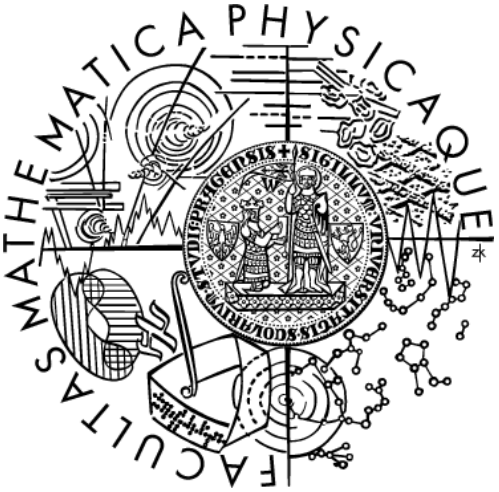
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Jaroslav Hájek

Některé aspekty nespojité Galerkinovy metody pro řešení konvektivně difuzních rovnic

Katedra numerické matematiky
Vedoucí diplomové práce: Prof. RNDr. Miloslav Feistauer, DrSc.
Studijní obor: Výpočtová matematika

Na tomto místě bych rád poděkoval prof. Miloslavu Feistauerovi za pomoc a vedení při studiu a přípravě práce. Také bych rád poděkoval Martinu Mádlíkovi za ochotu při řešení technických problémů s numerickými výpočty.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 1. září 2006 Jaroslav Hájek

Contents

Název: Některé aspekty nespojité Galerkinovy metody pro řešení konvektivně-difúzních rovnic
Autor: Jaroslav Hájek
Katedra: Katedra numerické matematiky
Vedoucí diplomové práce: Prof. RNDr. Miloslav Feistauer, CSc.
e-mail vedoucího: feist@karlin.mff.cuni.cz
Abstrakt: Práce se zabývá numerickým řešením smíšených úloh pro konvektivně - difúzní parciální diferenciální rovnice. Pro tento účel jsou zde studovány a srovnávány tři metody: kombinovaná metoda konečných prvků a konečných objemů (FE-FV), nespojitá Galerkinova (DGFE) metoda přímek a časoprostorová nespojitá Galerkinova metoda. Kombinovaná FE-FV metoda používá po částech lineární konformní konečné prvky pro diskretizaci difúzních členů a po částech konstantní aproximaci konvektivních členů pomocí konečných objemů. Vztah mezi těmito dvěma aproximacemi udává takzvaný "lumping operator". V nespojité Galerkinově metodě přímek je semidiskretizace v prostoru provedena s pomocí po částech polynomiálních funkcí nad trojúhelníkovou sítí, obecně nespojitých na rozhraních mezi sousedními elementy. V časoprostorové nespojité Galerkinově metodě je přibližné řešení po částech polynomiální jak v čase, tak i v prostoru. Diskutujeme teoretické i praktické aspekty metod a pro každou z nich uvádíme numerické výsledky. Pro nespojitou Galerkinovu metodu přímek odvozujeme aposteriorní odhad chyby.
Klíčová slova: konvekce-difúze, Galerkinova metoda, aposteriorní odhad

Title: Some aspects of the discontinuous Galerkin method for convection-diffusion problems
Autor: Jaroslav Hájek
Department: Department of Numerical Mathematics
Supervisor: Prof. RNDr. Miloslav Feistauer, DrSc.
Supervisor's e-mail address: feist@karlin.mff.cuni.cz
Abstract: This work is concerned with the numerical solution of initial-boundary value problems for convection-diffusion partial differential equations. Three methods are studied and compared for this purpose: the combined finite element - finite volume (FE-FV) method, the discontinuous Galerkin finite element (DGFE) method of lines, and the space-time discontinuous Galerkin method. The combined FE-FV method uses piecewise linear conforming finite elements for the discretization of the diffusion terms and piecewise constant FV approximation of the convective terms. The relation between the FE and FV approximations is determined by the so-called lumping operator. In the DGFE method of lines, the space semidiscretization is carried out by piecewise polynomial functions constructed over a triangular mesh, in general discontinuous on interfaces between neighbouring elements. In the space-time DGFE method, the approximate solution is piecewise polynomial in space as well as in time. We discuss both theoretical and practical aspects of the methods, and present numerical results for each of them. For the DGFE method of lines we derive an a posteriori error estimate.
Keywords: convection-diffusion, Galerkin method, a posteriori estimate

1 Introduction

The convection-diffusion-reaction equation arises naturally in a number of practical problems from science and technology. It has also a lot of theoretical importance being a simple (even linear) model yet still having important properties of complex flows - convection dominance and boundary layers. As such, it is a natural model for the development of new techniques for solving convection-dominated problems. It is well known that applying classical finite elements to these problems gives rise to the so-called Gibbs phenomenon which degrades the numerical solutions or even makes them unusable. Several approaches have been developed to face this problem. We study those based on the finite volume method - the combined finite element - finite volume (FE-FV) method and the discontinuous Galerkin finite element (DGFE) method, the first being a combination of FV and FE, the latter a natural generalization.

2 Combined Finite Element - Finite Volume Method of Lines

The finite volume method (FVM) represents an efficient and robust method for the solution of conservation laws and inviscid compressible flow. This technique is based on expressing the balance of fluxes of conserved quantities through boundaries of control volumes, combined with approximate Riemann solvers. On the other hand, the finite element method (FEM), based on the concept of a weak solution defined with the aid of suitable test functions is quite natural for the solution of elliptic and parabolic problems. In the solution of nonlinear convection–diffusion problems, including viscous compressible flow, it is quite natural to try to employ the advantages of both FV and FE methods in such a way that the FVM is used for the discretization of inviscid Euler fluxes, whereas the FEM is applied to the approximation of viscous terms. This idea leads us to the *combined finite volume–finite element method* (FV–FE method) proposed in [24]. (Sometimes it is also called the mixed FV–FE method.) The analysis and applications of this method were investigated in [25], [23], [26], [1] [15]. The numerical computations for the system of compressible viscous flow ([18], [27], [15], [20], [36]) demonstrate that the combined FV–FE method is feasible and produces good numerical results for technically relevant problems. The idea of using a combination of the FV and FE methods appears also in [2], [29] and [30].

2.1 Continuous problem

Let $\Omega \subset \mathbb{R}^2$ be a bounded polygonal domain and $T > 0$. We consider the following initial-boundary value problem: find a solution of the equation

$$\frac{\partial u}{\partial t} + \sum_{s=1}^{2} \frac{\partial f_s(u)}{\partial x_s} = \varepsilon \, \Delta u + g \quad \text{in } \mathcal{Q}_T = \Omega \times (0, T) \tag{2.1}$$

with the initial condition

$$u(x, 0) = u^0(x), \ x \in \Omega, \tag{2.2}$$

and the boundary condition

$$u|_{\partial\Omega \times (0,t)} = 0. \tag{2.3}$$

We assume that the data have the following properties:

1. $f_s \in C^1(\mathbb{R})$, $f_s(0) = 0$, $s = 1, 2$,
2. $\varepsilon > 0$,
3. $g \in C([0, T]; L^2(\Omega))$,
4. $u^0 \in L^2(\Omega)$.

Let the functions f_s have a bounded derivative: $|f_s'| \le c_{f'}$. Then they satisfy the Lipschitz condition with the constant $c_l^* = c_{f'}$. The constant ε is the diffusion coefficient and the functions f_s are fluxes of the quantity u in the directions x_s.

We shall use the following notation:

$$(u, v) = \int_\Omega u \, v \, \mathrm{d}x, \quad u, \, v \in l^2(\Omega), \tag{2.4}$$

$$a(u, v) = \varepsilon \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}x, \quad u, \, v \in H^1(\Omega), \tag{2.5}$$

$$b(u, v) = \sum_{s=1}^{2} \int_\Omega \frac{\partial f_s(u)}{\partial x_s} \, v \, \mathrm{d}x, \quad u \in H^1(\Omega) \cap L^\infty(\Omega), \ v \in L^2(\Omega), \tag{2.6}$$

$$\tag{2.7}$$

2.2 Discrete problem

Let $\mathcal{T}_h$ be a partition of the closure $\overline{\Omega}$ of the domain Ω formed by a finite number of closed triangles K called *finite elements*. We number all elements in such a way that we can write $\mathcal{T}_h = \{K_i\}_{i \in I}$, where $I \subset Z^+ = \{0, 1, 2, \dots\}$ is a suitable index set. We assume that the triangulation $\mathcal{T}_h$ satisfies the following conditions:

$$\overline{\Omega} = \bigcup_{i \in I} K$$

and two different elements K_i, K_j are either disjoint or have a common vertex or a common side.

Further, we shall consider a mesh $\mathcal{D}_h = \{D_i\}_{i\in J}$ formed by closed triangles D_i, which will be called *finite volumes*. Symbol $J \subset Z^+$ denotes a suitable index set. We assume that the mesh $\mathcal{D}_h$ has the same properties as the triangulation $\mathcal{T}_h$. If two finite volumes $D_i, D_j \in \mathcal{D}_h$ have a common side, we call them neighbours. Then we set

$$\Gamma_{ij} = \partial D_i \cap \partial D_j = \Gamma_{ji} \tag{2.8}$$

and

$$s(i) = \{j \in J; j \neq i, D_j \text{ is a neighbour of } D_i\}. \tag{2.9}$$

The sides of finite volumes adjacent to the boundary $\partial\Omega$, which form this boundary, will be denoted by S_j and numbered by indices $j \in J_b \subset Z^- = \{-1, -2, \dots\}$. Thus, $J \cap J_b = \emptyset$ and $\partial\Omega = \bigcup_{j\in J_b} S_j$. For a finite volume D_i adjacent to the boundary $\partial\Omega$ we write

$$\begin{aligned} \gamma(i) &= \{j \in J_b; S_j \subset \partial\Omega \cap \partial D_i\}, \\ \Gamma_{ij} &= S_j, \qquad \text{for } j \in \gamma(i). \end{aligned} \tag{2.10}$$

If D_i is not adjacent to $\partial\Omega$, then we set $\gamma(i) = \emptyset$. Further, we put

$$S(i) = s(i) \cup \gamma(i). \tag{2.11}$$

Then

$$\partial D_i = \bigcup_{j\in S(i)} \Gamma_{ij}, \tag{2.12}$$

$$\partial D_i \cap \partial\Omega = \bigcup_{j\in\gamma(i)} \Gamma_{ij}, \tag{2.13}$$

$$|\partial D_i| = \sum_{j\in S(i)} |\Gamma_{ij}|, \tag{2.14}$$

where $|\partial D_i|$ is the length of ∂D_i and $|\Gamma_{ij}|$ is the length of the side Γ_{ij}. By $\boldsymbol{n}_{ij}$ we shall denote the unit outer normal to ∂K_i on the side Γ_{ij}.

For $k \in Z^+, K \in \mathcal{T}_h$ we denote by $P^k(K)$ the space of all polynomials on K of degree $\leq k$. In what follows the following finite element spaces

$$X_h = \{v_h \in C(\overline{\Omega});\ v_h|_K \in P^1(K)\ \forall K \in \mathcal{T}_h\}, \tag{2.15}$$

$$V_h = \{v_h \in X_h;\ v_h|_{\partial\Omega} = 0\} \tag{2.16}$$

and the finite volume space

$$Y_h = \{v_h \in L^2(\Omega);\ v_h|_{D_i} \in P^0(D_i)\ \forall i \in J\} \tag{2.17}$$

will be used.

The relation between the FE and FV spaces is given by the *lumping operator*

$$L_h :\ X_h \to Y_h$$

or, more general,

$$L_h :\ C(\overline{\Omega}) \to Y_h.$$

Let u be a classical solution of problem (2.1) – (2.3). We multiply equation (2.1) by a test function $v \in V_h$, integrate over Ω and apply Green's theorem. We obtain the identity

$$\left(\frac{\partial u}{\partial t}, v\right) + \sum_{i \in J} \int\limits_{D_i} \sum_{s=1}^{2} \frac{\partial f_s(u)}{\partial x_s}\, v \,\mathrm{d}x + a(u,v) = (g,v). \tag{2.18}$$

In order to approximate the terms with fluxes f_s, the test function v is replaced by $L_h v$:

$$\sum_{i \in J} \int\limits_{D_i} \sum_{s=1}^{2} \frac{\partial f_s(u)}{\partial x_s}\, v \,\mathrm{d}x \approx \sum_{i \in J} L_h v|_{D_i} \int\limits_{D_i} \sum_{s=1}^{2} \frac{\partial f_s(u)}{\partial x_s}\,\mathrm{d}x \tag{2.19}$$

If we apply Green's theorem to the right-hand side and approximate fluxes with the aid of a so-called numerical flux H, we get

$$\begin{aligned}\int\limits_{D_i} \sum_{s=1}^{2} \frac{\partial f_s(u)}{\partial x_s}\,\mathrm{d}x = \int\limits_{\partial D_i} \sum_{s=1}^{2} f_s(u)\, n_s \,\mathrm{d}S = \sum_{j \in S(i)} \int\limits_{\Gamma_{ij}} \sum_{s=1}^{2} f_s(u)\, n_s \,\mathrm{d}S \\ \approx \sum_{j \in S(i)} H(L_h u|_{D_i}, L_h u|_{D_j}, \boldsymbol{n}_{ij})\, |\Gamma_{ij}|\end{aligned} \tag{2.20}$$

For the faces $\Gamma_{ij} \subset \partial\Omega$ (i. e. $j \in \gamma(i)$) we use the boundary condition (2.3), on the basis of which we set $H(L_h u|_{D_i}, L_h u|_{D_j}, \boldsymbol{n}_{ij}) = 0$. As a result we obtain the approximation of the convective terms represented by the form

$$b_h(u,v) = \sum_{i \in J} L_h v|_{D_i} \sum_{j \in s(i)} H(L_h u|_{D_i}, L_h u|_{D_j}, \boldsymbol{n}_{ij})\, |\Gamma_{ij}|. \tag{2.21}$$

Now we define an approximate solution of problem (2.1) - (2.3) as a function $u_h \in C^1([0,T];\, V_h)$ satisfying the conditions

$$\left(\frac{\partial u_h}{\partial t}, v_h\right) + b_h(u_h, v_h) + a(u_h, v_h) = (g, v_h),\ \forall v_h \in V_h \tag{2.22}$$

$$u_h(0) = u_h^0 = \Pi_h u^0, \tag{2.23}$$

where Π_h is the operator of X_h-interpolation.

These are equivalent to a system of ordinary differential equations, which can be solved, e.g. by the Runge-Kutta method. In [21], a priori error estimate (theoretical order of convergence) was established for the above method with lumping operator defined as

$$L_h v|_{D_i} = \frac{1}{|D_i|} \int_{D_i} v \, \mathrm{d}x, \quad i \in J. \tag{2.24}$$

We give the main result:

Theorem 1 *Given the above assumtpions, the error $e_h = u - u_h$, where u is the exact solution of problem* (2.1) – (2.3) *satisfying*

$$u \in C([0,T];\, H^2(\Omega)). \tag{2.25}$$

and u_h is the approximate solution defined by (2.22)*, satisfies the inequalities*

$$\max_{t\in[0,T]} ||e_h||_{L^2(\Omega)} \leq C\, h \tag{2.26}$$

and

$$\sqrt{\varepsilon} \sqrt{\int_0^T |e_h(\vartheta)|^2_{H^1(\Omega)} \, \mathrm{d}\vartheta} \leq C\, h. \tag{2.27}$$

This theorem states that the order of convergence of combined FE-FV method is at least one. Here we shall verify the optimality of estimate (2.26). See section 5.2. For many practical applications, higher order schemes are desirable, as they generally lead to reduction in number of unknowns. The finite volume method, using a piecewise constant approximations, is uncapable to achieve higher-order convergence on general unstructured meshes. The discontinuous Galerkin finite element method, introduced in next section, addresses this problem.

3 Discontinuous Galerkin Method of Lines

The discontinuous Galerkin finite element method (DGFEM) is a generalization of both traditional FE and FV schemes. It uses a higher-order piecewise polynomial approximation (like FE scheme), but discontinuities are allowed on element boundaries and handled by the numerical flux (like in FV schemes).

The original DGFE method was introduced in [38] for the solution of a neutron transport linear equation and analyzed in [37],[35]. The DGFE techniques for the numerical solution of elliptic problems were developed in [5], [43]. Further, the DGFE method was

applied to nonlinear conservation laws ([12], [34]), compressible flow ([7], [8], [9], [19], [16], [31], [42]), and many other problems. Theoretical analysis of various types of the DGFE method applied to elliptic problems can be found, e.g. in [6], [3] and [4]. In [39], DGFE analysis is performed in the case of a parabolic problem with a nonlinear diffusion. In [33], analysis of *hp*-version of the DGFE method applied to stationary advection-diffusion-reaction equations is analyzed. A survey of DGFE methods and techniques can be found in [11] and [10]. It is common to use *space semidiscretization* for nonstationary problems, i.e. to discretize the equation in the space variables and consider time continuous. This approach results in a system of ordinary differential equations, for which very spohisticated solvers exist. In this section, we shall develop an a posteriori error estimate for the DGFE space semidiscretization applied to linear convection-diffusion equation. We build on the work presented in [41]. The estimate therein is derived with the inverse assumption (quasi-uniformity of the mesh). Although the authors claim that it can be done without it, it is actually used at several places of the proof. We try to get rid of this assumption rigorously, by introducing a stronger version of the "continuous reconstruction" lemma from [41]. We also track the contribution of problem parameters ε,γ_0 into the a posteriori estimate.

