FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Ondřej Bartoš

Discontinuous Galerkin method for the solution of boundary-value problems in non-smooth domains

Department of Numerical Mathematics

Supervisor of the master thesis: prof. RNDr. Miloslav Feistauer, DrSc., dr. h. c.
Study programme: Matematika
Study branch: Numerická a výpočtová matematika

Prague 2017

I would like to express my gratitude to prof. Miloslav Feistauer for his guidance, patience during consultations, and providing me with essential literature. I am also thankful to Filip Roskovec for allowing me to use his program and helping me with technical issues.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date .... signature of the author

Název práce: Nespojitá Galerkinova metoda pro řešení okrajových problémů v nehladkých oblastech

Autor: Ondřej Bartoš

Katedra: Katedra numerické matematiky

Vedoucí diplomové práce: prof. RNDr. Miloslav Feistauer, DrSc., dr. h. c.

Abstrakt: Tato práce se zabývá analýzou metody konečných prvků a nespojité Galerkinovy metody pro numerické řešení eliptické okrajové úlohy s nelineární Newtonovou okrajovou podmínkou ve dvourozměrné polygonální oblasti. Slabé řešení ztrácí regularitu v okolí hraničních singularit, které se vyskytují v okolí rohů a kořenů slabého řešení na hranách. Hlavní pozornost se věnuje odhadům chyby. Ukazuje se, že řád konvergence není snížen nelinearitou, pokud je slabé řešení nenulové na větší části hranice. Pokud je slabé řešení nulové na celé hranici, pak nelinearita zpomaluje pouze konvergenci funkčních hodnot, ale ne konvergenci gradientu. Stejné výsledky jsou odvozeny pro přibližná řešení získaná pomocí numerické integrace. Odvozené výsledky jsou potvrzeny numerickými výpočty.

Klíčová slova: nelineární eliptický problém, hraniční singularity, metoda konečných prvků, nespojitá Galerkinova metoda, numerická kvadratura

Title: Discontinuous Galerkin method for the solution of boundary-value problems in non-smooth domains

Author: Ondřej Bartoš

Department: Department of Numerical Mathematics

Supervisor: prof. RNDr. Miloslav Feistauer, DrSc., dr. h. c.

Abstract: This thesis is concerned with the analysis of the finite element method and the discontinuous Galerkin method for the numerical solution of an elliptic boundary value problem with a nonlinear Newton boundary condition in a two-dimensional polygonal domain. The weak solution loses regularity in a neighbourhood of boundary singularities, which may be at corners or at roots of the weak solution on edges. The main attention is paid to the study of error estimates. It turns out that the order of convergence is not dampened by the nonlinearity, if the weak solution is nonzero on a large part of the boundary. If the weak solution is zero on the whole boundary, the nonlinearity only slows down the convergence of the function values but not the convergence of the gradient. The same analysis is carried out for approximate solutions obtained with numerical integration. The theoretical results are verified by numerical experiments.

Keywords: nonlinear elliptic problem, boundary singularities, finite element method, discontinuous Galerkin method, numerical quadrature

Contents

Introduction

There are many numerical techniques for solving partial differential equations. The effectivity of the respective methods is often closely related to the properties of the equations in question. We are concerned with the study of the finite element method (FEM) and the discontinuous Galerkin method (DGM). We use them for the solution of an elliptic equation with a nonlinear Newton boundary condition in a bounded two-dimensional polygonal domain with numerical integration. Such boundary value problems have applications in science and engineering, see [14], [2]. We suppose that the nonlinear term has a general "polynomial" growth. This can be found in the modelling of electrolysis of aluminium with the aid of the stream function. The nonlinear boundary condition describes turbulent flow in a boundary layer ([26]). Similar nonlinearity appears in a radiation heat transfer problem ([24], [21]) or in nonlinear elasticity ([15], [16]). A parabolic equation with a nonlinear Newton boundary condition is solved with the use of finite elements in [6] and [28], but the growth of the nonlinearity is only linear.

The paper [8] deals with the problem arising in the investigation of the electrolytical producing of aluminium. The problem is discretized by piecewise linear conforming triangular elements. The effect of the numerical integration applied to this problem is investigated in [9]. Using monotone operator theory in [13] and assuming regularity of the weak solution, the paper [10] gives error estimates. The paper [11] investigates this problem using discontinuous Galerkin method and piecewise polynomial functions, but does not consider the effect of numerical integration.

In this thesis we study an elliptic boundary value problem with nonlinear Newton boundary condition in a polygonal domain. The goal is to analyse both FEM and DGM used on conforming shape regular meshes with piecewise polynomial functions and the effect of numerical integration while considering the actual regularity of the weak solution. In Chapter 1 the boundary value problem is introduced, the weak solution is defined and some regularity results are derived in the neighbourhood of boundary edges. In chapter 2 the Galerkin approximation of the weak solution (approximate solution found with the aid of exact integration) is introduced with the aid of FEM. It turns out that the order of convergence changes based on whether the exact weak solution is zero on the boundary or not. Chapter 3 shows that the same error estimates mostly hold for the approximate solution found with numerical integration. Chapter 4 introduces the DG method and derives the same results. Chapter 5 confirms the theoretically found error estimates with numerical experiments.

1. Continuous problem

1.1 Classical formulation

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with a Lipschitz continuous boundary $\partial\Omega$. We consider a boundary value problem with a non-linear Newton boundary condition: find $u : \overline{\Omega} \to \mathbb{R}$ such that

$$-\Delta u = f \quad \text{in } \Omega, \tag{1.1}$$

$$\frac{\partial u}{\partial n} + \kappa |u|^\alpha u = \varphi \quad \text{on } \partial\Omega, \tag{1.2}$$

with given functions $f : \Omega \to \mathbb{R}$, $\varphi : \partial\Omega \to \mathbb{R}$ and constants $\kappa > 0$, $\alpha \geq 0$. By a classical solution of (1.1) with boundary conditions (1.2) we refer to a function $u \in C^2(\overline{\Omega})$ satisfying (1.1) pointwise at every point in Ω and satisfying (1.2) at every point on $\partial\Omega$ such that the outer normal unit vector n is defined.

We will not be concerned with the classical solution directly but we will instead define a weak solution. This can be obtained formally by multiplying (1.1) by a function v, integrating over Ω, using Green's theorem, and applying boundary condition (1.2). It will immediately follow that the classical solution, if it exists, is also a weak solution. Further, we will show that the weak solution exists and is unique. This will also imply that there exists at most one classical solution. We will then use smoothness of f, φ and $\partial\Omega$ to prove the regularity of the weak solution.

1.2 Function spaces

To define a weak solution we need to introduce some function spaces.

We will refer to the set of real numbers by $\mathbb{R}$, the set of positive integers by $\mathbb{N}$, and the set of non-negative integers by $\mathbb{N}_0$.

For a bounded domain (open connected set) Ω, $C^k(\overline{\Omega})$ denotes the set of all k-times continuously differentiable functions in Ω such that all of their partial derivatives of order up to k can be continuously extended on the closure $\overline{\Omega}$ of Ω. The space $C^{k,\lambda}(\overline{\Omega})$ contains functions from $C^k(\overline{\Omega})$ such that all of their k-th partial derivatives are Hölder continuous on $\overline{\Omega}$ with parameter λ. The space of all infinitely smooth functions with a compact support in the domain Ω is denoted by $C_c^\infty(\Omega)$ and the space of all infinitely smooth functions such that all of their partial derivatives can be continuously extended on $\overline{\Omega}$ is denoted by $C^\infty(\overline{\Omega})$.

For $p \in [1, \infty)$ we consider Lebesgue spaces $L^p(\Omega)$, $L^p(\partial\Omega)$ of classes of measurable functions which are equal almost everywhere with respect to the Lebesgue measure. These are Banach spaces with norms

$$\|f\|_{0,p,\Omega} = \left(\int_\Omega |f|^p \mathrm{d}x \right)^{\frac{1}{p}},$$

$$\|f\|_{0,p,\partial\Omega} = \left(\int_{\partial\Omega} |f|^p \mathrm{d}S \right)^{\frac{1}{p}}.$$

For $p \in [1, \infty)$ and $k \in \mathbb{N}$ we consider Sobolev spaces $W^{k,p}(\Omega)$, $W^{k,p}(\partial\Omega)$ with seminorms

$$|f|_{k,p,\Omega} = \left(\sum_{|\beta|=k} \int_\Omega \left| D^\beta f \right|^p \mathrm{d}x \right)^{\frac{1}{p}},$$

$$|f|_{k,p,\partial\Omega} = \left(\sum_{|\beta|=k} \int_{\partial\Omega} \left| D^\beta f \right|^p \mathrm{d}S \right)^{\frac{1}{p}},$$

where $\beta = (\beta_1, \beta_2)$ is a multi-index with $|\beta| = \beta_1 + \beta_2$, and norms

$$\|f\|_{k,p,\Omega} = \left(\sum_{|\beta|\le k} \int_\Omega \left| D^\beta f \right|^p \mathrm{d}x \right)^{\frac{1}{p}},$$

$$\|f\|_{k,p,\partial\Omega} = \left(\sum_{|\beta|\le k} \int_{\partial\Omega} \left| D^\beta f \right|^p \mathrm{d}S \right)^{\frac{1}{p}}.$$

The derivative $D^\beta f$ in the definition of Sobolev spaces is considered in a weak sense, that is, $D^\beta f$ is a weak derivative of f in Ω if $D^\beta f \in L^1_{loc}(\Omega)$ is a locally integrable function satisfying

$$\int_\Omega f D^\beta \varphi \mathrm{d}x = (-1)^{|\beta|} \int_\Omega D^\beta f \varphi \mathrm{d}x, \quad \forall \varphi \in C_c^\infty(\Omega).$$

The Sobolev spaces of functions which are zero on boundary $\partial\Omega$ are denoted by $W_0^{1,p}(\Omega)$. For $p \in [1, \infty)$, $k \in \mathbb{N}_0$ and $s \in (0, 1)$ the fractional Sobolev-Slobodetskii spaces $W^{k+s,p}(\Omega)$ use a seminorm

$$|f|_{k+s,p,\Omega} = \left(\sum_{|\beta|=k} \int_{\Omega\times\Omega} \frac{\left| D^\beta f(x) - D^\beta f(y) \right|^p}{|x-y|^{n+sp}} \mathrm{d}x\mathrm{d}y \right)^{\frac{1}{p}},$$

where n is the dimension of Ω, and they use a norm

$$\|f\|_{k+s,p,\Omega} = \left(\|f\|^p_{k,p,\Omega} + |f|^p_{k+s,p,\Omega} \right)^{\frac{1}{p}}.$$

We also denote $W^{k,2}(\Omega) = H^k(\Omega)$ and $W^{0,p}(\Omega) = L^p(\Omega)$. If a function f belongs to a Sobolev space $W^{k,p}(\Omega)$, then its β-th derivative belongs to a space $W^{k-|\beta|,p}(\Omega)$ for any $|\beta| \le k$ and the norms satisfy inequality

$$\|D^\beta f\|_{k-|\beta|,p,\Omega} \le \|f\|_{k,p,\Omega}.$$

The following continuous embeddings known as Sobolev embeddings hold for domains $\Omega \subset R^n$ with Lipschitz continuous boundaries (Section 5.6 in [7])

$$\begin{aligned} &W^{1,p}(\Omega) \hookrightarrow L^{\frac{np}{n-p}}(\Omega), && p \in [1, n), \\ &W^{1,n}(\Omega) \hookrightarrow L^q(\Omega), && q \in [1, \infty), \\ &W^{1,p}(\Omega) \hookrightarrow C^{0,1-\frac{n}{p}}(\overline{\Omega}), && p \in (n, \infty), \\ &W^{n,1}(\Omega) \hookrightarrow C(\overline{\Omega}), && \end{aligned} \tag{1.3}$$

and the following continuous trace embeddings also hold for domains with Lipschitz continuous boundaries (Section 5.5 in [7] or Theorems 1.4.4.1 and 1.5.1.1 in [17])

$$
\begin{aligned}
&W^{1,p}(\Omega) \hookrightarrow L^{\frac{(n-1)p}{n-p}}(\partial\Omega), && p \in [1, n),\\
&W^{1,n}(\Omega) \hookrightarrow L^{q}(\partial\Omega), && q \in [1, \infty),\\
&W^{1,p}(\Omega) \hookrightarrow C^{0,1-\frac{n}{p}}(\partial\Omega), && p \in (n, \infty),\\
&W^{n,1}(\Omega) \hookrightarrow C(\partial\Omega).
\end{aligned}
\tag{1.4}
$$

If $G \subset \partial\Omega$, then by $|G|$ we denote the one-dimensional measure defined on $\partial\Omega$ of the set G. Due to Poincaré inequality (5.8.1 in [7]) on spaces $W_0^{1,p}(\Omega)$ the seminorm is in fact equivalent to a norm.

Theorem 1.1 (Poincaré inequality)**.** *Let Ω be a domain with a Lipschitz continuous boundary. Let $u \in W^{1,p}(\Omega)$. Let $G \subset \partial\Omega$ with $|G| > 0$. Then there exists a constant $c_P > 0$ dependent on Ω, G and p such that*

$$\|u\|_{1,p,\Omega} \leq c_P(|u|_{1,p,\Omega} + \|u\|_{0,p,G}). \tag{1.5}$$

In what follows, we will use c to refer to a general positive constant which can change for different equations.

1.3 Weak solution

Suppose that

$$f \in L^2(\Omega), \quad \varphi \in L^2(\partial\Omega). \tag{1.6}$$

We introduce the following forms for u, $v \in H^1(\Omega)$

$$
\begin{aligned}
b(u, v) &= \int_\Omega \nabla u \cdot \nabla v \mathrm{d}x,\\
d(u, v) &= \kappa \int_{\partial\Omega} |u|^\alpha uv \mathrm{d}S,\\
L^\Omega(v) &= \int_\Omega f v \mathrm{d}x,\\
L^{\partial\Omega}(v) &= \int_{\partial\Omega} \varphi v \mathrm{d}S,\\
L(v) &= L^\Omega(v) + L^{\partial\Omega}(v),\\
a(u, v) &= b(u, v) + d(u, v).
\end{aligned}
\tag{1.7}
$$

Definition 1.2. *We say that a function $u : \Omega \to \mathbb{R}$ is a weak solution of* (1.1) *if*

$$
\begin{aligned}
&u \in H^1(\Omega),\\
&a(u, v) = L(v) \quad \forall v \in H^1(\Omega).
\end{aligned}
\tag{1.8}
$$

If we take the classical solution $u \in C^2(\overline{\Omega}) \subset H^1(\Omega)$ of (1.1)-(1.2), multiply the equation (1.1) by a function $v \in H^1(\Omega)$, integrate over Ω, use Green's theorem and the boundary conditions (1.2), we will obtain the condition (1.8).

$$\int_\Omega f v \mathrm{d}x = \int_\Omega -\Delta u v \mathrm{d}x = -\int_{\partial\Omega} \frac{\partial u}{\partial n} v \mathrm{d}S + \int_\Omega \nabla u \cdot \nabla v \mathrm{d}x$$

$$-\int_{\partial\Omega} \frac{\partial u}{\partial n} v \mathrm{d}S = \int_{\partial\Omega} \left(\kappa \left|u\right|^{\alpha} u - \varphi\right) v \mathrm{d}S$$

$$\int_{\Omega} \nabla u \cdot \nabla v \mathrm{d}x + \int_{\partial\Omega} \kappa \left|u\right|^{\alpha} uv \mathrm{d}S = \int_{\Omega} fv \mathrm{d}x + \int_{\partial\Omega} \varphi v \mathrm{d}S$$

The weak solution is a generalization of the classical solution. To show existence and uniqueness of the weak solution we need to know some properties of the form a. Note that

$$a(u, u-v) - a(v, u-v) = \int_{\Omega} |\nabla u - \nabla v|^2 \mathrm{d}x + \kappa \int_{\partial\Omega} (|u|^{\alpha} u - |v|^{\alpha} v)(u-v) \mathrm{d}S. \quad (1.9)$$

Let $g > 0$ and $\alpha \geq 0$. We define a function $y : \mathbb{R} \to \mathbb{R}$:

$$y(\zeta) = \left(\left|\zeta + g\right|^{\alpha} (\zeta + g) - \left|\zeta - g\right|^{\alpha} (\zeta - g)\right)(2g), \quad \xi \in \mathbb{R}.$$

Then the function y is decreasing in $(-\infty, 0)$ and increasing in $(0, \infty)$, and

$$\min_{\zeta \in \mathbb{R}} y(\zeta) = 4g^{\alpha+2}.$$

For $\xi, \eta \in \mathbb{R}$ let us set $2g = |\xi - \eta|$. Then

$$\left(\left|\eta\right|^{\alpha} \eta - \left|\xi\right|^{\alpha} \xi\right)(\eta - \xi) \geq 2^{-\alpha} \left|\eta - \xi\right|^{\alpha+2}, \quad (1.10)$$

and the following lemma follows, see [10].

Lemma 1.3. *Let $u, v \in H^1(\Omega)$. Then*

$$a(u, u-v) - a(v, u-v) \geq |u - v|^2_{1,2,\Omega} + \kappa 2^{-\alpha} \|u - v\|^{\alpha+2}_{\alpha+2,0,\partial\Omega}. \quad (1.11)$$

In [10] and [9] most of the following theorem was also proven.

Theorem 1.4.

a) L is a continuous linear functional on $H^1(\Omega)$.

b) The functional $a(u, \cdot)$ from $H^1(\Omega)$ into $\mathbb{R}$ is a continuous linear functional for every $u \in H^1(\Omega)$.

c) a is uniformly monotone, with

$$a(u, u-v) - a(v, u-v) \geq \varrho(\|u - v\|_{1,2,\Omega}) \quad (1.12)$$

for all $u, v \in H^1(\Omega)$, where

$$\varrho(t) = \begin{cases} C_0 \kappa 2^{-\alpha} t^{\alpha+2} & \text{for} \quad 0 \leq t \leq 2\kappa^{-1/\alpha}, \\ C_0 t^2 & \text{for} \quad t \geq 2\kappa^{-1/\alpha}, \end{cases} \quad (1.13)$$

for $\alpha = 0$, we set $\kappa^{-1/\alpha} = 0$.

d) The functional $a(\cdot, v)$ from $H^1(\Omega)$ into $\mathbb{R}$ is continuous for every $v \in H^1(\Omega)$ in the following sense: There exists a positive constant $C_1 > 0$ independent of v such that

$$|a(u, v) - a(w, v)| \leq C_1 \left(1 + \|u\|^{\alpha}_{1,2,\Omega} + \|w\|^{\alpha}_{1,2,\Omega}\right) \|u - w\|_{1,2,\Omega} \|v\|_{1,2,\Omega}, \quad (1.14)$$

for all $u, v \in H^1(\Omega)$.

e) The form a is coercive in the following sense: There exists a positive constant $C_2 > 0$ such that

$$a(u,u) \geq C_2 \|u\|^2_{1,2,\Omega} \tag{1.15}$$

holds for all $u \in H^1(\Omega)$ such that $\|u\|_{1,2,\Omega} \geq 1$.

Proof. It remains to prove the part d).

$$|a(u,v) - a(w,v)| \leq \left| \int_\Omega \nabla(u-w) \cdot \nabla v \mathrm{d}x \right| + \left| \kappa \int_{\partial\Omega} (|u|^\alpha u - |w|^\alpha w) v \mathrm{d}S \right|$$

Hölder inequality used on the first term yields

$$\left| \int_\Omega \nabla(u-w) \cdot \nabla v \mathrm{d}x \right| \leq |u-w|_{1,2,\Omega} |v|_{1,2,\Omega}.$$

Without loss of generality let $u \leq w$, then the second term can be rearranged using

$$|u|^\alpha u - |w|^\alpha w = \int_u^w \frac{\mathrm{d}}{\mathrm{d}t} (|t|^\alpha t) \, \mathrm{d}t = (\alpha+1) \int_u^w |t|^\alpha \, \mathrm{d}t.$$

Since the function $|t|^\alpha$ of $t \in \mathbb{R}$ is monotone in $(-\infty, 0)$ and in $(0, \infty)$ and its global minimum is reached for $t = 0$, it follows that

$$|t|^\alpha \leq (|u|^\alpha + |w|^\alpha), \quad t \in [u,w].$$

Take any $p_1, p_2, p_3 > 1$ such that $1/p_1 + 1/p_2 + 1/p_3 = 1$. Then using these relations and Hölder inequality gives us

$$\begin{aligned} \left| \kappa \int_{\partial\Omega} (|u|^\alpha u - |w|^\alpha w) v \mathrm{d}S \right| &\leq \kappa(\alpha+1) \int_{\partial\Omega} (|u|^\alpha + |w|^\alpha) |u-w| |v| \, \mathrm{d}S \\ &\leq \kappa(\alpha+1) \left(\|u\|^\alpha_{0,\alpha p_1,\partial\Omega} + \|w\|^\alpha_{0,\alpha p_1,\partial\Omega} \right) \|u-w\|_{0,p_2,\partial\Omega} \|v\|_{0,p_3,\partial\Omega}. \end{aligned}$$

The trace embedding (1.4) completes the proof of (1.14). □

We can define an operator $\mathcal{A} : H^1(\Omega) \to (H^1(\Omega))^*$ by $\langle \mathcal{A}u, v \rangle = a(u,v)$ for all $u, v \in H^1(\Omega)$. It follows from monotone operator theory, see [13], [30] and properties in Theorem 1.4 that problem (1.8) has exactly one solution.

1.4 Regularity

In the error estimates later in this text we will assume that the weak solution u belongs to a space of smoother functions than only $H^1(\Omega)$.

From this point on we will assume that the domain Ω is polygonal with N edges $\Gamma_1, \ldots, \Gamma_N$. To describe functions defined on the boundary $\partial\Omega$ we will use the trace theorem. Consider an operator T which takes functions from $C^\infty(\overline{\Omega})$ and assigns their restriction on $\partial\Omega$ and also possibly some partial derivatives in the outer normal direction on each edge in $\partial\Omega$

$$T : u \mapsto \left\{ u, \frac{\partial u}{\partial n}, \ldots, \frac{\partial^l u}{\partial n^l} \right\}.$$

It can be shown for $p \in [1,\infty)$ and $k \in \mathbb{N}$ that T has a unique continuous linear extension (also denoted by T) into a Cartesian product of Sobolev spaces on edges of the domain Ω

$$T : W^{k,p}(\Omega) \to \prod_{j=1}^{N} \prod_{i=0}^{k-1} W^{k-i-\frac{1}{p},p}(\Gamma_j).$$

This in particular means that there exists a constant c dependent only on Ω, k, p such that

$$\|u|_{\partial\Omega}\|_{k-\frac{1}{p},p,\partial\Omega} = \|u\|_{k-\frac{1}{p},p,\partial\Omega} \le c\|u\|_{k,p,\Omega}, \quad u \in W^{k,p}(\Omega). \tag{1.16}$$

The trace operator T does not map onto $\prod_{i=0}^{k-1} W^{k-i-\frac{1}{p},p}(\partial\Omega)$ because there might possibly occur jumps in derivatives in normal direction at the corners of $\partial\Omega$. However, these derivatives satisfy some compatibility conditions given in [11].

Let us consider a Neumann boundary value problem

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, \\ \frac{\partial u}{\partial n} &= g \quad \text{on } \partial\Omega, \end{aligned} \tag{1.17}$$

with a weak solution $u \in H^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v \mathrm{d}x = \int_\Omega f v \mathrm{d}x + \int_{\partial\Omega} g v \mathrm{d}S, \quad v \in H^1(\Omega). \tag{1.18}$$

Then the following theorem holds, see [18] Corollary 4.438 and [25] Corollary 8.3.3.

Lemma 1.5. *Let $u \in H^1(\Omega)$ be the weak solution given by* (1.18)*, let $f \in W^{k-2,q}(\Omega)$, $g \in W^{k-1-1/q,q}(\partial\Omega)$, where $k \ge 2$, $q > 1$, $\frac{2}{q} > k - \frac{\pi}{\omega_0}$, and ω_0 is the largest interior angle at boundary corners of $\partial\Omega$. Then $u \in W^{k,q}(\Omega)$.*

We can consider a weak solution $u \in H^1(\Omega)$ of (1.8). If we knew that the trace of $|u|^\alpha u$ was in the same space as g in Lemma 1.5, we could use this theorem for $g = \varphi + T(|u|^\alpha u)$ to obtain some regularity of u.

Lemma 1.6. *If $u \in H^1(\Omega)$, then $|u|^\alpha u \in W^{1,q}(\Omega)$ with $q = 2 - \varepsilon$, where $\varepsilon > 0$ is arbitrarily small, and thus $|u|^\alpha u|_{\partial\Omega} \in W^{1-1/q,q}(\partial\Omega)$.*

Proof. It follows from Sobolev embedding (1.3) that $H^1(\Omega) \hookrightarrow L^{(\alpha+1)q}(\Omega)$ and thus $u \in L^{(\alpha+1)q}(\Omega)$ or $|u|^\alpha u \in L^q(\Omega)$ for any $q \in [1,\infty)$. It remains to estimate the first weak derivatives of $|u|^\alpha u$ using Hölder inequality

$$\begin{aligned} \int_\Omega |\nabla (|u|^\alpha u)|^q \mathrm{d}x &= (\alpha+1)^q \int_\Omega |u|^{\alpha q} |\nabla u|^q \mathrm{d}x \\ &\le (\alpha+1)^q \left\| |u|^{\alpha q} \right\|_{0,s,\Omega} \left\| |\nabla u|^q \right\|_{0,s',\Omega}, \end{aligned} \tag{1.19}$$

where $\frac{1}{s} + \frac{1}{s'} = 1$. By choosing $s' = \frac{2}{2-\varepsilon} > 1$ we find that $qs' = 2$ and $\alpha q s < \infty$, and $|u|^\alpha u \in W^{1,q}(\Omega)$. □

Remark. It also directly follows from embedding (1.3) that $W^{1,q}(\Omega) \hookrightarrow C(\overline{\Omega})$ for $q > 2$ and the computation in this proof gives us: If $u \in W^{1,q}(\Omega)$, then also $|u|^\alpha u \in W^{1,q}(\Omega)$.

