

ANALYSIS OF THE FEM AND DGM FOR AN ELLIPTIC PROBLEM WITH A NONLINEAR NEWTON BOUNDARY CONDITION *

MILOSLAV FEISTAUER [†], ONDŘEJ BARTOŠ , FILIP ROSKOVEC , AND
ANNA-MARGARETE SÄNDIG[‡]

Abstract. The paper is concerned with the numerical analysis of an elliptic equation in a polygon with a nonlinear Newton boundary condition, discretized by the finite element or discontinuous Galerkin methods. Using the monotone operator theory, it is possible to prove the existence and uniqueness of the exact weak solution and the approximate solution. The main attention is paid to the study of error estimates. To this end, the regularity of the weak solution is investigated and it is shown that due to the boundary corner points, the solution loses regularity in a vicinity of these points. It comes out that the error estimation depends essentially on the opening angle of the corner points and on the parameter defining the nonlinear behaviour of the Newton boundary condition. Theoretical results are compared with numerical experiments confirming a nonstandard behaviour of error estimates.

Key words. elliptic equation, nonlinear Newton boundary condition, monotone operator method, finite element method, discontinuous Galerkin method, regularity and singular behaviour of the solution, error estimation

AMS subject classifications. 65N15, 65N30

1. Introduction. Let $\Omega \subset \mathbb{R}^2$ be a bounded polygonal domain with boundary $\partial\Omega$. We consider a boundary value problem with a non-linear Newton boundary condition: find $u : \bar{\Omega} \rightarrow \mathbb{R}$ such that

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

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

with given functions $f : \Omega \rightarrow \mathbb{R}$, $\varphi : \partial\Omega \rightarrow \mathbb{R}$ and constants $\kappa > 0$, $\alpha \geq 0$.

Such boundary value problems have applications in science and engineering. We can mention modelling of electrolysis of aluminium with the aid of the stream function ([11]), radiation heat transfer problem ([9], [10]) or nonlinear elasticity ([6], [7]). For example, by [2] our problem describes deformation of a flat plate with a nonlinear elastic support on the boundary.

In this paper we are concerned with the application of the finite element method (FEM) and the discontinuous Galerkin method (DGM) applied to the numerical solution of problem (1.1)-(1.2). Main attention is paid to a survey of error estimation. Detailed results are contained in the thesis [3] and the forthcoming paper [5].

2. Weak solution. In what follows we use the standard notation $L^p(\omega)$, $W^{k,p}(\omega)$, $H^k(\omega)$ for the Lebesgue and Sobolev spaces over a set ω . See, e.g., [12].

*This work was supported by Grant No. 17-01747S of the Czech Science Foundation.

[†]Charles University, Faculty of Mathematics and Physics, Sokolovská 83, 186 75 Praha 8, Czech Republic (feist@karlin.mff.cuni.cz, Ondra.Bartosh@seznam.cz, roskovec@gmail.com).

[‡]IANS, University Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany (Anna.Saendig@mathematik.uni-stuttgart.de)

Suppose that $f \in L^2(\Omega)$, $\varphi \in L^2(\partial\Omega)$. We introduce the following forms for $u, v \in H^1(\Omega)$:

$$\begin{aligned} b(u, v) &= \int_{\Omega} \nabla u \cdot \nabla v \, dx, & d(u, v) &= \kappa \int_{\partial\Omega} |u|^\alpha uv \, dS, & L^\Omega(v) &= \int_{\Omega} f v \, dx, \\ L^{\partial\Omega}(v) &= \int_{\partial\Omega} \varphi v \, dS, & L(v) &= L^\Omega(v) + L^{\partial\Omega}(v), & A(u, v) &= b(u, v) + d(u, v). \end{aligned} \quad (2.1)$$

DEFINITION 2.1. *We say that a function $u : \Omega \rightarrow \mathbb{R}$ is a weak solution of problem (1.1)-(1.2), if*

$$u \in H^1(\Omega), \quad A(u, v) = L(v) \quad \forall v \in H^1(\Omega). \quad (2.2)$$

Let us note that for $u, v \in H^1(\Omega)$

$$A(u, u-v) - A(v, u-v) = \int_{\Omega} |\nabla u - \nabla v|^2 \, dx + \kappa \int_{\partial\Omega} (|u|^\alpha u - |v|^\alpha v)(u-v) \, dS. \quad (2.3)$$

The next section will be devoted to the analysis of the numerical solution of problem (2.2) by the finite element method and the discontinuous Galerkin method. In the analysis of error estimation, the regularity of the weak solution plays an important role. In [5], the following result is proven.