3.1 Continuous problem

Let $\Omega \subset \mathbb{R}^d$ be a polygonal (for $d = 2$) or polyhedral (for $d = 3$) domain with a Lipschitz boundary $\partial\Omega$ and let $T > 0$. We set $\mathcal{Q}_T = \Omega \times (0, T)$. We consider the following initial-boundary value problem: Find $u\colon \mathcal{Q}_T \to \mathbb{R}$ such that

$$\frac{\partial u}{\partial t} + \boldsymbol{v} \cdot \nabla u - \varepsilon \Delta u + cu = g \qquad \text{in } \mathcal{Q}_T, \tag{3.28}$$

$$u = u_D \qquad \text{on } \partial\Omega^- \times (0, T), \tag{3.29}$$

$$\varepsilon \frac{\partial u}{\partial \boldsymbol{n}} = u_N \qquad \text{on } \partial\Omega^+ \times (0, T), \tag{3.30}$$

$$u(x, 0) = u^0(x), \qquad x \in \Omega. \tag{3.31}$$

We assume that $\partial\Omega = \partial\Omega^- \cup \partial\Omega^+$ and

$$\boldsymbol{v}(x, t) \cdot \boldsymbol{n}(x) < 0 \qquad \text{on } \partial\Omega^-, \tag{3.32}$$

$$\boldsymbol{v}(x, t) \cdot \boldsymbol{n}(x) > 0 \qquad \text{on } \partial\Omega^+ \quad \forall t \in (0, T). \tag{3.33}$$

By $\boldsymbol{n}(x)$ we mean the unit outer normal to $\partial\Omega$, $\partial\Omega^-$ is the inflow and $\partial\Omega^+$ is the outflow part of the boundary. In the case $\varepsilon = 0$ we put $u_N = 0$ and ignore the Neumann condition (3.30).

3.1.1 Assumptions on data

We assume that the data satisfy the following conditions:

1. $g \in C([0,T);\, L^2(\Omega))$;
2. $u^0 \in L^2(\Omega)$;
3. u_D is the trace of some $u^* \in C([0,T);\, H^1(\Omega)) \cap L^\infty(\mathcal{Q}_T)$ on $\partial\Omega^- \times (0,T)$;
4. $\boldsymbol{v} \in C([0,T);\, W^{1,\infty}(\Omega))$, $|\boldsymbol{v}|, \|\nabla \boldsymbol{v}\| \le C_{\boldsymbol{v}}$ a.e. in $\mathcal{Q}_T$;
5. $c \in C([0,T);\, L^\infty(\Omega))$, $|c(x,t)| \le C_c$ a.e. in $\mathcal{Q}_T$;
6. $c - \operatorname{div} \boldsymbol{v}/2 \ge \gamma_0 \ge 0$ in $\mathcal{Q}_T$ with a constant γ_0;
7. $u_N \in C([0,T);\, L^2(\partial\Omega^+)$;
8. $\varepsilon \ge 0$.

Assumption 6 is not very restrictive since we can use the transformation $u = e^{\alpha t} w$, $\alpha =$ const. to get a transformed equation for w:

$$\frac{\partial w}{\partial t} + \boldsymbol{v} \cdot \nabla w - \varepsilon \Delta w + (c + \alpha) w = g e^{-\alpha t}.$$

The condition 6 now becomes

$$c + \alpha - \frac{1}{2} \operatorname{div} \boldsymbol{v} \ge \gamma_0 > 0$$

which is satisfied with α large enough.

The weak formulation is derived as in [22]. We set

$$V = \{\varphi \in H^1(\Omega);\ \varphi|_{\partial\Omega^-} = 0\}.$$

The *weak solution* to (3.28)-(3.31) is then a function satisfying the conditions

$$u - u^* \in L^2(0,T;\, V), u \in L^\infty(\mathcal{Q}_T), \tag{3.34a}$$

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega u\varphi \,\mathrm{d}x + \varepsilon \int_\Omega \nabla u \cdot \nabla \varphi \,\mathrm{d}x + \int_{\partial\Omega^+} (\boldsymbol{v} \cdot \boldsymbol{n})\, u\varphi \,\mathrm{d}S \\ &\quad - \int_\Omega u \operatorname{div}(\varphi \boldsymbol{v}) \,\mathrm{d}x + \int_\Omega cu\varphi \,\mathrm{d}x = \int_\Omega g\varphi \,\mathrm{d}x + \int_{\partial\Omega^+} u_N \varphi \,\mathrm{d}S \\ &\quad \text{for all } \varphi \in V \text{ in the sense of distributions on } (0,T), \end{aligned} \tag{3.34b}$$

$$u(0) = u^0 \text{ in } \Omega. \tag{3.34c}$$

We shall assume the existence of u and its sufficient regularity, namely

$$\frac{\partial u}{\partial t} \in L^1(0,T;\, H^{p+1}(\Omega)), \quad u \in L^1(0,T;\, H^{p+1}(\Omega)) \cap L^2(0,T;\, H^{p+1}(\Omega)), \tag{3.35}$$

where p is the given degree of polynomial approximation defined in the next subsection. It is possible to show that such a solution satisfies equation (3.28) pointwise (almost everywhere) and $u \in C([0,T);\, H^{p+1}(\Omega))$. If $\varepsilon > 0$, then it is possible to show the uniqueness of the solution and a stronger regularity $\partial u/\partial t \in L^2(\mathcal{Q}_T)$.

3.2 Discretization

Let $\mathcal{T}_h = \{K_i;\ i \in I\}$ ($\subset \mathbb{N}$ is an index set) be a standard triangulation of $\overline{\Omega}$ formed by i finite number of closed triangles ($d = 2$) or tetrahedra ($d = 3$). If two elements K_i, K_j share a $(d-1)$-dimensional face, we call them *neighbours*. In this case we put $\Gamma_{ij} = \Gamma_{ji} = \partial K_i \cap \partial K_j$. For $i \in I$ we set

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

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

$$\gamma(i) = \{j \in I_b;\ S_j \text{ is a face of } K_i\}.$$

Obviously for K_i not containing boundary faces $\gamma(i) = \emptyset$. We further set $\Gamma_{ij} = S_j$ for $j \in \gamma(i)$. By definition, $s(i) \cap \gamma(i) = \emptyset$ for all $i \in I$. Writing $S(i) = s(i) \cup \gamma(i)$, we have

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

For $K \in \mathcal{T}_h$ we denote by h_K the diamater of K and by ρ_K the radius of the largest inscribed ball. We set $h_{\max} = \max_{i \in I} h_{K_i}$. In the following we shall consider a family of triangulations distinguished by the parameter $h \in (0, h_0)$. From now on, by the term "global constant" we mean a positive constant that is independent of $\varepsilon, \gamma_0, h, \mathcal{T}_h$ or other indices or elements. It may (and usually does) depend on p and other global constants. We shall assume the uniform shape-regularity of $\mathcal{T}_h$: there exists a global constant $C_{\mathcal{T}}$ such that

$$\frac{h_K}{\rho_K} \le C_{\mathcal{T}} \qquad \forall K \in \mathcal{T}_h. \tag{3.36}$$

We introduce the so-called broken Sobolev space

$$H^k(\Omega, \mathcal{T}_h) = \{\varphi;\ \varphi|_K \in H^k(K) \qquad \forall K \in \mathcal{T}_h\} \tag{3.37}$$

and define the seminorm

$$|\varphi|_{H^k(\Omega,\mathcal{T}_h)} = \left(\sum_{i \in I} |\varphi|^2_{H^k(K)} \right)^{\frac{1}{2}}. \tag{3.38}$$

For $\varphi \in H^1(\Omega, \mathcal{T}_h)$ and $i \in I, j \in s(i)$ we shall use the notation

$$\varphi|_{\Gamma_{ij}} = \text{ the trace of } \varphi|_{K_i} \quad \text{on } \Gamma_{ij}, \tag{3.39}$$

$$\varphi|_{\Gamma_{ji}} = \text{ the trace of } \varphi|_{K_j} \quad \text{on } \Gamma_{ji}, \tag{3.40}$$

$$\langle\varphi\rangle_{\Gamma_{ij}} = \frac{1}{2}(\varphi|_{\Gamma_{ij}} + \varphi|_{\Gamma_{ji}}), \tag{3.41}$$

$$[\varphi]_{\Gamma_{ij}} = \varphi|_{\Gamma_{ij}} - \varphi|_{\Gamma_{ji}}, \tag{3.42}$$

$$\boldsymbol{n}_{ij} = \text{ the unit outer normal to } \partial K_i \text{ on the face } \Gamma_{ij}. \tag{3.43}$$

In the sequel we shall often omit the notation $|_{\Gamma_{ij}}$ in the above, if it can be deduced, e.g., from the integration domain of an integral. If $\varphi \in H^1(\Omega, \mathcal{T}_h)$ and we prescribe a boundary condition φ_D for φ, then we extend the above notation in such a way that in the case $\Gamma_{ij} = \Gamma_{ji} \subset \partial\Omega$ formulae (3.39)-(3.42) are used with $\varphi|\Gamma_{ji} = \varphi_D|_{\Gamma_{ji}}$. Further, for $i \in I$ we set

$$\partial K_i^-(t) = \{x \in \partial K_i;\ \boldsymbol{v}(x,t) \cdot \boldsymbol{n}(x) < 0\}, \tag{3.44}$$

$$\partial K_i^+(t) = \{x \in \partial K_i;\ \boldsymbol{v}(x,t) \cdot \boldsymbol{n}(x) > 0\}, \tag{3.45}$$

where $\boldsymbol{n}$ denotes the unit outer normal. In the following we shall not emphasize the dependence of ∂K_i^+ and ∂K_i^- on time by notation.

The derivation of the discrete problem also closely follows [22]. On the basis of (3.28), (3.35) and Green's theorem we find that the exact solution satisfies the following identity for $\varphi \in H^2(\Omega, \mathcal{T}_h)$:

$$\left(\frac{\partial u(t)}{\partial t}, \varphi\right) + a_h(u(t),\varphi) + b_h(u(t),\varphi) + c_h(u(t),\varphi) + \ + \varepsilon J_h^\sigma(u(t),\varphi) = l_h(\varphi)(t). \tag{3.46}$$

The forms in (3.46) are defined as follows:

$$(u,\varphi) = \int_\Omega u\varphi \,\mathrm{d}x, \tag{3.47}$$

$$\begin{aligned} a_h(u,\varphi) = {} & \varepsilon \sum_{i\in I} \int_{K_i} \nabla u \cdot \nabla\varphi \,\mathrm{d}x \\ & - \varepsilon \sum_{i\in I} \sum_{j\in s(i), j<i} \int_{\Gamma_{ij}} (\langle\nabla u\rangle \cdot \boldsymbol{n}_{ij}[\varphi] - \langle\nabla\varphi\rangle \cdot \boldsymbol{n}_{ij}[u]) \,\mathrm{d}S \\ & - \varepsilon \sum_{i\in I} \int_{\partial K_i^- \cap \partial\Omega} ((\nabla u \cdot \boldsymbol{n})\varphi - (\nabla\varphi \cdot u)u) \,\mathrm{d}S, \end{aligned} \tag{3.48}$$

$$\begin{aligned} b_h(u,\varphi) = {} & \sum_{i\in I} \int_{K_i} (\boldsymbol{v} \cdot \nabla u)\,\varphi \,\mathrm{d}x - \sum_{i\in I} \int_{\partial K_i^- \cap \partial\Omega} (\boldsymbol{v}\cdot\boldsymbol{n})\, u\varphi \,\mathrm{d}S \\ & - \sum_{i\in I} \int_{\partial K_i^- \setminus \partial\Omega} (\boldsymbol{v}\cdot\boldsymbol{n})[u]\varphi \,\mathrm{d}S, \end{aligned} \tag{3.49}$$

$$c_h(u,\varphi) = \int_\Omega cu\varphi \,\mathrm{d}x, \tag{3.50}$$

$$J_h^\sigma(u,\varphi) = \sum_{i\in I, j\in s(i)} \int_{\Gamma_{ij}} \sigma[u][\varphi] \,\mathrm{d}S + \sum_{i\in I} \int_{\partial K_i^- \cap \partial\Omega} \sigma u\varphi \,\mathrm{d}S, \tag{3.51}$$

$$
\begin{aligned}
l_h(\varphi)(t) = &\int_\Omega g(t)\varphi \mathrm{d}x + \sum_{i\in I}\int_{\partial K_i^+\cap\partial\Omega} u_N(t)\varphi \,\mathrm{d}S \\
&+ \varepsilon\sum_{i\in I}\int_{\partial K_i^-\cap\partial\Omega} \sigma u_D(t)\varphi \,\mathrm{d}S + \varepsilon\sum_{i\in I}\int_{\partial K_i^-\cap\partial\Omega} u_D(t)(\nabla\varphi\cdot\boldsymbol{n})\,\mathrm{d}S \\
&- \sum_{i\in I}\int_{\partial K_i^-\cap\partial\Omega} (\boldsymbol{v}\cdot\boldsymbol{n}) u_D(t)\varphi\,\mathrm{d}S,
\end{aligned}
\tag{3.52}
$$

where $\sigma|_{\Gamma_{ij}} = 1/\operatorname{diam}(\Gamma_{ij})$.

The approximate solution will be sought in the space $C^1([0,T);\, S_h)$,where S_h is the finite element space

$$
S_h = \{\varphi \in L^2(\Omega); \varphi|_K \in P^p(K) \quad \forall K \in \mathcal{T}_h\}.
$$

For the sake of a posteriori error estimate we shall also need an additional restriction on u_D:

$$
u_D \in S_h(\partial\Omega^-) = \{\varphi \in C(\partial\Omega^-);\ \forall j \in I_b \varphi|_{S_j} \in P^p(S_j)\},
\tag{3.53}
$$

where $P^p(S_j)$ denotes the set of restrictions on s_j of all polynomials of degree $\le p$. This condition is somewhat restrictive but is often satisfied in practice. Note that it is not needed for the discrete problem. The *DGFE discrete problem* now reads: Find a finite element function u_h such that

$$
u_h \in C^1([0,T); S_h),
\tag{3.54a}
$$

$$
\left(\frac{\partial u_h(t)}{\partial t}, \varphi_h\right) + a_h(u_h(t),\varphi_h) + b_h(u_h(t),\varphi_h) + c_h(u_h(t),\varphi_h) + \varepsilon\, J_h^\sigma(u_h(t),\varphi_h) = l_h(\varphi_h)(t),
\tag{3.54b}
$$

$$
(u_h(0),\varphi_h) = (u^0,\varphi_h)
\tag{3.54c}
$$

for all $\varphi_h \in S_h$. In the case $\varepsilon = 0$ we allow $p = 0$ (the finite volume method for hyperbolic problems).

3.3 Auxiliary results

Lemma 1 *There exists a global constant C_Π and a mapping $\Pi\colon H^{p+1}(K) \to P^p(K), p \ge 1$, such that*

$$
\|\Pi v - v\|_{L^2(K)} \le C_\Pi h_K^{p+1} |v|_{H^{p+1}(K)},
\tag{3.55a}
$$

$$
|\Pi v - v|_{H^1(K)} \le C_\Pi h_K^{p} |v|_{H^{p+1}(K)},
\tag{3.55b}
$$

$$
|\Pi v - v|_{H^2(K)} \le C_\Pi h_K^{p-1} |v|_{H^{p+1}(K)} \qquad (\text{if } p \ge 1),
\tag{3.55c}
$$

for all $v \in H^{p+1}(K), K \in \mathcal{T}_h, h \in (0, h_0)$. The operator Π can be chosen as the $L^2(K)$-projection on $P^p(K)$.

Proof. See [22], Lemma 4.1.

Lemma 2 *There exists a global constant C_M such that*

$$\|v\|^2_{L^2(\partial K)} \leq C_M \left(\|v\|_{L^2(K)} |v|_{H^1(K)} + h_K^{-1} \|v\|^2_{L^2(K)} \right) \qquad \forall K \in \mathcal{T}_h,\ v \in H^1(K), h \in (0, h_0). \tag{3.56}$$

Proof. See [17].