Using these tools, it can be shown that the following regularity result holds, see [11].

Theorem 1.7. *Let $u \in H^1(\Omega)$ be a weak solution of* (1.8) *in a polygonal domain Ω. Let $f \in L^q(\Omega)$, $\varphi \in W^{1-1/q,q}(\partial\Omega)$, where*

$$\begin{aligned} q &= 1 + \frac{\pi}{2\omega_0 - \pi} - \varepsilon < 2 \quad for\ \omega_0 > \pi, \\ q &= 1 + \frac{\pi}{2\omega_0 - \pi} - \varepsilon > 2 \quad for\ \frac{\pi}{2} < \omega_0 < \pi, \\ &q \text{ is arbitrary} \qquad\qquad for\ \omega_0 \le \frac{\pi}{2}, \end{aligned} \tag{1.20}$$

and $\varepsilon > 0$ is arbitrarily small. Then $u \in W^{2,q}(\Omega)$.

Since all inner angles ω in Ω are less than 2π, it follows that $q > \frac{4}{3}$.

Lemma 1.8. *Let $u \in W^{k,q}(\Omega)$. Let $\beta = (\beta_1, \beta_2)$ be a multi-index with $\beta_1, \beta_2 \in \mathbb{N}_0$ such that $|\beta| = \beta_1 + \beta_2 \le k$. Then*

$$D^\beta \left(|u|^\alpha u\right) = \frac{\partial^{|\beta|}(|u|^\alpha u)}{\partial x_1^{\beta_1} \partial x_2^{\beta_2}}$$

can be expressed as a finite sum of terms of a form

$$c|u|^{\alpha+1-J} \prod_{j=1}^{J} D^{\gamma_j} u, \tag{1.21}$$

where c is a constant dependent on α and β, $J \in \mathbb{N}$ and γ_j, $j = 1, \ldots, J$ are multi-indices such that $\sum_{j=1}^{J} \gamma_j = \beta$. If $\alpha \in \mathbb{N}$, then $D^\beta(|u|^\alpha u)$ only contains terms with nonnegative exponent $\alpha + 1 - J$.

Proof. Let k, q be given. We will proceed using induction on $|\beta|$.

When $|\beta| = 0$ the only possible term has $J = 0$ and $c = 1$.

If $|\beta| = 1$, then $c = \alpha + 1$, $J = 1$, and either $\gamma_1 = (1, 0)$ or $\gamma_1 = (0, 1)$.

Suppose that that the lemma holds for all multi-indices with length smaller than $|\beta|$. In particular, we have

$$D^\beta \left(|u|^\alpha u\right) = \frac{\partial \left(D^{\beta'} \left(|u|^\alpha u\right)\right)}{\partial x_i}$$

for some $i \in \{1, 2\}$ and β' such that $|\beta| = |\beta'| + 1$. Then we only need to apply $\frac{\partial}{\partial x_i}$ to terms $c|u|^{\alpha+1-J'} \prod_{j=1}^{J'} D^{\gamma'_j} u$, which have $\sum_{j=1}^{J'} \gamma'_j = \beta'$. If the partial derivative $\frac{\partial}{\partial x_i}$ is applied to any factor in $\prod_{j=1}^{J'} D^{\gamma'_j} u$, then the resulting term does have the desired form with $J = J'$, one of the multi-indices γ'_j increased, and $\sum_{j=1}^{J} \gamma_j = \beta$. If $\frac{\partial}{\partial x_i}$ is applied to $|u|^{\alpha+1-J'}$, then the resulting term has $J = J' + 1$, $\sum_{j=1}^{J'} \gamma'_j + \gamma_{J'+1} = \beta$, where $\gamma_{J'+1}$ is either $(1, 0)$ or $(0, 1)$ depending on x_i, and therefore also has the desired form.

Suppose that $\alpha \in \mathbb{N}$. Then the exponent $\alpha + 1 - J$ in $|u|^{\alpha+1-J}$ is integer for any J. The only possibility to obtain a negative exponent in the induction step would be to apply $\frac{\partial}{\partial x_i}$ to $|u|^{\alpha+1-J'}$ for J' such that $\alpha + 1 - J' \in [0, 1)$, i.e. $\alpha + 1 - J' = 0$. But then $\frac{\partial |u|^0}{\partial x_i} = 0$ and the constant c would in fact be zero. □

Lemma 1.9. *Let $u \in W^{k,q}(\Omega)$, where $k \geq 2$, $q > 1$. Let Ω be a polygonal domain. Let $\alpha+1 \geq k$ or $\alpha \in \mathbb{N}_0$. Then $|u|^\alpha u|_{\partial\Omega} \in W^{k-1/q,q}(\partial\Omega)$ and there holds an estimate*

$$\| |u|^\alpha u\|_{k-1/q,q,\partial\Omega} \leq c\|u\|_{k,q,\Omega}^{\alpha+1} \tag{1.22}$$

with a constant $c > 0$ dependent on Ω, k, q, α.

Proof. We will prove that $|u|^\alpha u \in W^{k,q}(\Omega)$. Consider any multi-index $\beta = (\beta_1, \beta_2)$ such that $|\beta| = \beta_1 + \beta_2 \leq k$. Our goal is to show that

$$D^\beta\left(|u|^\alpha u\right) = \frac{\partial^{|\beta|}(|u|^\alpha u)}{\partial x_1^{\beta_1}\partial x_2^{\beta_2}} \in L^q(\Omega).$$

The expression $D^\beta(|u|^\alpha u)$ is a sum of several terms of the form (1.21) given in Lemma 1.8. Due to the triangle inequality in Lebesgue spaces, we only need to show that all of these terms belong to the space $L^q(\Omega)$ and are estimated by the right-hand side of (1.22). The assumption $\alpha+1 \geq k$ or $\alpha \in \mathbb{N}_0$ guarantees that the exponents $\alpha+1-J$ in (1.21) are nonnegative for all terms. Since $u \in W^{k,q}(\Omega) \hookrightarrow C^{k-2}(\overline{\Omega})$, we can trivially estimate terms which only have derivatives of orders up to $k-2$

$$c\left\| |u|^{\alpha+1-J}\prod_{j=1}^{J} D^{\gamma_j}u\right\|_{0,q,\Omega} \leq c\,\|u\|_{k,q,\Omega}^{\alpha+1}.$$

Consider the term

$$c|u|^\alpha D^\beta u.$$

Since $u \in W^{k,q}(\Omega) \hookrightarrow C(\overline{\Omega})$ and $D^\beta u \in L^q(\Omega)$, we have

$$\left\| |u|^\alpha\, D^\beta u\right\|_{0,q,\Omega}^q = \int_\Omega |u|^{\alpha q}|D^\beta u|^q \mathrm{d}x \leq \|u\|_{C(\overline{\Omega})}^{\alpha q}\int_\Omega |D^\beta u|^q \mathrm{d}x \leq c\|u\|_{k,q,\Omega}^{\alpha q+q}.$$

The only remaining terms are

$$c|u|^{\alpha-1}\prod_{j=1}^{2} D^{\gamma_j}u,$$

where γ_1 has length 1 and γ_2 has length $k-1$. If $k \geq 3$, then we again estimate

$$\left\| |u|^{\alpha-1}\prod_{j=1}^{2} D^{\gamma_j}u\right\|_{0,q,\Omega}^q \leq \|u\|_{C(\overline{\Omega})}^{(\alpha-1)q}\, \| |\nabla u| \|_{C(\overline{\Omega})}\int_\Omega |D^{\gamma_2}u|^q \mathrm{d}x \leq c\|u\|_{k,q,\Omega}^{\alpha q+q}.$$

If $k = 2$, γ_2 has also length 1, and we can use embedding (1.3)

$$\left\| |u|^{\alpha-1}\prod_{j=1}^{2} D^{\gamma_j}u\right\|_{0,q,\Omega}^q \leq \|u\|_{C(\overline{\Omega})}^{(\alpha-1)q}\, \| |\nabla u| \|_{0,2q,\Omega}^{2q} \leq c\|u\|_{k,q,\Omega}^{\alpha q+q},$$

where the last inequality holds because

- $W^{1,q}(\Omega) \hookrightarrow C(\overline{\Omega})$ for $q > 2$,
- $H^1(\Omega) \hookrightarrow L^4(\Omega)$ for $q = 2$,

- $W^{1,q}(\Omega) \hookrightarrow L^{2q}(\Omega)$ for $q \in [1,2)$ as $\frac{2q}{2-q} \geq 2q$.

Thus $|u|^\alpha u \in W^{k,q}(\Omega)$ and its trace satisfies (1.22). $\square$

Functions in $W^{2,q}(\Omega)$ are continuous. Therefore, it is possible to distinguish on which parts of the boundary $\partial\Omega$ is the weak solution u nonzero.

Lemma 1.10. *Let $u \in W^{k,q}(\Omega)$, where $k \geq 2$, $q > 1$ and Ω is a polygonal domain. Let $\alpha + 1 < k$. Let G be a closed subset of $\partial\Omega$. If $|G| > 0$ and $|u| \geq \varepsilon > 0$ on G, then $|u|^\alpha u|_G \in W^{k-1/q,q}(G)$.*

Proof. Function u is continuous in Ω. Therefore, we can find an open neighbourhood of G in Ω denoted by Ω_G such that $|u| \geq \varepsilon > 0$ in Ω_G. We can proceed similarly to the proof of Lemma 1.9. This time we cannot guarantee that the exponents $\alpha + 1 - J$ are non-negative. The lowest possible negative exponent is $\alpha + 1 - k$ and the same arguments as in the proof of Lemma 1.9 lead to an estimate

$$\left\| |u|^\alpha u \right\|_{k,q,\Omega_G} \leq c \left(\|u\|_{k,q,\Omega_G}^{\alpha+1} + \varepsilon^{\alpha+1-k} \|u\|_{k,q,\Omega_G}^{k} \right), \tag{1.23}$$

where c is dependent also on Ω_G and thus possibly on both G and u. $\square$

Remark. If we knew beforehand that u was sufficiently distant from zero on sufficiently large part of the boundary $\partial\Omega$, we could obtain estimate similar to (1.22) for arbitrary $\alpha \geq 0$.

Using Lemma 1.5 it was shown in [11] that the following lemma holds.

Lemma 1.11. *Let the assumptions of Theorem 1.7 be satisfied and let $f \in W^{k,q}(\Omega)$ for $k \geq 1$. Then $u \in W^{k+2,q}(\Omega_0)$, where Ω_0 is a subdomain of Ω with smooth boundary and $\overline{\Omega_0} \subset \Omega$.*

A similar regularity result to Lemma 1.5 holds for domains with smooth boundary, see [17] Theorem 2.5.1.1.

Lemma 1.12. *Let Ω_2 be a domain with smooth boundary, let $f \in W^{k-2,q}(\Omega_2)$, $g \in W^{k-1-1/q,q}(\partial\Omega_2)$, where $k \in \mathbb{N}$, $k \geq 2$, $q > 1$. Let $u \in H^1(\Omega_2)$ be a weak solution of a problem*

$$\int_{\Omega_2} \nabla u \cdot \nabla v \mathrm{d}x = \int_{\Omega_2} f v \mathrm{d}x + \int_{\partial\Omega_2} g v \mathrm{d}S, \quad v \in H^1(\Omega). \tag{1.24}$$

Then $u \in W^{k,q}(\Omega)$.

By using Lemma 1.12, Lemma 1.9 and Lemma 1.10, we can improve regularity results in Lemma 1.11 to include nonsingular parts of the boundary $\partial\Omega$.

Theorem 1.13. *Let Ω be a polygonal domain. Let $k \in \mathbb{N}$, $k \geq 2$, $q \geq 1$. Let $\alpha \geq 0$ and $\kappa > 0$ the be constants, and $f \in W^{k-2,q}(\Omega)$ and $\varphi \in W^{k-1-\frac{1}{q},q}(\partial\Omega)$ the functions from problem* (1.1)-(1.2). *Let u be a weak solution defined in* (1.8). *Let G be a closed subset of $\partial\Omega$ not containing any boundary vertices of Ω. If $\alpha \notin \mathbb{N}_0$ and $\alpha + 1 < k$, let u restricted to G satisfy $|u| \geq \varepsilon$ for some constant $\varepsilon > 0$. Let $\Omega_1 \subset \Omega$ be a domain with smooth boundary such that $\partial\Omega \cap \partial\Omega_1 \subset G$, i.e. Ω_1 shares boundary with Ω only in G. Then $u \in W^{k,q}(\Omega_1)$.*

Proof. It follows from Theorem 1.7 that $u \in W^{2,p}(\Omega)$ for some $p > \frac{4}{3}$. Due to embedding (1.3), we have $W^{2,p}(\Omega) \hookrightarrow C(\overline{\Omega})$ and the weak solution u is continuous. Therefore, we can find domain Ω_2 with smooth boundary such that $\Omega \supset \Omega_2 \supset \Omega_1$, $\partial\Omega_2$ does not contain boundary vertices of $\partial\Omega$, Ω_2 is some neighbourhood of $\overline{\Omega_1}$ in Ω, and $|u| \geq \frac{\varepsilon}{2}$ on $G_2 = \partial\Omega \cap \partial\Omega_2$, $\alpha \notin \mathbb{N}_0$ and $\alpha + 1 < k$.

We will proceed using induction on k. The base case is Theorem 1.7. Let us suppose that $u \in W^{k-1,q}(\Omega_2)$, we will show that $u \in W^{k,q}(\Omega_1)$.

It follows from Lemma 1.9 and Lemma 1.10 for $k \geq 3$ and from Lemma 1.6 for $k = 2$ that $|u|^\alpha u \in W^{k-1-\frac{1}{q},q}(G_2)$. Let us define $\psi \in C^\infty(\overline{\Omega})$ in the following way:

- $\psi = 1$ in Ω_1
- $\psi = 0$ in $\Omega \setminus \overline{\Omega_2}$
- $\frac{\partial\psi}{\partial n} = 0$ on G_2

Since we have $\varphi \in W^{k-1-\frac{1}{q},q}(G_2)$, $|u|^\alpha u \in W^{k-1-\frac{1}{q},q}(G_2)$, and ψ is smooth, we also have

$$\frac{\partial(\psi u)}{\partial n} = \psi \frac{\partial u}{\partial n} = \psi \left(-\kappa |u|^\alpha u + \varphi\right) \in W^{k-1-\frac{1}{q},q}(G_2). \tag{1.25}$$

As $\psi u = 0$ on $\partial\Omega_2 \setminus G_2$, we have $\frac{\partial(\psi u)}{\partial n} \in W^{k-1-\frac{1}{q},q}(\partial\Omega_2)$. We also have

$$-\Delta(\psi u) = -u\Delta\psi - 2\nabla\psi \cdot \nabla u - \psi\Delta u \in W^{k-2,q}(\Omega_2), \tag{1.26}$$

since $u\Delta\psi \in W^{k-1,q}(\Omega_2)$, $\nabla\psi \cdot \nabla u \in W^{k-2,q}(\Omega_2)$, and$-\psi\Delta u = \psi f \in W^{k-2,q}(\Omega_2)$. Therefore, $\psi u \in W^{k,q}(\Omega_2)$ follows from Lemma 1.12. Because $u = \psi u$ in Ω_1, we have $u \in W^{k,q}(\Omega_1)$. □

We can conclude that if the right-hand side functions f, φ from (1.1)-(1.2) are smooth enough, the weak solution u defined in (1.8) belongs to Sobolev spaces:

- $W^{2,q}$ in a neighbourhood of boundary corners, where q is dependent on the inner angles at the corners,
- $W^{\lfloor\alpha\rfloor+2,q}$ in a neighbourhood of roots of u on the boundary $\partial\Omega$ sides (not in any corner) for noninteger $\alpha \geq 0$,
- $W^{k,q}$ in the rest of Ω.

2. Finite element discretization

2.1 Discretization

We assume that the domain $\Omega \subset \mathbb{R}^2$ is polygonal. We construct its triangulation $\mathcal{T}_h$ consisting of a finite number of closed triangles T. We will consider only conforming triangulations satisfying the following conditions:

$$\begin{aligned}&\overline{\Omega} = \bigcup_{T\in\mathcal{T}_h} T,\\ &\text{if } T_1, T_2 \in \mathcal{T}_h,\ T_1 \neq T_2,\ \text{then } T_1 \cap T_2 = \emptyset,\ \text{or } T_1 \cap T_2\\ &\quad \text{is either a common vertex or a common side of } T_1 \text{ and } T_2.\end{aligned} \tag{2.1}$$

We say that $T \in \mathcal{T}_h$ is a boundary triangle, if T has a side $S \subset \partial\Omega$ and we denote the set of all sides $S \subset \partial\Omega$ by s_h. Then $\bigcup_{S\in s_h} S = \partial\Omega$. For simplicity, we assume that each boundary triangle has only one boundary edge S and thus can be referred to as T_S. If a triangle is not a boundary triangle we call it an inner triangle.

By h_T and ρ_T we denote the length of the maximal side of T and the radius of the maximal circle inscribed into T, respectively. We further set

$$h = \max_{T\in\mathcal{T}_h} h_T \tag{2.2}$$

Let us consider a shape regular system of triangulations $\{\mathcal{T}_h\}_{h\in(0,h_0)}$ of the domain Ω: there exists $C_R > 0$ such that

$$\frac{h_T}{\rho_T} \leq C_R \quad \forall T \in \mathcal{T}_h\ \forall h \in (0, h_0). \tag{2.3}$$

2.2 Galerkin approximation

We will seek a Galerkin approximation and an approximate solution in a space of continuous piecewise polynomial functions of order $r \in \mathbb{N}_0$ and then use a Lagrange interpolation as a tool to show the order of its convergence.

Definition 2.1. *Let $r \in \mathbb{N}$. Let $T \in \mathcal{T}_h$ be a triangle. We denote the space of polynomials in x_1, x_2 on T of degree at most r by $P_r(T)$*

$$P_r(T) = \left\{ p_T : T \to \mathbb{R}; p_T(x_1, x_2) = \sum_{\substack{i,j\in\mathbb{N}_0\\ i+j\leq r}} a_{i,j} x_1^i x_2^j, a_{i,j} \in \mathbb{R} \right\}. \tag{2.4}$$

Let S be a side of a triangle and let F be an affine mapping of an interval $[0, 1]$ onto S. Then we denote the set of polynomials of degree at most r on S by $P_r(S)$:

$$P_r(S) = \left\{ p \circ F^{-1};\ p \,\text{is a polynomial of degree}\ \leq r \,\text{on}\, [0, 1] \right\}.$$

Let $\mathcal{T}_h$ be a triangulation of Ω. The set of all continuous piecewise polynomial functions of degree at most r is

$$H_h^r = \left\{ v_h \in C(\overline{\Omega});\ v_h|_T \in P_r(T),\ T \in \mathcal{T}_h \right\}, \tag{2.5}$$

and the space of all piece-wise polynomial function of degree at most r is

$$S_h^r = \left\{ v_h \in L^2(\Omega);\ v_h|_T \in P_r(T),\ T \in \mathcal{T}_h \right\}. \tag{2.6}$$

Definition 2.2. *Let $\hat{T}$ be a closed reference triangle with vertices $(0,0)$, $(1,0)$, $(0,1)$, let $u \in C(\hat{T})$, let $r \in \mathbb{N}$, let $x_1, \ldots, x_{(r+1)(r+2)/2} \in T$ be pairwise different nodes in $\hat{T}$. Then we call a projection π of $C(\hat{T})$ onto $P_r(\hat{T})$ a Lagrange interpolation of order r if $u(x_\mu) = \pi u(x_\mu)$ for all $\mu = 1, \ldots, \frac{(r+1)(r+2)}{2}$.*

Let $u \in C(\overline{\Omega})$, let $r \in \mathbb{N}$, let π be a Lagrange interpolation of order r on a reference triangle $\hat{T}$ with vertices $(0,0)$, $(1,0)$, $(0,1)$. Let $\mathcal{T}_h$ be a triangulation of Ω, let F_T be an affine mapping of $\hat{T}$ onto T for each $T \in \mathcal{T}_h$. Then we call a projection π_h of $C(\overline{\Omega})$ onto H_h^r a piecewise Lagrange interpolation, if π_h restricted to T is given by $F_T \circ \pi \circ F_T^{-1}$ for all $T \in \mathcal{T}_h$.

Remark. The choice of $\frac{(r+1)(r+2)}{2}$ nodes is necessary for the existence and uniqueness of the interpolation, but the nodes must also satisfy some other conditions, see [4]. If there are $r+1$ nodes on each side of a triangle, then the interpolated function πu on that side will be given only by the values of u on that side, i.e. the piecewise Lagrange interpolation π_h will preserve continuity of u.

Now we can define a Galerkin approximation U_h of a solution u.

Definition 2.3. *We say that $U_h \in H_r^h$ is a Galerkin approximation of a weak solution $u \in H^1(\Omega)$ given by (1.8) if*

$$a(U_h, v_h) = L(v_h) \quad \forall v_h \in H_h^r. \tag{2.7}$$

Since H_r^h is a subset of $H^1(\Omega)$, it follows that the form a has all the properties in Theorem 1.4, and existence and uniqueness of the Galerkin approximation again follow from monotone operator theory in [13] and [30]. We can further improve the monotonicity of a by assuming that one of the functions in question is not close to zero on a part of $\partial\Omega$. More precisely:

$$\begin{aligned} &G \subset \partial\Omega, \quad |G| > 0, \\ &|u| \geq \varepsilon > 0 \quad \text{on } G. \end{aligned} \tag{2.8}$$

Theorem 2.4. *Let $u \in H^1(\Omega)$ and let the conditions (2.8) hold. Then there exists a constant $C_3 = C_3(\Omega, G, \varepsilon) > 0$ such that*

$$a(u, u-v) - a(v, u-v) \geq C_3 \|u - v\|_{1,2,\Omega}^2 \quad \forall v \in H^1(\Omega). \tag{2.9}$$

Proof. Since $|u|^\alpha - |v|^\alpha$ and $u^2 - v^2$ have the same sign, it follows that $(|u|^\alpha - |v|^\alpha)(u^2 - v^2) \geq 0$, or equivalently $|u|^\alpha u^2 + |v|^\alpha v^2 \geq |u|^\alpha v^2 + |v|^\alpha u^2$. Thus, we can write

$$\begin{aligned} 2(|u|^\alpha u - |v|^\alpha v)(u - v) &= |u|^\alpha (2u^2 - 2uv) + |v|^\alpha (2v^2 - 2uv) \\ &\geq |u|^\alpha (u^2 - 2uv + v^2) + |v|^\alpha (v^2 - 2uv + u^2) = (|u|^\alpha + |v|^\alpha)(u - v)^2. \end{aligned} \tag{2.10}$$

From this and equation (1.9) it directly follows that

$$a(u, u-v) - a(v, u-v) \geq |u - v|_{H^1(\Omega)}^2 + \frac{1}{2}\kappa\, |G|\, \varepsilon^\alpha \|u - v\|_{0,2,G}^2.$$

The existence of a constant C_3 from the statement of this theorem follows from Poincaré inequality (1.5). □

Under the conditions (2.8), we can redefine ϱ from (1.12), (1.13) as

$$\varrho(t) = C_3 t^2 \quad t \in [0, \infty), \tag{2.11}$$

but numerical experiments indicate that if u is close to zero on a large part of the boundary $\partial\Omega$, ϱ behaves as in (1.13), that is, C_3 is very small.

2.3 Error estimation

Our next goal is to show how fast the Galerkin approximation U_h converges to the weak solution u. Let us suppose that $u \in W^{k,q}(\Omega)$ for some $k \in \mathbb{N}$, $k \geq 2$, $q \geq 1$. Due to the embedding $W^{k,q}(\Omega) \hookrightarrow C(\overline{\Omega})$, the piecewise Lagrange interpolation is well-defined.