THEOREM 2.2. *Let $u \in H^1(\Omega)$ be a weak solution of (2.2) in a polygonal domain Ω . By ω_0 we denote the largest interior angle at corners on the boundary. Let $f \in L^q(\Omega)$, $\varphi \in W^{1-1/q, q}(\partial\Omega)$, where*

$$\begin{aligned} q &= 1 + \frac{\pi}{2\omega_0 - \pi} - \varepsilon < 2 & \text{for } \omega_0 > \pi, \\ q &= 1 + \frac{\pi}{2\omega_0 - \pi} - \varepsilon > 2 & \text{for } \frac{\pi}{2} < \omega_0 < \pi, \\ q &\geq 1 \text{ is arbitrary} & \text{for } \omega_0 \leq \frac{\pi}{2}, \end{aligned} \quad (2.4)$$

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

It is obvious that $4/3 < q < \infty$.

3. Discretization. In what follows we are concerned with the discretization of problem (2.2) by the finite element method and the discontinuous Galerkin method. To this end, in Ω we construct a system of triangulations \mathcal{T}_h , $h \in (0, \bar{h})$, with $\bar{h} > 0$, consisting of a finite number of closed triangles T with standard properties, see [4]. If $T \in \mathcal{T}_h$, then by h_T and ρ_T we denote the diameter of T and the radius of the largest circle inscribed into T . We assume that this system of triangulations \mathcal{T}_h is shape regular:

$$\frac{h_T}{\rho_T} \leq C_R \quad \forall T \in \mathcal{T}_h \quad \forall h \in (0, \bar{h}). \quad (3.1)$$

The approximate solution is sought in the space

$$H_h^r = \{v_h \in C(\bar{\Omega}); v_h|_T \in P_r(T), T \in \mathcal{T}_h\}, \quad (3.2)$$

in the case of the FEM discretization and in

$$S_h^r = \{v_h \in L^2(\Omega); v_h|_T \in P_r(T), T \in \mathcal{T}_h\}, \quad (3.3)$$

in the case of the DGM. Here $r \geq 1$ is an integer and $P_r(T)$ denotes the space of piecewise polynomial functions on T of degree $\leq r$.

Because of the DGM discretization we denote the set of all faces of all elements $T \in \mathcal{T}_h$ by \mathcal{F}_h and we further distinguish between 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 $\mathcal{F}_h^I = \mathcal{F}_h \setminus \mathcal{F}_h^B$. For an integer $k \geq 1$, a number $q \geq 1$ and a triangulation \mathcal{T}_h we define the broken Sobolev space

$$W^{k,q}(\Omega, \mathcal{T}_h) = \{v \in L^2(\Omega); v|_T \in W^{k,q}(T), T \in \mathcal{T}_h\} \quad (3.4)$$

and put $H^k(\Omega, \mathcal{T}_h) = W^{k,2}(\Omega, \mathcal{T}_h)$. For functions $v \in W^{k,p}(\Omega, \mathcal{T}_h)$ and inner faces $\Gamma \in \mathcal{F}_h^I$, we introduce the notation

$$\begin{aligned} v|_{\Gamma}^{(L)} &= \text{trace of } v|_{T_{\Gamma}^{(L)}} \text{ on } \Gamma, & v|_{\Gamma}^{(R)} &= \text{trace of } v|_{T_{\Gamma}^{(R)}} \text{ on } \Gamma, \\ \langle v \rangle_{\Gamma} &= \frac{1}{2}(v|_{\Gamma}^{(L)} + v|_{\Gamma}^{(R)}), & [v]_{\Gamma} &= v|_{\Gamma}^{(L)} - v|_{\Gamma}^{(R)}. \end{aligned} \quad (3.5)$$

Here $T_{\Gamma}^{(L)}$ and $T_{\Gamma}^{(R)}$ are elements adjacent to Γ . By n_{Γ} we denote the outer unit normal vector to $T_{\Gamma}^{(L)}$ on Γ .