Lemma 3 *There exists a global constant C_I such that*

$$|v|_{H^1(K)} \leq C_I h_K^{-1} \|v\|_{L^2(K)} \qquad \forall K \in \mathcal{T}_h,\ v \in P^p(K). \tag{3.57}$$

Lemma 4 *Let Q_1, Q_2 be a pair of nonzero quadratic forms on a real finite-dimensional space X with Q_1 positive semidefinite, and let $\ker Q_1 \subset \ker Q_2$. Then there exists a constant $C > 0$ such that*

$$C\, Q_1(x) \geq Q_2(x) \quad \forall x \in X. \tag{3.58}$$

Proof. Let $Y = (\ker Q_1)^\perp$. Then Q_1 is positive definite on Y, i.e. for some $\gamma > 0$

$$Q_1(y) \geq \gamma \, \|y\|^2 \quad \forall y \in Y. \tag{3.59}$$

As Q_2 is continuous,

$$Q_2(x) \leq c \, \|x\|^2 \quad \forall x \in X \tag{3.60}$$

with some $c > 0$. Writing any $x \in X$ as $x = y + \bar{y}$, $y \in Y$, $\bar{y} \in \ker Q_2$, we get

$$\frac{c}{\gamma} Q_1(x) = \frac{c}{\gamma} Q_1(y) \geq c\|y\|^2 \geq Q_2(y) = Q_2(x) \tag{3.61}$$

which concludes the proof.

Lemma 5 *Let $\hat{\Gamma}$ be a reference face and let $\hat{\phi}_l$, $l = 1, \ldots, \binom{p+1}{d-1}$, be a basis of the space $P^p(\hat{\Gamma})$. Then there exists a global constant $C_{\hat{\Gamma}}$ such that for any Γ_{ij}, $i \in I$, $j \in S(i)$ and any $v \in P^p(\Gamma_{ij})$ we have*

$$C_{\hat{\Gamma}} \|v\|^2_{L^2(\Gamma_{ij})} \geq \operatorname{diam}(\Gamma_{ij})^{d-1} \sum_{l=1}^{\binom{p+1}{d-1}} \beta_l^2, \tag{3.62}$$

where β_l are given by the (unique) decomposition

$$v = \sum_{l=1}^{\binom{p+1}{d-1}} \beta_l \phi_l, \quad \phi_l = \hat{\phi}_l \circ F_{ij}^{-1} \tag{3.63}$$

and $F_{ij} \colon \hat{\Gamma} \to \Gamma_{ij}$ is an affine one-to-one mapping.

Proof. We use Lemma 4 with the following pair of quadratic forms in $\beta_l, l = 1, \ldots, \binom{p+1}{d-1}$:

$$Q_1 = \| \sum_{l=1}^{\binom{p+1}{d-1}} \beta_l \hat{\phi}_l \|^2_{L^2(\hat{\Gamma}_{ij})}, \tag{3.64}$$

$$Q_2 = \sum_{l=1}^{\binom{p+1}{d-1}} \beta_l^2. \tag{3.65}$$

Apparently Q_1 is positive semidefinite, and since $\hat{\phi}_l$ form a basis, both Q_1 and Q_2 are regular (i.e. have trivial kernel). The conditions of Lemma 4 are thus satisfied. To complete the proof it suffices to notice that

$$\| \sum_{l=1}^{\binom{p+1}{d-1}} \beta_l \hat{\phi}_l \|^2_{L^2(\hat{\Gamma}_{ij})} = \|v \circ F_{ij}\|_{L^2(\hat{\Gamma}_{ij})^2} = \frac{|\Gamma_{ij}|}{|\hat{\Gamma}|} \|v\|^2_{L^2(\Gamma_{ij})} \tag{3.66}$$

and, by shape regularity (3.36)

$$c\,|\Gamma_{ij}| \geq \operatorname{diam}(\Gamma_{ij})^{d-1} \tag{3.67}$$

with a global constant c.

To be able to rigorously derive an a posteriori error estimate without an assumption of the quasi-uniformity of the mesh (or the so-called inverse assumption), we need a more general version of Lemma 6 in [41] about the properties of the *continuous reconstruction* - construction of a globally continuous approximate solution from a piecewise continuous one. The following lemma proves the *continuous reconstruction* operator to have "enough locality".

We define the continuous piecewise polynomial space as $\hat{S}_h = C(\overline{\Omega}) \cap S_h$.

Lemma 6 *Let there exist a global constant k_s such that any vertex, edge or face of the mesh is shared by no more than k_s elements. Then there exists a linear operator*

$$\mathcal{C} \colon S_h \times S_h(\partial\Omega^-) \to \hat{S}_h$$

with the following property: If w_i, $i \in I$, are positive numbers satisfying the condition

$$w_i \leq k_w w_j \quad \forall i \in I, j \in s(i) \tag{3.68}$$

with a global constant k_w, then there exists a global constant C such that any $u \in S_h(\Omega)$, $\hat{u} = \mathcal{C}(u, u_D)$ satisfy

$$\hat{u}|_{\partial\Omega^-} = u_D \tag{3.69}$$

and

$$\sum_{i \in I} w_i \|u - \hat{u}\|^2_{L^2(K_i)} \leq C \sum_{i \in I} w_i h_{K_i} \|[u]\|^2_{L^2(\partial K_i)} \tag{3.70}$$

with $[u]$ on $\Gamma_{ij} = S_j$ given by the extended notation introduced in subsection 3.2.

Proof. Let $\{\phi_l\}_{l=1}^{\dim \hat{S}_h}$ be a standard node-based basis for the conforming (continuous) finite element space $\hat{S}_h$ and $\{x_l\}_{l=1}^{\dim \hat{S}_h}$ be the setof the corresponding nodes, i.e. $\mathrm{span}\{\phi_l\} = \hat{S}_h$, $\phi_l(x_m) = \delta_{lm}$. We denote $I_l = \{i;\ K_i \subset \mathrm{supp}\,\phi_l\}$. We can (uniquely) decompose u as

$$u = \sum_{l=1}^{\dim \hat{S}_h} \sum_{i \in I_l} \alpha_{li}\, \phi_l|_{K_i}. \tag{3.71}$$

To deal with the boundary condition, we shall denote $L_b^- = \{l;\ x_l \in \overline{\partial\Omega^-}\}$ and decompose also u_D as

$$u_D = \sum_{l \in L_b^-} \alpha_{lb}\, \phi_l|_{\partial\Omega^-}. \tag{3.72}$$

We also denote

$$\gamma(i)^- = \{j \in \gamma(i);\ \Gamma_{ij} \subset \partial\Omega^-\}. \tag{3.73}$$

For simplicity of the notation, we shall treat b as a special index from an extended index set $I^b = I \cup \{b\}$ and allow the expression α_{li} to become α_{lb} for $i = b$. We shall also denote

$$I_l^b = \begin{cases} I_l & l \notin L_b^- \\ I_l \cup \{b\} & l \in L_b^- \end{cases}. \tag{3.74}$$

Now we define $\hat{u}$ as

$$\hat{u} = \sum_{l=1}^{\dim \hat{S}_h} \bar{\alpha}_l \phi_l, \tag{3.75}$$

where we set

$$\bar{\alpha}_l = \begin{cases} \frac{\sum_{i \in I_l} \alpha_{li}}{\# I_l} & l \notin L_b^- \\ \alpha_{lb} & l \in L_b^- \end{cases}. \tag{3.76}$$

For any $i \in I$, let us denote $L_i = \{l;\ K_i \subset \mathrm{supp}\,\phi_l\}$. We can write

$$\begin{aligned} \sum_{i \in I} w_i \|u - \hat{u}\|^2_{L^2(K_i)} &= \sum_{i \in I} w_i \Big\| \sum_{l \in L_i} (\alpha_{li} - \bar{\alpha}_l) \phi_l|_{K_i} \Big\|^2_{L^2(K_i)} \\ &\le \sum_{i \in I} w_i \# L_i \sum_{l \in L_i} \|(\alpha_{li} - \bar{\alpha}_l)\phi_l|_{K_i}\|^2_{L^2(K_i)} \\ &\le k_s \sum_{l=1}^{\dim \hat{S}_h} \sum_{i \in I_l} w_i \|(\alpha_{li} - \bar{\alpha}_l)\phi_l|_{K_i}\|^2_{L^2(K_i)} \\ &\le k_1 \sum_{l=1}^{\dim \hat{S}_h} \left(\max_{i \in I_l} h_{K_i}^d w_i \right) \sum_{i \in I_l} (\alpha_{li} - \bar{\alpha}_l)^2, \end{aligned} \tag{3.77}$$

where k_1 is a global constant. Now we use Lemma 5 to estimate

$$\begin{aligned}
&\sum_{i\in I} w_i h_{K_i} \|[u]\|^2_{L^2(\partial K_i)\setminus\partial\Omega^+} = \sum_{i\in I} w_i h_{K_i} \sum_{j\in s(i)\cap\gamma(i)^-} \|[u]\|^2_{L^2(\Gamma_{ij})} \\
&\geq k_2 \sum_{i\in I} w_i h_{K_i} \left(\sum_{j\in s(i)} \sum_{l\in L_i\cap L_j} \operatorname{diam}(\Gamma_{ij})^{d-1} (\alpha_{li}-\alpha_{lj})^2 \right. \\
&\left. + \sum_{j\in\gamma(i)^-} \sum_{l\in L_i\cap L_b^-} \operatorname{diam}(\Gamma_{ij})^{d-1} (\alpha_{li}-\alpha_{lb})^2 \right) \\
&= k_2 \sum_{l=1}^{\dim \hat{S}_h} \sum_{i\in I_l} w_i h_{K_i} \left(\sum_{j\in s(i)} \operatorname{diam}(\Gamma_{ij})^{d-1} (\alpha_{li}-\alpha_{lj})^2 \right. \\
&\left. + \sum_{j\in\gamma(i)^-} \operatorname{diam}(\Gamma_{ij})^{d-1} (\alpha_{li}-\alpha_{lb})^2 \right) \\
&\geq k_3 \sum_{l=1}^{\dim \hat{S}_h} \left(\min_{i\in I_l} h^d_{K_i} w_i \right) \sum_{i\in I_l} \left(\sum_{j\in s(i)} (\alpha_{li}-\alpha_{lj})^2 + \#\gamma(i)^- (\alpha_{li}-\alpha_{lb})^2 \right),
\end{aligned} \tag{3.78}$$

where k_2, k_3 are global constants. Because of the fact that $\#I_l^b \leq k_s$, there exists a global constant k_4 such that

$$\min_{i\in I_l^b} h^d_{K_i} w_i \geq k_4 \max_{i\in I_l^b} h^d_{K_i} w_i. \tag{3.79}$$

Now we use Lemma 4. We see that to conclude the proof we must show that for some global constant k_5

$$Q_l^1(\alpha_{li},\, i\in I_l^b) \leq k_5\, Q_l^2(\alpha_{li},\, i\in I_l^b), \quad l=1,\ldots,\dim\hat{S}_h, \tag{3.80}$$

where

$$Q_l^1(\alpha_{li},\, i\in I_l^b) = \sum_{i\in I_l} (\alpha_{li}-\bar{\alpha}_l)^2, \tag{3.81}$$

$$Q_l^2(\alpha_{li},\, i\in I_l^b) = \sum_{i\in I_l} \left(\sum_{j\in s(i)} (\alpha_{li}-\alpha_{lj})^2 + \#\gamma(i)^- (\alpha_{li}-\alpha_{lb})^2 \right). \tag{3.82}$$

As both Q_l^1 and Q_l^2 are sums of squares of linear terms, they are positive semidefinite quadratic forms. To use Lemma 4 we must also show that $\ker Q_l^1 \subset \ker Q_l^2$. Since

$$\bigcup_{i\in I_l} K_i = \operatorname{supp}\phi_l$$

and $\operatorname{supp}\phi_l$ is closed and connected, the oriented graphs

$$G_l = \left(I_l^b;\ \{(i,j);\ i \in I_l, j \in I_l \cap s(i)\} \cup \{(i,b);\ i \in I_l, \gamma(i)^- \neq \emptyset\}\right)$$

are weakly connected. Thus

$$\begin{aligned} &Q_l^2(\alpha_{li},\, i \in I_l^b) = 0 \Leftrightarrow \forall (i,j) \in G_l \quad \alpha_{li} = \alpha_{lj} \\ &\Rightarrow \forall i \in I_l, j \in I_l^b \quad \alpha_{li} = \alpha_{lj} \Rightarrow Q_l^1(\alpha_{li},\, i \in I_l^b) = 0. \end{aligned} \tag{3.83}$$

Now we can employ Lemma 4 for (3.80), but this time the quadratic forms Q_l^1, Q_l^2 depend on the mesh. Although they are mesh dependent, they depend only on the graph G_l and not on the shape or size of elements K_i. Due to the fact that $\#I_l^b \leq k_s$, there are only finitely many graphs and one can always pick the largest constant k_5 that satisfies (3.80). This concludes the proof.

3.4 A posteriori error estimate

Since we shall not be interested in estimating the error due to initial condition, we shall assume that $e(0) = 0$, i.e. $u \in S_h$. For simplicity of notation we introduce the following norm over a subset B of either $\partial\Omega$ or ∂K_i for some $i \in I$:

$$\|\varphi\|_{\boldsymbol{v},B} = \|\sqrt{|\boldsymbol{v}\cdot\boldsymbol{n}|}\,\varphi\|_{L^2(B)}, \tag{3.84}$$

where $\boldsymbol{n}$ is the corresponding outer unit normal. Moreover, from now on we shall not emphasize the dependence of e on time by notation, i.e. we shall write simply e instead of $e(t)$. We denote

$$\rho_I = \frac{\partial e}{\partial t} + \boldsymbol{v}\cdot\nabla e - \varepsilon\Delta e + ce = \frac{\partial u_h}{\partial t} + \boldsymbol{v}\cdot\nabla u_h - \varepsilon\Delta u_h + cu_h - g \tag{3.85}$$

the *interior residual*. Moreover, on each edge we define $\rho_{B0} = [e] = [u_h]$, $\rho_{B1} = [\nabla e]\cdot\boldsymbol{n} = [\nabla u_h]\cdot\boldsymbol{n}$. These quantities are easily computable once the approximate solution is computed. The main result of this section follows:

Theorem 2 *There exists a global constant C such that*

$$\begin{aligned} &\frac{1}{2}\|e\|^2_{L^\infty(0,T;L^2(\Omega))} + \varepsilon\,|e|^2_{L^2(0,T;H^1(\Omega))} + \gamma_0\|e\|^2_{L^2(0,T;L^2(\Omega))} \\ &\leq C\sum_{i\in I}\left(\varepsilon^{-1}h_{K_i}^2\|\rho_I\|^2_{L^2(0,T;L^2(K_i))} + \gamma_0^{-1}h_{K_i}\Big\|\frac{\partial\rho_{B0}}{\partial t}\Big\|^2_{L^2(0,T;L^2(\partial K\setminus\partial\Omega^+))}\right. \\ &+(\varepsilon h_{K_i}^{-1} + \varepsilon^{-1}h_{K_i} + \gamma_0^{-1}h_{K_i} + 1)\|\rho_{B0}\|^2_{L^2(0,T;L^2(\partial K_i\setminus\partial\Omega^+))} + \\ &\left.\varepsilon h_{K_i}\|\rho_{B1}\|^2_{L^2(0,T;L^2(\partial K_i\setminus\partial\Omega^-))} + \gamma_0^{-1}h_{K_i}\|\rho_{B0}\|^2_{L^\infty(0,T;L^2(\partial K_i\setminus\partial\Omega^+))}\right). \end{aligned} \tag{3.86}$$

Proof. To derive the a posteriori estimate, we subtract (3.46) with $\varphi = \varphi_h$ from (3.54b) to get the so-called *Galerkin orthogonality*:

$$\left(\frac{\partial e(t)}{\partial t}, \varphi\right) + a_h(e(t), \varphi_h) + b_h(e(t), \varphi) + c_h(e(t), \varphi) + +\varepsilon J_h^\sigma(e(t), \varphi) = 0, \tag{3.87}$$

where $e(t) = u_h(t) - u(t)$. From now on, we shall write simply e instead of $e(t)$. Now we define $\bar{e}$ to be the piecewise constant L^2-projection of e. We set $\varphi = e$ in (3.46) and $\varphi_h = \bar{e}$ in (3.54b) and integrate with respect to time from 0 to τ to obtain

$$\begin{aligned}
&\int_0^\tau \left(\left(\frac{\partial e}{\partial t}, e\right) + a_h(e, e) + b_h(e, e) + c_h(e, e) + \varepsilon J_h^\sigma(e, e) \right) \mathrm{d}t = \\
&= \int_0^\tau \left(\left(\frac{\partial e}{\partial t}, e - \bar{e}\right) + a_h(e, e - \bar{e}) + b_h(e, e - \bar{e}) \right. \\
&\left. + c_h(e, e - \bar{e}) + \varepsilon J_h^\sigma(e, e - \bar{e}) \right) \mathrm{d}t.
\end{aligned} \tag{3.88}$$

