Cockburn, G. E. Karniadakis, and C.-W. Shu, editors, Discontinuous Galerkin Methods. Theory, Computation and Applications, Lecture Notes in Computational Science and Engineering 11, 113–123, (Springer, Berlin, 2000).

[18] P. Houston, C. Schwab, and E. Süli, Discontinuous hp-finite element methods for advection-diffusion problems, SIAM J. Numer. Anal., 39 (2002) 2133–2163.

[19] A. Kufner, O. John, and S. Fučík, Function Spaces, (Academia, Prague, 1977).

[20] J. L. Lions, Quelques méthodes de résolution des problèmes aux limites non linéaires, (Dunod, Gauthier-Villars, Paris, 1969).

[21] K. Rektorys, The Method of Discretization in Time and Partial Differential Equations, (Reidel, Dodrecht, 1982).

[22] B. Rivière and M. F. Wheeler, A discontinuous Galerkin method applied to nonlinear parabolic equations, In B. Cockburn, G. E. Karniadakis, and C.-W. Shu, editors, Discontinuous Galerkin methods. Theory, Computation and Applications. Lecture Notes in Computational Science and Engineering 11, 231-244, (Berlin, Springer, 2000).

[23] R. Verfürth, A note on polynomial approximation in Sobolev spaces, $\mathrm{M}^2\mathrm{AN}$, 33 (1999) 715–719.

List of preprints released in this series

All preprints available on http://www.karlin.mff.cuni.cz/ms-preprints/
E-mail contact to series administrator: knm-prep@karlin.mff.cuni.cz

[**MATH-kma-2004/123**] Petr Holicky: *Extension of Borel measurable maps and ranges of Borel bimesurable maps* (05.01.2004)

[**MATH-kma-2004/124**] Jiri Spurny: *Affine Baire-one functions on Choquet simplexes* (25.01.2004)

[**MATH-kma-2004/125**] Ludek Zajicek: *On σ-porous sets in abstract spaces (a partial survey)* (25.01.2004)

[**MATH-kma-2004/126**] R. Chill, E. Fasangova, G. Metafune, D. Pallara: *The sector of analycity of the Ornstein-Uhlenbeck semigroup on L^p spaces with respect to invariant measure* (25.01.2004)

[**MATH-kma-2004/127**] Petr Hajek, Michal Johanis: *Isomorphic Embeddings and Harmonic Behaviour of Smooth Operators* (25.01.2004)

[**MATH-kma-2004/128**] Tomas Barta: *On a Viscoelastic Problem* (11.02.2004)

[**MATH-alg-2004/1**] J.Jezek: *The ordering of commutative terms* (18.02.2004)

[**MATH-alg-2004/2**] A.Drapal: *On left conjugacy closed loops with a nucleus of index two* (20.02.2004)

[**MATH-alg-2004/3**] R.Freese, J.Jezek, P.Jipsen, P.Markovic, M.Maroti, R.McKenzie: *The variety generated by order algebras* (20.02.2004)

[**MATH-alg-2004/4**] J.Jezek: *A note on idempotent modifications of groups* (20.02.2004)

[**MATH-alg-2004/5**] J.Jezek: *One-element extensions in the variety generated by tournaments* (20.02.2004)

[**MATH-kma-2004/129**] Etienne Matheron, Miroslav Zeleny: *Rudin-like sets and hereditary families of compact sets* (17.03.2004)

[**MATH-kma-2004/130**] Dalibor Prazak: *Reducing the 2d Navier-Stokes equations to a system of delayed ODEs* (20.04.2004)

[**MATH-MU-2004/1**] J. Rataj, M. Zahle: *General normal cycles and Lipschitz manifolds of bounded curvature* (23.04.2004)

[**MATH-kma-2004/131**] Ludek Zajicek, Miroslav Zeleny: *Inscribing closed non-σ-lower porous sets into Suslin non-σ-lower porous sets* (15.05.2004)

[**MATH-kma-2004/132**] Petr Dostal, Jaroslav Lukes, Jiri Spurny: *Measure convex and measure extremal sets* (15.05.2004)

[**MATH-kma-2004/133**] Ondrej Kalenda, Jiri Spurny: *Extending Baire-one functions on topological spaces* (15.05.2004)

[**MATH-kma-2004/134**] Petr Holicky, Ludek Zajicek, Miroslav Zeleny: *A remark on a theorem of Solecki* (15.05.2004)

[**MATH-alg-2004/6**] J. Trlifaj and O. Kerner: *Tilting classes over wild hereditary algebras* (28.05.2004)

[**MATH-kma-2004/135**] Jakub Duda: *On a decomposition of superreflexive spaces* (03.06.2004)

[**MATH-kma-2004/136**] Jakub Duda: *Curves with finite turn* (03.06.2004)

[**MATH-kma-2004/137**] Jan Kalis: *On Whitney sets and their generalization* (03.06.2004)

[**MATH-kma-2004/138**] David Pavlica: *A d.c. C^1 function need not be the difference of convex C^1 functions* (03.06.2004)

[MATH-kma-2004/139] Bernd Kawohl, Jana Stara: *Observations on a nonlinear evolution equation* (23.07.2004)

[MATH-kma-2004/140] Salvatore Leonardi, Jiri Kottas, Jana Stara: *Hoelder regularity of the solutions of some class of elliptic systems in convex non-smooth domains* (23.07.2004)

[MATH-kma-2004/141] Miroslav Zeleny: *Descriptive properties of* σ*-porous sets* (06.08.2004)

[MATH-kma-2004/142] Maria J. Carro, Amiran Gogatishvili, Joaquim Martin, Lubos Pick: *Functional properties of rearrangement invariant spaces defined in terms of oscillations* (06.08.2004)