Theorem 2.5. *Let* $u \in H^1(\Omega)$ *be a weak solution of* (1.8)*, let* $U_h \in H_h^r$ *be the Galerkin approximation defined by* (2.7)*. Then*

$$\varrho_1(\|u - U_h\|_{1,2,\Omega}) \leq C \inf_{v_h \in H_h^r} \|u - v_h\|_{1,2,\Omega}, \tag{2.12}$$

where

$$\varrho_1(t) = \varrho(t)/t. \tag{2.13}$$

Proof. The Galerkin approximation (2.7) is given by the same formula as a weak solution defined in (1.8), but only on a restriction from $H^1(\Omega)$ to H_h^r. This means that Theorem 1.4 holds again and the Galerkin approximation exists and is unique. Then

$$\begin{aligned}\varrho(\|U_h\|_{1,2,\Omega}) &\leq a(U_h, U_h) = L(U_h) \\ &\leq c((\|f\|_{0,2,\Omega} + \|\varphi\|_{0,2,\partial\Omega})\, \|U_h\|_{1,2,\Omega},\end{aligned}$$

where we used trace theorem in the last inequality. This shows that $\varrho_1(\|U_h\|_{1,2,\Omega})$ is bounded independently of h and U_h is uniformly bounded. Another consequence of formulas (2.7) and (1.8) is that

$$a(u, v_h) = L(v_h) = a(U_h, v_h) \quad \forall v_h \in H_h^r,$$

which implies Galerkin orthogonality

$$a(u, v_h) - a(U_h, v_h) = 0.$$

Since $U_h \in H_h^r$, we have

$$a(u, U_h - v_h) - a(U_h, U_h - v_h) = 0$$

for all $v_h \in H_h^r$, which can be rearranged into

$$a(u, u - U_h) - a(U_h, u - U_h) = a(u, u - v_h) - a(U_h, u - v_h) \quad \forall v_h \in H_h^r. \tag{2.14}$$

Then, by (1.12), (2.14) and (1.14) for arbitrary $v_h \in H_h^r$, we have

$$\begin{aligned}\varrho(\|u - U_h\|_{1,2,\Omega}) &\leq a(u, u - U_h) - a(U_h, u - U_h) \\ &= |a(u, u - U_h) - a(U_h, u - U_h)| \\ &= |a(u, u - v_h) - a(U_h, u - v_h)| \\ &\leq C_1 \left(1 + \|u\|_{1,2,\Omega}^\alpha + \|U_h\|_{1,2,\Omega}^\alpha\right) \|u - U_h\|_{1,2,\Omega} \|u - v_h\|_{1,2,\Omega}.\end{aligned} \tag{2.15}$$

Dividing (2.15) by $\|u - U_h\|_{1,2,\Omega}$ and using the uniform boundedness of U_h, we obtain the sought inequality. ☐

In what follows, we will use Theorem 3.1.5 from [4]:

Theorem 2.6. *Let $r, m \in \mathbb{N}_0$, $p, q \geq 1$. Let the piecewise Lagrange interpolation π_h preserve polynomials of degree at most r. Let the the triangulation $\mathcal{T}_h$ be shape regular according to* (2.3)*. Let the following embeddings hold:*

$$W^{r+1,q}(T) \hookrightarrow C(T),$$
$$W^{r+1,q}(T) \hookrightarrow W^{m,p}(T).$$

Then there exists a constant $C_4 = C_4(\pi, C_R) > 0$ such that for all $T \in \mathcal{T}_h$ and $h \in (0, h_0)$ we have

$$|u - \pi_h u|_{m,p,T} \leq C_4 |u|_{r+1,q,T} h_T^{r+1-m+\frac{2}{p}-\frac{2}{q}} \quad \forall u \in W^{r+1,q}(T). \tag{2.16}$$

Note that the terms $\frac{2}{p} - \frac{2}{q}$ in the exponent of h_T correspond to the fact that we are considering a two-dimensional case and the theorem also holds for any other dimension. We will use this theorem mainly to estimate error of interpolation measured in H^1-norms or seminorms.

Corollary. Let $k \in \mathbb{N}$, $q \geq 1$. Let the piecewise Lagrange interpolation π_h preserve polynomials of degree $\leq r$. Set $\nu = \min(r, k)$. Let the triangulation $\mathcal{T}_h$ be shape-regular according to (2.3). Then there exists a constant $C_6 = C_6(\pi, C_R) > 0$ such that

$$\|u - \pi_h u\|_{1,2,T} \leq C_6 |u|_{\nu+1,q,T} h_T^{\nu+1-2/q} \quad \forall u \in W^{k+1,q}(T), \forall T \in \mathcal{T}_h, \forall h \in (0, h_0).$$

Lemma 2.7. *Let $\beta \geq 1$, $n \in \mathbb{N}$, $x_i \geq 0$, $w_i > 0$, $i = 1, \dots, n$. Then the following inequalities hold:*

$$\sum_{i=1}^{n} x_i^\beta \leq \left(\sum_{i=1}^{n} x_i \right)^\beta, \tag{2.17}$$

$$\left(\frac{\sum_i w_i x_i}{\sum_i w_i} \right)^\beta \leq \frac{\sum_i w_i x_i^\beta}{\sum_i w_i}. \tag{2.18}$$

Proof. We will prove (2.17) using induction. For $n = 1$ both sides are equal. Suppose that the inequality holds for $n - 1$. Then (2.17) holds for n, if we have $x_n = 0$. To complete the induction step, it suffices to show that the left-hand side has a lower or equal derivative with respect to x_n than the right-hand side.

The derivative of the left-hand side is

$$\beta x_n^{\beta-1},$$

and the derivative of the right-hand side is

$$\beta \left(\sum_{i=1}^{n} x_i \right)^{\beta-1}.$$

As $x_n \leq \sum_{i=1}^{n} x_i$ and $\beta x^{\beta-1}$ is a non-decreasing function, the inequality between the derivatives does hold. Therefore the inequality (2.17) holds for n and arbitrary $x_n \geq 0$.

Inequality (2.18) is known as Jensen's inequality, see Theorem 3.3 in [29]. □

Recall that the inverse of the function ϱ_1 arising from the monotonicity is

$$\varrho_1^{-1}(t) = \begin{cases} \left(\frac{t}{C_0\kappa 2^{-\alpha}}\right)^{\frac{1}{\alpha+1}} & \text{for} \quad 0 \le t \le 2\kappa^{-1/\alpha}, \\ \frac{t}{C_0} & \text{for} \quad t \ge 2\kappa^{-1/\alpha}, \end{cases}, \tag{2.19}$$

which can be replaced under conditions (2.8) by

$$\varrho_1^{-1}(t) = \frac{t}{C_3}, \tag{2.20}$$

as follows from (2.11).

Theorem 2.8. *Let the solution of* (1.8) *be* $u \in W^{k+1,q}(\Omega)$*, where* $W^{k+1,q}(\Omega) \hookrightarrow H^1(\Omega)$*. Then*

$$\|u - U_h\|_{1,2,\Omega} \le \begin{cases} \varrho_1^{-1}\left(c|u|_{k+1,q,\Omega}h^{\nu+1-\frac{2}{q}}\right), & q \in [1,2), \\ \varrho_1^{-1}\left(c|u|_{k+1,q,\Omega}h^{\nu}\right), & q \in [2,\infty). \end{cases} \tag{2.21}$$

Proof. Using Theorem 2.5 for $v_h = \pi_h u$ and Theorem 2.6, we obtain

$$\begin{aligned} \varrho_1(\|u - U_h\|_{1,2,\Omega}) &\le c\|u - \pi_h u\|_{1,2,\Omega} = c\left(\sum_{T\in\mathcal{T}_h} \|u - \pi_h u\|_{1,2,T}^2\right)^{1/2} \\ &\le c\left(\sum_{T\in\mathcal{T}_h} |u|_{k+1,q,T}^2 h_T^{2\nu+2-4/q}\right)^{1/2}. \end{aligned} \tag{2.22}$$

For $q < 2$, we use (2.17) with $\beta = \frac{2}{q}$, $x_i = |u|_{k+1,q,T}^q h_T^{q\nu+q-2}$:

$$\sum_{T\in\mathcal{T}_h} |u|_{k+1,q,T}^2 h_T^{2\nu+2-4/q} \le \left(\sum_{T\in\mathcal{T}_h} |u|_{k+1,q,T}^q h_T^{q\nu+q-2}\right)^{2/q},$$

$$\left(\sum_{T\in\mathcal{T}_h} |u|_{k+1,q,T}^2 h_T^{2\nu+2-4/q}\right)^{1/2} \le \left(\sum_{T\in\mathcal{T}_h} |u|_{k+1,q,T}^q h_T^{q\nu+q-2}\right)^{1/q} \le |u|_{k+1,q,\Omega}h^{\nu+1-2/q}.$$

Inequality (2.18) can be rewritten as

$$\left(\sum_i w_i x_i\right)^{\beta} \le \left(\sum_i w_i x_i^{\beta}\right)\left(\sum_i w_i\right)^{\beta-1}.$$

For $q \ge 2$ we use this inequality with $\beta = \frac{q}{2}$, $w_i = h_T^2$, $x_i = |u|_{k+1,q,T}^2 h_T^{2\nu-4/q}$:

$$\left(\sum_{T\in\mathcal{T}_h} |u|_{k+1,q,T}^2 h_T^{2\nu+2-4/q}\right)^{1/2} \le \left(\sum_{T\in\mathcal{T}_h} h_T^2|u|_{k+1,q,T}^q h_T^{q\nu-2}\right)^{\frac{1}{q}}\left(\sum_{T\in\mathcal{T}_h} h_T^2\right)^{\frac{1}{2}-\frac{1}{q}}. \tag{2.23}$$

Let o_T be a sum of lengths of sides of a triangle $T \in \mathcal{T}_h$, then we clearly have $h_T \le \frac{o_T}{2}$ and $|T| = \frac{1}{2}o_T\rho_T$. This together with the shape regularity condition (2.3) gives us

$$h_T^2 \le C_R h_T\rho_T \le C_R\frac{1}{2}o_T\rho_T = C_R|T|,$$

which can be summed over all $T \in \mathcal{T}_h$

$$\sum_{T \in \mathcal{T}_h} h_T^2 \leq C_R \, |\Omega| \, . \tag{2.24}$$

Combining (2.22), (2.23) and (2.24) gives us

$$\begin{aligned} \varrho_1(\|u - U_h\|_{1,2,\Omega}) &\leq \left(\sum_{T \in \mathcal{T}_h} h_T^2 |u|_{k+1,q,T}^q h_T^{q\nu - 2} \right)^{\frac{1}{q}} \left(\sum_{T \in \mathcal{T}_h} h_T^2 \right)^{\frac{1}{2} - \frac{1}{q}} \\ &\leq c(\Omega, C_R, p) |u|_{k+1,q,\Omega} h^\nu, \end{aligned}$$

and completes the proof. □

Remark. The arguments from this proof can be repeated for different values of ν on different parts of Ω. Considering what regularity was shown in Section 1.4, the error estimates would require triangles near vertices to have length at most ch^ν (or $ch^{\frac{\nu+1-2/q}{2-2/q}}$) and the triangles near roots of u on boundary edges for noninteger $\alpha > 0$ to have length at most $ch^{\frac{\nu}{\lfloor \alpha \rfloor + 1}}$ (or $ch^{\frac{\nu+1-2/q}{\lfloor \alpha \rfloor + 2 - 2/q}}$).

So far, we have shown that the $H^1(\Omega)$-norm of the error of the Galerkin approximation converges to zero with a rate of convergence r, if the approximation uses continuous piecewise polynomial functions of degree r, if the exact solution is sufficiently smooth and if the exact solution is distant from zero on a part of the boundary. Without the assumption (2.8), the order of convergence is divided by $\alpha + 1$. Next, we will show that if the exact solution is zero on the whole boundary, we can improve the estimate for the $H^1(\Omega)$-seminorm.

Theorem 2.9. *Let the weak solution $u \in W^{k+1,q}(\Omega) \hookrightarrow H^1(\Omega)$ given by (1.8) be zero on $\partial\Omega$. Then*

$$|u - U_h|_{1,2,\Omega} \leq \begin{cases} c|u|_{k+1,q,\Omega} h^{\nu+1-\frac{2}{q}}, & q \in [1, 2), \\ c|u|_{k+1,q,\Omega} h^\nu, & q \in [2, \infty), \end{cases} \tag{2.25}$$

where $\nu = \min(r, k)$ and r is the degree of used polynomials.

Proof. Neglecting a term in (1.11) gives us

$$|u - U_h|_{1,2,\Omega}^2 \leq a(u, u - U_h) - a(U_h, u - U_h),$$

using Galerkin orthogonality (2.14) for a piecewise Lagrange interpolation yields

$$a(u, u - U_h) - a(U_h, u - U_h) = a(u, u - \pi_h u) - a(U_h, u - \pi_h u).$$

Using the fact that $\pi_h u$ is also zero on $\partial\Omega$ and Hölder inequality gives us

$$\begin{aligned} a(u, u - \pi_h u) - a(U_h, u - \pi_h u) &= \int_\Omega \nabla(u - U_h) \cdot \nabla(u - \pi_h u) \mathrm{d}x \\ &\leq |u - U_h|_{1,2,\Omega} \, |u - \pi_h u|_{1,2,\Omega} \, . \end{aligned}$$

Dividing by $|u - U_h|_{1,2,\Omega}$ leads to an estimate

$$|u - U_h|_{1,2,\Omega} \leq |u - \pi_h u|_{1,2,\Omega} \, .$$

Using Theorem 2.6 for $H^1(T)$-seminorm instead of a norm and the same arguments as in the proof of Theorem 2.8 gives us the sought estimate. □

3. Discrete problem with numerical integration

3.1 Quadrature formula

In practical computation, integrals in the definition of the forms are evaluated by numerical integration. In this section, we are concerned with the analysis of the effect of numerical integration.

Consider a reference triangle $\hat{T}$ with vertices $(0,0)$, $(1,0)$, $(0,1)$. We approximate an integral of a continuous function $\hat{\psi}$ over $\hat{T}$ using values at M different points x_μ and M weights ω_μ, $\mu = 1,\dots,M$. Considering that the area of $\hat{T}$ is $1/2$, we then have

$$\int_{\hat{T}} \hat{\psi}\mathrm{d}x \approx \frac{1}{2}\sum_{\mu=1}^{M} \omega_\mu \hat{\psi}(x_\mu). \tag{3.1}$$

Any triangle T can be obtained using an affine function F_T, such that $F_T(\hat{T}) = T$. Then the nodes become $x_{T,\mu} = F(x_\mu)$, $\mu = 1,\dots,M$ and we obtain a quadrature formula for a function ψ defined on T

$$\int_T \psi\mathrm{d}x \approx |T|\sum_{\mu=1}^{M} \omega_\mu \psi(x_{T,\mu}), \quad T \in \mathcal{T}_h. \tag{3.2}$$

Analogically, we introduce numerical integration over edges S on $\partial\Omega$. As a reference element, we use the interval $[0,1]$ with m nodes x_μ and weights β_μ, $\mu = 1,\dots,m$. The quadrature formula on reference interval is

$$\int_0^1 \hat{\vartheta}\mathrm{d}S \approx \sum_{\mu=1}^{m} \beta_\mu \vartheta(x_\mu), \tag{3.3}$$

and the quadrature formula on edges is

$$\int_S \vartheta\mathrm{d}S \approx |S|\sum_{\mu=1}^{m} \beta_\mu \vartheta(x_{S,\mu}), \quad S \in s_h. \tag{3.4}$$

The errors of integration are

$$\begin{aligned}
E_T(\psi) &= \int_T \psi\mathrm{d}x - |T|\sum_{\mu=1}^{M} \omega_\mu \psi(x_{T,\mu}),\\
E_S(\vartheta) &= \int_S \vartheta\mathrm{d}S - |S|\sum_{\mu=1}^{m} \beta_\mu \vartheta(x_{S,\mu}),\\
E_\Omega(\psi) &= \int_\Omega \psi\mathrm{d}x - \sum_{T\in\mathcal{T}_h}|T|\sum_{\mu=1}^{M} \omega_\mu \psi(x_{T,\mu}) = \sum_{T\in\mathcal{T}_h} E_T(\psi),\\
E_{\partial\Omega}(\vartheta) &= \int_{\partial\Omega} \vartheta\mathrm{d}S - \sum_{S\in s_h}|S|\sum_{\mu=1}^{m} \beta_\mu \vartheta(x_{S,\mu}) = \sum_{S\in s_h} E_S(\vartheta).
\end{aligned} \tag{3.5}$$

The approximations of forms are defined as

$$\begin{aligned} d_d(u,v) &= \kappa \sum_{S\in s_h} |S| \sum_{\mu=1}^{m} \beta_\mu(|u|^\alpha uv)(x_{S,\mu}), \\ L_d^{\partial\Omega}(v) &= \sum_{S\in s_h} |S| \sum_{\mu=1}^{m} \beta_\mu(\varphi v)(x_{S,\mu}), \\ L_d^{\Omega}(v) &= \sum_{T\in \mathcal{T}_h} |T| \sum_{\mu=1}^{M} \omega_\mu(fv)(x_{T,\mu}). \end{aligned} \tag{3.6}$$

We assume that the form b will be evaluated exactly as its arguments will be polynomials of order $\leq 2r-2$. Furthermore, we again define forms

$$\begin{aligned} a_d(u,v) &= b(u,v) + d_d(u,v), \\ L_d(v) &= L_d^{\Omega}(v) + L_d^{\partial\Omega}(v). \end{aligned} \tag{3.7}$$

Definition 3.1. *Let E_T be the error of numerical quadrature on a triangle $T \in \mathcal{T}_h$. We say that a quadrature on triangles is exact for polynomials of degree $\leq R$, if $E_T(v_h) = 0$ for any $v_h \in \mathcal{P}_R(T)$, $T \in \mathcal{T}_h$.*

Let E_S be the error of numerical quadrature on an edge $S \in s_h$. We say that a quadrature on edges is exact for polynomials of degree $\leq R$, if $E_S(v_h) = 0$ for any $v_h \in \mathcal{P}_R(S)$, $S \in s_h$.

We will use error estimates from Theorems 7.36 and 7.37 in [5].

Theorem 3.2. *Let $S \in s_h$. Let the quadrature formula on edges be exact for polynomials of degree $\leq r+s_1-1$. Let $q, q' \in [1,\infty]$ be such that $1/q+1/q' = 1$ (we set $1/\infty = 0$). Then there exists a constant $c > 0$ such that for any $\varphi \in W^{s_1,q}(S)$, $v_h \in \mathcal{P}_r(S)$, we have:*

$$|E_S(\varphi v_h)| \leq c|S|^{s_1} |\varphi|_{s_1,q,S} \|v_h\|_{0,q',S}. \tag{3.8}$$

Let $T \in \mathcal{T}_h$, where $\mathcal{T}_h$ are shape regular triangulations. Let the quadrature formula on triangles be exact for polynomials of degree $\leq r+s_2-1$. Let $q, q' \in [1,\infty]$ be such that $1/q + 1/q' = 1$. Then there exists a constant $c > 0$ such that for any $f \in W^{s_2,q}(T)$, $v_h \in \mathcal{P}_r(T)$, we have:

$$|E_T(fv_h)| \leq c|h_T|^{s_2} |f|_{s_2,q,T} \|v_h\|_{0,q',T}. \tag{3.9}$$

The next theorem follows directly from these estimates.

Theorem 3.3. *Let the quadrature formula on edges be exact for polynomials of degree $\leq r+s_1-1$ on each $S \in s_h$, let $q \in (1,\infty)$. Then there exists a constant $c > 0$ such that for any $\varphi \in W^{s_1,q}(\partial\Omega)$, $v_h \in H_h^r$, we have:*

$$|E_{\partial\Omega}(\varphi v_h)| \leq ch^{s_1} |\varphi|_{s_1,q,\partial\Omega} \|v_h\|_{1,2,\Omega}. \tag{3.10}$$

Let the quadrature formula on triangles be exact for polynomials of degree $\leq r+s_2-1$ on each $T \in \mathcal{T}_h$, where $\mathcal{T}_h$ are shape regular, let $q \in (1,\infty)$. Then there exists a constant $c > 0$ such that for any $f \in W^{s_2,q}(\Omega)$, $v_h \in H_h^r$, we have:

$$|E_{\Omega}(fv_h)| \leq ch^{s_2} |f|_{s_2,q,\Omega} \|v_h\|_{1,2,\Omega}. \tag{3.11}$$

Proof. We have

$$|E_{\partial\Omega}(\varphi v_h)| \le \sum_{S\in s_h} |E_S(\varphi v_h)| \le ch^{s_1} \sum_{S\in s_h} |\varphi|_{s_1,q,S}\|v_h\|_{0,q',S}$$

By applying discrete Hölder inequality with parameters q and q', $\frac{1}{q}+\frac{1}{q'}=1$, we obtain

$$\sum_{S\in s_h} |\varphi|_{s_1,q,S}\|v_h\|_{0,q',S} \le |\varphi|_{s_1,q,\partial\Omega}\|v_h\|_{0,q',\partial\Omega}.$$

Finally, by applying trace embedding $H^1(\Omega) \hookrightarrow L^{q'}(\partial\Omega)$ on v_h, we obtain the first error estimate (3.10). Analogically, we also obtain

$$|E_\Omega(fv_h)| \le ch^{s_2} \sum_{T\in\mathcal{T}_h} |f|_{s_2,q,T}\|v_h\|_{0,q',T} \le ch^{s_2}|f|_{s_2,q,\Omega}\|v_h\|_{0,q',\Omega},$$

and we complete the proof of (3.11) with embedding $H^1(\Omega) \hookrightarrow L^{q'}(\Omega)$. □

Remark. We have only used the fact that v_h is continuous at the end by embedding inequality. If $v_h \in S_h^r$, we have:

$$|E_{\partial\Omega}(\varphi v_h)| \le ch^{s_1}|\varphi|_{s_1,q,\partial\Omega}\|v_h\|_{0,q',\partial\Omega}, \tag{3.12}$$
$$|E_\Omega(fv_h)| \le ch^{s_2}|f|_{s_2,q,\Omega}\|v_h\|_{0,q',\Omega}. \tag{3.13}$$

3.2 Approximate solution

Definition 3.4. *We call $u_h \in H_h^r$ an approximate solution of problem* (1.1)-(1.2) *if*

$$a_d(u_h, v_h) = L_d(v_h) \quad \forall v_h \in H_h^r. \tag{3.14}$$

In order to obtain error estimates of the approximate solution we need an analogy to monotonicity results for the new form a_d.

Theorem 3.5. *Let the quadrature* (3.4) *have at least* $r+1$ *nodes and only positive weights, i.e.*

$$m \ge r+1, \qquad \beta_\mu > 0, \quad \mu = 1,\dots,m. \tag{3.15}$$

Then there exists a constant $C_5 > 0$ such that the following inequality holds for every $u_h, v_h \in H_h^r$:

$$a_d(u_h, u_h - v_h) - a_d(v_h, u_h - v_h) \ge |u_h - v_h|^2_{1,2,\Omega} + C_5\|u_h - v_h\|^{\alpha+2}_{0,\alpha+2,\partial\Omega}. \tag{3.16}$$

Let $s_{h1} \subset s_h$ be a set of some boundary segments and denote $G_h = \bigcup s_{h1}$. If

$$|v_h| \ge \varepsilon_h > 0 \quad \text{on } G_h \tag{3.17}$$

holds for some $\varepsilon_h > 0$, then following inequality holds as well:

$$a_d(u_h, u_h - v_h) - a_d(v_h, u_h - v_h) \ge |u_h - v_h|^2_{1,2,\Omega} + C_6\|u_h - v_h\|^2_{0,2,G_h}. \tag{3.18}$$

Proof. We will proceed similarly to [10] Lemma 4.31. We have

$$a_d(u_h, u_h - v_h) - a_d(v_h, u_h - v_h) = |u_h - v_h|_{1,2,\Omega}^2 + \kappa \sum_{S \in s_h} |S| Z_S(u_h, v_h), \quad (3.19)$$

where

$$Z_S(u_h, v_h) = \sum_{\mu=1}^{m} \beta_\mu \left(|u_h|^\alpha u_h - |v_h|^\alpha v_h\right)(u_h - v_h)(x_{S,\mu}). \quad (3.20)$$

Now we apply either (1.10) and then we find that

$$Z_S(u_h, v_h) \geq 2^{-\alpha} \sum_{\mu=1}^{m} \beta_\mu |u_h - v_h|^{\alpha+2}(x_{S,\mu}),$$

or (2.10) to find that

$$Z_S(u_h, v_h) \geq \frac{1}{2} \sum_{\mu=1}^{m} \beta_\mu (u_h - v_h)^2 (|u_h|^\alpha + |v_h|^\alpha)(x_{S,\mu}).$$

We clearly have $Z_S(u_h, v_h) \geq 0$ and additionally, since each edge has more nodes than the degree of used polynomials r, $Z_S(u_h, v_h) > 0$ for $u_h \neq v_h$. Consider a function $\hat{\varphi_h} = \varphi_h \circ F_S$ defined on $[0,1]$, where $F_S : [0,1] \xrightarrow{\text{onto}} S$ is an affine mapping. Then expressions

$$\|\hat{\varphi_h}\|_{\alpha+2} = \left(\sum_{\mu=1}^{m} \beta_\mu |\hat{\varphi_h}|^{\alpha+2}(x_\mu)\right)^{\frac{1}{2+\alpha}}, \quad (3.21)$$

and

$$\|\hat{\varphi_h}\|_2 = \left(\sum_{\mu=1}^{m} \beta_\mu |\hat{\varphi_h}|^2(x_\mu)\right)^{\frac{1}{2}}, \quad (3.22)$$

are norms of a finite-dimensional space of polynomials of degree at most r defined on $[0,1]$. Using the fact that all norms are equivalent on a finite dimensional space, we conclude that there exist constants $\hat{C}_1$ and $\hat{C}_2$ such that

$$\|\hat{\varphi_h}\|_{\alpha+2} \geq \hat{C}_1 \|\hat{\varphi_h}\|_{0,\alpha+2,(0,1)}, \quad \|\hat{\varphi_h}\|_2 \geq \hat{C}_2 \|\hat{\varphi_h}\|_{0,2,(0,1)}. \quad (3.23)$$

This means that

$$\begin{aligned} Z_S(u_h, v_h) &\geq 2^{-\alpha} \|\hat{u_h} - \hat{v_h}\|_{\alpha+2}^{\alpha+2} \geq 2^{-\alpha} \hat{C_1}^{\alpha+2} \|\hat{u_h} - \hat{v_h}\|_{0,\alpha+2,(0,1)}^{\alpha+2} = \\ &= 2^{-\alpha} \hat{C_1}^{\alpha+2} \int_0^1 |\hat{u_h} - \hat{v_h}|^{\alpha+2} \mathrm{d}S = 2^{-\alpha} \hat{C_1}^{\alpha+2} |S|^{-1} \int_S |u_h - v_h|^{\alpha+2} \mathrm{d}S, \end{aligned} \quad (3.24)$$

and

$$\begin{aligned} Z_S(u_h, v_h) &\geq \frac{\varepsilon_h^\alpha}{2} \|\hat{u_h} - \hat{v_h}\|_2^2 \geq \frac{\varepsilon_h^\alpha}{2} \hat{C_2}^2 \|\hat{u_h} - \hat{v_h}\|_{0,2,(0,1)}^2 = \\ &= \frac{\varepsilon_h^\alpha}{2} \hat{C_2}^2 \int_0^1 |\hat{u_h} - \hat{v_h}|^2 \mathrm{d}S = \frac{\varepsilon_h^\alpha}{2} \hat{C_2}^2 |S|^{-1} \int_S |u_h - v_h|^2 \mathrm{d}S \end{aligned} \quad (3.25)$$

for $S \in s_{h1}$. Combining these inequalities with (3.19) gives the sought results. □

Remark. We will use this theorem for the piecewise Lagrange interpolation $v_h = \pi_h u$. We expect the existence of such G_h and ε_h to follow from the conditions (2.8). We have $G_h \subset G$ and $\varepsilon_h < \varepsilon$ and we would like this subset of the boundary $\partial\Omega$ and the positive constant to be independent of h. It follows from the considerations about a Lebesgue constant and C_L, C_l defined in (3.45) that for sufficiently smooth mesh refinement $\mathcal{T}_{h_0}$, some G_{h_0} and ε_{h_0} are bound to exist. Moreover, it also follows that for any further refinements $\mathcal{T}_h$, the constant ε_h will not decrease and the part of the boundary G_h will not decrease either, that is

$$\begin{aligned} &\varepsilon_h \geq \varepsilon_{h_0} > 0, \\ &\partial\Omega \supset G_h \supset G_{h_0}. \end{aligned} \tag{3.26}$$

The constant C_6 in (3.18) is therefore independent of h.