First, we shall estimate the terms on the left-hand side of equation (3.88). Obviously,

$$\left(\frac{\partial e}{\partial t}, e\right) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|e\|^2_{L^2(\Omega)}, \tag{3.89}$$

$$a_h(e, e) = \varepsilon |e|^2_{H^1(\Omega, \mathcal{T}_h)}. \tag{3.90}$$

Further, by (3.49), the relation $e\nabla e = 1/2\, \nabla e^2$ and Green's theorem, we have

$$\begin{aligned}
b_h(e, e) = \sum_{i \in I} \Bigg(&\int_{K_i} (\boldsymbol{v} \cdot \nabla e)\, e \,\mathrm{d}x - \int_{\partial K_i^- \cap \partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n}) e^2 \mathrm{d}S \\
&- \int_{\partial K_i^- \setminus \partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n})[e] e \,\mathrm{d}S \Bigg) = \sum_{i \in I} \Bigg\{ -\frac{1}{2} \int_{K_i} e^2 \operatorname{div} \boldsymbol{v} \,\mathrm{d}x \\
&+ \frac{1}{2} \int_{\partial K_i} (\boldsymbol{v} \cdot \boldsymbol{n})\, e^2 \,\mathrm{d}S - \int_{\partial K_i^- \cap \partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n})\, e^2 \,\mathrm{d}S \\
&- \int_{\partial K_i^- \setminus \partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n})\, e(e - e^-) \mathrm{d}S \Bigg\}.
\end{aligned}$$

Using the decomposition $\partial K = \partial K^- \cup \partial K^+$, we can write

$$\begin{aligned}
b_h(e, e) = \sum_{i \in I} \frac{1}{2} \Bigg\{ &- \int_{K_i} e^2 \operatorname{div} \boldsymbol{v} \,\mathrm{d}x - \int_{\partial K_i^- \cap \partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n}) e^2 \,\mathrm{d}S \\
&- \int_{\partial K_i^- \setminus \partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n})(e^2 - 2ee^-) \,\mathrm{d}S + \int_{\partial K_i^+ \cap \partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n}) e^2 \,\mathrm{d}S \\
&+ \int_{\partial K_i^+ \setminus \partial\Omega} (\boldsymbol{v} \cdot \boldsymbol{n}) e^2 \,\mathrm{d}S \Bigg\}.
\end{aligned}$$

Now, using the relation

$$\sum_{i\in I}\int_{\partial K_i^+\setminus\partial\Omega}(\boldsymbol{v}\cdot\boldsymbol{n})e^2\,\mathrm{d}S=-\sum_{i\in I}\int_{\partial K_i^-\setminus\partial\Omega}(\boldsymbol{v}\cdot\boldsymbol{n})(e^-)^2\,\mathrm{d}S,$$

(which follows from the fact that for each $i\in I$ if $\Gamma_{ij}\subset\partial K_i^+\setminus\partial\Omega$, then $j\in s(i)$ and $\Gamma_{ij}=\Gamma_{ji}\subset\partial K_j^-\setminus\partial\Omega$) we find that

$$\begin{aligned}b_h(e,e)&=\sum_{i\in I}\frac{1}{2}\left\{-\int_{\partial K_i}e^2\operatorname{div}\boldsymbol{v}\,\mathrm{d}x-\int_{\partial K_i^-\cap\partial\Omega}\mathrm{d}S\right.\\&\quad\left.-\int_{\partial K_i^-\setminus\partial\Omega}(\boldsymbol{v}\cdot\boldsymbol{n})(e^2-2ee^-+(e^-)^2)\,\mathrm{d}S+\int_{\partial K_i^+\cap\partial\Omega}(\boldsymbol{v}\cdot\boldsymbol{n})e^2\,\mathrm{d}S\right\}\\&=\frac{1}{2}\sum_{i\in I}\left(\|e\|^2_{\boldsymbol{v},\partial K_i^-\cap\partial\Omega}+\|[e]\|^2_{\boldsymbol{v},\partial K_i^-\setminus\partial\Omega}+\|e\|_{\boldsymbol{v},\partial K_i^+\cap\partial\Omega}\right)\\&\quad-\frac{1}{2}\int_\Omega e^2\operatorname{div}\boldsymbol{v}\,\mathrm{d}x\end{aligned}\tag{3.91}$$

and, thus, we can estimate

$$\begin{aligned}b_h(e,e)+c_h(e,e)\geq\frac{1}{2}\sum_{i\in I}&\left(\|e\|^2_{\boldsymbol{v},\partial K_i^-\cap\partial\Omega}+\|[e]\|^2_{\boldsymbol{v},\partial K_i^-\setminus\partial\Omega}+\|e\|^2_{\boldsymbol{v},\partial K_i^+\cap\partial\Omega}\right)\\&+\gamma_0\|e\|^2_{L^2(\Omega)}.\end{aligned}\tag{3.92}$$

Finally,

$$J_h^\sigma(e,e)=\sum_{i\in I}\sum_{j\in s(i)}\int_{\Gamma_{ij}}\sigma[e]^2\,\mathrm{d}S+\sum_{i\in I}\int_{\partial K_i^-\cap\partial\Omega}\sigma e^2\,\mathrm{d}S.\tag{3.93}$$

Let us look at the right-hand side of the equation (3.88). Integration by parts in $a_h(e,e)$ yields

$$\begin{aligned}a_h(e,e-\bar e)&=-\varepsilon\sum_{i\in I}\int_{K_i}\Delta e(e-\bar e)\,\mathrm{d}x\\&\quad+\varepsilon\sum_{i\in I}\sum_{j\in s(i),j<i}\int_{\Gamma_{ij}}([\nabla e]\cdot\boldsymbol{n}_{ij}\langle e-\bar e\rangle+\langle\nabla e\rangle\cdot\boldsymbol{n}_{ij}[e-\bar e])\,\mathrm{d}S\\&\quad-\varepsilon\sum_{i\in I}\left(\int_{\partial K_i^+\cap\partial\Omega}(\nabla e\cdot\boldsymbol{n})(e-\bar e)+\int_{\partial K_i^-\cap\partial\Omega}(\nabla e\cdot\boldsymbol{n})e\,\mathrm{d}S\right).\end{aligned}\tag{3.94}$$

Using this relation, we see that the right-hand side of (3.88) can be rewritten in the following way:

$$\begin{aligned}&\int_0^\tau\left(\frac{\partial e}{\partial t},e-\bar e\right)+a_h(e,e-\bar e)+b_h(e,e-\bar e)+c_h(e,e-\bar e)+\varepsilon J_h^\sigma(e,e-\bar e)\,\mathrm{d}t\\&=\int_0^\tau(T_1+T_2+T_3+T_4+T_5)\,\mathrm{d}t,\end{aligned}\tag{3.95}$$

where

$$T_1 = \sum_{i\in I}\int_{K_i}\left(\frac{\partial e}{\partial t} + \boldsymbol{v}\cdot\nabla e - \varepsilon\Delta e + ce\right)(e-\bar{e})\,\mathrm{d}x, \tag{3.96}$$

$$T_2 = \sum_{i\in I}\sum_{j\in s(i),j<i}\int_{\Gamma_{ij}}\varepsilon[\nabla e]\cdot\boldsymbol{n}_{ij}\langle e-\bar{e}\rangle\,\mathrm{d}S + \sum_{i\in I}\int_{\partial K_i^+\cap\partial\Omega}\varepsilon(\nabla e\cdot\boldsymbol{n})(e-\bar{e})\,\mathrm{d}S, \tag{3.97}$$

$$T_3 = \sum_{i\in I}\sum_{\in s(i),j<i}\int_{\Gamma_{ij}}\varepsilon\langle\nabla e\rangle\cdot\boldsymbol{n}_{ij}[e])\,\mathrm{d}S + \sum_{i\in I}\int_{\partial K_i^-\cap\partial\Omega}\varepsilon(\nabla e\cdot\boldsymbol{n})e\,\mathrm{d}S, \tag{3.98}$$

$$T_4 = -\sum_{i\in I}\int_{\partial K_i^-\cap\partial\Omega}(\boldsymbol{v}\cdot\boldsymbol{n})\,e(e-\bar{e})\,\mathrm{d}S - \sum_{i\in I}\int_{\partial K_i^-\setminus\partial\Omega}(\boldsymbol{v}\cdot\boldsymbol{n})[e](e-\bar{e})\,\mathrm{d}S, \tag{3.99}$$

$$T_5 = \varepsilon J_h^\sigma(e, e-\bar{e}). \tag{3.100}$$

In the following we shall estimate the terms T_i. Whenever a constant c_i appears, we mean that there exists a global constant c_i such that the inequality holds, whereas for α_j we mean there exists a global constant α_j^{up} such that the inequality holds for all $\alpha_j \in (0, \alpha_j^{\mathrm{up}})$.

We begin with the estimation of T_1. We use Cauchy-Schwarz inequality

$$\int_{K_i}\rho_I(e-\bar{e})\,\mathrm{d}x \le \|\rho_I\|_{L^2(K_i)}\|e-\bar{e}\|_{L^2(K_i)} \tag{3.101}$$

and employ Young's inequality and Lemma 1 to get

$$T_1 \le c_1\left(\varepsilon\alpha_1|e|^2_{H^1(\Omega;\mathcal{T}_h)} + \frac{1}{\varepsilon\alpha_1}\sum_{i\in I}h_{K_i}^2\int_{K_i}\rho_I^2\,\mathrm{d}x\right). \tag{3.102}$$

To estimate the term T_2, we use

$$\begin{aligned}&\sum_{j\in s(i),j<i}\int_{\Gamma_{ij}}\varepsilon[\nabla e]\cdot\boldsymbol{n}_{ij}\langle e-\bar{e}\rangle\,\mathrm{d}S + \int_{\partial K_i^+\cap\partial\Omega}\varepsilon(\nabla e\cdot\boldsymbol{n})(e-\bar{e})\,\mathrm{d}S\\ &\le \varepsilon\|\rho_{B1}\|_{L^2(\partial K_i\setminus\partial\Omega^-)}\|e-\bar{e}\|_{L^2(\partial K_i)}.\end{aligned} \tag{3.103}$$

Lemma 2 with Lemma 1 gives

$$\|e-\bar{e}\|_{L^2(\partial K_i)} \le \sqrt{C_M(C_\Pi + C_\Pi^2)}\,h_{K_i}^{1/2}|e|_{H^1(K_i)} \tag{3.104}$$

and thus, using Young's inequality, we get

$$T_2 \le c_2\left(\varepsilon\alpha_2|e|^2_{H^1(\Omega;\mathcal{T}_h)} + \frac{\varepsilon}{\alpha_2}\sum_{i\in I}h_{K_i}\|\rho_{B1}\|^2_{\partial K_i\setminus\partial\Omega^-}\right). \tag{3.105}$$

Similarly, we proceed for T_4 and T_5:

$$\begin{aligned} &-\int_{\partial K_i^- \cap \partial\Omega} (\boldsymbol{v}\cdot\boldsymbol{n})\, e(e-\bar{e})\,\mathrm{d}S - \int_{\partial K_i^- \setminus \partial\Omega} (\boldsymbol{v}\cdot\boldsymbol{n})[e](e-\bar{e})\,\mathrm{d}S \\ &\le C_{\boldsymbol{v}} \|\rho_{B0}\|_{L^2(\partial K_i \setminus \partial\Omega^+)} \|e-\bar{e}\|_{L^2(\partial K_i)}. \end{aligned} \tag{3.106}$$

Hence, using (3.104) and Young's inequality,

$$T_4 \le c_4 \left(\varepsilon\alpha_4 |e|^2_{H^1(\Omega;\mathcal{T}_h)} + \frac{1}{\varepsilon\alpha_4} \sum_{i\in I} h_{K_i} \|\rho_{B0}\|^2_{\partial K_i \setminus \partial\Omega^+} \right). \tag{3.107}$$

Finally,

$$\begin{aligned} &\sum_{j\in s(i)} \int_{\Gamma_{ij}} \sigma[e][e-\bar{e}]\,\mathrm{d}S + \int_{\partial K_i^- \cap \partial\Omega} \sigma e\,(e-\bar{e})\,\mathrm{d}S, \\ &\le \frac{C_{\mathcal{T}}}{2} h_{K_i}^{-1} \|\rho_{B0}\|_{L^2(\partial K_i \setminus \partial\Omega^+)} \|e-\bar{e}\|_{L^2(\partial K_i)}. \end{aligned} \tag{3.108}$$

Again, from (3.104) and Young's inequality,

$$T_5 \le c_5 \left(\varepsilon\alpha_5 |e|^2_{H^1(\Omega;\mathcal{T}_h)} + \frac{\varepsilon}{\alpha_5} \sum_{i\in I} h_{K_i}^{-1} \|\rho_{B0}\|^2_{\partial K_i \setminus \partial\Omega^+} \right). \tag{3.109}$$

The term T_3 is more challenging because $\langle\nabla e\rangle\cdot\boldsymbol{n}$ cannot be easily handled using the trace inequality (Lemma 2). To be able to proceed, we employ the *continuous reconstruction* $\hat{u}_h = \mathcal{C}(u, u_D)$ according to Lemma 6. We shall set $\hat{\xi} = u_h - \hat{u}_h$ and use the Galerkin orthogonality (3.87) with $\varphi_h = \hat{\xi}$ to get

$$0 = \int_0^\tau T_6 + T_7 + T_8 + T_9 + T_{10} + T_{11}\,\mathrm{d}t, \tag{3.110}$$

where

$$T_6 = \left(\frac{\partial e}{\partial t}, \hat{\xi} \right) \tag{3.111}$$

$$T_7 = \sum_{i\in I} \int_{K_i} (\varepsilon \nabla e\cdot\nabla\hat{\xi} + (\boldsymbol{v}\cdot\nabla e)\hat{\xi} + ce\hat{\xi})\,\mathrm{d}S, \tag{3.112}$$

$$T_8 = -\varepsilon \sum_{i\in I} \sum_{\in s(i), j<i} \int_{\Gamma_{ij}} \langle\nabla e\rangle\cdot\boldsymbol{n}_{ij}[\hat{\xi}]\,\mathrm{d}S - \varepsilon \sum_{i\in I} \int_{\partial K_i^- \cap \partial\Omega} \nabla e\cdot\boldsymbol{n}\hat{\xi}\,\mathrm{d}S, \tag{3.113}$$

$$T_9 = \varepsilon \sum_{i\in I} \sum_{\in s(i), j<i} \int_{\Gamma_{ij}} \langle\nabla\hat{\xi}\rangle\cdot\boldsymbol{n}_{ij}[e]\,\mathrm{d}S + \varepsilon \sum_{i\in I} \int_{\partial K_i^- \cap \partial\Omega} \nabla\hat{\xi}\cdot\boldsymbol{n}e\,\mathrm{d}S, \tag{3.114}$$

$$T_{10} = -\sum_{i\in I} \int_{\partial K_i^- \cap \partial\Omega} (\boldsymbol{v}\cdot\boldsymbol{n})\, e\hat{\xi}\,\mathrm{d}S - \sum_{i\in I} \int_{\partial K_i^- \setminus \partial\Omega} (\boldsymbol{v}\cdot\boldsymbol{n})[e]\hat{\xi}\,\mathrm{d}S, \tag{3.115}$$

$$T_{11} = \varepsilon J_h^\sigma(e, \hat{\xi}). \tag{3.116}$$

Integrating (3.111) over $(0,\tau)$ with respect to t and using integration by parts yields

$$\int_0^\tau T_6 \,\mathrm{d}t = \int_0^\tau \left(\frac{\partial e}{\partial t}, \hat{\xi}\right) \,\mathrm{d}t = \left(e(\tau), \hat{\xi}(\tau)\right) - \int_0^\tau \left(e, \frac{\partial \hat{\xi}}{\partial t}\right) \,\mathrm{d}t. \tag{3.117}$$

Now, using (3.111) and Young's inequality we get

$$\begin{aligned}\int_0^\tau T_6 \,\mathrm{d}t \le c_6 \Bigg(&\alpha_6 \left(\frac{1}{2}\|e(\tau)\|^2_{L^2(\Omega)} + \gamma_0 \int_0^\tau \|e\|^2_{L^2(\Omega)} \,\mathrm{d}t\right) \\ &+ \frac{1}{\alpha_6}\left(2\|\hat{\xi}(\tau)\|^2_{L^2(\Omega)} + \frac{1}{\gamma_0}\int_0^\tau \|\frac{\partial \hat{\xi}}{\partial t}\|^2_{L^2(\Omega)} \,\mathrm{d}t\right)\Bigg).\end{aligned} \tag{3.118}$$

We notice that $[\hat{\xi}] = \rho_{B0}$ and $[\nabla\hat{\xi}]\cdot\boldsymbol{n} = \rho_{B1}$ on edges. We see that the term T_8 cancels with T_3. We estimate the remaining terms:

$$\begin{aligned}\int_{K_i} (\varepsilon \nabla e\cdot\nabla\hat{\xi} + (\boldsymbol{v}\cdot\nabla e)\hat{\xi} + ce\hat{\xi})\,\mathrm{d}x \le{}& \varepsilon |e|_{H^1(K_i)}|\hat{\xi}|_{H^1(K_i)} \\ &+ C_{\boldsymbol{v}}|e|_{H^1(K_i)}\|\hat{\xi}\|_{L^2(K_i)} + C_c\|e\|_{L^2(K_i)}\|\hat{\xi}\|_{L^2(K_i)}.\end{aligned} \tag{3.119}$$

By Lemma 3 we get

$$|\hat{\xi}|_{H^1(K_i)} \le C_I h_{K_i}^{-1}\|\hat{\xi}\|_{L^2(K_i)} \tag{3.120}$$

and thus

$$T_7 \le c_7\left(\alpha_7\left(\gamma_0\|e\|^2_{L^2(\Omega)} + \varepsilon|e|^2_{H^1(\Omega;\mathcal{T}_h)}\right) + \frac{1}{\alpha_7}\sum_{i\in I}(\varepsilon h_{K_i}^{-2} + \varepsilon^{-1} + \gamma_0^{-1})\|\hat{\xi}\|^2_{L^2(K_i)}\right). \tag{3.121}$$