In the FEM we use the forms defined by (2.1). In the case of the DGM for $u, v \in H^2(\Omega, \mathcal{T}_h)$ we introduce their analogies. Namely, we set

$$b_h(u, v) = \sum_{T \in \mathcal{T}_h} \int_T \nabla u \cdot \nabla v \, dx - \sum_{\Gamma \in \mathcal{F}_h^I} \int_{\Gamma} (n_{\Gamma} \cdot \langle \nabla u \rangle [v] + \theta n_{\Gamma} \cdot \langle \nabla v \rangle [u]) \, dS. \quad (3.6)$$

The parameter θ can be chosen as 1, 0, -1, which leads to symmetric, incomplete and non-symmetric versions of the diffusion forms denoted by SIPG, IIPG, NIPG, respectively. Further, we introduce the interior penalty form

$$J_h(u, v) = \sum_{\Gamma \in \mathcal{F}_h^I} \frac{C_W}{h_{\Gamma}} \int_{\Gamma} [u][v] \, dS \quad (3.7)$$

with a parameter C_W . The form d is again defined by (2.1). Finally, we set

$$a_h(u, v) = b_h(u, v) + J_h(u, v), \quad (3.8)$$

$$A_h(u, v) = a_h(u, v) + d(u, v). \quad (3.9)$$

DEFINITION 3.1. *We say that a function u_h is a FEM approximate solution of problem (2.2), if*

$$u_h \in H_h^r, \quad A(u_h, v_h) = L(v_h) \quad \forall v_h \in H_h^r. \quad (3.10)$$

The function U_h is a DGM approximate solution, if

$$U_h \in S_h^r, \quad A_h(U_h, v_h) = L(v_h) \quad \forall v_h \in S_h^r. \quad (3.11)$$

The error of the FEM will be estimated in the standard norm $\|\cdot\|_{1,2,\Omega}$ and seminorm $|\cdot|_{1,2,\Omega}$ of the Sobolev space $H^1(\Omega)$. For the analysis of the DGM we introduce the seminorm

$$|v|_h = \left(\sum_{T \in \mathcal{T}_h} \int_T |\nabla v|^2 \, dx + J_h(v, v) \right)^{\frac{1}{2}}, \quad (3.12)$$

and the norm

$$\|v\| = \left(|v|_h^2 + \|v\|_{0,2,\Omega}^2 \right)^{\frac{1}{2}}. \quad (3.13)$$

By $\|\cdot\|_{0,2,\Omega}$ we denote the norm in $L^2(\Omega)$.

4. Properties of the forms A and A_h . In what follows, by the symbols C_0, C_1, C_2, \dots , we denote constants independent of the exact and approximate solutions and of h . Proofs of the following results are rather technical. We refer to [5].

LEMMA 4.1. *There exists a constant $C_0 > 0$ independent of $u, v \in H^1(\Omega)$, $u_h, v_h \in S_h^r$ and $h \in (0, \bar{h})$ such that*

$$A(u, u-v) - A(v, u-v) \geq |u-v|_{1,2,\Omega}^2 + C_0 \|u-v\|_{0,\alpha+2,\partial\Omega}^{\alpha+2} \quad \forall u, v \in H^1(\Omega). \quad (4.1)$$

Moreover, if the constant C_W from the definition of the penalty form J_h satisfies the conditions

$$C_W > 0, \text{ for } \theta = -1 \text{ (NIPG)}, \quad (4.2)$$

$$C_W > 4C_M(1 + C_I), \text{ for } \theta = 1 \text{ (SIPG)}, \quad (4.3)$$

$$C_W > C_M(1 + C_I), \text{ for } \theta = 0 \text{ (IIPG)}, \quad (4.4)$$

then $\forall u_h, v_h \in S_h^r, \forall h \in (0, \bar{h})$

$$A_h(u_h, u_h - v_h) - A_h(v_h, u_h - v_h) \geq \frac{1}{2} |u_h - v_h|_h^2 + C_0 \|u - v\|_{0,\alpha+2,\partial\Omega}^{\alpha+2}. \quad (4.5)$$

Similarly as in [5] we can prove the monotonicity and continuity of the forms A and A_h .