We have an analogy to Lemma 1.3 and hence, by using Hölder inequality

$$\|u\|_{0,\alpha+2,\partial\Omega} |\Omega|^{\frac{1}{2}-\frac{1}{\alpha+2}} \geq \|u\|_{0,2,\partial\Omega},$$

and Poinacaré inequality (1.5), Theorem 1.4 c) follows and we have an inequality

$$a_d(u_h, u_h - v_h) - a_d(v_h, u_h - v_h) \geq \widetilde{\varrho}(\|u_h - v_h\|_{1,2,\Omega}),$$

where $\widetilde{\varrho}$ is given by the same formula as ϱ but with different constants.

Lemma 3.6. *Let the assumptions of Theorem 3.5 hold. Then*

$$a_d(u_h, u_h - v_h) - a_d(v_h, u_h - v_h) \geq \widetilde{\varrho}(\|u_h - v_h\|_{1,2,\Omega}), \tag{3.27}$$

where

$$\widetilde{\varrho}(t) = C_{0d} t^{\alpha+2}. \tag{3.28}$$

If the condition (3.17) *holds, then we can redefine* $\widetilde{\varrho}$ *as*

$$\widetilde{\varrho}(t) = C_{1d} t^2. \tag{3.29}$$

Uniform monotonicity of the form a_d on the finite dimensional space H_h^r guarantees the existence and the uniqueness of the approximate solution u_h given by (3.14), see [13], [30].

Denote

$$R(t) = \widetilde{\varrho}(t)/t \tag{3.30}$$

and let R_{-1} be the inverse of R. It holds that

$$R_{-1}(t) = \left(\frac{t}{C_{0d}}\right)^{\frac{1}{\alpha+1}}, \tag{3.31}$$

which can be replaced under the condition (3.17) by

$$R_{-1}(t) = \frac{t}{C_{1d}}. \tag{3.32}$$

Then we have the following abstract error estimate.

Theorem 3.7. *Let u_h be the approximate solution of problem* (3.14) *and let u be the weak solution defined by* (1.8)*. If $v_h \in H_h^r$, then*

$$\begin{aligned}\|u-u_h\|_{1,2,\Omega} \leq &\|u-v_h\|_{1,2,\Omega} \\ &+ R_{-1}\Bigg(C_1\|u-v_h\|_{1,2,\Omega}(1+\|u\|_{1,2,\Omega}^{\alpha}+\|v_h\|_{1,2,\Omega}^{\alpha}) \\ &+ \sup_{0\neq w_h\in H_h^r}\frac{|a(v_h,w_h)-a_d(v_h,w_h)|}{\|w_h\|_{1,2,\Omega}} + \sup_{0\neq w_h\in H_h^r}\frac{|L(w_h)-L_d(w_h)|}{\|w_h\|_{1,2,\Omega}}\Bigg).\end{aligned} \tag{3.33}$$

Proof. The monotonicity of a_h from Theorem 3.5 gives us

$$\widetilde{\varrho}(\|u_h-v_h\|_{1,2,\Omega}) \leq a_d(u_h,u_h-v_h)-a_d(v_h,u_h-v_h).$$

By using

$$\begin{aligned} a_d(u_h,u_h-v_h) &= L_d(u_h-v_h), \\ L(u_h-v_h) &= a(u,u_h-v_h), \end{aligned}$$

and adding and subtracting the same terms, we get

$$\begin{aligned} a_d(u_h,u_h-v_h)-a_d(v_h,u_h-v_h) &= [L_d(u_h-v_h)-L(u_h-v_h)] \\ +[a(u,u_h-v_h)-a(v_h,u_h-v_h)] &+ [a(v_h,u_h-v_h)-a_d(v_h,u_h-v_h)]. \end{aligned}$$

The first bracket can be estimated directly using an inequality from a definition of a norm of a dual operator:

$$|L_d(u_h-v_h)-L(u_h-v_h)| \leq \sup_{0\neq w_h\in H_h^r}\frac{|L(w_h)-L_d(w_h)|}{\|w_h\|_{1,2,\Omega}}\|u_h-v_h\|_{1,2,\Omega}.$$

The second bracket can be estimated using the continuity (1.14) of the form a:

$$\begin{aligned} |a(u,u_h-v_h)-a(v_h,u_h-v_h)| \leq &C_1\left(1+\|u\|_{1,2,\Omega}^{\alpha}+\|v_h\|_{1,2,\Omega}^{\alpha}\right) \\ &\|u-v_h\|_{1,2,\Omega}\,\|u_h-v_h\|_{1,2,\Omega}\,. \end{aligned}$$

The third bracket can be estimated similarly to the first bracket:

$$|a(v_h,u_h-v_h)-a_d(v_h,u_h-v_h)| \leq \sup_{0\neq w_h\in H_h^r}\frac{|a(v_h,w_h)-a_d(v_h,w_h)|}{\|w_h\|_{1,2,\Omega}}\|u_h-v_h\|_{1,2,\Omega}.$$

Combining these estimates with the definition of R in (3.30) gives us

$$\begin{aligned} R(\|u_h-v_h\|_{1,2,\Omega}) \leq &\sup_{0\neq w_h\in H_h^r}\frac{|L(w_h)-L_d(w_h)|}{\|w_h\|_{1,2,\Omega}} \\ &+ C_1\|u-v_h\|_{1,2,\Omega}(1+\|u\|_{1,2,\Omega}^{\alpha}+\|v_h\|_{1,2,\Omega}^{\alpha}) + \sup_{0\neq w_h\in H_h^r}\frac{|a(v_h,w_h)-a_d(v_h,w_h)|}{\|w_h\|_{1,2,\Omega}}. \end{aligned} \tag{3.34}$$

By using triangle inequality $\|u-u_h\|_{1,2,\Omega} \leq \|u-v_h\|_{1,2,\Omega}+\|u_h-v_h\|_{1,2,\Omega}$, we arrive at (3.33). ☐

Recall that

$$L(w_h) - L_d(w_h) = E_\Omega(f w_h) + E_{\partial\Omega}(\varphi w_h)$$

represents the error of integration of terms derived from the right hand sides of (1.1) and (1.2). This error can be estimated using (3.11) and (3.10) from Theorem 3.3:

$$|L(w_h) - L_d(w_h)| \le c \left(h^{s_2} |f|_{s_2,q,\Omega} + h^{s_1} |\varphi|_{s_1,q,\partial\Omega} \right) \|w_h\|_{1,2,\Omega} . \tag{3.35}$$

The term

$$a(v_h, w_h) - a_d(v_h, w_h) = E_{\partial\Omega}(|v_h|^\alpha v_h w_h)$$

is the error of integration of the non-linear term on the boundary $\partial\Omega$. It cannot be estimated directly using (3.10), because the continuous piecewise polynomial function v_h may have jumps in its derivatives at vertices of boundary triangles and some derivatives of $|v_h|^\alpha v_h$ may become nonintegrable near the roots of v_h in case of noninteger parameter α. Using (3.8) and repeating arguments from the proof of Theorem 3.3 on separate parts of the boundary will lead to an estimate similar to (3.10). But first, we will need to prove boundedness of $|v_h|^\alpha v_h$ on the boundary $\partial\Omega$ in a norm of some Sobolev space.

For the purpose of error estimation of the term

$$\|u - v_h\|_{1,2,\Omega}(1 + \|u\|_{1,2,\Omega}^\alpha + \|v_h\|_{1,2,\Omega}^\alpha),$$

we will need v_h to be a function from the space H_h^r such that $\|v_h\|_{1,2,\Omega}$ is uniformly bounded based on u and $\|u - v_h\|_{1,2,\Omega}$ converges to zero with some rate. Therefore, we will set $v_h = \pi_h u$, where π_h is the continuous piecewise Lagrange interpolation operator.

3.3 Boundedness of interpolated function

In this section we are concerned with estimating derivatives of functions on the boundary of Ω. As we are aiming to use (3.8), it is sufficient to consider one fixed segment $S \in s_h$ of a boundary triangle T_S of a triangulation $\mathcal{T}_h$. Let F be an affine mapping of $I = [0, |S|]$ onto S. Set $\widetilde{v_h} = v_h \circ F$ for a function $v_h \in H_h^r$. Function $\widetilde{v_h}$ is therefore a polynomial of degree r defined on interval $[0, |S|]$.

Remark. Since $|F'| = 1$, it follows that functions f measured in arbitrary $W^*(S)$-norm or in any $C^*(S)$-norm give the same result as functions $\widetilde{f} = f \circ F$ measured in the corresponding $W^*(I)$-norm or the $C^*(I)$-norm. This affine transformation is only used for simplicity as $S \subset \mathbb{R}^2$ uses two coordinates and $I \subset \mathbb{R}$ uses only one coordinate. Therefore all derivatives of $\widetilde{f}$ are only with respect to one variable.

Let us begin by expressing the actual terms which appear after using chain rule on derivatives of $|\widetilde{v_h}|^\alpha \widetilde{v_h}$. We will proceed similarly to Lemma 1.8.

Lemma 3.8. *Let $\widetilde{v_h}$ be a polynomial of degree r on the interval $[0, |S|]$. Let $\alpha \ge 0$ and $\beta \in \mathbb{N}$. Then*

$$(|\widetilde{v_h}|^\alpha \widetilde{v_h})^{(\beta)}$$

can be expressed as a finite sum of terms of the form

$$c\, |\widetilde{v_h}|^{\alpha+1-J} \prod_{j=1}^{J} \widetilde{v_h}^{(\gamma_j)}, \tag{3.36}$$

where c is a constant dependent on α and β, $J \in \mathbb{N}$ and $\gamma_j \in$, $j = 1, \dots, J$ are positive integers satisfying $\sum_{j=1}^{J} \gamma_j = \beta$. If $\alpha \in \mathbb{N}$, then $(|\widetilde{v_h}|^\alpha \, \widetilde{v_h})^{(\beta)}$ only contains terms where the exponent $\alpha + 1 - J$ is non-negative.

Proof. We will use induction on β. When $\beta = 1$, the only term has $c = \alpha + 1$, $J = 1$ and $\gamma_1 = 1$.

Suppose that that the lemma holds for $\beta - 1$. Then we only need to apply $\frac{\mathrm{d}}{\mathrm{d}x}$ to terms $c\,|\widetilde{v_h}|^{\alpha+1-J'} \prod_{j=1}^{J'} \widetilde{v_h}^{(\gamma_j')}$, which have $J' \in \mathbb{N}$ and $\sum_{j=1}^{J'} \gamma_j' = \beta - 1$. If the derivative $\frac{\mathrm{d}}{\mathrm{d}x}$ is applied to any factor in $\prod_{j=1}^{J'} \widetilde{v_h}^{(\gamma_j')}$, then the resulting term does have the desired form with $J = J'$, a single index γ_j' increased by one, the remaining indices unchanged, and $\sum_{j=1}^{J} \gamma_j = \beta$. If $\frac{\mathrm{d}}{\mathrm{d}x}$ is applied to $|\widetilde{v_h}|^{\alpha+1-J'}$, then the resulting term has $J = J' + 1$, $\sum_{j=1}^{J'} \gamma_j' + \gamma_{J'+1} = \beta$, where $\gamma_{J'+1} = 1$, and therefore also has the desired form.

Suppose that $\alpha \in \mathbb{N}$. Then the exponent $\alpha + 1 - J$ in $|u|^{\alpha+1-J}$ is integer for any J. The only possibility to obtain a negative exponent from the induction step would be to apply $\frac{\mathrm{d}}{\mathrm{d}x}$ to $|\widetilde{v_h}|^{\alpha+1-J'}$ for J' such that $\alpha + 1 - J' \in [0, 1)$, i.e. $\alpha + 1 - J' = 0$. But then $\left(|\widetilde{v_h}|^0\right)' = 0$ and the constant c would in fact be zero. $\square$

To estimate integrals of terms of the form (3.36), the most straight-forward way is to estimate most of its factors in L^∞-norm and take them out of the integral. As we assume $u \in W^{r+1,q}(\Omega)$ and the embeddings

$$\begin{aligned} &W^{r+1,q}(\Omega) \hookrightarrow C^r\left(\overline{\Omega}\right), \quad q > 2, \\ &H^{r+1}(\Omega) \hookrightarrow C^{r-1,\lambda}\left(\overline{\Omega}\right), \quad \lambda \in [0, 1), \\ &W^{r+1,q}(\Omega) \hookrightarrow C^{r-1,2-\frac{2}{q}}\left(\overline{\Omega}\right), \quad q \in [1, 2), \end{aligned} \tag{3.37}$$

follow from (1.3), we can approach estimating of lower derivatives by considerations applying to continuous functions rather than using properties of Sobolev spaces.

Lemma 3.9. *Let $u \in C^r(T_S)$, let $\pi_h u$ be its Lagrange interpolation of order r using $r + 1$ nodes at the sides of T_S. Let F be an affine mapping of $I = [0, |S|]$ onto S. Let $\widetilde{u} = u \circ F$ and $\widetilde{\pi_h u} = (\pi_h u) \circ F$. Then for any $i \in \{0, \dots, r\}$ holds an estimate*

$$|\widetilde{\pi_h u}|_{i,\infty,I} \leq \sum_{j=i}^{r} |S|^{j-i} \, |\widetilde{u}|_{j,\infty,I} \,. \tag{3.38}$$

Proof. We will use induction on i downward from r to 0.

Let $i = r$. Our goal is to show that

$$|\widetilde{\pi_h u}|_{r,\infty,I} \leq |\widetilde{u}|_{r,\infty,I} \,.$$

Function $\widetilde{\pi_h u}$ is a polynomial of degree r and its r-th derivative is a constant. Therefore, it is sufficient to prove that $\widetilde{\pi_h u}^{(r)} = \widetilde{u}^{(r)}(t)$ for some $t \in I$ or that a continuous function $(\widetilde{\pi_h u} - \widetilde{u})^{(r)}$ has a root. The Lagrange interpolation is exact at all nodes and thus $\widetilde{\pi_h u} - \widetilde{u}$ has $r + 1$ roots in I. It follows from Rolle's theorem that $(\widetilde{\pi_h u} - \widetilde{u})'$ has r roots in I and repeating this argument r times gives us a root of $(\widetilde{\pi_h u} - \widetilde{u})^{(r)}$ in I.

Let the inequality (3.38) hold for $i+1$. Take arbitrary $t, t_0 \in I$. Considering that $|[t_0,t]| \le |I| = |S|$, we have

$$\left|\widetilde{\pi_h u}^{(i)}(t)\right| = \left|\widetilde{\pi_h u}^{(i)}(t_0) + \int_{t_0}^{t} \widetilde{\pi_h u}^{(i+1)}(\tau)\mathrm{d}\tau\right| \le \left|\widetilde{\pi_h u}^{(i)}(t_0)\right| + |S| \left|\widetilde{\pi_h u}\right|_{i+1,\infty,I}.$$

Using (3.38) for $i+1$ and the definition of L^∞-norm, we have

$$\begin{aligned}\left|\widetilde{\pi_h u}\right|_{i,\infty,I} &\le \left|\widetilde{\pi_h u}^{(i)}(t_0)\right| + |S| \sum_{j=i+1}^{r} |S|^{j-(i+1)} \left|\widetilde{u}\right|_{j,\infty,I} \\ &= \left|\widetilde{\pi_h u}^{(i)}(t_0)\right| + \sum_{j=i+1}^{r} |S|^{j-i} \left|\widetilde{u}\right|_{j,\infty,I}.\end{aligned}$$

To complete the induction step, it suffices to find some $t_0, t_1 \in I$ such that $\widetilde{\pi_h u}^{(i)}(t_0) = \widetilde{u}^{(i)}(t_1)$. Take $i+1$ of the $r+1$ nodes of interpolation. Construct a polynomial v of degree at most i such that $\widetilde{u}$, $\widetilde{\pi_h u}$ and v are equal at these nodes. Functions $\widetilde{u} - v$ and $\widetilde{\pi_h u} - v$ have $i+1$ roots in I and they both belong to a space $C^i(I)$. By Rolle's theorem, there are t_0 and t_1 such that $\left(\widetilde{\pi_h u} - v\right)^{(i)}(t_0) = \left(\widetilde{u} - v\right)^{(i)}(t_1) = 0$. This together with the fact that $v^{(i)}$ is a constant completes the proof. ☐

The case of $u \in C^{r-1,2-\frac{2}{q}}(T_S)$ for $q \in (1,2)$ (and for $q = 2$) is almost identical.

Lemma 3.10. *Let $u \in C^{r-1,\lambda}(T_S)$, where $\lambda \in (0,1)$, let $\pi_h u$ be its Lagrange interpolation of order r using $r+1$ nodes at the sides of T_S. Let F be an affine mapping of $I = [0,|S|]$ onto S. Let $\widetilde{u} = u \circ F$ and $\widetilde{\pi_h u} = (\pi_h u) \circ F$. Then there exists a constant $c > 0$ such that for any $i \in \{0,\dots,r\}$ holds an estimate*

$$\left|\widetilde{\pi_h u}\right|_{i,\infty,I} \le c\,|S|^{r-1+\lambda-i} \left|\widetilde{u}^{(r-1)}\right|_{C^{0,\lambda}(I)} + \sum_{j=i}^{r-1} |S|^{j-1+\lambda-i} \left|\widetilde{u}\right|_{j,\infty,I}. \tag{3.39}$$

Proof. Again, we will use induction on i from r down to 0.

Let $i = r$. Our goal is to show that

$$\left|\widetilde{\pi_h u}\right|_{r,\infty,I} \le c\,|S|^{-1+\lambda} \left|\widetilde{u}^{(r-1)}\right|_{C^{0,\lambda}(I)}.$$

Let v be a Taylor polynomial of the function $\widetilde{u}$ of degree $r-1$ at point 0, that is v is a polynomial of degree $\le r-1$ and it has the same derivatives of orders up to $r-1$ at the point 0 as $\widetilde{u}$. Function $\widetilde{u} - v$ has the same r-th derivative as $\widetilde{u}$ and also has the same seminorm (Hölder constant) $\left|\widetilde{u}^{(r-1)}\right|_{C^{0,\lambda}(I)}$. Its interpolation $\widetilde{\pi_h u} - v$ also has the r-th derivative unchanged. We only need to show that

$$\left|\widetilde{\pi_h u} - v\right|_{r,\infty,I} \le c\,|S|^{-1+\lambda} \left|(\widetilde{u} - v)^{(r-1)}\right|_{C^{0,\lambda}(I)},$$

where $(\widetilde{u} - v)$ satisfies $(\widetilde{u} - v)^{(j)}(0) = 0$ for all $j = 0,\dots,r-1$.

It follows from $(\widetilde{u} - v)^{(r-1)}(0) = 0$ and the definition of the Hölder continuity that

$$\left|\widetilde{u} - v\right|_{r-1,\infty,I} \le |S|^{\lambda} \left|\widetilde{u}^{(r-1)}\right|_{C^{0,\lambda}(I)}.$$

Since $(\widetilde{u}-v)^{(r-2)} = 0$ (if $r \geq 2$), it follows that

$$|\widetilde{u}-v|_{r-2,\infty,I} \leq |S|^{1+\lambda} \left|\widetilde{u}^{(r-1)}\right|_{C^{0,\lambda}(I)} .$$

Repeating this argument yields

$$|\widetilde{u}-v|_{0,\infty,I} \leq |S|^{r-1+\lambda} \left|\widetilde{u}^{(r-1)}\right|_{C^{0,\lambda}(I)} .$$

Consider an affine transformation of $\widetilde{u}-v$ and $\widetilde{\pi_h u}-v$ from $I = [0,|S|]$ onto $[0,1]$. Denote the resulting functions by $\widehat{u-v}$ and $\widehat{\pi_h u - v}$. The function $\widehat{u-v}$ is also bounded in L^∞-norm by $|S|^{r-1+\lambda} |\widetilde{u}|_{C^{0,\lambda}(I)}$. The interpolation $\widehat{\pi_h u - v}$ of $\widehat{u-v}$ is therefore bounded by

$$\left|\widehat{\pi_h u - v}\right|_{0,\infty,[0,1]} \leq c\,|S|^{r-1+\lambda} |\widetilde{u}|_{C^{0,\lambda}(I)} ,$$

where $c > 0$ is a constant dependent only on the choice of nodes of interpolation on the reference interval $[0,1]$. The space of polynomials of degree $\leq r$ on $[0,1]$ is a finite-dimensional space. Every seminorm on a finite-dimensional space can be estimated from above by any norm. Taking a seminorm $|\cdot|_{r,\infty,[0,1]}$ and a norm $|\cdot|_{0,\infty,[0,1]}$ thus yields

$$\left|\widehat{\pi_h u - v}\right|_{r,\infty,[0,1]} \leq c\,|S|^{r-1+\lambda} |\widetilde{u}|_{C^{0,\lambda}(I)} ,$$

where $c > 0$ is again some constant dependent only on π. Since affine transformation from I onto $[0,1]$ multiplies the r-th derivative by $|S|^r$, we have

$$|S|^r \left|\widetilde{\pi_h u} - v\right|_{r,\infty,I} = \left|\widehat{\pi_h u - v}\right|_{r,\infty,[0,1]} \leq c\,|S|^{r-1+\lambda} |\widetilde{u}|_{C^{0,\lambda}(I)} .$$

This is (3.39) for $i = r$.

Since functions in the space $C^{r-1,\lambda}(I)$ are also in $C^j(I)$ for $j = 0,\ldots,r-1$, the whole induction step in the proof of the previous lemma works here too and we again have

$$|\widetilde{\pi_h u}|_{i,\infty,I} \leq |\widetilde{u}|_{i,\infty,I} + |S|\,|\widetilde{\pi_h u}|_{i+1,\infty,I} . \tag{3.40}$$

Combining (3.40) and the inequality (3.39) for $i+1$ gives (3.39) for i. ☐

To estimate the interpolation in a norm of Sobolev spaces we use a one-dimensional corrolary of Theorem 2.6 or Theorem 3.1.5 in [4]. The interpolation preserves polynomials of degree up to $k \leq r$. We further set $m = k+1$ and $p = q$ in Theorem 2.6.

Corollary. Let the piecewise Lagrange interpolation π_h preserve polynomials of degree $\leq r$. Let the restriction of the interpolated function $\pi_h u$ on any side of a triangle be given only by the values of u on that side (that is, let it have $r+1$ nodes on every side of a triangle). Let $k \in \mathbb{N}_0$, $r \leq k$, $q \geq 1$. Then there exists a constant $C(\pi)$ such that

$$|\widetilde{u} - \widetilde{\pi_h u}|_{k+1,q,I} \leq C\,|\widetilde{u}|_{k+1,q,I} \quad \forall \widetilde{u} \in W^{k+1,q}(I),$$

and it follows form triangle inequality that we also have

$$|\widetilde{\pi_h u}|_{k+1,q,I} \leq (C+1)\,|\widetilde{u}|_{k+1,q,I} \quad \forall \widetilde{u} \in W^{k+1,q}(I). \tag{3.41}$$

When we use polynomials of degree r and consider only numerical quadrature for boundary nonlinear terms satisfying (3.15), we expect the order of convergence in H^1-norm to be r. But we need in addition to the regularity of the exact weak solution u an upper bound for the r-th derivative of $(|\widetilde{\pi_h u}|^\alpha \widetilde{\pi_h u})$. We need to have some upper estimate for all terms of the form

$$c\,|\widetilde{\pi_h u}|^{\alpha+1-J} \prod_{j=1}^{J} \widetilde{\pi_h u}^{(\gamma_j)}, \quad \sum_{j=1}^{J} \gamma_j = r. \tag{3.42}$$

If α is an integer, then all exponents $\alpha+1-J$ in powers of $|\widetilde{\pi_h u}|$ are non-negative (those that are negative are in terms multiplied by $c=0$) and we only need an upper estimate of $|\widetilde{\pi_h u}|$. The lowest possible exponent is $\alpha+1-r$ and therefore in a case of $\alpha \geq r-1$, we also only need an upper estimate.