By Lemma 2 and Lemma 3

$$\|\nabla\hat{\xi}\|_{L^2(\partial K_i)} \le \sqrt{C_M(C_I^2 + C_I^3)}\, h_{K_i}^{-3/2}\|\hat{\xi}\|_{L^2(K_i)} \tag{3.122}$$

and thus

$$\begin{aligned}T_9 &\le \varepsilon \sum_{i\in I}\|\nabla\hat{\xi}\|_{L^2(\partial K_i)}\|\rho_{B0}\|_{L^2(\partial K_i\setminus\partial\Omega^+)} \\ &\le \sum_{i\in I}\frac{1}{2}\left(\varepsilon\, h_{K_i}^{-2}\|\hat{\xi}\|^2_{L^2(K_i)} + \varepsilon\, h_{K_i}^{-1}\|\rho_{B0}\|^2_{L^2(\partial K_i\setminus\partial\Omega^+)}\right).\end{aligned} \tag{3.123}$$

We have

$$\begin{aligned}&-\int_{\partial K_i^-\cap\partial\Omega}(\boldsymbol{v}\cdot\boldsymbol{n})\, e\hat{\xi}\,\mathrm{d}S - \int_{\partial K_i^-\setminus\partial\Omega}(\boldsymbol{v}\cdot\boldsymbol{n})[e]\hat{\xi}\,\mathrm{d}S \\ &\le C_{\boldsymbol{v}}\|\rho_{B0}\|_{L^2(\partial K_i\setminus\partial\Omega^+)}\|\hat{\xi}\|_{L^2(\partial K_i)}.\end{aligned} \tag{3.124}$$

By Lemma 2 and Lemma 3 we have

$$\|\hat{\xi}\|_{L^2(\partial K_i)} \leq \sqrt{C_M(1+C_I)}\, h_{K_i}^{-1/2} \|\hat{\xi}\|_{L^2(K_i)} \tag{3.125}$$

and thus

$$T_{10} \leq c_{10} \left(\sum_{i \in I} h_{K_i}^{-1} \|\hat{\xi}\|^2_{L^2(K_i)} + \|\rho_{B0}\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)} \right). \tag{3.126}$$

Finally,

$$T_{11} \leq c_{11} \left(\sum_{i \in I} \varepsilon h_{K_i}^{-1} \|\rho_{B0}\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)} \right). \tag{3.127}$$

Noticing that (from the linearity of operator $\mathcal{C}$, see Lemma 6)

$$\frac{\partial \mathcal{C}(u_h, u_D)}{\partial t} = \mathcal{C}(\frac{\partial u_h}{\partial t}, \frac{\partial u_D}{\partial t}) = \frac{\partial \hat{u_h}}{\partial t},$$

using Lemma 6, we can estimate

$$\sum_{i \in I} h_{K_i}^{-2} \|\hat{\xi}\|^2_{L^2(K_i)} \leq k_1 \sum_{i \in I} h_{K_i}^{-1} \|\rho_{B0}\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)}, \tag{3.128}$$

$$\sum_{i \in I} h_{K_i}^{-1} \|\hat{\xi}\|^2_{L^2(K_i)} \leq k_2 \sum_{i \in I} \|\rho_{B0}\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)}, \tag{3.129}$$

$$\|\hat{\xi}\|^2_{L^2(\Omega)} \leq k_3 \sum_{i \in I} h_{K_i} \|\rho_{B0}\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)}, \tag{3.130}$$

$$\|\frac{\partial \hat{\xi}}{\partial t}\|^2_{L^2(\Omega)} \leq k_4 \sum_{i \in I} h_{K_i} \|\frac{\partial \rho_{B0}}{\partial t}\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)}. \tag{3.131}$$

Substituting these estimates into (3.118), (3.121), (3.123) and (3.126), we get

$$\begin{aligned} \int_0^\tau T_6 \,\mathrm{d}t \leq \hat{c}_6 \Bigg(& \alpha_6 \left(\frac{1}{2} \|e(\tau)\|^2_{L^2(\Omega)} + \gamma_0 \int_0^\tau \|e\|^2_{L^2(\Omega)} \,\mathrm{d}t \right) \\ & + \frac{1}{\alpha_6} \Bigg(2 \sum_{i \in I} h_{K_i} \|\rho_{B0}(\tau)\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)} \\ & + \frac{1}{\gamma_0} \int_0^\tau \sum_{i \in I} h_{K_i} \|\frac{\partial \rho_{B0}}{\partial t}\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)} \,\mathrm{d}t \Bigg) \Bigg), \end{aligned} \tag{3.132}$$

$$\begin{aligned} T_7 \leq \hat{c}_7 \Bigg(& \alpha_7 \left(\gamma_0 \|e\|^2_{L^2(\Omega)} + \varepsilon |e|^2_{H^1(\Omega; \mathcal{T}_h)} \right) \\ & + \frac{1}{\alpha_7} \sum_{i \in I} (\varepsilon h_{K_i}^{-1} + \varepsilon^{-1} h_{K_i} + \gamma_0^{-1} h_{K_i}) \|\rho_{B0}\|^2_{L^2(\partial K_i \setminus \partial\Omega^+)} \Bigg), \end{aligned} \tag{3.133}$$

$$T_9 \le \hat{c}_9 \sum_{i\in I} \varepsilon\, h_{K_i}^{-1} \|\rho_{B0}\|^2_{L^2(\partial K_i\setminus\partial\Omega^+)}, \tag{3.134}$$

$$T_{10} \le \hat{c}_{10} \sum_{i\in I} \|\rho_{B0}\|^2_{L^2(\partial K_i\setminus\partial\Omega^+)}. \tag{3.135}$$

Plugging together (3.89)-(3.93) and (3.102)-(3.121), we see that

$$\begin{aligned}
&\frac{1}{2}\|e(\tau)\|^2_{L^2(\Omega)} + \int_0^\tau (\varepsilon|e|^2_{H^1(\Omega;\mathcal{T}_h)} + \gamma_0\|e\|^2_{L^2(\Omega)})\,\mathrm{d}t \le \int_0^\tau \sum_{j=1}^{11} T_j\,\mathrm{d}t \\
&\le c\left[\alpha\left(\frac{1}{2}\|e(\tau)\|^2_{L^2(\Omega)} + \int_0^\tau (\varepsilon|e|^2_{H^1(\Omega;\mathcal{T}_h)} + \gamma_0\|e\|^2_{L^2(\Omega)})\,\mathrm{d}t\right)\right. \\
&+\frac{1}{\alpha}\int_0^\tau \sum_{i\in I}\left(\varepsilon^{-1}h^2_{K_i}\|\rho_I\|^2_{L^2(K_i)} + \gamma_0^{-1}h_{K_i}\Big\|\frac{\partial\rho_{B0}}{\partial t}\Big\|^2_{L^2(\partial K_i\setminus\partial\Omega^+)}\right. \\
&+(\varepsilon h_{K_i}^{-1} + \varepsilon^{-1}h_{K_i} + \gamma_0^{-1}h_{K_i} + 1)\|\rho_{B0}\|^2_{L^2(\partial K_i\setminus\partial\Omega^+)} \\
&\left.\left.+\varepsilon h_{K_i}\|\rho_{B1}\|^2_{L^2(\partial K\setminus\partial\Omega^-)}\right)\,\mathrm{d}t + \frac{1}{\alpha}\sum_{i\in I}\gamma_0^{-1}h_{K_i}\|\rho_{B0}(\tau)\|^2_{L^2(\partial K\setminus\partial\Omega^+)}\right].
\end{aligned} \tag{3.136}$$

The a posteriori error estimate immediatelly follows from this by choosing α sufficiently small and taking the maximum of both sides over $\tau \in (0,T)$.

4 Space-Time DGFEM

In the previous section, we described a space semidiscretization of a convection-diffusion problem by DGFEM, resulting in a system of ordinary differential equations. This approach, called the *method of lines*, is straightforward and efficient, because the resulting system can be handled using sophisticated ODE solvers like `LSODE`. On the other hand, it does not easily allow changing the mesh or the domain. Unlike classical FEM, DGFEM can be used to fully discretize the problem in the entire space-time domain. This can be done in two manners:

1. Time-uniform: Space-time elements have the form $K_i \times (t_j, t_{j+1})$ where $K_i \in \mathcal{T}_h$ form a space triangulation and $0 = t_0 < t_1 < \ldots < t_N = T$.

2. Nonuniform: Space-time elements are general polygons in $d+1$-dimensional space.

The first method suffices to allow changing the mesh through time, whereas the latter also allows changing the domain.

4.1 Time discretization

We discretize the problem (3.28)-(3.31) also in time using the discontinuous Galerkin method. For this purpose, we consider a partition $0 = t_0 < t_1 < ... < t_M = T$ of the time interval $[0, T]$ and denote $I_m = (t_{m-1}, t_m), \bar{I}_m = [t_{m-1}, t_m], \tau_m = t_m - t_{m-1}, m = 1, ..., M$. We have

$$[0, T] = \bigcup_{i=1}^{M} \bar{I}_m, \qquad I_m \cap I_n = \emptyset \text{ for } m \neq n. \tag{4.137}$$

On each interval, we define

$$S_{h,m} = \{\varphi \in L^2(\Omega); \varphi|_K \in P^p(K) \quad \forall K \in \mathcal{T}_{h,m}\}. \tag{4.138}$$

For a function φ defined on $\cup_{i=1}^{M}(t_{m-1}, t_m)$ we introduce the following notation:

$$\varphi_m^{\pm} = \varphi(t_m \pm) = \lim_{t \to t_m \pm} \varphi(t) \tag{4.139}$$

$$\{\varphi\}_m = \varphi_m^+ - \varphi_m^-. \tag{4.140}$$

For each time interval $I_m, m = 1, ..., M$, we shall consider, in general, a different triangulation $\mathcal{T}_{h,m} = \{K_i\}_{i \in i_{h,m}}$ of the domain Ω. For each $m = 1, \ldots, M$ we define the forms on $S_{h,m}$ with the use of the forms (3.47)-(3.52):

$$A_{h,m}(u, \varphi) = a_h(u, \varphi) + b_h(u, \varphi) + c_h(u, \varphi) + \varepsilon J_h^{\sigma}(u, \varphi), \tag{4.141}$$

$$l_{h,m}(u, \varphi) = l_h(u, \varphi), \tag{4.142}$$

where in the definition of the forms $a_h, b_h, c_h, J_h^{\sigma}, l_h$ we set $\mathcal{T}_h = \mathcal{T}_{h,m}$. Let $q \geq 0$ be an integer. We define approximate solution as a function

$$U(x, t) \in S_{h,\tau} = \left\{ \varphi \in L^2(Q_T); \varphi|_{I_m} = \sum_{i=0}^{q} t^i \varphi_i, \ \varphi_i \in S_{h,m}, \ m = 1, \ldots, M \right\}, \tag{4.143}$$

satisfying

$$\begin{aligned} &\sum_{m=1}^{M} \int_{I_m} ((U', \varphi) + A_{h,m}(U, \varphi)) \, \mathrm{d}t + \sum_{m=2}^{M} (\{U\}_{m-1}, \varphi_{m-1}^+) + (U_0^+, \varphi_0^+) \\ &= \sum_{m=1}^{M} \int_{I_m} l_{h,m}(\varphi) \, \mathrm{d}t + (u_0, \varphi_0^+) \quad \forall \varphi \in S_{h,\tau}. \end{aligned} \tag{4.144}$$

If we denote

$$\begin{aligned} B(u, v) &= \sum_{m=1}^{M} \int_{I_m} ((u', v) + A_{h,m}(u, v)) \, dt + \sum_{m=2}^{M} (\{u\}_{m-1}, v_{m-1}^+) + (u_0^+, v_0^+), \\ L(v) &= \sum_{m=1}^{M} \int_{I_m} l_{h,m}(v) \, \mathrm{d}t + (u_0, v_0^+), \end{aligned} \tag{4.145}$$

we can write (4.144) in the form

$$B(U,\varphi) = L(\varphi) \qquad \forall \varphi \in S_{h,\tau}. \tag{4.146}$$

We shall assume *shape regularity*, namely that (3.36) holds with $\mathcal{T}_h = \mathcal{T}_{h,m}$ for all $m = 1,\dots,M$ and the constant $C_{\mathcal{T}}$ is independent of τ. Moreover, we shall assume that there exist constants $C_S, \hat{C}_S$ such that

$$\frac{1}{\hat{C}_S} h_K \le \tau_m \le C_S h_K, \qquad K \in \mathcal{T}_h, \quad m = 1,\dots,M. \tag{4.147}$$

We shall also assume that $u \in \mathcal{H}$, where

$$\mathcal{H} = H^{q+1}(0,T;H^1(\Omega)) \cap C(0,T;H^{p+1}(\Omega)). \tag{4.148}$$

Under these assumptions one can prove the following result:

Theorem 3 *Let u be the exact solution of problem* (3.28) – (3.31) *satisfying the condition $u \in \mathcal{H}$, and let U denote the approximate solution obtained with the aid of method (4.144). Then there exists a constant C independent of h, τ and ε such that the error $e = U - u$ satisfies the estimate*

$$\sum_{m=1}^{M} \int_{I_m} (\|e\|^2_{L^2(\Omega)} + \varepsilon |e|^2_{H^1(\Omega,\mathcal{T}_{h,m})})\,\mathrm{d}t \le Ch^{2p}\{|u|^2_{L^2(0,T;H^{p+1}(\Omega))} + |u|^2_{C([0,T];H^{p+1}(\Omega))}\}$$
$$+ C\tau^{2q}\{|u|^2_{H^{q+1}(0,T;L^2(\Omega))} + |u|^2_{H^{q+1}(0,T;H^1(\Omega))}\}. \tag{4.149}$$

This estimate is also valid for $\varepsilon = 0$, i. e. in the hyperbolic case.

5 Numerical experiments

5.1 Implementation issues

Computer programs were created for each of these methods (combined FV-FE, DGFE method of lines and space-time DGFEM). The first two were programmed in Fortran, for the last one, we used the `FreeFEM++` environment [32]. All code examples in this chapter are also in Fortran.

5.1.1 The lumping operator of the combined FE-FV method

In section 2 we described the combined FE-FV method on triangular meshes. Although the discrete problem formulation (2.22) is itself not complicated, the evaluation of the lumping operator (2.24) is not easy. As meshes do not change frequently, it is, of course, best to assemble a sparse matrix corresponding to this linear operator. Two problems arise:

1. For a given $D_j \in \mathcal{D}_h$, determine all $K_i \in \mathcal{T}_h$ such that $D_j \cap K_i \neq \emptyset$.

2. Given a basis function λ on K_i, evaluate

$$\frac{1}{|D_j|} \int_{K_i \cap D_j} \lambda(x) \, \mathrm{d}x. \tag{5.150}$$

The naive way to solve 1 (testing all pairs of triangles) would yield a quadratic complexity algorithm, too slow for practical problems. If the meshes $\mathcal{D}_h$ and $\mathcal{T}_h$ are completely independent, special data structures, such as *kd-trees*, allowing a fast "proximity" search, need to be involved. If the FV mesh is derived from the FE mesh in a suitable way, we would usually have this "proximity information" in advance, i.e. for each $j \in J$ we can efficiently give a subset $N_j \subset I$ such that

$$D_j \subset \bigcup_{i \in N_j} K_i.$$

Still, the latter of the above problems may be a difficult task, because two triangles may overlap in many different ways (given that the vertex order matters in a computer program). Although sophisticated libraries exist for polygon intersections, it turns out that the combined FE-FV method is not very sensitive to the lumping operator and an approximate evaluation suffices. Here, the following strategy is suggested:

First, evaluate (5.150) approximately by a quasi-Monte Carlo approach. This means that we uniformly distribute N points $p_1, p_2, \ldots, p_N$ in K_i, then determine the set

$$L = \{l;\ p_l \text{ is inside } D_j\}.$$

This can be easily done by a single matrix-vector multiplication: If (x_i, y_i), $i = 1, 2, 3$, are the coordinates of vertices of D_j, we write

$$M_\Delta = \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix}$$

A point (x, y) is inside D_j if and only if

$$(x, y, 1) M_\Delta^{-1} > 0.$$

(here we mean that all three components of the resulting row vector must be positive). The approximation we use is

$$\frac{1}{|D_j|} \int_{K_i \cap D_j} \lambda(x) \, \mathrm{d}x \approx \frac{|K_i|}{|D_j|} \frac{\sum_{l \in L} \lambda(p_l)}{N}. \tag{5.151}$$

We then enforce the conservativity of the lumping operator by scaling the rows of the resulting sparse matrix so that each row sums up to 1 (i.e. divide each row by its sum). By conservativity we mean the following property:

$$\int_\Omega L_h v = \int_\Omega v \quad \forall v \in X_h. \tag{5.152}$$

The implementation of the quasi-Monte Carlo is straightforward and simple, and if a small enough N turns out to suffice, it may be even more efficient than an exact evaluation.