THEOREM 4.2. *The following results hold:*

a) *The forms A and A_h are uniformly monotone. Namely, we have*

$$A(u, u-v) - A(v, u-v) \geq \varrho(\|u-v\|_{1,2,\Omega}) \quad \forall u, v \in H^1(\Omega), \quad (4.6)$$

where

$$\varrho(t) = \begin{cases} C_1 t^{\alpha+2} & \text{for } 0 \leq t \leq 1, \\ C_1 t^2 & \text{for } t \geq 1, \end{cases} \quad (4.7)$$

with the constant C_1 depending on C_0, κ and α . If C_W satisfies (4.2)-(4.4), then

$$A_h(u_h, u_h - v_h) - A_h(v_h, u_h - v_h) \geq \varrho(\|u_h - v_h\|) \quad \forall u_h, v_h \in S_h^r \quad \forall h \in (0, \bar{h}), \quad (4.8)$$

where the function ϱ is again defined by (4.7).

b) *The forms A and A_h are continuous: There exists a constant $C_2 > 0$ such that $\forall u, v, w \in H^1(\Omega)$*

$$|A(u, v) - A(w, v)| \leq C_2 (1 + \|u\|_{1,2,\Omega}^\alpha + \|w\|_{1,2,\Omega}^\alpha) \|u-w\|_{1,2,\Omega} \|v\|_{1,2,\Omega}. \quad (4.9)$$

Further, if C_W satisfies (4.2)-(4.4), then

$$\begin{aligned} |A_h(u, w) - A_h(v, w)| &\leq C_2 \left\{ \|u-v\| + R_h(u-v, q) \right. \\ &\quad \left. + G_h(u-v) \left(\|u\|_{1,2,\Omega}^\alpha + \|v\|^\alpha \right) \right\} \|w\|, \end{aligned} \quad (4.10)$$

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

$$R_h(\phi, q) = \left(C_M \sum_{T \in \mathcal{T}_h} h_T |\phi|_{1,q',T} |\phi|_{2,q,T} \right)^{1/2}, \quad (4.11)$$

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

$$R_h(\phi, q) = \left(C_M \sum_{T \in \mathcal{T}_h} h_T |\phi|_{1,2,T} |\phi|_{2,2,T} \right)^{1/2}, \quad (4.12)$$

for $\phi \in W^{2,q}(\Omega, \mathcal{T}_h)$, $q \geq 2$. If $s \geq 3$, $q > 1$ and $u \in W^{s,q}(\Omega)$, then R_h is defined by (4.12). Moreover,

$$G_h(\phi) = \left(C_M \sum_{T \in \mathcal{T}_h} \left(\|\phi\|_{0,2,T}^2 h_T^{-1} + |\phi|_{1,2,T} \|\phi\|_{0,2,T} \right) \right)^{1/2}. \quad (4.13)$$

5. Error estimates. The basis for the error estimation is an abstract error estimate. Using the results formulated in Theorem 4.2, using approach from [3] and [5], it is possible to prove the following result:

THEOREM 5.1. *Let $u \in H^1(\Omega)$ be a weak solution of (2.2). There exists a constant $C_3 > 0$ such that if $u_h \in H_h^r$ is the FEM approximate solution defined by (3.10), then*

$$\|u - u_h\|_{1,2,\Omega} \leq \varrho_1^{-1} (C_3 \|u - v_h\|_{1,2,\Omega}) \quad \forall v_h \in H_h^r \quad \forall h \in (0, \bar{h}), \quad (5.1)$$

where

$$\varrho_1(t) = \varrho(t)/t, \quad (5.2)$$

and ϱ_1^{-1} is its inverse. In the case of the DGM we have

$$\begin{aligned} \| \|u - U_h\| \| \leq \varrho_1^{-1} (C_3 (\| \|u - v_h\| \| + R_h(u - v_h; q) + G_h(u - v_h) (\| \|u\|_{1,2,\Omega}^\alpha + \| \|v_h\| \|^\alpha))) \\ + \| \|u - v_h\| \|, \quad \forall v_h \in S_h^r, \quad \forall h \in (0, \bar{h}), \end{aligned} \quad (5.3)$$

where U_h is the approximate solution satisfying (3.11). The function $\varrho_1(t)$ is again defined by (5.2).