Lemma 3.11. *Let $u \in W^{r+1,q}(\Omega)$, where $r \in \mathbb{N}_0$, $q>1$. Let T_S be a boundary triangle of the triangulation $\mathcal{T}_h$, and $I=[0,|S|]$. Let π_h be a continuous piecewise Lagrange interpolation of order r that uses $r+1$ nodes on the sides of triangles. Let F be the affine transformation of I onto S and let $\widetilde{\pi_h u} = (\pi_h u|_S) \circ F$. Let $i \in \mathbb{N}_0$, $i \leq r$, and let $\alpha \geq 0$ be the constant from (1.2). Let either $\alpha \in \mathbb{N}_0$ or $\alpha \geq i-1$. Then $|\widetilde{\pi_h u}|^\alpha \widetilde{\pi_h u} \in W^{i,q}(I)$ and there exists a constant $c = c(\pi,\alpha,r,i,q) > 0$ such that*

$$\left| |\widetilde{\pi_h u}|^\alpha \widetilde{\pi_h u} \right|_{i,q,I} \leq c\, \|u\|^\alpha_{k+1,q,\Omega} \|\tilde{u}\|_{i,q,I}\,. \tag{3.43}$$

Proof. Due to triangle inequality in Lebesgue spaces, we only need to estimate terms of the form given in (3.42) (with $\sum_{j=1}^{J}\gamma_j = i$) by the right hand-side of (3.43). The assumption $\alpha \geq i-1$ or $\alpha \in \mathbb{N}_0$ guarantees that the exponents $\alpha+1-J$ in (3.42) are non-negative for all terms that need to be estimated. Since we have an embedding (3.37), all derivatives of orders up to $r-1$ can be estimated in L^∞-norm by $\|\tilde{u}\|_{k+1,q,\Omega}$ due to (3.39) for $q \in (1,2]$ and all derivatives of orders up to r due to (3.38) for $q \in (2,\infty)$.

Let us take a term of the form (3.42):

$$c\,|\widetilde{\pi_h u}|^{\alpha+1-J} \prod_{j=1}^{J} \widetilde{\pi_h u}^{(\gamma_j)}, \quad \sum_{j=1}^{J} \gamma_j = i.$$

Write the seminorm as an integral

$$\left| |\widetilde{\pi_h u}|^{\alpha+1-J} \prod_{j=1}^{J} \widetilde{\pi_h u}^{(\gamma_j)} \right|_{0,q,I} = \left(\int_I |\widetilde{\pi_h u}|^{(\alpha+1-J)q} \prod_{j=1}^{J} \left| \widetilde{\pi_h u}^{(\gamma_j)} \right|^q \mathrm{d}S \right)^{\frac{1}{q}}.$$

All terms that are continuous can be simply taken out of the integral and give us some upper bound for the seminorm. Without loss of generality assume that γ_J is the largest order of derivative. Suppose for the moment that all other factors are continuous and can be estimated in the following way:

$$\left\| \widetilde{\pi_h u}^{(\gamma_j)} \right\|_{0,\infty,I} \leq c\,\|\tilde{u}\|_{C^r(I)} \leq c\,\|u\|_{C^r(\partial\Omega)} \leq c\,\|u\|_{r+1,q,\Omega}\,, \tag{3.44}$$

with replacing the C^r-norm by the $C^{r-1,\lambda}$-norm if $q \in (1,2]$. Then we have an estimate

$$\left(\int_I |\widetilde{\pi_h u}|^{(\alpha+1-J)q} \prod_{j=1}^{J} \left| \widetilde{\pi_h u}^{(\gamma_j)} \right|^q \mathrm{d}S \right)^{\frac{1}{q}} \leq c\,\|u\|^{(\alpha+1-J)+(J-1)}_{r+1,q,\Omega} \left(\int_I \left| \widetilde{\pi_h u}^{(\gamma_J)} \right|^q \mathrm{d}S \right)^{\frac{1}{q}}.$$

Using (3.41) gives an estimate of the last remaining part

$$\left(\int_I \left|\widetilde{\pi_h u}^{(\gamma_J)}\right|^q \mathrm{d}S\right)^{\frac{1}{q}} = |\widetilde{\pi_h u}|_{\gamma_J,q,I} \le c\,|\widetilde{u}|_{\gamma_J,q,I} \le c\,\|\widetilde{u}\|_{i,q,I}\,.$$

Combining these estimates yields

$$\left| |\widetilde{\pi_h u}|^{\alpha+1-J} \prod_{j=1}^{J} \widetilde{\pi_h u}^{(\gamma_j)} \right|_{0,q,I} \le c\,\|u\|^{\alpha}_{r+1,q,\Omega}\,\|\widetilde{u}\|_{i,q,I}\,.$$

The assumption that all factors in $|\widetilde{\pi_h u}|^{\alpha+1-J}\prod_{j=1}^{J}\widetilde{\pi_h u}^{(\gamma_j)}$ besides $\widetilde{\pi_h u}^{(\gamma_J)}$ can be estimated by (3.44) follows from (3.38) and (3.39) if

- $\gamma_J \le r-1$,
- $\gamma_J = r$ and $q > 2$,
- $\gamma_J = i$ (in this case $J = 1$ and there are no other factors with derivatives).

Since we have $\gamma_J \le i \le r$ one of these cases always holds and we have in fact completed the proof. □

If neither $\alpha \in \mathbb{N}_0$ nor $\alpha \ge r-1$ and we are still trying to use the estimate (3.8) of order r, we need to obtain some positive lower bounds on $\widetilde{\pi_h u}$. These estimates can be derived with some aid from the Lebesgue constants if we include an assumption that $\max_I |\widetilde{u}|$ and $\min_I |\widetilde{u}|$ are relatively close, see Chapter 3 in [23].

Consider a fixed Lagrange interpolation π_h of order r preserving polynomials of degree $\le r$ on the boundary. More precisely: the nodes of interpolation on the reference triangle $\hat{T}$ are in one fixed position for all triangles $T \in \mathcal{T}_h$ and there are $r+1$ nodes of interpolation on every side of this triangle. Take an arbitrary function $\widetilde{u} \in C(I)$ such that $\|\widetilde{u}\|_{0,\infty,I} \le 1$. Then there exists a constant Λ_π such that $\|\widetilde{\pi_h u}\|_C \le \Lambda_\pi$ for all such $\widetilde{u}$. It can be defined as

$$\Lambda_\pi = \max_{\substack{\widetilde{u}\in C(I)\\ \|\widetilde{u}\|_{0,\infty,I}\le 1}} \|\widetilde{\pi_h u}\|_{0,\infty,I}.$$

Considering that $\widetilde{\pi_h u}$ is given by a finite $(r+1)$ amount of values of $\widetilde{u}$ and the interpolation operator π_h is linear, the maximum in the definition of Λ_π can be found by taking functions which have either 1 or -1 at each node (that is 2^{r+1} combinations). If we further consider that rescaling a function from one interval onto another with a linear substitution will not change the function's extremes, we see that this constant Λ_π is shared for all segments in s_h for all triangulations $\{T_h\}, h > 0$.

If we now take a function $\widetilde{u} \in C(I)$ which is bounded by $a+b$ from above and by $a-b$ from below for some $a \in \mathbb{R}$ and $b > 0$, it follows that the interpolated function $\widetilde{\pi_h u} \in P_r(I)$ is bounded by $a + \Lambda_\pi b$ from above and by $a - \Lambda_\pi b$ from below. Suppose that the values of $\widetilde{u}$ are in $[C_L, 1]$ for some constant $C_L \in (0,1)$. Then we have $a = \frac{1}{2}(1+C_L)$ and $b = \frac{1}{2}(1-C_L)$, and $\widetilde{\pi_h u}$ is estimated from below by

$$\frac{1}{2}(C_L+1) - \frac{\Lambda_\pi}{2}(1-C_L) = \frac{1}{2}\left(C_L(\Lambda_\pi+1) - (\Lambda_\pi - 1)\right),$$

k	Λ_π	C_L
1	1.000000	0.000000
2	1.250000	0.111111
3	1.422919	0.174549
4	1.559490	0.218594

Table 3.1: Values of optimal Lebesgue constants for polynomials of degrees up to 4 and the corresponding constants C_L

which is zero for the choice $C_L = \frac{\Lambda_\pi - 1}{\Lambda_\pi + 1}$. Then, from the conditions

$$C_L = \frac{\Lambda_\pi - 1}{\Lambda_\pi + 1}, \quad C_l \in (C_L, 1), \quad \frac{\min_I |\tilde{u}|}{\max_I |\tilde{u}|} \geq C_l, \tag{3.45}$$

follows the lower bound estimate

$$\min_I |\widetilde{\pi_h u}| \geq \frac{1}{2} \left(C_l(\Lambda_\pi + 1) - (\Lambda_\pi - 1) \right) \max_I |\tilde{u}| \,.$$

Therefore, we have an estimate in L^∞-norm for a negative power of the interpolated function

$$\left\| |\widetilde{\pi_h u}|^{-\gamma} \right\|_{0,\infty,I} \leq \left(\frac{2}{C_l(\Lambda_\pi + 1) - (\Lambda_\pi - 1)} \right)^\gamma \|\tilde{u}\|^\gamma_{0,\infty,I} \,, \quad \gamma > 0. \tag{3.46}$$

If the triangulation is refined by dividing some triangles into smaller ones, the maximum of $|u|$ on any new segment is bounded from above by the old maximum and the new minimum is bounded from below by the old minimum. Thus the new segment also satisfies the conditions (3.45) and the constant C_l might even be increased.

Choosing linearly transformed Chebyshev nodes for the interpolation π gives an estimate for the Lebesgue constant

$$\Lambda_\pi = \frac{2}{\pi} \left(\log(r+1) + 0.7219 + \log \frac{8}{\pi} \right) + O\left(r^{-2} \right), \tag{3.47}$$

where r is the degree of interpolation and $\gamma = 0.577215$ is the Euler-Mascheroni constant, see [12], [20]. The explicit formula for optimal Lebesgue constant is known for $k \leq 3$, see [27]. Using the optimal Lebesgue constants for $r \leq 4$ (formulas (3.3) and (7.4) in [27]) give us some possible values for the constant C_L in (3.45).

Lemma 3.12. *Let $u \in W^{r+1,q}(\Omega)$, let $S \in s_h$ be a boundary segment such that $u|_S$ is non-zero and does not change sign and furthermore, let $\frac{\min_S |u|}{\max_S |u|} \geq C_l$. Suppose that $C_l > C_L$, where C_L is defined above. Let $\tilde{u}$ be the affine transformation of $u|_S$ onto $I = [0, |S|]$ as defined above. Then there exists a constant $c = c(\pi, \alpha, r, q) > 0$ such that*

$$\left| |\widetilde{\pi_h u}|^\alpha \, \widetilde{\pi_h u} \right|_{r,q,I} \leq c \left(\|u\|^\alpha_{r+1,q,\Omega} + \|u\|^{2r-2-\alpha}_{r+1,q,\Omega} \right) \|\tilde{u}\|_{r,q,I} \,. \tag{3.48}$$

Proof. We can proceed similarly as we did in the proof of Lemma 3.11. The only new concern is that now the exponents $\alpha+1-J$ in the terms of the form

$$c\,|\widetilde{\pi_h u}|^{\alpha+1-J}\prod_{j=1}^{J}\widetilde{\pi_h u}^{(\gamma_j)},\quad \sum_{j=1}^{J}\gamma_j=r$$

can be negative. Whereas we previously used an estimate (3.44) for nonnegative $\alpha+1-J$, we now use

$$\left\||\widetilde{\pi_h u}|^{\alpha+1-J}\right\|_{0,\infty,I}\leq\left(\frac{2}{C_l(\Lambda_{\pi_h}+1)-(\Lambda_{\pi_h}-1)}\right)^{J-\alpha-1}\|\widetilde{u}\|_{0,\infty,I}^{J-\alpha-1}\leq c\,\|u\|_{r+1,q,\Omega}^{J-\alpha-1}, \tag{3.49}$$

for negative $\alpha+1-J<0$. This estimate leads to an inequality

$$\left||\widetilde{\pi_h u}|^{\alpha+1-J}\prod_{j=1}^{J}\widetilde{\pi_h u}^{(\gamma_j)}\right|_{0,q,I}\leq c\,\|u\|_{r+1,q,\Omega}^{(J-\alpha-1)+(J-1)}\,\|\widetilde{u}\|_{r,q,I}.$$

Since $\alpha+1<J\leq r$ holds for negative exponents $\alpha+1-J<0$, the exponent $2J-2-\alpha$ is between α and $2r-2-\alpha$. It follows that

$$\|u\|_{r+1,q,\Omega}^{(J-\alpha-1)+(J-1)}\leq\|u\|_{r+1,q,\Omega}^{\alpha}+\|u\|_{r+1,q,\Omega}^{2r-2-\alpha},$$

and the inequality (3.48) holds. □

3.4 Error estimation

The purpose of this section is to estimate the error of quadrature on the boundary $\partial\Omega$ denoted by $E_{\partial\Omega}(|\pi_h u|^{\alpha}(\pi_h u)w_h)$. We can divide the boundary segments $S\in s_h$ into three disjoint sets $s_h=s_{h0}\cup s_{h1}\cup s_{h2}$.

- s_{h0} contains segments S with $u|_S=0$. Then also $\pi_h u|_S=0$ and the quadrature is exact there, i.e. $E_S(|\pi_h u|^{\alpha}(\pi_h u)w_h)=0$.
- If $\alpha+1\geq r$ or $\alpha\in\mathbb{N}_0$, then s_{h1} contains all segments not in s_{h0}. If $\alpha\notin\mathbb{N}_0$ and $\alpha+1<r$, then s_{h1} contains all segments not in s_{h0} satisfying $\frac{\min_S|u|}{\max_S|u|}\geq C_l$, where C_l is given by (3.45). Then combining (3.48) (or (3.43)) and (3.8) gives us an error estimate of order r.
- s_{h2} contains the remaining segments, i.e. for $\alpha\notin\mathbb{N}_0$ and $\alpha+1<r$, s_{h2} contains segments satisfying $\frac{\min_S|u|}{\max_S|u|}<C_l$ and u is not identically zero on S. Let us set $h_2=\max\{|S|\,;S\in s_{h2}\}$ (or $h_2=0$ if there are no segments in s_{h2}). Combining (3.43) and (3.8) gives us an error estimate of order

$$r_2=\lfloor\alpha\rfloor+1. \tag{3.50}$$

Theorem 3.13. *Let the weak solution u given in* (1.8) *belong to $W^{r+1,q}(\Omega)$ and let the right-hand side functions belong to spaces $f\in W^{r,q}(\Omega)$ and $\varphi\in W^{r,q}(\partial\Omega)$. Let $\{\mathcal{T}_h\}_{h\in(0,h_0)}$ be a shape regular system of triangulations of Ω according to* (2.3)*. Let its boundary segments s_h be divided according to the cases above for a* *piecewise continuous Lagrange interpolation π_h of order r with $r+1$ nodes on sides of triangles. Let the approximate solution be given by* (3.14)*. Let the quadrature formulas on edges and on triangles be exact for polynomials of degree $\leq 2r-1$ and let the quadrature formula on edges satisfy* (3.15)*. Then there exist constants $c_1 = c_1(u, r, q, \Omega, \pi) > 0$, $c_2 = c_2(u, r, q, \Omega, \pi, \alpha) > 0$, $c_3 = c_3(u, r, q, \Omega, \pi, \alpha) > 0$, $c_4 = c_4(f, \varphi, r, \Omega, \pi) > 0$ such that*

$$\|u - u_h\|_{1,2,\Omega} \leq c_1 h^{r+1-\frac{2}{q}} + R_{-1}\left(c_2 h^{r+1-\frac{2}{q}} + c_3\left(h^r + h_2^{r_2}\right) + c_4 h^r\right), \tag{3.51}$$

if $q \in (1, 2)$ and

$$\|u - u_h\|_{1,2,\Omega} \leq c_1 h^r + R_{-1}\left(c_2 h^r + c_3\left(h^r + h_2^{r_2}\right) + c_4 h^r\right), \tag{3.52}$$

if $q \geq 2$, where R_{-1} is defined in (3.31)-(3.32)*.*

Proof. It was proven in Theorem 3.7 that the error $\|u - u_h\|_{1,2,\Omega}$ is bounded from above by

$$\|u - \pi_h u\|_{1,2,\Omega} + R_{-1}\Bigg(c\|u - \pi_h u\|_{1,2,\Omega}(1 + \|u\|^{\alpha}_{1,2,\Omega} + \|\pi_h u\|^{\alpha}_{1,2,\Omega}) + \sup_{0 \neq w_h \in H_h^r} \frac{|a(\pi_h u, w_h) - a_d(\pi_h u, w_h)|}{\|w_h\|_{1,2,\Omega}} + \sup_{0 \neq w_h \in H_h^r} \frac{|L(w_h) - L_d(w_h)|}{\|w_h\|_{1,2,\Omega}}\Bigg).$$

Estimation of $\|u - \pi_h u\|_{1,2,\Omega}$ by

$$\|u - \pi_h u\|_{1,2,\Omega} \leq |u|_{r+1,q,\Omega}\, h^{r+1-2/q}$$

for $q \in (1, 2)$ and by

$$\|u - \pi_h u\|_{1,2,\Omega} \leq c\, |u|_{r+1,q,\Omega}\, h^r$$

for $q \geq 2$ was done in the proof of Theorem 2.8. Inequality

$$\|\pi_h u\|_{1,2,\Omega} \leq c\, \|u\|_{1,2,\Omega}$$

follows from (3.41) if consider that $\|\tilde{\cdot}\|_{W^*(I)} = \|\cdot\|_{W^*(S)}$.

Since the quadrature formulas are exact for polynomials of degree $\leq 2r-1$ and

$$L(w_h) - L_d(w_h) = E_\Omega(f w_h) + E_{\partial\Omega}(\varphi w_h),$$

it follows from Theorem 3.3 that

$$\sup_{0 \neq w_h \in H_h^r} \frac{|L(w_h) - L_d(w_h)|}{\|w_h\|_{1,2,\Omega}} \leq c h^r \left(|f|_{r,q,\Omega} + |\varphi|_{r,q,\partial\Omega}\right).$$

Finally, we have

$$|a(\pi_h u, w_h) - a_d(\pi_h u, w_h)| = \sum_{S \in s_h} E_S\left(|\pi_h u|^{\alpha}(\pi_h u) w_h\right).$$

Errors on the segments $s_h = s_{h0} \cup s_{h1} \cup s_{h2}$ are estimated separately using Theorem 3.2. Since $u = \pi_h u = 0$ on segments $S \in s_{h0}$, we have

$$\sum_{S \in s_{h0}} E_S\left(|\pi_h u|^{\alpha}(\pi_h u) w_h\right) = 0.$$

On segments $S \in s_{h1}$, we can use the estimate

$$\left|E_S\left(|\pi_h u|^\alpha (\pi_h u) w_h\right)\right| \le c\,|S|^r\,\left||\pi_h u|^\alpha \pi_h u\right|_{r,q,S} \|w_h\|_{0,q',S}\,,$$

for $\frac{1}{q}+\frac{1}{q'}=1$, and then either (3.48) or (3.43)

$$\left||\pi_h u|^\alpha \pi_h u\right|_{r,q,S} \le c\left(\|u\|^\alpha_{r+1,q,\Omega}\left(+\|u\|^{2r-2-\alpha}_{r+1,q,\Omega}\right)\right)\|u\|_{r,q,S}\,,$$

which yields

$$\left|E_S\left(|\pi_h u|^\alpha (\pi_h u) w_h\right)\right| \le c(u) h^r \|u\|_{r,q,S} \|w_h\|_{0,q',S}\,.$$

Summing over all $S \in s_{h1}$, using discrete Hölder inequality and trace embedding, we conclude that

$$\begin{aligned}\left|\sum_{S\in s_{h1}} E_S\left(|\pi_h u|^\alpha (\pi_h u) w_h\right)\right| &\le \sum_{S\in s_{h1}} c(u) h^r \|u\|_{r,q,S}\|w_h\|_{0,q',S}\\ &\le c(u)h^r \|u\|_{r,q,\bigcup s_{h1}}\|w_h\|_{0,q',\bigcup s_{h1}}\\ &\le c(u)h^r\|u\|_{r+1,q,\Omega}\|w_h\|_{1,2,\Omega}\,.\end{aligned}$$

On segments $S \in s_{h2}$ we can similarly use the estimate

$$\left|E_S\left(|\pi_h u|^\alpha \pi_h u w_h\right)\right| \le c\,|S|^{r_2}\left||\pi_h u|^\alpha \pi_h u\right|_{r_2,q,S}\|w_h\|_{0,q',S}\,,$$

followed by (3.43)

$$\left||\pi_h u|^\alpha \pi_h u\right|_{r_2,q,S} \le c\,\|u\|^\alpha_{r+1,q,\Omega}\|u\|_{r_2,q,S}\,,$$

which leads to

$$\left|\sum_{S\in s_{h2}} E_S\left(|\pi_h u|^\alpha \pi_h u w_h\right)\right| \le c(u) h_2^{r_2}\|u\|_{r+1,q,\Omega}\|w_h\|_{1,2,\Omega}\,.$$

Combining these estimates yields the inequalities (3.51) and (3.52). □

Remark. In order to obtain the regularity $u \in W^{r+1,q}$ at least in the set Ω_1 from Theorem 1.13, we need $f \in W^{r-1,q}(\Omega)$ and $\varphi \in W^{r-1/q,q}(\partial\Omega)$. But the computations in this theorem needed $f \in W^{r,q}(\Omega)$ and $\varphi \in W^{r,q}(\partial\Omega)$. If f and φ had only the regularity which was necessary to obtain the regularity of the weak solution u, the order of convergence in this theorem would be decreased by 1.

Note that the function R_{-1} defined in (3.31)-(3.32) is linear, if the exact solution u is sufficiently distant from zero on a large part of the boundary $\partial\Omega$. Our theoretical estimates for the order of convergence in the H^1-norm are divided by $\alpha+1$ only if the exact solution is zero on most of the boundary. Similarly to the Galerkin approximation, we can improve the estimate for the rate of convergence in H^1-seminorm by omitting the denominator $\alpha+1$, if the exact solution u is zero on the whole boundary $\partial\Omega$. In this case, we also need to assume that the right-hand side integrals are evaluated exactly, that is

$$\int_\Omega f v_h \mathrm{d}x, \quad \int_{\partial\Omega}\varphi v_h \mathrm{d}S,$$

can be evaluated exactly for the given functions f, φ from (1.1)-(1.2), and $v_h \in H_h^r$. Whereas

$$\int_{\partial\Omega}|v_h|^\alpha v_h w_h, \quad v_h, w_h \in H_h^r$$

is evaluated with numerical quadrature. The argument is similar to Theorem 2.9.

Theorem 3.14. *Let the weak solution* $u \in W^{r+1,q}(\Omega)$ *given in* (1.8) *be zero on* $\partial\Omega$*. Let an approximate solution* $u_h \in H_h^r$ *be given by*

$$a_d(u_h, v_h) = L(v_h), \quad \forall v_h \in H_h^r, \tag{3.53}$$

where a_h *and* L *are defined in* (3.7) *and* (1.7)*. Let the quadrature formula on edges satisfy* (3.15)*. Then*

$$|u - u_h|_{1,2,\Omega} \leq \begin{cases} c\,|u|_{r+1,q,\Omega}\, h^{r+1-\frac{2}{q}}, & q \in [1, 2), \\ c\,|u|_{r+1,q,\Omega}\, h^r, & q \in [2, \infty). \end{cases} \tag{3.54}$$

Proof. Neglecting the second term on the right-hand side of (3.16) gives us

$$|u_h - \pi_h u|^2_{1,2,\Omega} \leq a_d(u_h, u_h - \pi_h u) - a_d(\pi_h u, u_h - \pi_h u).$$

The definitions of solutions u_h and u yield

$$a_d(u_h, u_h - \pi_h u) = L(u_h - \pi_h u) = a(u, u_h - \pi_h u).$$

Using the fact that u is zero on $\partial\Omega$ and thus the integral of $|u|^\alpha\, u(u_h - \pi_h u)$ on the boundary is evaluated exactly, we obtain

$$a(u, u_h - \pi_h u) = a_d(u, u_h - \pi_h u).$$

By Hölder inequality we finally get

$$\begin{aligned} a_d(u, u_h - \pi_h u) - a_d(\pi_h u, u_h - \pi_h u) &= \int_\Omega \nabla(u - \pi_h u) \cdot \nabla(u_h - \pi_h u) \mathrm{d}x \\ &\leq |u - \pi_h u|_{1,2,\Omega}\, |u_h - \pi_h u|_{1,2,\Omega}\,. \end{aligned}$$

Dividing

$$|u_h - \pi_h u|^2_{1,2,\Omega} \leq |u - \pi_h u|_{1,2,\Omega}\, |u_h - \pi_h u|_{1,2,\Omega}$$

by $|u_h - \pi_h u|_{1,2,\Omega}$ leads to and estimate

$$|u_h - \pi_h u|_{1,2,\Omega} \leq |u - \pi_h u|_{1,2,\Omega}\,.$$

Triangle inequality further gives us

$$|u - u_h|_{1,2,\Omega} \leq |u - \pi_h u|_{1,2,\Omega} + |\pi_h u - u_h|_{1,2,\Omega} \leq 2\,|u - \pi_h u|_{1,2,\Omega}\,.$$

The arguments from the proof of Theorem 2.8 give us the sought estimate. □

We have shown theoretically that using numerical integration for evaluating forms in the definition of the approximate solution will not decrease the order of convergence which was derived in section 2.3. In case of noninteger $\alpha > 0$ and the degree of used polynomials $r > \alpha + 1$, it might be necessary to refine the triangulation $\mathcal{T}_h$ near the roots of the exact solution u on the boundary $\partial\Omega$. These refined triangles T_S, $S \in s_{h2}$ would require their size to be $h_{T_S} \leq c h^{\frac{r}{\lfloor\alpha\rfloor+1}}$. Also note that the estimates in section 2.3 only required the solution to be regular, but the estimates in this chapter near the boundary edges were only able to use the same regularity that we were able to prove in section 1.4. Numerical experiments did not require this refinement to converge with the derived order of convergence.