5.1.2 Computational efficiency of DGFEM vs. FEM

If we directly compare the continuous piecewise linear FEM to piecewise linear DGFEM on the same mesh, the DGFEM gives roughly three times as much degrees of freedom in the discretization. From this point of view, DGFEM may seem computationally inferior. Nevertheless, DGFEM has also computational advantages over the FEM, although these are less obvious. Roughly speaking, the greater "locality" of the DGFEM results not only in more degrees of freedom (DOFs), but also in neater equation structure. We explain this in more detail below:

The "heart" of most PDE solvers are sparse matrices and their manipulation, especially multiplying a vector by a sparse matrix. The sparsity structure of the resulting matrices is given by the topology of the mesh. Since the basis functions have localized support, one can determine a common sparsity structure which every sparse matrix resulting from discretization of a bilinear form must fit into. In particular, for continuous finite elements, a matrix element a_{ij} can ever be nonzero only if the supports of i-th and j-th basis functions overlap - this can only happen, if they belong to triangles that share a vertex of the mesh. If we shall operate with more bilinear forms (and, thus, sparse matrices) on the same mesh, which is often needed, it is advantageous to exploit this common sparsity in some way. Most common practice for continuous finite elements is to use a storage scheme for general sparse matrices. One of the most popular ones is the Compressed Sparse Row (CSR) format [40]. This format is usually represented by the dimensions `m,n` of the matrix, number of its nonzero elements `nnz` and three arrays

```
real,dimension(nnz):: a
integer,dimension(nnz):: ja
integer,dimension(n+1):: ip
```

where all the nonzero elements are packed row-wise consecutively in the array `a`. Further, `ja` are the corresponding column indices, and `ip` are pointers to these arrays such that the `i`-th row is given by indices `ip(i):ip(i+1)-1`. For several matrices on a mesh, a common sparsity structure can be chosen so that the arrays `ja` and `ip` are shared amongst matrices, to save space and allow efficient adding and subtracting of the matrices (reduces to vector operations). A simple matrix-vector multiplication $y = Ax$ for the CSR storage is performed like this:

```
integer:: i,pl,pu
do i=1,m
  pl = ia(i)
  pu = ia(i+1)-1
  y(i) = dot_product(a(pl:pu),x(ja(pl:pu)))
end do
```

Although the computation is straightforward and parallelizable (the cycles are completely independent), it often executes slowly on modern high-performance processors. The main trouble here is caused by the non-local indexing `x(ja(pu:pl))`, because the index sequence in `ja` may have many jumps and thus fairly often cause "cache mishits", making the processor wait for memory traffic instead of doing useful computations. Although improved formats have been suggested for general sparse matrices to address this issue, such as the jagged diagonal (JAD) format [40], we shall see that this can be naturally overcome in DGFEM exploiting a *block structure* of the arising sparse matrices.

Suppose we have `ndf` degrees of freedom per triangle, there are `nel` triangles in total, and we have an array

```
integer,dimension(3,nel):: eln
```

such that `eln(:,i)` gives the indices of neighbours of the `i`-th element, i.e. those triangles with a common face (zero or negative value indicates a boundary face). If we construct basis functions

$$\phi_{ik}; \quad i = 1 \ldots \texttt{ndf}, k = 1 \ldots \texttt{nel},$$

we see that in for a DGFE integral form a_h involving volume and face integrals element $a_h(\phi_{ik}, \phi_{jl})$ can be nonzero only if triangles K_k and K_l share a common face. A general DGFE square matrix `A` (corresponding to a bilinear form a_h) and vector `x` (either corresponding to a DGFE function or a linear form) can thus be represented as

```
real,dimension(ndf,ndf,0:3,nel):: A
real,dimension(ndf,nel):: x
```

where `A(i,j,0,K)` $= a_h(\phi_{iK}, \phi_{jK})$ corresponds `i`-th and `j`-th DOF of the `K`-th element and `A(i,j,q,k)` $= a_h(\phi_{iK}, \phi_{jL})$ corresponds to `i`-th DOF of `K`-th element and `j`-th DOF of `L`-th element, with

$$L = \texttt{eln(q,K)}.$$

The matrix-vector multiplication then proceeds as follows:

```
integer:: K,L,q
do K=1,n
  y(:,K) = matmul(A(:,:,0,K),x(:,K))
  do q=1,3
```

```
    L = eln(q,K)
    if (L > 0) y(:,K) = y(:,K) + matmul(A(:,:,q,K),x(:,L))
  end do
end do
```

The non-local indexing still exists, because the index `L` is being read from memory, but now it addresses blocks instead of single elements. This means that unlike the simple CSR case, where one multiplication is performed per one non-local memory reference (and thus potential cache mishit), now one *block* multiplication and thus `ndf`^{2} mutliplications are performed per reference, which often improves the speed of the whole algorithm by a factor of ten or more on modern high-performance processors. Since individual cycles of the outermost loop are completely independent, the algorithm is still well parallelizable. A similar success is not achievable for continuous finite elements, because then the DOFs cannot be partitioned by element.

Another great advantage of the DGFEM arises for evolution (i.e., time-dependent) problems solved by explicit time-stepping methods such as the Euler method, because the mass matrix (i.e. the matrix corresponding to L^2-scalar product form) turns out to be block-diagonal and can be easily inverted. On the contrary, for continuous finite elements, one must either employ sparse direct solvers (e.g. MUMPS) or use "mass lumping" (and degrade the order of the method).

5.1.3 Time discretization

As we have already seen, the transfer from DGFE method of lines to space-time DGFEM is simple and straightforward. However, (4.144) might give us the feeling that we need to solve a very large system of equations, for all the unknowns in all time levels simultaneously. Fortunately, this is not the case. If we choose φ such that $\varphi|_{I_m} = 0$ for $m \neq \bar{m}$, then for $\bar{m} > 1$, (4.144) becomes

$$\int_{I_{\bar{m}}} ((U', \varphi) + A_{h,\bar{m}}(U, \varphi)) \, \mathrm{d}t + (U^+_{\hat{m}-1}, \varphi^+_{\hat{m}-1}) = \int_{I_{\bar{m}}} l_{h,\bar{m}}(\varphi) \, \mathrm{d}t + (U^-_{\hat{m}-1}, \varphi^+_{\hat{m}}) \quad \forall \varphi \in S_{h,\bar{m}} \tag{5.153}$$

and for $\bar{m} = 1$

$$\int_{I_1} ((U', \varphi) + A_{h,1}(U, \varphi)) \, \mathrm{d}t + (U^+_0, \varphi^+_0) = \int_{I_0} l_{h,1}(\varphi) \, \mathrm{d}t + (u_0, \varphi^+_0) \quad \forall \varphi \in S_{h,\bar{m}}. \tag{5.154}$$

These equations enable us to determine separately $U|I_{\hat{m}}$ from $U|I_{\hat{m}-1}$ and $U|I_1$ from the initial condition u_0. The implementation thus reduces to repeatedly solving the following problem:

Given two meshes $\mathcal{T}_h, \mathcal{T}_h^-$ and $u^- \in S_h^-$, determine $u_0, \ldots, u_q \in S_h$ such that

$$U = \sum_{i=0}^{q} t^i u_i$$

satisfies

$$\int_0^\tau ((U', \varphi) + A_h(U, \varphi))\ \mathrm{d}t + (u_0, \varphi_0) = \int_0^\tau l_h(\varphi)\,\mathrm{d}t + (u^-, \varphi_0) \tag{5.155}$$

for any

$$\varphi = \sum_{i=0}^{q} t^i \varphi_i, \quad \varphi_i \in S_h.$$

Given a suitable set of basis functions ϕ_k for S_h, we can write

$$u_i = \sum_k w_i^k \phi_k, \quad \varphi_i = \sum_k y_i^k \phi_k.$$

Then (5.155) transforms into

$$\int_0^\tau \sum_{i=0}^{q}\sum_{j=0}^{q} (i t^{i-1+j} \mathrm{y}_j^T \mathrm{M} \mathrm{w}_i + t^{i+j} \mathrm{y}_j^T \mathrm{B} \mathrm{w}_i)\,\mathrm{d}t + \mathrm{y}_0^T \mathrm{M} \mathrm{w}_0 = \int_0^\tau \sum_{j=0}^{q} t^j \mathrm{y}_j^T \mathrm{f}(t)\,\mathrm{d}t + \mathrm{y}_0^T \mathrm{d}, \tag{5.156}$$

where is w_i is the vector of w_i^k, y_i the vector of y_i^k, M is the mass matrix,

$$M_{kl} = (\phi_l, \phi_k),$$

B is the matrix of the form A_h,

$$B_{kl} = A_h(\phi_l, \phi_k),$$

$\mathrm{f}(t)$ is the vector corresponding to the linear form l_h at time t, i.e.

$$f^l(t) = l_h(\phi_l)(t),$$

and

$$d^l = (u^-, \phi_l). \tag{5.157}$$

Identity (5.156) should hold for any choice of y_i vectors, thus we can eliminate them to get

$$\int_0^\tau \sum_{i=0}^{q} (i t^{i-1} \mathrm{M} \mathrm{w}_i + t^i \mathrm{B} \mathrm{w}_i)\,\mathrm{d}t + \mathrm{M} \mathrm{w}_0 = \int_0^\tau \mathrm{f}(t)\,\mathrm{d}t + \mathrm{d} \tag{5.158}$$

and

$$\int_0^\tau \sum_{i=0}^{q} (i t^{j+i-1} \mathrm{M} \mathrm{w}_i + t^i \mathrm{B} \mathrm{w}_i)\,\mathrm{d}t = \int_0^\tau t^j \mathrm{f}(t)\,\mathrm{d}t \tag{5.159}$$

for $j = 1, \ldots, q$. By integrating the left-hand sides, we get

$$\sum_{i=1}^{q} \tau^i \mathrm{M} \mathrm{w}_i + \sum_{i=0}^{q} \frac{\tau^{i+1}}{i+1} \mathrm{B} \mathrm{w}_i + \mathrm{M} \mathrm{w}_0 = \int_0^\tau \mathrm{f}(t)\,\mathrm{d}t + \mathrm{d} \tag{5.160}$$

and

$$\sum_{i=1}^{q} \frac{i\tau^{j+i}}{j+i}\mathrm{Mw}_i + \frac{\tau^{i+j+1}}{i+j+1}\mathrm{Bw}_i = \int_0^\tau t^j \mathrm{f}(t)\,\mathrm{d}t \tag{5.161}$$

for $j = 1, \ldots, q$. This is a square linear system of $(q+1)N$ equations ($N = \dim S_h$), which can be solved, e.g.,by an iterative linear solver. If a direct solver has to be used, however, there is a better strategy - treat the matrix as a $(q+1) \times (q+1)$ matrix of $N \times N$ blocks, and perform *block gaussian elimination* on this matrix first. This is especially advantageous if banded structure is exploited for the blocks.

As we have already mentioned, the solution proceeds by advancing over the intervals $I_1, I_2, \ldots, I_M$ and repeatedly solving (5.155) with $\tau = t_m - t_{m-1}$, $l_h(t) = l_{h,m}(t + t_{m-1})$ and $u^- = U^-_{m-1}$, i.e. we "shift" the problem from interval (t_{m-1}, t_m) to $(0, \tau)$ and take the new initial condition from the previous step. If the mesh did not change, i.e. $\mathcal{T}_h = \mathcal{T}_h^-$, we can see from (5.157) that

$$\mathrm{d} = \mathrm{Mw}_-,$$

where

$$u^- = \sum_k w_-^k \phi_k$$

and w_-^k are the elements of the vector w_-. If the mesh *did* change, $u^- \in \mathcal{T}_h^- \neq \mathcal{T}_h$, we can substitute the L^2-projection of u^- onto S^h, $\Pi_{S^h} u^-$, into (5.155) to get

$$\int_0^\tau \left((U', \varphi) + A_h(U, \varphi)\right)\,\mathrm{d}t + (u_0, \varphi_0) = \int_0^\tau l_h(\varphi)\,\mathrm{d}t + (\Pi_{S_h} u^-, \varphi_0) \tag{5.162}$$

Thus, we can roughly say that *switching the mesh means L^2-projecting the current state onto the new mesh and continuing integration with the new mesh* (of course, we need to update the sparse matrices). This might suggest a similar strategy what to do when changing the mesh for other time-stepping methods: project the necessary data from the old mesh to the new mesh via L^2-projection and continue. However, such an approach is purely heuristic here.

5.2 Numerical results

5.2.1 Combined FE-FV method

We verified the estimates (2.26), (2.27) by numerical experiments. We applied the combined FV-FE method to the *scalar 2D viscous Burgers equation*

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x_1} + u\frac{\partial u}{\partial x_2} - \varepsilon\Delta u = g \tag{5.163}$$

with $\varepsilon = 0.1$ in the space-time domain $\mathcal{Q}_T = \Omega \times (0, 1)$, $\Omega = (-1, 1)^2$, equipped with Dirichlet boundary condition $u|_{\partial\Omega} = 0$, and initial condition $u|_{t=0} = 0$. The right-hand

side g is chosen so that it conforms to the exact solution

$$u_{\text{ex}} = (1 - e^{-2t})(1 - x_1^2)^2(1 - x_2^2)^2.$$

The time discretization is carried by a semiimplicit Euler scheme:

$$\left(\frac{u_h^k - u_h^{k-1}}{\tau}, v_h\right) + b_h(u_h^{k-1}, v_h) + a_h(u_h^k, v_h) = (g^{k-1}, v_h), \tag{5.164}$$

which should have better stability properties than a purely explicit scheme with no added computational cost, because the FE mass and stiffness matrices share their sparsity structure. As we want to examine the error of the space discretization, we *overkill* the time step so that the time discretization error is negligible. The numerical flux we use is given by the formula

$$H(u, v, \boldsymbol{n}) = \begin{cases} u^2(n_1 + n_2)/2 & \text{if } (u + v)(n_1 + n_2) > 0 \\ v^2(n_1 + n_2)/2 & \text{if } (u + v)(n_1 + n_2) < 0 \end{cases}. \tag{5.165}$$

In each computation we consider the FE mesh primary and derive the FV mesh from it. We successively refine the FE mesh and for each refinement we evaluate the so-called *experimental order of convergence* (EOC^1,EOC^2) defined as follows:

$$\text{EOC}^1 = \frac{\log e^1_{h'} - \log e^1_h}{\log h' - \log h} \tag{5.166}$$

$$\text{EOC}^2 = \frac{\log e^2_{h'} - \log e^2_h}{\log h' - \log h} \tag{5.167}$$

where h' refers to the refined FE mesh and h to the original one. e^1_h stands for the $L^\infty(L^2)$ error (see (2.26)), e^2_h for the $L^2(H^1)$ error from (2.27). We also consider two different methods of deriving the secondary FV mesh: The first method (Method 1 in the first table) consists in simply copying the FE mesh, in the second method (Method 2 in the second table) we create an interior FV node as a center of each FE triangle, add the FE boundary nodes and triangulate these nodes by means of Delaunay triangulation. Two of these secondary FV meshes are shown in Figures 7,7. The approximate construction of the lumping operator described in subsection 5.1.1 is used in the second case.

The results are given in Appendix in Tables 1, 2.

5.2.2 DGFE method of lines

In this subsection we present some numerical experiments concerning the derived a posteriori error estimates for the DGFE method of lines. We solve the equation (3.28) (convection-diffusion) with $\Omega = (0, 1)^2$, $T = 0.8$, $\boldsymbol{v} = (1, 1)^T$, $c = 0.2$ and $\varepsilon = 0.005$. The boundary and initial conditions were chosen in such a way that they conform to the exact solution

$$u_{\text{ex}} = \frac{x + y}{2} + (1 - e^{-2t})(1 - ye^{v_1(x-1)/\nu})(1 - xe^{v_2(y-1)/\nu}),$$

where ν is a parameter determining the steepness of the boundary layer in the solution. We evaluate the true squared error

$$\text{err} = \|e\|^2_{L^\infty(0,T;L^2(\Omega))} + \varepsilon\,|e|^2_{L^2(0,T;H^1(\Omega))} \tag{5.168}$$

over the interval $T \in (0, T_{\max})$. We also evaluate the a posteriori error estimate

$$\text{est} = E_1 + E_2 + E_4 + E_4, \tag{5.169}$$

where

$$E_1 = \sum_{i\in I} \varepsilon^{-1} h^2_{K_i} \|\rho_I\|^2_{L^2(0,T;L^2(K_i))}, \tag{5.170}$$

$$E_2 = \sum_{i\in I} \gamma_0^{-1} h_{K_i} \big\|\frac{\partial \rho_{B0}}{\partial t}\big\|^2_{L^2(0,T;L^2(\partial K_i\setminus\partial\Omega^+))}, \tag{5.171}$$

$$E_3 = \sum_{i\in I} (\varepsilon h_{K_i}^{-1} + \varepsilon^{-1} h_{K_i} + \gamma_0^{-1} h_{K_i} + 1)\|\rho_{B0}\|^2_{L^2(0,T;L^2(\partial K_i\setminus\partial\Omega^+))}, \tag{5.172}$$

$$E_4 = \sum_{i\in I} \varepsilon h_{K_i} \|\rho_{B1}\|^2_{L^\infty(0,T;L^2(\partial K_i\setminus\partial\Omega^-))}. \tag{5.173}$$

$$\tag{5.174}$$

The last term from (3.86) is omitted for simplicity, because

$$\|\rho_{B0}\|_{L^\infty(0,T;L^2(\partial K_i\setminus\partial\Omega^+))} \le k\big\|\frac{\partial \rho_{B0}}{\partial t}\big\|_{L^2(0,T;L^2(\partial K_i\setminus\partial\Omega^+))} \tag{5.175}$$

with a global constant k, and thus it can be included in E_2. We define the *effectivity index* as

$$\text{EI} = \sqrt{\frac{\text{est}}{\text{err}}}.$$

The complete results (program output for different meshes) are given in Tables 3-9. Test starts with a coarse uniform mesh (Table 3). This mesh was subsequently refined uniformly (Tables 4-6) and (independently) adaptively (Tables 7-9) via the method of anisotropic mesh adaptation [14]. The computation proceeded by forward Euler method with time step τ. The error and a posteriori estimate were evaluated every N_{err} steps and the integrals dumped every N_{dump} steps. We also tracked the relative contributions of E_i ((5.170)-(5.173)) to `est`, given in percents in the columns `E1-E4`.