Now we can derive error estimates in terms of the size h of triangulations \mathcal{T}_h . To this end, it is necessary to introduce suitable H_h^r - and S_h^r -interpolations. Here we apply the Lagrangian interpolation denoted by π_h defined elementwise (cf. e.g. [4]). From the interpolation theory in [4] we get the following result:

LEMMA 5.2. *Let us assume that $s, m \geq 0$ be integers and $p, q \geq 1$, the piecewise Lagrange interpolation π_h preserve polynomials of degree at most r , the triangulation \mathcal{T}_h be shape regular according to (3.1) and the following embeddings hold:*

$$W^{\mu,q}(T) \hookrightarrow C(T), \quad W^{\mu,q}(T) \hookrightarrow W^{m,p}(T),$$

where $\mu = \min(r + 1, s)$. Then there exists a constant $C_4 = C_4(\pi, C_R) > 0$ such that for all $T \in \mathcal{T}_h$ and $h \in (0, \bar{h})$ we have

$$|u - \pi_h u|_{m,p,T} \leq C_4 |u|_{\mu,q,T} h_T^{\mu - m + \frac{2}{p} - \frac{2}{q}} \quad \forall u \in W^{s,q}(T). \quad (5.4)$$

The application of Theorem 5.1 and Lemma 5.2 combined with Jensen's inequality (Theorem 3.3 in [13]) yields the sought error estimates.

THEOREM 5.3. *Let the solution of (2.2) be $u \in W^{s,q}(\Omega)$, $\mu = \min(r+1, s)$ and $W^{\mu,q}(\Omega) \hookrightarrow H^1(\Omega)$. Then for the FEM approximate solution u_h defined by (3.10) the error estimate*

$$\|u - u_h\|_{1,2,\Omega} \leq \begin{cases} \varrho_1^{-1} \left(C_5 |u|_{\mu,q,\Omega} h^{\mu - \frac{2}{q}} \right), & q \in [1, 2), \\ \varrho_1^{-1} \left(C_5 |u|_{\mu,q,\Omega} h^{\mu-1} \right), & q \in [2, \infty). \end{cases} \quad (5.5)$$

holds for all $h \in (0, \bar{h})$.

In the case of the DGM we obtain the following results. (See [5].)

THEOREM 5.4. *Let $u \in W^{s,q}(\Omega)$, where $q > \frac{4}{3}$ for $s = 2$ and $q > 1$ for $s \geq 3$ be the weak solution given by (2.2), let U_h be the discontinuous Galerkin approximation of degree r given by (3.11) and let C_W satisfy (4.2)-(4.4). Let us set $\mu = \min(r+1, s)$. Then*

$$\| \|u - U_h\| \| \leq \rho_1^{-1} \left(C_6 (\|u\|_{1,2,\Omega}) h^{\mu-2/q} |u|_{\mu,q,\Omega} \right) + C_7 h^{\mu-2/q} |u|_{\mu,q,\Omega}, \quad h \in (0, \bar{h}), \quad (5.6)$$

for $q \in (1, 2)$. If $q \geq 2$, then

$$\| \|u - U_h\| \| \leq \rho_1^{-1} \left(C_6 (\|u\|_{1,2,\Omega}) h^{\mu-1} |u|_{\mu,q,\Omega} \right) + C_7 h^{\mu-1} |u|_{\mu,q,\Omega}, \quad h \in (0, \bar{h}). \quad (5.7)$$

6. Numerical experiments. In order to verify the obtained theoretical results, some numerical experiments are presented. They were realized with the aid of the FEniCS software [1]. We explore the reduction of the order of convergence caused by the nonlinearity and find out how it affects different norms. In both experiments we discretize the problem by the FEM and by the SIPG variant of the DGM. 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

$$EOC = \frac{\log \|e_{h_{l-1}}\| - \log \|e_{h_l}\|}{\log h_{l-1} - \log h_l}, \quad l = 1, 2, \dots, 5. \quad (6.1)$$

We evaluate the experimental order of convergence separately for the H^1 -seminorm and L^2 -norm for the FEM, and $|\cdot|_h$ -seminorm and L^2 -norm for the SIPG variant of DG method. The discrete problems (3.10) and (3.11) represent nonlinear systems for $\alpha > 0$. They are solved by a dampened Newton method with tolerance on the residual 10^{-9} .