Combining Theorem 2.8, ϱ_1^{-1} given in (2.20), and Theorem 2.9 suggested that the Galerkin approximation given in (2.7) should always converge to the exact weak solution defined in (1.8) in the H^1-seminorm with a rate of convergence of r. The same conclusions can be drawn from Theorem 3.13, R_{-1} given in (3.32) and Theorem 3.14 in this Chapter, which takes into account the effect of numerical integration. This theoretical result is in agreement with the numerical experiments.

4. Discontinuous Galerkin method

4.1 Discretization

A similar analysis can be carried out by the discretization using piecewise polynomial but in general discontinuous functions. Consider a polygonal domain $\Omega \subset \mathbb{R}^2$ with a conforming triangulation $\mathcal{T}_h$ satisfying (2.1). We denote the set of all faces by $\mathcal{F}_h$ and we further distinguish the set of all boundary faces

$$\mathcal{F}_h^B = \{\Gamma \in \mathcal{F}_h;\, \Gamma \subset \partial\Omega\}\,,$$

and the set of all inner faces

$$\mathcal{F}_h^I = \mathcal{F}_h \setminus \mathcal{F}_h^B.$$

The set $\mathcal{F}_h^B$ was denoted by s_h in the previous chapters concerning FEM. For every inner face $\Gamma \in \mathcal{F}_h^I$ we choose an arbitrary but fixed unit vector n_Γ orthogonal to Γ. Then there are two neighbouring triangles $T_\Gamma^{(L)}$, $T_\Gamma^{(R)} \in \mathcal{T}_h$ such that $\Gamma = T_\Gamma^{(L)} \cap T_\Gamma^{(R)}$ and we choose $T_\Gamma^{(L)}$ to be the one with outer normal n_Γ (thus making n_Γ an inner normal to $T_\Gamma^{(R)}$). For any boundary face $\Gamma \subset \partial\Omega$, by $T_\Gamma^{(L)}$ we denote the element from $\mathcal{T}_h$ adjacent to Γ and we set n_Γ to be the outer normal.

For $k \in \mathbb{N}$, $q \geq 1$ and a triangulation $\mathcal{T}_h$ we define a broken Sobolev space

$$W^{k,q}(\Omega, \mathcal{T}_h) = \left\{ v \in L^2(\Omega);\, v|_T \in W^{k,q}(T)\, \forall T \in \mathcal{T}_h \right\} \tag{4.1}$$

and also $H^k(\Omega, \mathcal{T}_h) = W^{k,2}(\Omega, \mathcal{T}_h)$.

For functions $v \in W^{k,p}(\Omega, \mathcal{T}_h)$ and inner faces $\Gamma \in \mathcal{F}_h^I$, we introduce notation

$$\begin{aligned} v|_\Gamma^{(L)} &= \text{trace of}\, v|_{T_\Gamma^{(L)}} \,\text{on}\, \Gamma, \quad & v|_\Gamma^{(R)} &= \text{trace of}\, v|_{T_\Gamma^{(R)}} \,\text{on}\, \Gamma, \\ \langle v \rangle_\Gamma &= \frac{1}{2}\left(v|_\Gamma^{(L)} + v|_\Gamma^{(R)} \right), \quad & [v]_\Gamma &= v|_\Gamma^{(L)} - v|_\Gamma^{(R)}, \end{aligned} \tag{4.2}$$

for the left trace, the right trace, the mean value of traces and the jump of v on Γ. Even though the value of $[v]_\Gamma$ depends on the choice of the orientation of n_Γ, the vector $[v]_\Gamma n_\Gamma$ is independent of this orientation.

The approximate solution will be sought in a space of discontinuous piecewise polynomial functions of degree $r \in \mathbb{N}$:

$$S_h^r = \left\{ v_h \in L^1(\Omega);\, v_h|_T \in P_r(T),\, T \in \mathcal{T}_h \right\}. \tag{4.3}$$

We suppose that the set of triangulations $\{T_h\}_{h>0}$ is shape regular in accordance with (2.3). Due to Theorem 1.7 the weak solution u given by (1.8) belongs to the space $W^{2,q}(\Omega)$ for some $q > \frac{4}{3}$ dependent on the largest inner angle in Ω. From Sobolev embedding (1.3), it follows that u is in fact continuous with $[u]_\Gamma = 0$. In this section we give another definition for a (discontinuous) weak solution with different forms. This newly defined formulation will be again satisfied by the same previously defined weak solution. Therefore, only the numerical methods for finding it will differ.

The weak solution can be derived by taking the equation (1.1) for the classical solution u in Ω, multiplying it by a test function $v \in W^{2,q}(\Omega, \mathcal{T}_h)$, $q > 1$ integrating over an element $T \in \mathcal{T}_h$, using Green's theorem and boundary condition (1.2), summing over all elements $T \in \mathcal{T}_h$ and possibly adding some terms which are zero for the weak solution.

$$\int_T fv\mathrm{d}x = \int_T -\Delta uv\mathrm{d}x = -\int_{\partial T} \frac{\partial u}{\partial n} v\mathrm{d}S + \int_T \nabla u \cdot \nabla v\mathrm{d}x,$$

$$-\int_{\partial\Omega} \frac{\partial u}{\partial n} v\mathrm{d}S = \int_{\partial\Omega} \left(\kappa \left|u\right|^\alpha u - \varphi\right) v\mathrm{d}S,$$

$$\frac{\partial u}{\partial n}\Big|_{T_\Gamma^{(L)}} v|_{T_\Gamma^{(L)}} + \frac{\partial u}{\partial n}\Big|_{T_\Gamma^{(R)}} v|_{T_\Gamma^{(R)}} = n_\Gamma \cdot \langle \nabla u \rangle v|_{T_\Gamma^{(L)}} - n_\Gamma \cdot \langle \nabla u \rangle v|_{T_\Gamma^{(R)}} = n_\Gamma \cdot \langle \nabla u \rangle [v],$$

$$\int_\Omega fv\mathrm{d}x = \int_{\partial\Omega} \left(\kappa \left|u\right|^\alpha u - \varphi\right) v\mathrm{d}S - \sum_{\Gamma \in \mathcal{F}_h^I} \int_\Gamma n_\Gamma \cdot \langle \nabla u \rangle [v]\mathrm{d}S + \sum_{T \in \mathcal{T}_h} \int_T \nabla u \cdot \nabla v\mathrm{d}x.$$

These considerations lead to the following form defined for functions $u \in W^{2,q}(\Omega), q > \frac{4}{3}$, $v \in (\Omega, \mathcal{T}_h)$:

$$\begin{aligned} b_h(u,v) = &\sum_{T \in \mathcal{T}_h} \int_T \nabla u \cdot \nabla v\mathrm{d}x \\ &- \sum_{\Gamma \in \mathcal{F}_h^I} \int_\Gamma \left(n_\Gamma \cdot \langle \nabla u \rangle [v] + \theta n_\Gamma \cdot \langle \nabla v \rangle [u]\right) \mathrm{d}S. \end{aligned} \tag{4.4}$$

This form represents the left-hand side integration over elements $T \in \mathcal{T}_h$. The terms multiplied by a parameter θ are zero for the weak solution. This parameter can be chosen as $1, 0, -1$, which leads to symmetric, incomplete and non-symmetric versions of the diffusion forms (4.9)-(4.11) denoted by SIPG, IIPG, NIPG, respectively. Further, we introduce the interior penalty form

$$J_h(u,v) = \sum_{\Gamma \in \mathcal{F}_h^I} \int_\Gamma \sigma[u][v]\mathrm{d}.S \tag{4.5}$$

The coefficient σ is given by

$$\sigma = \frac{C_W}{h_\Gamma}, \tag{4.6}$$

where h_Γ is the length of the face Γ and $C_W > 0$ will be specified later. The form d represents the nonlinear boundary term and d_d is the form d evaluated using numerical quadrature formula (3.4):

$$d(u,v) = \kappa \sum_{\Gamma \in \mathcal{F}_h^B} \int_\Gamma \left|u\right|^\alpha uv\mathrm{d}S = \kappa \int_{\partial\Omega} \left|u\right|^\alpha uv\mathrm{d}S, \tag{4.7}$$

$$d_d(u,v) = \kappa \sum_{\Gamma \in \mathcal{F}_h^B} |\Gamma| \sum_{\mu=1}^m \beta_\mu \left(\left|u\right|^\alpha uv\right)(x_{\Gamma,\mu}). \tag{4.8}$$

Combining these forms into a_h, A_h, and A_h^d will allow us to define the Galerkin approximation and the approximate solution using relatively simple formulae.

$$a_h(u,v) = b_h(u,v) + J_h(u,v), \tag{4.9}$$

$$A_h(u,v) = a_h(u,v) + d(u,v), \tag{4.10}$$

$$A_{dh}(u,v) = a_h(u,v) + d_d(u,v). \tag{4.11}$$

Definition 4.1. *We call $u_h \in S_h^r$ the discontinuous Galerkin approximation of the weak solution given in* (1.8) *if*

$$A_h(u_h, v_h) = L(v_h) \quad \forall v_h \in S_h^r. \tag{4.12}$$

Definition 4.2. *We call $u_{hd} \in S_h^r$ the discontinuous approximate solution of the problem* (1.1)-(1.2) *if*

$$A_{dh}(u_{dh}, v_h) = L_d(v_h) \quad \forall v_h \in S_h^r. \tag{4.13}$$

The broken Sobolev space $H^1(\Omega, \mathcal{T}_h)$ and its subspace S_h^r are Banach spaces with seminorms

$$|v|_{H^1(\Omega,\mathcal{T}_h)} = \left(\sum_{T\in\mathcal{T}_h} \int_T |\nabla v|^2 \, \mathrm{d}x \right)^{\frac{1}{2}}, \tag{4.14}$$

$$|v|_h = \left(\sum_{T\in\mathcal{T}_h} \int_T |\nabla v|^2 \, \mathrm{d}x + J_h(v,v) \right)^{\frac{1}{2}}, \tag{4.15}$$

and a norm

$$|||v||| = \left(|v|_h^2 + \|v\|_{L^2(\Omega)}^2 \right)^{\frac{1}{2}}. \tag{4.16}$$

Let us summarize some basic properties used in the analysis of discontinuous Galerkin method, see Theorems 4.1 and 4.4 in [3], Sections 2.5.1, 2.5.2, and 2.6.3 in [5] and Lemma 5.3 and Lemma 5.4 in [11].

Lemma 4.3. *There exists a constant $c = c(q) > 0$ such that the following inequalities hold for all $q \geq 1$, $v_h \in S_h^r$, $h > 0$:*

$$\|v_h\|_{0,q,\Omega} \leq c|||v_h|||, \tag{4.17}$$

$$\|v_h\|_{0,q,\partial\Omega} \leq c|||v_h|||. \tag{4.18}$$

There exists a constant $C_I > 0$ such that the following inverse inequality holds for all $v_h \in P_r(T)$, $T \in \mathcal{T}_h$, $h > 0$:

$$|v_h|_{1,2,T} \leq C_I h_T^{-1} \|v_h\|_{0,2,T}. \tag{4.19}$$

There exists a constant $C_M > 0$ such that the following multiplicative trace inequality holds for all $v \in H^1(T)$, $T \in \mathcal{T}_h$, $h > 0$:

$$\|v\|_{0,2,\partial T}^2 \leq C_M \left(\|v\|_{0,2,T} |v|_{1,2,T} + h_T^{-1} \|v\|_{0,2,T}^2 \right), \tag{4.20}$$

and the following multiplicative trace inequality holds for all $v \in W^{1,q}(T)$, $T \in \mathcal{T}_h$, $h > 0$, $q \in \left(\frac{4}{3}, 2\right)$, $\frac{1}{q} + \frac{1}{q'} = 1$:

$$\|v\|_{0,2,\partial T}^2 \leq C_M \left(\|v\|_{0,q',T} |v|_{1,q,T} + h_T^{-1} \|v\|_{0,2,T}^2 \right). \tag{4.21}$$

Let the constant C_W in (4.6) *satisfy the following conditions for θ from* (4.4)*:*

$$C_W > 0, \text{ for } \theta = -1 \text{ (NIPG)}, \tag{4.22}$$

$$C_W > 4C_M(1 + C_I), \text{ for } \theta = 1 \text{ (SIPG)}, \tag{4.23}$$

$$C_W > C_M(1 + C_I), \text{ for } \theta = 0 \text{ (IIPG)}, \tag{4.24}$$

then the form a_h is coercive in the following way:

$$a_h(v_h, v_h) \geq \frac{1}{2} |v_h|_h^2 \tag{4.25}$$

holds for all $v_h \in S_h^r$, $h > 0$.

4.2 Monotonicity and continuity

Deriving error estimates will require us to have lower estimates of the forms in the definition of the DG solution, which will follow from monotonicity, upper estimates following from continuity, and combining them with Galerkin orthogonality and interpolation error estimates.

The continuity of the form A_h was proven in [11] Lemma 5.6.

Lemma 4.4. *For $q > \frac{4}{3}$ there exists a constant $c > 0$ such that*

$$|A_h(u,w) - A_h(v,w)| \leq c\Big\{ |||u - v||| + R_h(u - v, q) + G_h(u - v)\left(\|u\|_{1,2,\Omega}^{\alpha} + |||v|||^{\alpha} \right) \Big\} |||w||| \tag{4.26}$$

holds for all $u \in W^{2,q}(\Omega, \mathcal{T}_h)$, $v, w \in S_h^r$, $h > 0$, where

$$R_h(\phi, q) = \left(C_M \sum_{T \in \mathcal{T}_h} h_T |\phi|_{1,q',T} |\phi|_{2,q,T} \right)^{1/2} \tag{4.27}$$

for $\phi \in W^{2,q}(\Omega, \mathcal{T}_h)$, $q \in \left(\frac{4}{3}, 2\right)$, $\frac{1}{q} + \frac{1}{q'} = 1$ and

$$R_h(\phi, q) = \left(C_M \sum_{T \in \mathcal{T}_h} h_T |\phi|_{1,2,T} |\phi|_{2,2,T} \right)^{1/2}, \tag{4.28}$$

for $\phi \in W^{2,q}(\Omega, \mathcal{T}_h)$, $q \geq 2$. Moreover,

$$G_h(\phi) = \left(C_M \sum_{T \in \mathcal{T}_h} \left(\|\phi\|_{0,2,T}^2 h_T^{-1} + |\phi|_{1,2,T} \|\phi\|_{0,2,T} \right) \right)^{1/2} \tag{4.29}$$

for $\phi \in H^1(\Omega, \mathcal{T}_h)$.

Lemma 4.5. *Let the constant C_W satisfy conditions (4.22)-(4.24). Then the form A_h is uniformly monotone on the space S_h^r in the following way: There exists a continuous increasing function $\rho : [0, \infty) \to [0, \infty)$ such that*

$$A_h(u_h, u_h - v_h) - A_h(v_h, u_h - v_h) \geq \rho\left(|||u_h - v_h||| \right), \quad u_h, v_h \in S_h^r, h > 0. \tag{4.30}$$

This lemma was proven in [11] Lemma 5.7 with a monotone function

$$\rho(t) = \begin{cases} ct^{\alpha+2} & \text{for} \quad t \in [0,1], \\ ct^2 & \text{for} \quad t \in [1, \infty). \end{cases} \tag{4.31}$$

It follows from monotone operator theory, see [13], [30], that there exists exactly one discontinuous Galerkin approximation $u_h \in S_h^r$ defined in (4.12).

We improve the monotonicity with condition

$$|v_h| \geq \varepsilon > 0 \text{ on } G \subset \partial\Omega, \; |G| > 0, \; h > 0, \tag{4.32}$$

which follows from (3.45) and (2.8) for sufficiently refined triangulation $\mathcal{T}_h$. Note that G and ε might be smaller than those given in (2.8), but they are not decreasing as $\mathcal{T}_h$ is refined further. Combining the coercivity (4.25) with (4.32) allows us to redefine ρ.

Lemma 4.6. *Let C_W satisfy conditions* (4.22)-(4.24) *and let the condition* (4.32) *hold. Then* (4.30) *holds for a function*

$$\rho(t) = ct^2, \quad t \in [0, \infty). \tag{4.33}$$

Proof. It follows from the definition (4.10) of A_h, linearity of a_h, (4.25), Theorem 2.4, and Poincaré inquality (1.5) that

$$\begin{aligned}
&A_h(u_h, u_h - v_h) - A_h(v_h, u_h - v_h)\\
&\quad = a_h(u_h - v_h, u_h - v_h) + d(u_h, u_h - v_h) - d(v_h, u_h - v_h)\\
&\quad \geq \frac{1}{2}\left|u_h - v_h\right|_h^2 + c\left\|u_h - v_h\right\|_{0,2,G}^2\\
&\quad \geq c|||u_h - v_h|||^2.
\end{aligned}$$

□

The same results can be obtained if the boundary terms are evaluated with numerical quadrature.

Lemma 4.7. *Let the quadrature formula* (3.4) *used in the evaluation of the form d_d (and thus A_{dh}) use at least $r+1$ nodes and have all weights positive, i.e. let it satisfy* (3.15)*. Let C_W satisfy conditions* (4.22)-(4.24)*. Then the form A_{dh} is uniformly monotone on S_h^r in the following way: There exists a continuous increasing function $\tilde{\rho} : [0, \infty) \to [0, \infty)$ such that*

$$A_{dh}(u_h, u_h - v_h) - A_{dh}(v_h, u_h - v_h) \geq \tilde{\rho}\left(|||u_h - v_h|||\right), \quad u_h, v_h \in S_h^r, h > 0, \tag{4.34}$$

where

$$\tilde{\rho}(t) = \begin{cases} ct^{\alpha+2} & \text{for} \quad t \in [0, 1],\\ ct^2 & \text{for} \quad t \in [1, \infty). \end{cases} \tag{4.35}$$

If the condition (4.32) *also holds, then* (4.34) *holds for a function*

$$\tilde{\rho}(t) = ct^2, \quad t \in [0, \infty). \tag{4.36}$$

Proof. Using the definition (4.11) of A_{dh} and (4.25), we have

$$\begin{aligned}
&A_{dh}(u_h, u_h - v_h) - A_{dh}(v_h, u_h - v_h)\\
&\quad = a_h(u_h - v_h, u_h - v_h) + d_d(u_h, u_h - v_h) - d_d(v_h, u_h - v_h)\\
&\quad \geq \frac{1}{2}\left|u_h - v_h\right|_h^2 + \kappa \sum_{S \in s_h} |S| Z_S(u_h, v_h),
\end{aligned}$$

where

$$Z_S(u_h, v_h) = \sum_{\mu=1}^{m} \beta_\mu \left(|u_h|^\alpha u_h - |v_h|^\alpha v_h\right)(u_h - v_h)(x_{S,\mu}). \tag{4.37}$$

The inequalities

$$Z_S(u_h, v_h) \geq c\left\|u_h - v_h\right\|_{0,\alpha+2,\Omega}^{\alpha+2}, \tag{4.38}$$

and

$$Z_S(u_h, v_h) \geq c\left\|u_h - v_h\right\|_{0,2,\Omega}^2, \tag{4.39}$$

for (4.32) were shown in the proof of Theorem 3.5. Combining these inequalities gives us (4.34) with (4.35) and (4.36). □

We again define monotone functions ρ_1 and $\tilde{\rho}_1$ for the error estimates as

$$\rho_1(t) = \frac{\rho(t)}{t}, \quad \tilde{\rho}_1(t) = \frac{\tilde{\rho}(t)}{t}, \qquad t \geq 0, \tag{4.40}$$

and ρ_1^{-1} and $\tilde{\rho}_1^{-1}$ are their inverses. The estimates above show that ρ_1 and $\tilde{\rho}_1$ are either linear or grow with $t^{\alpha+1}$, and ρ_1^{-1} and $\tilde{\rho}_1^{-1}$ are thus either linear or grow with $t^{\frac{1}{\alpha+1}}$.

The last step needed before deriving error estimates is to express in some simple terms the interpolation errors $u - \pi_h u$ evaluated in functions $R_h(u - \pi_h u, q)$ and $G_h(u - \pi_h u)$ from the continuity of A_h.

Lemma 4.8. *Let π_h be a piecewise Lagrange interpolation of order $r \in \mathbb{N}$, let the system of triangulations $\{\mathcal{T}_h\}$, $h > 0$ be shape regular in accordance with* (2.3)*, let $u \in W^{k,q}(\Omega)$, $k \in \mathbb{N}$, $k \geq 2$, $\nu = \min(r+1, k)$. Then*

$$R_h(u - \pi_h u, q) \leq c \left(\sum_{T \in \mathcal{T}_h} h_T^{2(\nu - 2/q)} \, |u|_{\nu,q,T}^2 \right)^{1/2}, \tag{4.41}$$

where $c > 0$ is independent of $q > \frac{4}{3}$ and $h > 0$. There also exists $c > 0$ independent of $h > 0$ such that

$$G_h(u - \pi_h u) \leq c \left(\sum_{T \in \mathcal{T}_h} h_T^{2\nu + 1 - 4/q} \, |u|_{\nu,q,T}^2 \right)^{1/2}. \tag{4.42}$$

Proof. Both estimates follow from applying (2.16) to (4.27)-(4.29). Let us show, for example, (4.27).

We have $q \in \left(\frac{4}{3}, 2\right)$ and $\frac{1}{q} + \frac{1}{q'} = 1$, (2.16) gives us:

$$|u - \pi_h u|_{1,q',T} \leq c \, |u|_{\nu,q,T} \, h_T^{\nu - 1 + 2\left(\frac{1}{q'} - \frac{1}{q}\right)},$$

$$|u - \pi_h u|_{2,q,T} \leq c \, |u|_{\nu,q,T} \, h_T^{\nu - 2}.$$

Using this in the right-hand side of (4.27) yields

$$\left(C_M \sum_{T \in \mathcal{T}_h} h_T \, |u - \pi_h u|_{1,q',T} \, |u - \pi_h u|_{2,q,T} \right)^{1/2}$$

$$\leq c \left(\sum_{T \in \mathcal{T}_h} h_T \, |u|_{\nu,q,T}^2 \, h_T^{\nu - 1 + 2\left(1 - \frac{2}{q}\right) + \nu - 2} \right)^{1/2},$$

which is (4.41) for $q \in \left(\frac{4}{3}, 2\right)$. □

4.3 Error estimation

Let us begin by estimating the error of DG approximation.