5.2.3 The Space-Time DGFEM

In this section we present some numerical experiments with the space-time DGFE method described and analyzed in previous sections. We solve equation (3.28) in $Q_T = (0,1)^2 \times (0,1)$ with $v_1 = v_2 = 1,\ c = 0.5$ and two choices of ε: $\varepsilon = 0.005$ (parabolic case) and $\varepsilon = 0$

(hyperbolic case). The right-hand side g, boundary and initial conditions are chosen in such way that they conform to the exact solution

$$u_{ex}(x_1, x_2, t) = (1 - e^{-t}) \left(2x + 2y - xy + 2(1 - e^{v_1(x_1-1)/\nu})(1 - e^{v_2(x_2-1)/\nu})\right),$$

where $\nu = 0.05$ is a constant determining the steepness of the boundary layer in the exact solution. The problem is solved on a sequence of non-nested nonuniform space meshes $\mathcal{T}_{h_1}$, $\mathcal{T}_{h_2}, \ldots$, kept unchanged on all time levels. We inspect the experimental order of convergence (EOC) with respect to τ and h, which are varied simultaneously due to condition (4.147). For successive pairs (τ, h) and (τ', h') we evaluate the experimental order of convergence in space and time defined as

$$\mathrm{EOC}_{\mathrm{space}} = \frac{\log\left(\|e_{\tau' h'}\|_{L^2(Q_T)}\right) - \log\left(\|e_{\tau h}\|_{L^2(Q_T)}\right)}{\log h' - \log h},$$

$$\mathrm{EOC}_{\mathrm{time}} = \frac{\log\left(\|e_{\tau' h'}\|_{L^2(Q_T)}\right) - \log\left(\|e_{\tau h}\|_{L^2(Q_T)}\right)}{\log \tau' - \log \tau},$$

where $e_{\tau h} = u_{ex} - U$ is the error of the method, when the exact solution u_{ex} is approximated by the DG approximate solution U computed with the aid of a space triangulation of size h and a time interval partition of size τ. Moreover, we compute the global experimental order of convergence with the aid of additional data sets with halved time step and fitting a general nonlinear model of the form

$$\|e_{\tau h}\|_{L^2(Q_T)} \approx C_1\, h^r + C_2\, \tau^s$$

through the data via the method of nolinear least squares, using the `MINPACK` package [44]. The results are shown in Tables 10 – 13 in Appendix.

The space-time DGFE computations were carried out with the aid of the `FreeFEM++` modelling environment from [32], which was adapted to the DGFE space-time discretization. The time integrals were evaluated by quadrature formulae exact for polynomials of degree 5 and 9 in the case of elements linear in time and quadratic in time, respectively. The quadrature formulae used for the integration over triangles and their sides were exact for polynomials of degree 5 both for linear and quadratic elements. The nonsymmetric linear problem was solved in each time step by the multifrontal direct solver `UMFPACK` ([13]).

6 Conclusion

We have studied approaches for solving the convection-diffusion equation, especially the DGFE method of lines, for which we developed a posteriori error-estimate improving the one given in [41]. We also discussed some practical implementation aspects. While the DGFE method itself holds much promise and is definitely worth further studying,

practical uses of explicit a posteriori error estimates for DGFEM are questionable due to the presence of unknown constants, like the constant C in (3.86). Note that the proofs of all estimates are "constructive" and potentially allow C to be explicitly evaluated. Nevertheless, given that the estimates are crude in a number of places, the resulting value would probably be practically useless. Without the possibility to guess C better (e.g., from a more accurate error approximation), explicit a posteriori error estimates are often used as refinement indicators. Specifically, the error contribution on a given element is considered proportional to the contribution to the a posteriori estimate of that element (the bracketed expression in (3.86)). Again, without further analysis there is no theoretical justification for this assumption. Moreover, there are many other quite successful techniques for mesh adaptation.

7 Appendix

$\#I$	h	e_h^1	EOC^1	e_h^2	EOC^2
128	3.54E-01	6.57E-02	-	1.09E-01	-
512	1.77E-01	2.95E-02	1.16	5.58E-02	0.97
2048	8.84E-02	1.40E-02	1.08	2.81E-02	0.99
8192	4.42E-02	6.87E-03	1.03	1.41E-02	0.99
32768	2.21E-02	3.40E-03	1.02	7.05E-03	1.00
131072	1.11E-02	1.69E-03	1.01	3.53E-03	1.00
Average			1.06		0.99

Table 1: Method 1

$\#I$	h	e_h^1	EOC^1	e_h^2	EOC^2
128	3.54E-01	7.50E-02	-	1.13E-01	-
512	1.77E-01	4.57E-02	0.71	6.18E-02	0.87
2048	8.84E-02	1.78E-02	1.36	3.01E-02	1.04
8192	4.42E-02	1.18E-02	0.59	1.62E-02	0.89
32768	2.21E-02	4.37E-03	1.43	7.56E-03	1.10
131072	1.11E-02	2.99E-03	0.55	4.12E-03	0.88
Average			0.93		0.96

Table 2: Method 2

T	err	est	EI	E1	E2	E3	E4
0.040	2.045E-10	1.129E-07	23.49	62.0	24.2	0.8	13.0
0.080	1.598E-09	7.232E-07	21.28	75.9	7.1	1.0	16.0
0.120	5.019E-09	2.185E-06	20.86	79.0	3.3	1.1	16.6
0.160	1.137E-08	4.881E-06	20.72	80.2	1.8	1.1	16.9
0.200	2.084E-08	8.889E-06	20.65	80.8	1.2	1.1	17.0
0.240	3.431E-08	1.459E-05	20.62	81.1	0.8	1.1	17.0
0.280	5.132E-08	2.178E-05	20.60	81.2	0.6	1.1	17.1
0.320	7.296E-08	3.092E-05	20.59	81.4	0.4	1.1	17.1
0.360	9.815E-08	4.156E-05	20.58	81.4	0.3	1.1	17.1
0.400	1.283E-07	5.428E-05	20.57	81.5	0.3	1.1	17.1
0.440	1.617E-07	6.839E-05	20.57	81.5	0.2	1.1	17.1
0.480	2.001E-07	8.463E-05	20.56	81.6	0.2	1.1	17.1
0.520	2.415E-07	1.021E-04	20.56	81.6	0.2	1.1	17.1
0.560	2.878E-07	1.217E-04	20.56	81.6	0.1	1.1	17.2
0.600	3.366E-07	1.423E-04	20.56	81.6	0.1	1.1	17.2
0.640	3.903E-07	1.649E-04	20.56	81.6	0.1	1.1	17.2
0.680	4.459E-07	1.884E-04	20.56	81.6	0.1	1.1	17.2
0.720	5.062E-07	2.139E-04	20.56	81.6	0.1	1.1	17.2
0.760	5.680E-07	2.400E-04	20.55	81.7	0.1	1.1	17.2
0.800	6.343E-07	2.680E-04	20.55	81.7	0.1	1.1	17.2

$\tau = T_{\max}/10000$, $N_{\text{err}} = 25$, $N_{\text{dump}} = 500$
$\#I = 512$, $h_{\max} = .884E-01$, $h_{\max}/h_{\min} = .100E+01$

Table 3: Case `uni1`

T	err	est	EI	E1	E2	E3	E4
0.040	4.021E-08	3.061E-05	27.59	8.9	86.9	2.3	1.9
0.080	2.770E-07	7.946E-05	16.94	26.5	61.4	6.5	5.6
0.120	8.549E-07	1.661E-04	13.94	40.9	40.8	9.8	8.5
0.160	1.883E-06	3.050E-04	12.73	50.1	27.5	11.9	10.5
0.200	3.443E-06	5.070E-04	12.13	55.9	19.2	13.2	11.7
0.240	5.591E-06	7.799E-04	11.81	59.6	14.0	14.0	12.4
0.280	8.367E-06	1.129E-03	11.62	62.0	10.6	14.5	12.9
0.320	1.179E-05	1.558E-03	11.49	63.6	8.2	14.9	13.3
0.360	1.588E-05	2.068E-03	11.41	64.8	6.5	15.1	13.5
0.400	2.063E-05	2.659E-03	11.35	65.7	5.3	15.3	13.7
0.440	2.603E-05	3.332E-03	11.31	66.3	4.4	15.5	13.8
0.480	3.207E-05	4.084E-03	11.29	66.8	3.7	15.6	13.9
0.520	3.874E-05	4.914E-03	11.26	67.2	3.1	15.6	14.0
0.560	4.601E-05	5.820E-03	11.25	67.5	2.7	15.7	14.1
0.600	5.386E-05	6.797E-03	11.23	67.8	2.4	15.7	14.1
0.640	6.227E-05	7.844E-03	11.22	68.0	2.1	15.8	14.2
0.680	7.121E-05	8.958E-03	11.22	68.1	1.8	15.8	14.2
0.720	8.065E-05	1.013E-02	11.21	68.3	1.6	15.8	14.2
0.760	9.056E-05	1.137E-02	11.20	68.4	1.5	15.9	14.3
0.800	1.009E-04	1.266E-02	11.20	68.5	1.3	15.9	14.3

$\tau = T_{\max}/50000$, $N_{\text{err}} = 100$, $N_{\text{dump}} = 2500$
$\#I = 2375$, $h_{\max} = .441E-01$, $h_{\max}/h_{\min} = .176E+01$

Table 4: Case `uni2`

T	err	est	EI	E1	E2	E3	E4
0.040	1.827E-08	8.428E-06	21.48	11.8	82.1	2.3	3.7
0.080	1.314E-07	2.415E-05	13.56	31.6	52.6	5.9	10.0
0.120	4.119E-07	5.423E-05	11.47	45.0	32.5	8.3	14.2
0.160	9.150E-07	1.039E-04	10.66	52.7	21.0	9.6	16.6
0.200	1.681E-06	1.772E-04	10.27	57.2	14.3	10.4	18.1
0.240	2.740E-06	2.768E-04	10.05	59.9	10.3	10.9	18.9
0.280	4.110E-06	4.048E-04	9.92	61.7	7.7	11.2	19.5
0.320	5.803E-06	5.622E-04	9.84	62.9	5.9	11.4	19.8
0.360	7.826E-06	7.497E-04	9.79	63.7	4.7	11.5	20.1
0.400	1.018E-05	9.674E-04	9.75	64.3	3.8	11.6	20.3
0.440	1.286E-05	1.215E-03	9.72	64.7	3.1	11.7	20.4
0.480	1.585E-05	1.492E-03	9.70	65.1	2.6	11.7	20.5
0.520	1.916E-05	1.798E-03	9.69	65.4	2.2	11.8	20.6
0.560	2.277E-05	2.132E-03	9.67	65.6	1.9	11.8	20.7
0.600	2.667E-05	2.492E-03	9.67	65.7	1.7	11.8	20.7
0.640	3.085E-05	2.878E-03	9.66	65.9	1.5	11.9	20.8
0.680	3.529E-05	3.288E-03	9.65	66.0	1.3	11.9	20.8
0.720	3.999E-05	3.722E-03	9.65	66.1	1.2	11.9	20.8
0.760	4.492E-05	4.177E-03	9.64	66.2	1.0	11.9	20.9
0.800	5.008E-05	4.654E-03	9.64	66.2	0.9	11.9	20.9

$\tau = T_{\max}/200000$, $N_{\text{err}} = 250$, $N_{\text{dump}} = 10000$
$\#I = 4433$, $h_{\max} = .316E-01$, $h_{\max}/h_{\min} = .188E+01$

Table 5: Case `uni3`

T	err	est	EI	E1	E2	E3	E4
0.040	4.864E-09	1.204E-06	15.74	18.0	72.0	2.3	7.6
0.080	3.603E-08	4.161E-06	10.75	39.5	38.8	5.0	16.8
0.120	1.141E-07	1.037E-05	9.54	50.5	21.7	6.4	21.5
0.160	2.546E-07	2.098E-05	9.08	55.9	13.3	7.1	23.8
0.200	4.692E-07	3.685E-05	8.86	58.7	8.8	7.4	25.0
0.240	7.663E-07	5.859E-05	8.74	60.4	6.2	7.6	25.7
0.280	1.151E-06	8.661E-05	8.67	61.5	4.6	7.8	26.2
0.320	1.628E-06	1.212E-04	8.63	62.1	3.5	7.9	26.5
0.360	2.197E-06	1.624E-04	8.60	62.6	2.8	7.9	26.7
0.400	2.859E-06	2.103E-04	8.58	63.0	2.2	8.0	26.8
0.440	3.614E-06	2.649E-04	8.56	63.2	1.8	8.0	26.9
0.480	4.459E-06	3.259E-04	8.55	63.4	1.5	8.0	27.0
0.520	5.393E-06	3.933E-04	8.54	63.6	1.3	8.0	27.1
0.560	6.411E-06	4.668E-04	8.53	63.7	1.1	8.1	27.1
0.600	7.512E-06	5.463E-04	8.53	63.8	1.0	8.1	27.2
0.640	8.691E-06	6.314E-04	8.52	63.9	0.9	8.1	27.2
0.680	9.945E-06	7.219E-04	8.52	63.9	0.8	8.1	27.2
0.720	1.127E-05	8.175E-04	8.52	64.0	0.7	8.1	27.3
0.760	1.266E-05	9.180E-04	8.51	64.0	0.6	8.1	27.3
0.800	1.412E-05	1.023E-03	8.51	64.1	0.6	8.1	27.3

$\tau = T_{\max}/400000$, $N_{\text{err}} = 500$, $N_{\text{dump}} = 20000$
$\#I = 16384$, $h_{\max} = .156E-01$, $h_{\max}/h_{\min} = .100E+01$

Table 6: Case `uni4`

T	err	est	EI	E1	E2	E3	E4
0.040	3.928E-09	1.133E-06	16.98	24.9	63.7	1.9	9.6
0.080	2.916E-08	4.456E-06	12.36	47.9	30.1	3.6	18.4
0.120	9.241E-08	1.180E-05	11.30	57.7	15.8	4.4	22.2
0.160	2.064E-07	2.458E-05	10.91	62.1	9.4	4.7	23.8
0.200	3.805E-07	4.381E-05	10.73	64.3	6.2	4.8	24.7
0.240	6.215E-07	7.026E-05	10.63	65.6	4.3	4.9	25.2
0.280	9.339E-07	1.044E-04	10.57	66.4	3.2	5.0	25.5
0.320	1.320E-06	1.466E-04	10.54	66.9	2.4	5.0	25.7
0.360	1.782E-06	1.970E-04	10.51	67.2	1.9	5.0	25.8
0.400	2.320E-06	2.555E-04	10.50	67.5	1.5	5.0	25.9
0.440	2.932E-06	3.222E-04	10.48	67.7	1.3	5.1	26.0
0.480	3.618E-06	3.969E-04	10.47	67.8	1.1	5.1	26.0
0.520	4.376E-06	4.793E-04	10.47	67.9	0.9	5.1	26.1
0.560	5.202E-06	5.693E-04	10.46	68.0	0.8	5.1	26.1
0.600	6.096E-06	6.665E-04	10.46	68.1	0.7	5.1	26.1
0.640	7.052E-06	7.706E-04	10.45	68.2	0.6	5.1	26.2
0.680	8.070E-06	8.813E-04	10.45	68.2	0.5	5.1	26.2
0.720	9.145E-06	9.983E-04	10.45	68.2	0.5	5.1	26.2
0.760	1.028E-05	1.121E-03	10.45	68.3	0.4	5.1	26.2
0.800	1.146E-05	1.250E-03	10.44	68.3	0.4	5.1	26.2

$\tau = T_{\max}/100000$, $N_{\text{err}} = 250$, $N_{\text{dump}} = 5000$
$\#I = 3359$, $h_{\max} = .885E-01$, $h_{\max}/h_{\min} = .924E+01$

Table 7: Case `ad1`

T	err	est	EI	E1	E2	E3	E4
0.040	2.190E-09	1.726E-06	28.08	38.3	52.5	1.4	7.8
0.080	1.619E-08	7.884E-06	22.07	63.5	21.4	2.3	12.8
0.120	5.123E-08	2.211E-05	20.77	72.2	10.6	2.6	14.6
0.160	1.143E-07	4.718E-05	20.31	75.8	6.2	2.7	15.3
0.200	2.107E-07	8.515E-05	20.10	77.6	4.0	2.8	15.6
0.240	3.440E-07	1.375E-04	19.99	78.5	2.8	2.8	15.8
0.280	5.168E-07	2.052E-04	19.93	79.2	2.0	2.9	16.0
0.320	7.306E-07	2.889E-04	19.89	79.5	1.5	2.9	16.0
0.360	9.861E-07	3.889E-04	19.86	79.8	1.2	2.9	16.1
0.400	1.283E-06	5.053E-04	19.84	80.0	1.0	2.9	16.1
0.440	1.622E-06	6.378E-04	19.83	80.2	0.8	2.9	16.2
0.480	2.002E-06	7.861E-04	19.82	80.3	0.7	2.9	16.2
0.520	2.421E-06	9.500E-04	19.81	80.3	0.6	2.9	16.2
0.560	2.878E-06	1.129E-03	19.81	80.4	0.5	2.9	16.2
0.600	3.372E-06	1.322E-03	19.80	80.5	0.4	2.9	16.2
0.640	3.901E-06	1.529E-03	19.80	80.5	0.4	2.9	16.2
0.680	4.464E-06	1.749E-03	19.79	80.5	0.3	2.9	16.2
0.720	5.059E-06	1.982E-03	19.79	80.6	0.3	2.9	16.2
0.760	5.684E-06	2.226E-03	19.79	80.6	0.3	2.9	16.2
0.800	6.337E-06	2.482E-03	19.79	80.6	0.2	2.9	16.2

$\tau = T_{\max}/100000$, $N_{\text{err}} = 250$, $N_{\text{dump}} = 5000$
$\#I = 2457$, $h_{\max} = .861E-01$, $h_{\max}/h_{\min} = .136E+02$

Table 8: Case `ad2`

T	err	est	EI	E1	E2	E3	E4
0.040	3.455E-07	7.742E-04	47.34	5.8	91.7	2.2	0.2
0.080	1.878E-06	1.769E-03	30.69	18.8	73.3	7.0	0.9
0.120	5.199E-06	3.327E-03	25.30	32.7	54.1	11.8	1.5
0.160	1.074E-05	5.690E-03	23.02	43.5	39.1	15.4	2.0
0.200	1.882E-05	9.041E-03	21.92	51.2	28.6	17.8	2.4
0.240	2.966E-05	1.351E-02	21.34	56.4	21.5	19.5	2.6
0.280	4.339E-05	1.920E-02	21.03	60.1	16.5	20.7	2.8
0.320	6.010E-05	2.615E-02	20.86	62.7	13.0	21.5	2.9
0.360	7.980E-05	3.441E-02	20.76	64.6	10.4	22.0	3.0
0.400	1.025E-04	4.397E-02	20.71	66.0	8.5	22.4	3.0
0.440	1.281E-04	5.485E-02	20.69	67.1	7.1	22.8	3.1
0.480	1.566E-04	6.700E-02	20.68	67.9	6.0	23.0	3.1
0.520	1.879E-04	8.041E-02	20.69	68.6	5.1	23.2	3.1
0.560	2.219E-04	9.504E-02	20.70	69.1	4.4	23.3	3.2
0.600	2.584E-04	1.108E-01	20.71	69.5	3.8	23.4	3.2
0.640	2.974E-04	1.278E-01	20.73	69.9	3.4	23.5	3.2
0.680	3.388E-04	1.457E-01	20.74	70.2	3.0	23.6	3.2
0.720	3.823E-04	1.648E-01	20.76	70.4	2.7	23.7	3.2
0.760	4.280E-04	1.847E-01	20.78	70.7	2.4	23.7	3.2
0.800	4.757E-04	2.056E-01	20.79	70.8	2.2	23.8	3.2

$\tau = T_{\max}/600000$, $N_{\text{err}} = 800$, $N_{\text{dump}} = 30000$
$\#I = 23235$, $h_{\max} = .285E-01$, $h_{\max}/h_{\min} = .210E+02$

Table 9: Case `ad3`

h	τ	$\|e_{\tau h}\|_{L^2(Q_T)}$	$\text{EOC}_{\text{space}}$	EOC_{time}
0.2838	0.2500	4.5853E-02	-	-
0.2172	0.2000	3.5474E-02	0.96	1.15
0.1540	0.1667	2.2387E-02	1.34	2.52
0.1035	0.1000	1.2945E-02	1.38	1.07
0.0768	0.0769	5.3557E-03	2.95	3.36
0.0532	0.0526	2.3742E-03	2.22	2.14
0.0398	0.0400	1.3345E-03	1.98	2.10
0.0270	0.0270	5.2577E-04	2.40	2.38
0.0223	0.0222	2.7946E-04	3.30	3.23
0.0144	0.0145	1.1835E-04	1.98	2.01
Global order of convergence			2.07	2.11

Table 10: $\varepsilon = 0.005$, $p = 1$, $q = 1$ (parabolic case)

h	τ	$\|e_{\tau h}\|_{L^2(Q_T)}$	$\mathrm{EOC}_{\mathrm{space}}$	$\mathrm{EOC}_{\mathrm{time}}$
0.2838	0.2500	2.0470E-02	-	-
0.2172	0.2000	1.0103E-02	2.64	3.16
0.1540	0.1667	4.3992E-03	2.42	4.56
0.1035	0.1000	1.6821E-03	2.42	1.88
0.0768	0.0769	4.9668E-04	4.08	4.65
0.0532	0.0526	1.6550E-04	3.00	2.90
0.0398	0.0400	7.7630E-05	2.61	2.76
0.0270	0.0270	2.7654E-05	2.66	2.63
Global order of convergence			2.89	2.78

Table 11: $\varepsilon = 0.005$, $p = 2$, $q = 2$ (parabolic case)

h	τ	$\|e_{\tau h}\|_{L^2(Q_T)}$	$\mathrm{EOC}_{\mathrm{space}}$	$\mathrm{EOC}_{\mathrm{time}}$
0.2838	0.2500	4.9212E-02	-	-
0.2172	0.2000	3.8843E-02	0.89	1.06
0.1540	0.1667	2.5997E-02	1.17	2.20
0.1035	0.1000	1.5581E-02	1.29	1.00
0.0768	0.0769	6.9089E-03	2.72	3.10
0.0532	0.0526	3.2904E-03	2.02	1.95
0.0398	0.0400	1.8620E-03	1.96	2.07
0.0270	0.0270	7.5458E-04	2.32	2.30
0.0223	0.0222	4.1924E-04	3.07	3.00
0.0144	0.0145	1.7556E-04	2.01	2.04
Global order of convergence			1.95	1.99

Table 12: $\varepsilon = 0$, $p = 1$, $q = 1$ (hyperbolic case)

h	τ	$\|e_{\tau h}\|_{L^2(Q_T)}$	$\mathrm{EOC}_{\mathrm{space}}$	$\mathrm{EOC}_{\mathrm{time}}$
0.2838	0.2500	2.3451E-02	-	-
0.2172	0.2000	1.2484E-02	2.36	2.83
0.1540	0.1667	6.1746E-03	2.05	3.86
0.1035	0.1000	2.6342E-03	2.14	1.67
0.0768	0.0769	8.0848E-04	3.95	4.50
0.0532	0.0526	2.6400E-04	3.05	2.95
0.0398	0.0400	1.0761E-04	3.09	3.27
0.0270	0.0270	2.7962E-05	3.47	3.44
Global order of convergence			2.87	2.98

Table 13: $\varepsilon = 0$, $p = 2$, $q = 2$ (hyperbolic case)

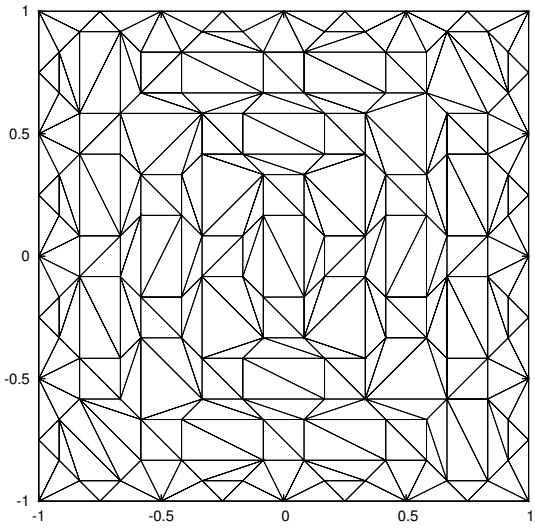


Figure 1: FV mesh for case 1

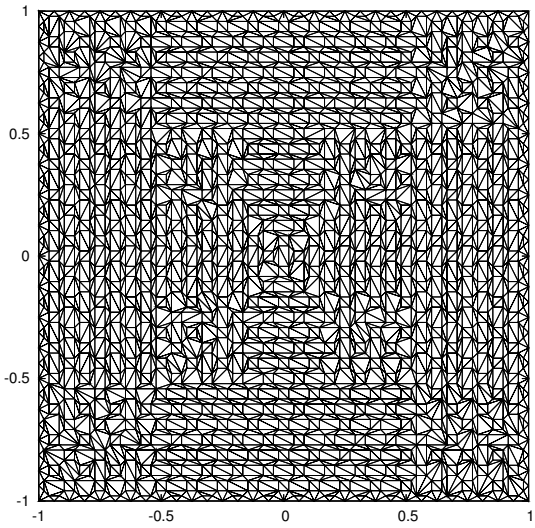


Figure 2: FV mesh for case 3

References

[1] P. Angot, V. Dolejší, M. Feistauer, J. Felcman: Analysis of a combined barycentric finite volume – nonconforming finite element method for nonlinear convection-diffusion problems. Appl. Math. 43 (1998), 263–310.

[2] P. Arminjon, A. Madrane, A mixed finite volume/finite element method for 2-dimensional compressible Navier–Stokes equations on unstructured grids, In: Hyperbolic Problems: Theory, Numerics, Applications, 11–20, Birkhäuser, 1999, M. Fey and R. Jeltsch (eds.), Volume I,Basel

[3] D.N. Arnold, F. Brezzi, B. Cockburn and D. Marini, Discontinuos Galerkin methods for elliptic problems, in Discontinuous Galerkin methods. Theory, Computation and Applications. Lecture Notes in Computational Science and Engineering 11 (Eds. B.Cockburn et al.), Springer, Berlin, 2000, 89–101.

[4] D.N. Arnold, F. Brezzi, B. Cockburn, and D. Marini, Unified analysis of discontinuous Galerkin methods for elliptic problems, SIAM J. Numer. Anal., 39 (2001), 1749–1779.

[5] D.N. Arnold: An interior penalty finite element method with discontinuous elements, SIAM J. Numer. Anal., 19 (1982), 742–760.

[6] I. Babuška, C. E. Baumann, and J. T. Oden, A discontinuous *hp* finite element method for diffusion problems, 1-D analysis, Comput. Math. Appl., 37 (1999), 103–122.

[7] 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.

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

[9] 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.

[10] B. Cockburn, G. E. Karniadakis, and C.–W. Shu (Eds.), Discontinuous Galerkin Methods, Lecture Notes in Computational Science and Engineering 11. Springer, Berlin, 2000.

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

[12] 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.

[13] T.A. Davis and I. S. Duff: A combined unifrontal/multifrontal method for unsymmetric sparse matrices, ACM Transactions on Mathematical Software 25 (1999), 1–19.

[14] V. Dolejší: Anisotropic mesh adaptation for finite volume and finite element methods on triangular meshes. Computing and Visualisation in Science, 1 (1998), 165-178

[15] V. Dolejší, M. Feistauer, J. Felcman, A. Kliková, Error Estimates for Barycentric Finite Volumes Combined with Nonconforming Finite Elements Applied to Nonlinear Convection–Diffusion Problems, Appl. Math., 47 (2002), 301–340.

[16] V. Dolejší and M. Feistauer, A semi-implicit discontinuous Galerkin finite element method for the numerical solution of inviscid compressible flow, J. Comput. Phys., 198 (2004), 727–746.

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

[18] V. Dolejší, M. Feistauer, J. Felcman: Numerical Simulation of Compresssible Viscous Flow through Cascades of Profiles. ZAMM, 76 (1996), 301-304.

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

[20] V. Dolejší, Sur des méthodes combinant des volumes finis et des éléments finis pour le calcul d'écoulements compressibles sur des maillages non structurés, Charles Univeristy Prague and Université Mediterannée Aix–Marseille II, 1998.

[21] M. Feistauer, T. Gallouet, J. Hájek, R. Herbin: Combined triangular FV - triangular FE method for nonlinear convection-diffusion problems. Submitted to ACCM.

[22] M. Feistauer, K. Švadlenka: Discontinuous Galerkin method of lines for solving nostationary singularly perturbed linear problems. J. Numer. Math., 12 (2004)

[23] M. Feistauer, J. Felcman, M. Lukáčová, 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.

[24] M. Feistauer, J. Felcman, M. Lukáčová, Combined finite element–finite volume solution of compressible flow, J. Comput. Appl. Math., 63 (1995), 179-199.

[25] M. Feistauer, J. Felcman, M. Lukáčová, On the Convergence of a Combined Finite Volume–Finite Element Method for Nonlinear Convection–Diffusion Problems, Numer. Methods Partial Differ. Equations, 13 (1997), 163-190.

[26] M. Feistauer, J. Slavík, P. Stupka: Convergence of the combined finite element - finite volume method for nonlinear convection - diffusion problems. Explicit schemes. Numer Methods Partial Differential Eq 15 (1999), 215-235.

[27] M. Feistauer, J. Felcman: Theory and applications of numerical schemes for nonlinear convection-diffusion problems and compressible Navier-Stokes equations. The Mathematics of Finite Elements and Applications. Highlights 1996 (J. R. Whiteman, ed.), Wiley, Chichester, 1997, 175-194.

[28] M. Feistauer, J. Hájek, K. Švadlenka, Space-Time discontinuous Galerkin method for solving nostationary convection-diffusion-reaction problems. to appear in East-West Journal of Numerical Mathematics, 2006.

[29] J.M. Ghidaglia, F. Pascal: Footbridges finite volume-finite elements. C. R. Acd. Sci., Ser. I, Math. 328, vol. 8 (1999), 711–716.

[30] J.M. Ghidaglia, F. Pascal: Footbridges between finite volume-finite elements with applications to CFD. Technical report, University Paris-Sud, 2001 (preprint)

[31] R. Hartmann and P. Houston, Adaptive discontinuous Galerkin finite element methods for the compressible Euler equations, Technical Report 2001-42 (SFB 359), IWR Heidelberg.

[32] F. Hecht, O. Pironneau, and A. Le Hyaric: `www.freefem.org/ff++`

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

[34] J. Jaffre, C. Johnson, and A. Szepessy: Convergence of the discontinuous Galerkin finite element method for hyperbolic conservation laws, Math. Models Methods Appl. Sci., 5 (1995), 367–386.

[35] C. Johnson and J. Pitkäranta: An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation, Math. Comp., 46 (1986), 1–26.

[36] A. Kliková: Finite volume-finite element solution of compressible flow. PhD. Dissertation, Charles University Prague, Faculty of Mathematics and Physics, 2000.

[37] P. Le Saint and P.–A. Raviart: On a finite element method for solving the neutron transport equation, in Mathematical Aspects of Finite Elements in Partial Differential Equations (Ed. C. de Boor), Academic Press, 1974, 89–145.

[38] W.H. Reed and T. R. Hill: Triangular mesh methods for the neutron transport equation, Technical Report LA–UR–73–479, Los Alamos Scientific Laboratory, 1973.

[39] B. Rivière and M. F. Wheeler, A discontinuous Galerkin method applied to nonlinear parabolic equations, in Discontinuous Galerkin methods. Theory, Computation and Applications. Lecture Notes in Computational Science and Engineering 11 (Eds. B.Cockburn et al.), Springer, Berlin, 2000, 231–244.

[40] Y. Saad. SPARSKIT: A Basic Tool Kit for Sparse Matrix Computation. Tech. Report CSRD TR 1029, Center for Supercomputing Research and Development, University of Illinois at Urbana Champaign, 1990.

[41] S. Sun, M. F. Wheeler: $L^2(H^1)$ norm a posteriori error estimation for discontinuous Galerkin approximations of reactive transport problems. Journal of Scientific Computing, 22-23 (2005)

[42] J.J.W. van der Vegt and H. Van der Ven, Space-time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flow, part I. General formulation, J. Comput. Phys., 182 (2002), 546–585.

[43] M.F. Wheeler: An elliptic collocation-finite element method with interior penalties, SIAM J. Numer. Anal., 15 (1978), 152–161.

[44] `www.netlib.org/minpack`