Theorem 4.9. *Let $u \in W^{r+1,q}(\Omega)$, $q > \frac{4}{3}$ be the weak solution given by* (1.8)*, let u_h be the discontinuous Galerkin approximation of order r given by* (4.12)*, let the system of triangluations $\mathcal{T}_h$, $h > 0$ be shape regular in accordance with* (2.3)*, let C_W satisfy* (4.22)-(4.24)*. Then there exist constants $c_1, c_2 > 0$ independent of u, h, such that*

$$\|\|u - u_h\|\| \le \rho_1^{-1}\left(c_1 h^{r+1-2/q} |u|_{r+1,q,\Omega}\left(1 + h^{1/2}\|u\|_{1,2,\Omega}^{\alpha}\right)\right) + c_2 h^{r+1-2/q}|u|_{r+1,q,\Omega} \tag{4.43}$$

for $q \in \left(\frac{4}{3}, 2\right)$, and

$$\|\|u - u_h\|\| \le \rho_1^{-1}\left(c_1 h^{r} |u|_{r+1,q,\Omega}\left(1 + h^{1/2}\|u\|_{1,2,\Omega}^{\alpha}\right)\right) + c_2 h^{r}|u|_{r+1,q,\Omega} \tag{4.44}$$

for $q \ge 2$, where ρ_1 was given in (4.40)*.*

Proof. Let π_h be a continuous piecewise Lagrange interpolation operator of order r. Then $\|\|u\|\| = \|u\|_{1,2,\Omega}$ and $\|\|\pi_h u\|\| = \|\pi_h u\|_{1,2,\Omega}$. By virtue of (4.30) and the definitions of solutions u_h and u,

$$\begin{aligned}\rho\left(\|\|u_h - \pi_h u\|\|\right) &\le A_h(u_h, u_h - \pi_h u) - A_h(\pi_h u, u_h - \pi_h u)\\ &= L(u_h - \pi_h u) - A_h(\pi_h u, u_h - \pi_h u)\\ &= A_h(u, u_h - \pi_h u) - A_h(\pi_h u, u_h - \pi_h u).\end{aligned}$$

This relation and Lemma 4.4 give us

$$\begin{aligned}\rho_1\left(\|\|u_h - \pi_h u\|\|\right) \le c\Bigg(&\|u - \pi_h u\|_{1,2,\Omega} + R_h(u - \pi_h u, q)\\ &+ G_h(u - \pi_h u)\left(\|u\|_{1,2,\Omega}^{\alpha} + \|\pi_h u\|_{1,2,\Omega}^{\alpha}\right)\Bigg).\end{aligned}$$

Lemma 4.8 and inequality

$$\|\pi_h u\|_{1,2,\Omega} \le c\|u\|_{1,2,\Omega}$$

following from (2.16) imply

$$\begin{aligned}\rho_1\left(\|\|u_h - \pi_h u\|\|\right) \le c\Bigg(&\|u - \pi_h u\|_{1,2,\Omega} + \left(\sum_{T\in\mathcal{T}_h} h_T^{2(r+1-2/q)}|u|_{r+1,q,T}^2\right)^{1/2}\\ &+ \left(\sum_{T\in\mathcal{T}_h} h_T^{2r+3-4/q}|u|_{r+1,q,T}^2\right)^{1/2}\|u\|_{1,2,\Omega}^{\alpha}\Bigg).\end{aligned}$$

The term $\|u - \pi_h u\|_{1,2,\Omega}$ was estimated in Theorem 2.8 and the same ideas from its proof give us

$$\left(\sum_{T\in\mathcal{T}_h} h_T^{2(r+1-2/q)}|u|_{r+1,q,T}^2\right)^{1/2} \le \begin{cases} h^{r+1-2/q}|u|_{r+1,q,\Omega}, & q \in \left(\frac{4}{3}, 2\right),\\ ch^{r}|u|_{r+1,q,\Omega}, & q \ge 2.\end{cases}$$

Using this inequality yields

$$\rho_1\left(\|\|u_h - \pi_h u\|\|\right) \le \begin{cases} ch^{r+1-2/q}|u|_{r+1,q,\Omega}\left(1 + h^{1/2}\|u\|_{1,2,\Omega}^{\alpha}\right), & q \in \left(\frac{4}{3}, 2\right),\\ ch^{r}|u|_{r+1,q,\Omega}\left(1 + h^{1/2}\|u\|_{1,2,\Omega}^{\alpha}\right), & q \ge 2.\end{cases}$$

Triangle inequality

$$|||u - u_h||| \leq \|u - \pi_h u\|_{1,2,\Omega} + |||u_h - \pi_h u|||$$

and the estimate of $\|u - \pi_h u\|_{1,2,\Omega}$ from Theorem 2.8 completes the proof. $\square$

The discontinuous approximate solution can be estimated similarly with ideas from Theorem 3.7 and Theorem 3.13. Let us again separate the boundary faces $\mathcal{F}_h^B$ into three disjoint sets $\mathcal{F}_h^B = \mathcal{F}_{h0}^B \cup \mathcal{F}_{h1}^B \cup \mathcal{F}_{h2}^B$.

- $\mathcal{F}_{h0}^B$ contains segments Γ with $u|_\Gamma = 0$. Then also $\pi_h u|_\Gamma = 0$ and the quadrature is exact there, i.e. $E_\Gamma(|\pi_h u|^\alpha (\pi_h u) w_h) = 0$.
- If $\alpha + 1 \geq r$ or $\alpha \in \mathbb{N}_0$, then $\mathcal{F}_{h1}^B$ contains all segments not in $\mathcal{F}_{h0}^B$. If $\alpha \notin \mathbb{N}_0$ and $\alpha + 1 < r$, then $\mathcal{F}_{h1}^B$ contains all segments not in $\mathcal{F}_{h0}^B$ satisfying $\frac{\min_\Gamma |u|}{\max_\Gamma |u|} \geq C_l$, where C_l is given by (3.45). Then combining (3.48) (or (3.43)) and (3.8) gives us an error estimate of order r.
- $\mathcal{F}_{h2}^B$ contains the remaining segments, i.e. for $\alpha \notin \mathbb{N}_0$ and $\alpha + 1 < r$, $\mathcal{F}_{h2}^B$ contains segments satisfying $\frac{\min_\Gamma |u|}{\max_\Gamma |u|} < C_l$ and u is not identically zero on Γ. Let us set $h_2 = \max\left\{|\Gamma|\,; \Gamma \in \mathcal{F}_{h2}^B\right\}$ (or $h_2 = 0$ if there are no segments in $\mathcal{F}_{h2}^B$). Combining (3.43) and (3.8) gives us an error estimate of order $r_2 = \lfloor \alpha \rfloor + 1$.

Theorem 4.10. *Let $u \in W^{r+1,q}(\Omega)$, $q > \frac{4}{3}$ be the weak solution given by* (1.8)*, let $f \in W^{r,q}(\Omega)$ and $\varphi \in W^{r,q}(\partial\Omega)$, let u_{dh} be the discontinuous Galerkin approximation of order r given by* (4.13)*, let C_W satisfy* (4.22)-(4.24)*, let the system of triangluations $\mathcal{T}_h$, $h > 0$ be shape regular in accordance with* (2.3)*, let the quadrature formula* (3.4) *used in the evaluation of the form d_d (and thus A_{dh}) satisfy* (3.15)*, let the quadrature formulas on edges and on triangles be exact for polynomials of degree $\leq 2r - 1$, let the boundary faces $\mathcal{F}_h^B$ be divided into $\mathcal{F}_{h0}^B$, $\mathcal{F}_{h1}^B$, $\mathcal{F}_{h2}^B$ as above. Then there exist constants $c_1, c_2, c_3 > 0$ independent of h such that*

$$|||u - u_{dh}||| \leq c_1 h^{r+1-\frac{2}{q}} + \tilde{\rho}_1^{-1}\left(c_2 h^{r+1-\frac{2}{q}} + c_3\left(h^r + h_2^{r_2}\right)\right), \tag{4.45}$$

if $q \in \left(\frac{4}{3}, 2\right)$, and

$$|||u - u_{dh}||| \leq c_1 h^r + \tilde{\rho}_1^{-1}\left(c_2 h^r + c_3\left(h^r + h_2^{r_2}\right)\right), \tag{4.46}$$

if $q \geq 2$, where $\tilde{\rho}_1$ was given in (4.40)*.*

Proof. Let π_h be a continuous piecewise Lagrange interpolation operator of order r. Inequality (4.34) gives us

$$\tilde{\rho}\left(|||u_{dh} - \pi_h u|||\right) \leq A_{dh}(u_{dh}, u_{dh} - \pi_h u) - A_{dh}(\pi_h u, u_{dh} - \pi_h u). \tag{4.47}$$

Using

$$A_{dh}(u_{dh}, u_{dh} - \pi_h u) = L_d(u_{dh} - \pi_h u),$$

$$L(u_{dh} - \pi_h u) = A_h(u, u_{dh} - \pi_h u),$$

and adding and subtracting the same terms, we get

$$\begin{aligned}A_{dh}(u_{dh}, u_{dh} - \pi_h u) - A_{dh}(\pi_h u, u_{dh} - \pi_h u) = &[L_d(u_{dh} - \pi_h u) - L(u_{dh} - \pi_h u)]\\ &+ [A_h(u, u_{dh} - \pi_h u) - A_h(\pi_h u, u_{dh} - \pi_h u)]\\ &+ [A_h(\pi_h u, u_{dh} - \pi_h u) - A_{dh}(\pi_h u, u_{dh} - \pi_h u)].\end{aligned}$$

The first bracket can be estimated with (3.13) and (3.12):

$$\begin{aligned}&|L_d(u_{dh} - \pi_h u) - L(u_{dh} - \pi_h u)|\\ &\qquad \leq ch^r \left(|f|_{r,q,\Omega} \|v_h\|_{0,q',\Omega} + |\varphi|_{r,q,\partial\Omega} \|v_h\|_{0,q',\partial\Omega} \right),\end{aligned}$$

which can be further estimated with (4.17) and (4.18) to obtain

$$|L_d(u_{dh} - \pi_h u) - L(u_{dh} - \pi_h u)| \leq ch^r |||v_h|||.$$

The third bracket can be estimated similarly to find

$$|A_h(\pi_h u, u_{dh} - \pi_h u) - A_{dh}(\pi_h u, u_{dh} - \pi_h u)| \leq ch^r \left| |\pi_h u|^\alpha \pi_h u \right|_{r,q,\partial\Omega} |||v_h|||,$$

which was already estimated in the proof of Theorem 3.13. The second bracket was estimated in the proof of Theorem 4.9 by

$$ch^{r+1-2/q} |u|_{r+1,q,\Omega} \left(1 + h^{1/2} \|u\|^\alpha_{1,2,\Omega}\right), \quad q \in \left(\frac{4}{3}, 2\right),$$

$$ch^r |u|_{r+1,q,\Omega} \left(1 + h^{1/2} \|u\|^\alpha_{1,2,\Omega}\right), \quad q \geq 2.$$

Combining these inequalities, using triangle inequality

$$|||u - u_h||| \leq \|u - \pi_h u\|_{1,2,\Omega} + |||u_h - \pi_h u|||$$

and the estimate of $\|u - \pi_h u\|_{1,2,\Omega}$ from Theorem 2.8 completes the proof. □

If the exact weak solution is zero on the whole boundary, we can again similarly improve the estimate of the order of convergence in the $|\cdot|_h$-seminorm.

Theorem 4.11. *Let the weak solution $u \in W^{r+1,q}(\Omega)$, $q > \frac{4}{3}$ given by* (1.8) *be zero on $\partial\Omega$, let u_h be the discontinuous Galerkin approximation of order r given by* (4.12)*, let the system of triangluations $\mathcal{T}_h$, $h > 0$ be shape regular in accordance with* (2.3)*, let C_W satisfy* (4.22)-(4.24)*. Then there exists a constant $c > 0$ independent of u, h, such that*

$$|u - u_h|_h \leq \begin{cases} ch^{r+1-2/q} |u|_{r+1,q,\Omega}, & q \in \left(\frac{4}{3}, 2\right), \\ ch^r |u|_{r+1,q,\Omega}, & q \geq 2. \end{cases} \tag{4.48}$$

Proof. Let $\pi_h u$ be a piecewise continuous Lagrange interpolation of u. Using (4.25), the definitions of norms a_h, A_h, J_h in (4.9), (4.10), (4.5), Hölder inequality, and the definitions of the solutions u and u_h in (1.8), (4.12), we have

$$
\begin{aligned}
&\frac{1}{2}\left|u_h-\pi_h u\right|_h^2 \\
&=\frac{1}{2}\left(\sum_{T\in\mathcal{T}_h}\int_T\left|\nabla(u_h-\pi_h u)\right|^2\mathrm{d}x+\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma\sigma[u_h-\pi_h u]^2\mathrm{d}S\right) \\
&\le a_h(u_h-\pi_h u,u_h-\pi_h u) \\
&\le A_h(u_h,u_h-\pi_h u)-A_h(\pi_h u,u_h-\pi_h u) \\
&=L(u_h-\pi_h u)-A_h(\pi_h u,u_h-\pi_h u) \\
&=A_h(u,u_h-\pi_h u)-A_h(\pi_h u,u_h-\pi_h u) \\
&=\sum_{T\in\mathcal{T}_h}\int_T\nabla(u-\pi_h u)\cdot\nabla(u_h-\pi_h u)\mathrm{d}x \\
&\quad+\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma\sigma[u-\pi_h u][u_h-\pi_h u]\mathrm{d}S \\
&\quad-\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma\left(n_\Gamma\cdot\langle\nabla(u-\pi_h u)\rangle[u_h-\pi_h u]+\theta n_\Gamma\cdot\langle\nabla(u_h-\pi_h u)\rangle[u-\pi_h u]\right)\mathrm{d}S \\
&\quad+\kappa\int_{\partial\Omega}\left(|u|^\alpha u-|\pi_h u|^\alpha\pi_h u\right)(u_h-\pi_h u)\mathrm{d}S.
\end{aligned}
\tag{4.49}
$$

Hölder inequality gives us

$$
\begin{aligned}
&\sum_{T\in\mathcal{T}_h}\int_T\nabla(u-\pi_h u)\cdot\nabla(u_h-\pi_h u)\mathrm{d}x+\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma\sigma[u-\pi_h u][u_h-\pi_h u]\mathrm{d}S \\
&\qquad\le\left|u-\pi_h u\right|_h\left|u_h-\pi_h u\right|_h.
\end{aligned}
$$

Since $u=\pi_h u=0$ on $\partial\Omega$,

$$
\kappa\int_{\partial\Omega}\left(|u|^\alpha u-|\pi_h u|^\alpha\pi_h u\right)(u_h-\pi_h u)\mathrm{d}S=0.
$$

Both u and $\pi_h u$ are continuous, therefore

$$
\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma\theta n_\Gamma\cdot\langle\nabla(u_h-\pi_h u)\rangle[u-\pi_h u]\mathrm{d}=0.
$$

Using Hölder inequality on $\int_\Gamma n_\Gamma\cdot\langle\nabla(u-\pi_h u)\rangle[u_h-\pi_h u]$ leads to

$$
\begin{aligned}
&\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma n_\Gamma\cdot\langle\nabla(u-\pi_h u)\rangle[u_h-\pi_h u]\mathrm{d}S \\
&\quad\le\left(\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma\sigma[u_h-\pi_h u]^2\mathrm{d}S\right)^{\frac{1}{2}}\left(\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma\sigma^{-1}\left(n_\Gamma\cdot\langle\nabla(u-\pi_h u)\rangle\right)^2\mathrm{d}S\right)^{\frac{1}{2}}.
\end{aligned}
$$

Using these inequalities in (4.49) and dividing it by $\left|u_h-\pi_h u\right|_h$ yields

$$
\left|u_h-\pi_h u\right|_h\le c\left(\left|u-\pi_h u\right|_h+\left(\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma\sigma^{-1}\left(n_\Gamma\cdot\langle\nabla(u-\pi_h u)\rangle\right)^2\mathrm{d}S\right)^{\frac{1}{2}}\right),
$$

which can be estimated by multiplicative trace inequalities (4.20) and (4.21) for $\frac{1}{q}+\frac{1}{q'}=1$, if $q\in\left(\frac{4}{3},2\right)$:

$$\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma \sigma^{-1}\left(n_\Gamma\cdot\langle\nabla(u-\pi_h u)\rangle\right)^2\mathrm{d}S \\ \le c\sum_{T\in\mathcal{T}_h}\left(h\left|u-\pi_h u\right|_{1,2,T}\left|u-\pi_h u\right|_{2,2,T}+\left|u-\pi_h u\right|_{1,2,T}^2\right),$$

$$\sum_{\Gamma\in\mathcal{F}_h^I}\int_\Gamma \sigma^{-1}\left(n_\Gamma\cdot\langle\nabla(u-\pi_h u)\rangle\right)^2\mathrm{d}S \\ \le c\sum_{T\in\mathcal{T}_h}\left(h\left|u-\pi_h u\right|_{1,q',T}\left|u-\pi_h u\right|_{2,q,T}+\left|u-\pi_h u\right|_{1,2,T}^2\right).$$

Triangle inequality $\left|u-u_h\right|_h\le\left|u-\pi_h u\right|_h+\left|u_h-\pi_h u\right|_h$ and the fact that

$$\left|u-\pi_h u\right|_h=\left|u-\pi_h u\right|_{1,2,\Omega},$$

lead to

$$\left|u-u_h\right|_h\le c\left(\left|u-\pi_h u\right|_{1,2,\Omega}+\left(R_h^2(u-\pi_h u,q)+\left|u-\pi_h u\right|_{1,2,\Omega}^2\right)^{1/2}\right).$$

Using $a^2+b^2\le(a+b)^2$, estimates of $\left|u-\pi_h u\right|_{1,2,\Omega}$ in Theorem 2.8, and (4.41) completes the proof. □

If we consider a problem, where the numerical quadrature is used only on the nonlinear boundary term, but the right-hand side integrals are evaluated exactly, we still obtain the same result with almost identical proof.

Theorem 4.12. *Let the weak solution $u\in W^{r+1,q}(\Omega)$, $q>\frac{4}{3}$ given by* (1.8) *be zero on $\partial\Omega$, let $u_{dh}\in S_h^r$ be the discontinuous Galerkin approximation given by*

$$A_{dh}(u_{dh},v_h)=L(v_h),\qquad \forall v_h\in S_h^r. \tag{4.50}$$

Let the system of triangluations $\mathcal{T}_h$, $h>0$ be shape regular in accordance with (2.3)*, let C_W satisfy* (4.22)-(4.24)*, let the quadrature formula* (3.4) *used in the evaluation of the form A_{dh} satisfy* (3.15)*. Then there exists a constant $c>0$ independent of u, h, such that*

$$\left|u-u_{dh}\right|_h\le\begin{cases} ch^{r+1-2/q}\left|u\right|_{r+1,q,\Omega}, & q\in\left(\frac{4}{3},2\right),\\ ch^r\left|u\right|_{r+1,q,\Omega}, & q\ge 2.\end{cases} \tag{4.51}$$

Proof. Let $\pi_h u$ be a piecewise continuous Lagrange interpolation of u. Inequality (4.25) gives us

$$\frac{1}{2}\left|u_{dh}-\pi_h u\right|_h^2\le a_h(u_{dh}-\pi_h u,u_{dh}-\pi_h u) \\ \le A_{dh}(u_{dh},u_{dh}-\pi_h u)-A_{dh}(\pi_h u,u_{dh}-\pi_h u).$$

Using the definitions of the solutions u and u_h and the fact that numerical quadrature is evaluated exactly for functions which are identically zero, we have

$$A_{dh}(u_{dh},u_{dh}-\pi_h u)=L(u_{dh}-\pi_h u)=A_h(u,u_{dh}-\pi_h u),$$

$$A_{dh}(\pi_h u, u_{dh} - \pi_h u) = A_h(\pi_h u, u_{dh} - \pi_h u).$$

Therefore,

$$\frac{1}{2} |u_{dh} - \pi_h u|_h^2 \leq A_h(u, u_{dh} - \pi_h u) - A_h(\pi_h u, u_{dh} - \pi_h u),$$

and the rest of this proof is identical to Theorem 4.11. □

We have proven theoretically that the results concerning the error estimates of FEM derived in Chapters 2 and 3 also hold for DG method under the same assumptions. The order of convergence depends on the degree of used polynomials, the regularity of the exact solution and it is again divided by $\alpha + 1$, if the weak solution is zero on the boundary. For the DG method, the approximate solution converges to the exact weak solution in the $|||\cdot|||$-norm instead of the H^1-norm. If the exact solution is zero on the boundary $\partial\Omega$, we have improved the order of convergence in the $|\cdot|_h$-seminorm instead of the H^1-seminorm.

5. Numerical experiments

In this chapter we present two numerical examples computed using the FEniCS software [1]. We explore the reduction of the order of convergence caused by the nonlinearity, how it affects different norms, and whether this changes, if the exact solution of problem (1.1)-(1.2) is zero on the whole boundary $\partial\Omega$. In both experiments we discretize the problem by the FEM and by the SIPG variant of the DG method. We use uniform triangular meshes with element diameters $h_l = \frac{h_0}{2^l}, l = 0, 1, \dots, 5$. The amount of degrees of freedom (DOF) is therefore expected to increase about four times with each refinement. Denoting the error of the discrete solution by $e_h = u - u_h$, we compute the experimental order of convergence (EOC) by

$$EOC = \frac{\log e_{h_{l-1}} - \log e_{h_l}}{\log h_{l-1} - \log h_l}, \qquad l = 1, 2, \dots, 5. \tag{5.1}$$

The discrete problems (2.7), (3.14), (4.12), (4.13) represent nonlinear systems for $\alpha > 0$. We solved this problem by a dampened Newton method with tolerance on the residual 10^{-9}.

5.1 Example 1 - solution is zero on the boundary

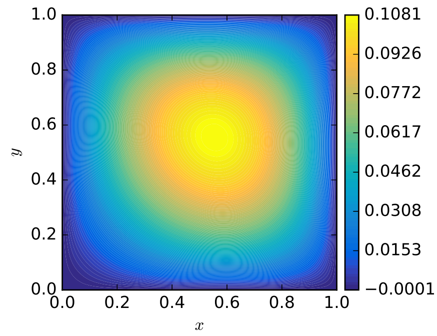
In the first experiment we consider the problem (1.1)-(1.2) on a unit square domain $\Omega = (0, 1)^2$. The data f and φ are chosen such that the exact solution is

$$u(x_1, x_2) = x_1(1 - x_1)x_2(1 - x_2)\left(x_1^2 + x_2^2\right)^{1/4}. \tag{5.2}$$

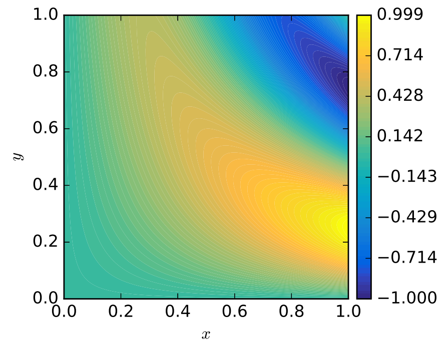
This function belongs to $W^{4,q}(\Omega)$, $q \in \left(1, \frac{4}{3}\right)$, or $H^{3.5-\delta}(\Omega)$, $\delta > 0$. Therefore, we expect $|e_h|_{1,2,\Omega} \approx O\left(h^{\min(2.5,r)}\right)$ and $\|e_h\|_{0,2,\Omega} \approx O\left(h^{\frac{\min(2.5,r)}{\alpha+1}}\right)$.

We have discretized the problem with FEM and SIPG variant of the DG method. For polynomials of degree $r = 2$ we have tried different values of the nonlinearity parameter $\alpha = 0.5, 1.0, 1.5, 2.0$, and for parameter $\alpha = 1.5$ we have tried FEM with polynomials of degrees $r = 1, 2, 3, 4$. The results shown in Table 5.1 and Table 5.2 also include the mesh element size $h = \max_{T\in\mathcal{T}_h} h_T$, the number of degrees of freedom and the number of Newton iterations.

The H^1-seminorm and $|\cdot|_h$-seminorm seem to behave as expected, i.e. their order of convergence is $\min(2.5, r)$. The most significant part of the error measured in H^1-norm (or $|||\cdot|||$-norm) was its L^2-norm. Our estimates for the L^2-norm give us an order of convergence $\frac{\min(2.5,r)}{\alpha+1}$, which would be $\frac{1}{\alpha+1}, \frac{2}{\alpha+1}, \frac{2.5}{\alpha+1}, \frac{2.5}{\alpha+1}$ for $r = 1, 2, 3, 4$, respectively. The EOC, however, suggests $\frac{2}{\alpha+1}, \frac{2.5}{\alpha+1}, \frac{2.5}{\alpha+1}, \frac{2.5}{\alpha+1}$ for $r = 1, 2, 3, 4$, respectively. The theoretical error estimate is therefore suboptimal.

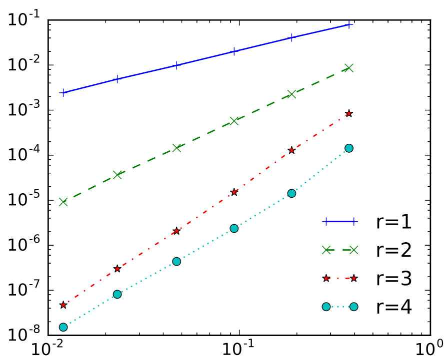


(a) Example 1 - function, which is zero on the whole boundary

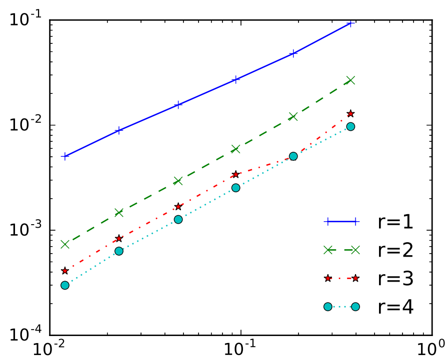


(b) Example 2 - smooth function, which is nonzero on a part of the boundary

Figure 5.1: The exact weak solutions of the discretized problems.

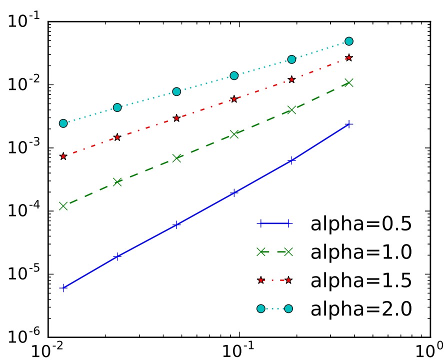


(a) error measured in $|\cdot|_{1,2,\Omega}$

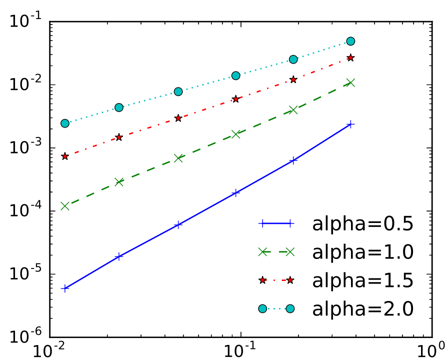


(b) error measured in $\|\cdot\|_{0,2,\Omega}$

Figure 5.2: Example 1 - EOC of FEM for $\alpha = 1.5$.



(a) error of FEM



(b) error of SIPG variant of DGM

Figure 5.3: Example 1 - EOC measured in $\|\cdot\|_{0,2,\Omega}$ for $r = 2$.

$\alpha = 1.5, \quad r = 1$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	49	4	9.3448e-02	–	7.9119e-02	–	1.2244e-01	–
0.188	161	6	4.8018e-02	0.96	4.0634e-02	0.96	6.2904e-02	0.96
0.094	577	6	2.7109e-02	0.82	2.0042e-02	1.02	3.3713e-02	0.90
0.047	2177	6	1.5600e-02	0.80	9.8458e-03	1.03	1.8447e-02	0.87
0.023	8449	6	8.8992e-03	0.81	4.8780e-03	1.01	1.0148e-02	0.86
0.012	33281	6	5.0395e-03	0.82	2.4321e-03	1.00	5.5957e-03	0.86
$\alpha = 1.5, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	161	3	2.6724e-02	–	8.6570e-03	–	2.8091e-02	–
0.188	577	6	1.2058e-02	1.15	2.2618e-03	1.94	1.2268e-02	1.20
0.094	2177	6	5.9243e-03	1.03	5.7373e-04	1.98	5.9520e-03	1.04
0.047	8449	6	2.9464e-03	1.01	1.4479e-04	1.99	2.9499e-03	1.01
0.023	33281	6	1.4700e-03	1.00	3.6421e-05	1.99	1.4704e-03	1.00
0.012	132097	6	7.3425e-04	1.00	9.1384e-06	1.99	7.3430e-04	1.00
$\alpha = 1.5, \quad r = 3$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	337	3	1.2840e-02	–	8.3916e-04	–	1.2867e-02	–
0.188	1249	6	4.9724e-03	1.37	1.2809e-04	2.71	4.9741e-03	1.37
0.094	4801	5	3.3908e-03	0.55	1.5021e-05	3.09	3.3908e-03	0.55
0.047	18817	6	1.6746e-03	1.02	2.0634e-06	2.86	1.6746e-03	1.02
0.023	74497	6	8.3301e-04	1.01	2.9962e-07	2.78	8.3301e-04	1.01
0.012	296449	3	4.1014e-04	1.02	4.7016e-08	2.67	4.1014e-04	1.02
$\alpha = 1.5, \quad r = 4$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	577	3	9.6870e-03	–	1.4266e-04	–	9.6880e-03	–
0.188	2177	6	5.0551e-03	0.94	1.4161e-05	3.33	5.0551e-03	0.94
0.094	8449	6	2.5318e-03	1.00	2.3612e-06	2.58	2.5318e-03	1.00
0.047	33281	6	1.2653e-03	1.00	4.3600e-07	2.44	1.2653e-03	1.00
0.023	132097	6	6.3245e-04	1.00	8.1398e-08	2.42	6.3245e-04	1.00
0.012	526337	4	2.9917e-04	1.08	1.5154e-08	2.43	2.9917e-04	1.08
$\alpha = 0.5, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	161	4	2.3779e-03	–	8.6544e-03	–	8.9752e-03	–
0.188	577	5	6.3232e-04	1.91	2.2617e-03	1.94	2.3485e-03	1.93
0.094	2177	4	1.9356e-04	1.71	5.7372e-04	1.98	6.0550e-04	1.96
0.047	8449	3	6.0476e-05	1.68	1.4479e-04	1.99	1.5691e-04	1.95
0.023	33281	3	1.8977e-05	1.67	3.6421e-05	1.99	4.1069e-05	1.93
0.012	132097	3	6.0396e-06	1.65	9.1384e-06	1.99	1.0954e-05	1.91
$\alpha = 1.0, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	161	4	1.0793e-02	–	8.6566e-03	–	1.3835e-02	–
0.188	577	6	3.9942e-03	1.43	2.2618e-03	1.94	4.5901e-03	1.59
0.094	2177	6	1.6433e-03	1.28	5.7373e-04	1.98	1.7406e-03	1.40
0.047	8449	5	6.8640e-04	1.26	1.4479e-04	1.99	7.0150e-04	1.31
0.023	33281	4	2.8784e-04	1.25	3.6421e-05	1.99	2.9014e-04	1.27
0.012	132097	3	1.1988e-04	1.26	9.1384e-06	1.99	1.2023e-04	1.27
$\alpha = 2.0, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_h$	EOC	$\|\|\|e\|\|\|$	EOC
0.375	161	3	4.8888e-02	–	8.6572e-03	–	4.9648e-02	–
0.188	577	6	2.5182e-02	0.96	2.2618e-03	1.94	2.5284e-02	0.97
0.094	2177	6	1.3928e-02	0.85	5.7373e-04	1.98	1.3940e-02	0.86
0.047	8449	6	7.7818e-03	0.84	1.4479e-04	1.99	7.7831e-03	0.84
0.023	33281	6	4.3594e-03	0.84	3.6421e-05	1.99	4.3595e-03	0.84
0.012	132097	6	2.4446e-03	0.83	9.1384e-06	1.99	2.4446e-03	0.83

Table 5.1: Example 1 - number of DOF and Newton iterations, discretization errors and convergence rates for $r = 1, 2, 3, 4$ and $\alpha = 0.5, 1.0, 1.5, 2.0$ in FEM.

h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_h$	EOC	$\|\|\|e\|\|\|$	EOC
$\alpha = 0.5, \quad r = 2$								
0.375	384	4	2.3711e-03	–	7.7517e-03	–	8.1062e-03	–
0.188	1536	5	6.3176e-04	1.91	2.0084e-03	1.95	2.1054e-03	1.94
0.094	6144	4	1.9354e-04	1.71	5.0545e-04	1.99	5.4124e-04	1.96
0.047	24576	3	6.0472e-05	1.68	1.2673e-04	2.00	1.4042e-04	1.95
0.023	98304	3	1.8994e-05	1.67	3.1764e-05	2.00	3.7009e-05	1.92
0.012	393216	3	5.9364e-06	1.68	7.9534e-06	2.00	9.9246e-06	1.90
$\alpha = 1.0, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_h$	EOC	$\|\|\|e\|\|\|$	EOC
0.375	384	4	1.0791e-02	–	7.7532e-03	–	1.3288e-02	–
0.188	1536	6	3.9941e-03	1.43	2.0084e-03	1.95	4.4706e-03	1.57
0.094	6144	6	1.6433e-03	1.28	5.0545e-04	1.99	1.7193e-03	1.38
0.047	24576	5	6.8640e-04	1.26	1.2673e-04	2.00	6.9800e-04	1.30
0.023	98304	4	2.8785e-04	1.25	3.1764e-05	2.00	2.8960e-04	1.27
0.012	393216	3	1.1989e-04	1.26	7.9534e-06	2.00	1.2015e-04	1.27
$\alpha = 1.5, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_h$	EOC	$\|\|\|e\|\|\|$	EOC
0.375	384	4	2.6723e-02	–	7.7536e-03	–	2.7825e-02	–
0.188	1536	6	1.2058e-02	1.15	2.0084e-03	1.95	1.2224e-02	1.19
0.094	6144	6	5.9243e-03	1.03	5.0545e-04	1.99	5.9459e-03	1.04
0.047	24576	6	2.9464e-03	1.01	1.2673e-04	2.00	2.9491e-03	1.01
0.023	98304	6	1.4700e-03	1.00	3.1764e-05	2.00	1.4703e-03	1.00
0.012	393216	6	7.3425e-04	1.00	7.9534e-06	2.00	7.3429e-04	1.00
$\alpha = 2.0, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_h$	EOC	$\|\|\|e\|\|\|$	EOC
0.375	384	3	4.8888e-02	–	7.7537e-03	–	4.9499e-02	–
0.188	1536	6	2.5182e-02	0.96	2.0084e-03	1.95	2.5262e-02	0.97
0.094	6144	6	1.3928e-02	0.85	5.0545e-04	1.99	1.3937e-02	0.86
0.047	24576	6	7.7818e-03	0.84	1.2673e-04	2.00	7.7828e-03	0.84
0.023	98304	6	4.3594e-03	0.84	3.1764e-05	2.00	4.3595e-03	0.84
0.012	393216	6	2.4446e-03	0.83	7.9534e-06	2.00	2.4446e-03	0.83

Table 5.2: Example 1 - number of DOF and Newton iterations, discretization errors and convergence rates for $r = 2$ and $\alpha = 0.5, 1.0, 1.5, 2.0$ in SIPG variant of DG method.

5.2 Example 2 - solution not identically zero on the boundary

In the second experiment, we again consider the problem (1.1)-(1.2) on a unit square domain $\Omega = (0,1)^2$. We prescribe the data f and φ in such a way that the exact solution is

$$u(x_1, x_2) = \frac{1}{4}(1 + x_1)^2 \sin(2\pi x_1 x_2). \tag{5.3}$$

This function was used in [19]. It is smooth, zero on boundary segments going through points $[0,1]$, $[0,0]$, $[1,0]$ and nonzero on segments going through points $[1,0]$, $[1,1]$, $[0,1]$. The expected order of convergence is r in all norms and seminorms considered and should not depend on the nonlinearity parameter α.

In a discretization of this problem, we have chosen $\alpha = 1.5$ and degrees of polynomials $r = 1,2,3$ for both FEM and SIPG variant of DG method. For FEM, we have also tried $r = 4$, and $\alpha = 0.5$. The order of convergence is not affected by boundary nonlinearity parameter α, which is in agreement with theoretical results. The H^1-seminorm and $|\cdot|_h$-seminorm converge with the predicted order of convergence r, but the L^2-norm converges faster with order $r+1$. The L^2-norm error estimate is again suboptimal, but in this case, the error is dominated by the H^1-seminorm or the $|\cdot|_h$-seminorm. Therefore the resulting order of convergence in H^1-norm or $|||\cdot|||$-norm is still r in accordance with the theoretical results.

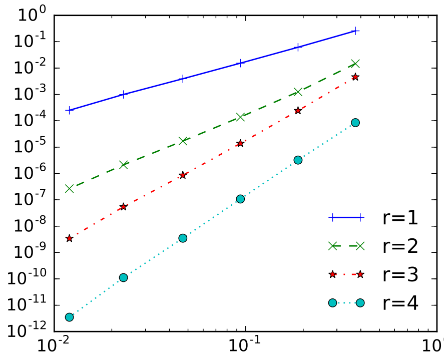
The numerical experiments confirmed that the theoretical error estimates in seminorms were optimal and that the order of convergence changes based on whether the exact solution is zero on the whole boundary. The numerical results, however, suggest that the order of convergence in L^2-norm is suboptimal. The theoretical results give us an order of convergence r (or $\frac{r}{\alpha+1}$), but the EOC is $r+1$ (or $\frac{r+1}{\alpha+1}$). This improvement only appeared when the exact solution belonged to the space $H^{r+1}(\Omega)$.

$\alpha = 1.5, \quad r = 1$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	49	6	2.5883e-01	–	9.5881e-01	–	9.9314e-01	–
0.188	161	5	6.1723e-02	2.07	5.3381e-01	0.84	5.3736e-01	0.89
0.094	577	4	1.5381e-02	2.00	2.8145e-01	0.92	2.8187e-01	0.93
0.047	2177	4	3.9289e-03	1.97	1.4421e-01	0.96	1.4426e-01	0.97
0.023	8449	3	9.9584e-04	1.98	7.2704e-02	0.99	7.2711e-02	0.99
0.012	33281	3	2.4986e-04	1.99	3.6390e-02	1.00	3.6391e-02	1.00
$\alpha = 1.5, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	161	6	1.4730e-02	–	2.3514e-01	–	2.3560e-01	–
0.188	577	4	1.2493e-03	3.56	5.8813e-02	2.00	5.8826e-02	2.00
0.094	2177	3	1.3819e-04	3.18	1.5173e-02	1.95	1.5173e-02	1.95
0.047	8449	3	1.6986e-05	3.02	3.8676e-03	1.97	3.8676e-03	1.97
0.023	33281	2	2.1254e-06	3.00	9.7489e-04	1.99	9.7489e-04	1.99
0.012	132097	2	2.6587e-07	3.00	2.4425e-04	2.00	2.4425e-04	2.00
$\alpha = 1.5, \quad r = 3$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	337	6	4.5914e-03	–	2.3116e-02	–	2.3568e-02	–
0.188	1249	3	2.4182e-04	4.25	3.4931e-03	2.73	3.5015e-03	2.75
0.094	4801	3	1.3800e-05	4.13	4.7873e-04	2.87	4.7893e-04	2.87
0.047	18817	2	8.5542e-07	4.01	6.2363e-05	2.94	6.2369e-05	2.94
0.023	74497	2	5.4140e-08	3.98	7.9229e-06	2.98	7.9231e-06	2.98
0.012	296449	2	3.4211e-09	3.98	9.9474e-07	2.99	9.9474e-07	2.99
$\alpha = 1.5, \quad r = 4$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	577	6	8.4789e-05	–	4.2824e-03	–	4.2832e-03	–
0.188	2177	3	3.2227e-06	4.72	3.2812e-04	3.71	3.2813e-04	3.71
0.094	8449	2	1.0740e-07	4.91	2.2035e-05	3.90	2.2036e-05	3.90
0.047	33281	2	3.4969e-09	4.94	1.4299e-06	3.95	1.4299e-06	3.95
0.023	132097	2	1.1140e-10	4.97	9.0809e-08	3.98	9.0809e-08	3.98
0.012	526337	2	3.5005e-12	4.99	5.6988e-09	3.99	5.6988e-09	3.99
$\alpha = 0.5, \quad r = 2$								
h	DOF	$iter$	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_{1,2,\Omega}$	EOC	$\|\|e\|\|_{1,2,\Omega}$	EOC
0.375	161	6	1.4072e-02	–	2.3527e-01	–	2.3569e-01	–
0.188	577	4	1.2379e-03	3.51	5.8815e-02	2.00	5.8828e-02	2.00
0.094	2177	4	1.3806e-04	3.16	1.5173e-02	1.95	1.5173e-02	1.95
0.047	8449	3	1.6989e-05	3.02	3.8676e-03	1.97	3.8676e-03	1.97
0.023	33281	3	2.1256e-06	3.00	9.7489e-04	1.99	9.7489e-04	1.99
0.012	132097	2	2.6588e-07	3.00	2.4425e-04	2.00	2.4425e-04	2.00

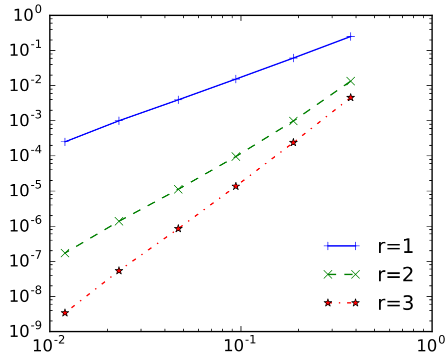
Table 5.3: Example 2 - number of DOF and Newton iterations, discretization errors and convergence rates for $r = 1, 2, 3, 4$ and $\alpha = 1.5, 0.5$ in FEM.

$\alpha = 1.5$, $r = 1$								
h	DOF	*iter*	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_h$	EOC	$\|\|\|e\|\|\|$	EOC
0.375	192	6	2.5073e-01	–	8.7620e-01	–	9.1137e-01	–
0.188	768	5	6.1030e-02	2.04	4.7862e-01	0.87	4.8249e-01	0.92
0.094	3072	4	1.5377e-02	1.99	2.4855e-01	0.95	2.4902e-01	0.95
0.047	12288	4	3.9457e-03	1.96	1.2692e-01	0.97	1.2698e-01	0.97
0.023	49152	3	1.0016e-03	1.98	6.3982e-02	0.99	6.3990e-02	0.99
0.012	196608	3	2.5142e-04	1.99	3.2043e-02	1.00	3.2044e-02	1.00
$\alpha = 1.5$, $r = 2$								
h	DOF	*iter*	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_h$	EOC	$\|\|\|e\|\|\|$	EOC
0.375	384	6	1.3432e-02	–	2.2029e-01	–	2.2069e-01	–
0.188	1536	4	9.8475e-04	3.77	5.4667e-02	2.01	5.4676e-02	2.01
0.094	6144	3	9.5957e-05	3.36	1.3884e-02	1.98	1.3884e-02	1.98
0.047	24576	3	1.1194e-05	3.10	3.5122e-03	1.98	3.5122e-03	1.98
0.023	98304	2	1.3773e-06	3.02	8.8228e-04	1.99	8.8229e-04	1.99
0.012	393216	2	1.7139e-07	3.01	2.2075e-04	2.00	2.2075e-04	2.00
$\alpha = 1.5$, $r = 3$								
h	DOF	*iter*	$\|\|e\|\|_{0,2,\Omega}$	EOC	$\|e\|_h$	EOC	$\|\|\|e\|\|\|$	EOC
0.375	640	6	4.5720e-03	–	2.7526e-02	–	2.7903e-02	–
0.188	2560	3	2.4012e-04	4.25	4.2359e-03	2.70	4.2427e-03	2.72
0.094	10240	3	1.3676e-05	4.13	5.7642e-04	2.88	5.7658e-04	2.88
0.047	40960	2	8.4847e-07	4.01	8.1035e-05	2.83	8.1039e-05	2.83
0.023	163840	2	5.3738e-08	3.98	1.0459e-05	2.95	1.0460e-05	2.95
0.012	655360	2	3.3983e-09	3.98	1.3431e-06	2.96	1.3431e-06	2.96

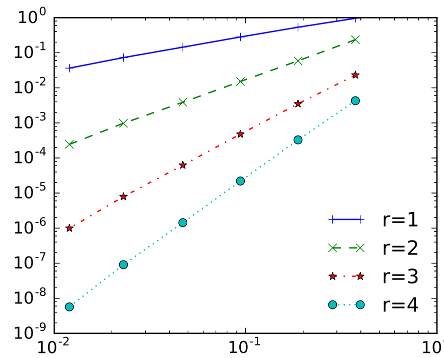
Table 5.4: Example 2 - number of DOF and Newton iterations, discretization errors and convergence rates for $\alpha = 1.5$ and $r = 1, 2, 3$ in SIPG variant of DG method.



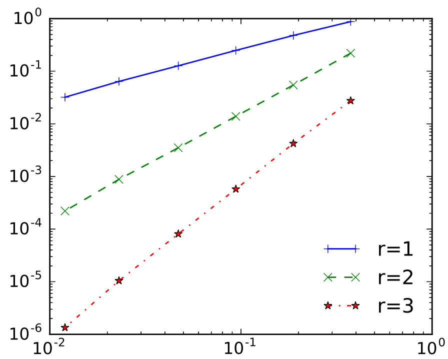
(a) error of FEM measured in $\|\cdot\|_{0,2,\Omega}$



(b) error of SIPG measured in $\|\cdot\|_{0,2,\Omega}$



(c) error of FEM measured in $|\cdot|_{1,2,\Omega}$



(d) error of SIPG measured in $|\cdot|_h$

Figure 5.4: Example 2 - EOC for $\alpha = 1.5$.

Conclusion

In this thesis we have analyzed finite element method (FEM) and discontinuous Galerkin method (DG) for elliptic problem in a polygonal domain with nonlinear Newton boundary condition. We have given some regularity properties of the weak solution in a neighbourhood of boundary edges. Then we have discretized the problem with FEM and derived abstract error estimates for the approximate solution. These estimates were further improved if the exact solution was distant from zero on a large part of the boundary to give the same estimates as if there was no nonlinearity on the boundary. If the exact solution was zero on the whole boundary, we have improved the seminorm part of the error estimates. We have then shown that the same results are obtained even if we consider that the integrals in the definition of the approximate solution are evaluated with numerical integration. Then we have discretized the problem with DG and we have derived the same error estimates even while considering numerical integration. The numerical experiments confirmed that order of convergence changes based on whether the exact solution is zero on the boundary and the seminorm part of the error estimates behaved exactly as predicted. It remains to prove an optimal $L^2(\Omega)$-error estimate.

Bibliography

[1] Alnæs M.M., Blechta J., Hake J., Johansson A., Kehlet B. Logg A., Richardson C., Ring J.,Rognes M.E., Wells G.N.: *The FEniCS Project Version 1.5.* Archive of Numerical Software, 2015.

[2] Bialecki R., Nowak A. J.: Boundary value problems in heat conduction with nonlinear material and nonlinear boundary conditions. *Appl. Math. Model.* 5 (1981), 417–421.

[3] Buffa A., Ortner C.: Compact embedding of broken Sobolev spaces and applications. *IMA J. Numer. Anal.* 29 (2009), 827–855.

[4] Ciarlet P. G.: *The Finite Element Method for Elliptic Problems.* North Holland, Amsterdam, 1978.

[5] Dolejší V., Feistauer M.: *Discontinuous Galerkin Method. Analysis and Applications to Compressible Flow.* Springer, Cham, 2015.

[6] Douglas J. Jr., Dupont T.: Galerkin methods for parabolic equations with nonlinear boundary conditions. Numer. Math. 20 (1973), 213–237.

[7] Evans L.C.: *Partial Differential Equations*, American Mathematical Society, 2, 2010

[8] Feistauer M., Kalis H., Rokyta M.: Mathematical modelling of an electrolysis process. *Comment. Math. Univ. Carolinae* 30 (1989), 465–477.

[9] Feistauer M., Najzar K.: Finite element approximation of a problem with a nonlinear Newton boundary condition. *Numer. Math.* 78 (1998), 403–425.

[10] Feistauer M., Najzar K., Sobotíková V.: Error Estimates for the Finite Element Solution of Elliptic Problems with Nonlinear Newton Boundary Conditions. *Numerical Functional Analysis and Optimization* 20 (1999), 835-851.

[11] Feistauer M., Roskovec F., Sändig A.-M.: Discontinuous Galerkin Method for an Elliptic Problem with Nonlinear Newton Boundary Conditions in a Polygon (2016)

[12] Fornberg B., *A practical guide to pseudospectral methods, Cambridge Monographs on Applied and Computational Mathematics*, 1. Cambridge University Press, Cambridge, 1996.

[13] Franců J.: Monotone operators. A survey directed to applications to differential equations. *Appl. Math.* 35 (1990), 257–301.

[14] Ganesh M., Graham I. G., Sivaloganathan J.: A pseudospectral three-dimensional boundary integral method applied to a nonlinear model problem from finite elasticity. *SIAM J. Numer. Anal.* 31 (1994), 1378–1414.

[15] Ganesh M., Steinbach O.: Boundary element methods for potential problems with nonlinear boundary conditions. *Mathematics of Computation* 70 (2000), 1031–1042.

[16] Ganesh M., Steinbach O.: Nonlinear boundary integral equations for harmonic problems. *Journal of Integral Equations and Applications* 11 (1999), 437–459.

[17] Grisvard P.: *Elliptic problems in nonsmooth domains*, Monographs and Studies in Mathematics 21, Pitman, Boston 1985.

[18] Grisvard P.: *Singularities in Boundary Value Problems.* Springer–Verlag, Berlin – Heidelberg – New York – London, 1992.

[19] Harriman K., Houston P., Senior B., Süli E.: *hp-Version Discontinuous Galerkin Methods with Interior Penalty for Partial Differential Equations with Nonnegative Characteristic Form*, Contemporary Mathematics Vol. 330, pp. 89-119, AMS, 2003.

[20] Hesthaven J. S., From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex, *SIAM J. Numer. Anal.* 35 (1998), 655–676.

[21] Křížek, M., Liu L., Neittaanmäki P.: Finite element analysis of a nonlinear elliptic problem with a pure radiation condition. In: Proc. Conf. devoted to the 70th birthday of Prof. J. Nečas, Lisbon, 1999.

[22] Lasis A., Süli E.: Poincaré-Type Inequalities for Broken Sobolev Spaces. Report no. 03/10, Oxford University Computing Laboratory, October 2003.

[23] Powell M. J. D., *Approximation theory and methods.* Cambridge University Press, Cambridge-New York, 1981.

[24] Liu L., Křížek, M.: Finite element analysis of a radiation heat transfer problem. *J. Comput. Math.* 16 (1998), 327–336.

[25] Maz'ya V., Rossmann J: *Elliptic Equations in Polyhedral Domains.* American Mathematical Society, Mathematical Surveys and Monographs, Vol.162, 2010.

[26] Moreau R., Ewans J. W.: An analysis of the hydrodynamics of alluminium reduction cells. *J. Electrochem. Soc.* 31 (1984), 2251–2259.

[27] Rack H.-J., Vajda R.: *Optimal cubic Lagrange interpolation: Extremal node systems with minimal Lebesgue constant.* Studia Universitatis Babeş-Bolyai Mathematica, 60 (2015): 151–171.

[28] Roubíček T.: A finite-element approximation of Stefan problems in heterogeneous media. Free boundary value problems. Proc. Conf., Oberwolfach/FRG 1989, Int. Ser. Numer. Math. 95 (1990), 267–275.

[29] Rudin W.: *Real and comples analysis*, McGraw-Hill, 1987

[30] Vainberg, M. M.: *Variational Methods for the Study of Nonlinear Operators.* Holden-Day, Inc., San Francisco, 1964.