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

$$u(x_1, x_2) = x_1(1-x_1)x_2(1-x_2) (x_1^2 + x_2^2)^{1/4}. \quad (6.2)$$

This function belongs to $W^{4,q}(\Omega)$, $q \in (1, \frac{4}{3})$. As $W^{4,q}(\Omega) \hookrightarrow H^3(\Omega)$ and $4 - 2/\frac{4}{3} = 2.5$, it follows from Theorems 5.3 and 5.4 that the EOC should be in the norms $\|\cdot\|_{1,2,\Omega}$ and $\| \cdot \|$ (at least) $\frac{\min(2.5, r)}{\alpha+1}$.

Table 6.1: 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.

$\alpha = 1.5, r = 1$								
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
$\alpha = 1.5, r = 2$								
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
$\alpha = 1.5, r = 3$								
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
$\alpha = 1.5, r = 4$								
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
$\alpha = 0.5, r = 2$								
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
$\alpha = 1.0, r = 2$								
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
$\alpha = 2.0, r = 2$								
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

We discretized the problem with FEM and SIPG variant of the DG method. For polynomials of degree $r = 2$ we tested different values of the nonlinearity parameter $\alpha = 0.5, 1.0, 1.5, 2.0$, and for parameter $\alpha = 1.5$ we tested FEM with polynomials of degrees $r = 1, 2, 3, 4$. The results shown in Table 6.1 and Table 6.2 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 EOC in H^1 -seminorm and $|\cdot|_h$ -seminorm are $\min(2.5, r)$, i.e. the error seems to be unaffected by the nonlinearity. 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 for $r = 1, 2$.

6.2. Example 2 - solution not identically zero on the boundary. In the second experiment, we again consider the problem (1.1)-(1.2) on the unit square

Table 6.2: 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.

$\alpha = 0.5, r = 2$								
h	DOF	iter	$\ e\ _{0,2,\Omega}$	EOC	$ e _h$	EOC	$\ e\ $	EOC
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
$\alpha = 1.0, r = 2$								
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
$\alpha = 1.5, r = 2$								
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
$\alpha = 2.0, r = 2$								
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

domain $\Omega = (0,1)^2$. We prescribe the data f and φ in such a way that the exact solution is $u(x_1, x_2) = \frac{1}{4}(1+x_1)^2 \sin(2\pi x_1 x_2)$. This function was used in [8]. It is smooth, zero on boundary segments going through the points $[0, 1]$, $[0, 0]$, $[1, 0]$ and nonzero on segments going through the points $[1, 0]$, $[1, 1]$, $[0, 1]$.

In this example we choose $\alpha = 1.5$ and polynomial degrees $r = 1, 2, 3$ for both the FEM and the SIPG variant of the DGM. For the FEM, we have also tried $r = 4$, and $\alpha = 0.5$. The EOC is not affected by the boundary nonlinearity parameter α . The H^1 -seminorm and $|\cdot|_h$ -seminorm converge with the order of convergence r , and the L^2 -norm converges faster with order $r + 1$. The error estimates in Theorems 5.3 and 5.4 are again suboptimal, but in this case, the error is dominated by the H^1 -seminorm or the $|\cdot|_h$ -seminorm.

7. Additional estimates. On the basis of the numerical experiments we come to the conclusion that the error estimates can be influenced by the behaviour of the exact solution on the boundary $\partial\Omega$, namely, if the exact solution u vanishes on the whole boundary and, on the other hand, if it is nonzero on a sufficiently large subset of the boundary. We present here some theoretical results derived for the FEM.

THEOREM 7.1. *Let the weak solution $u \in W^{s,q}(\Omega)$ given by (2.2) be zero on $\partial\Omega$. Let us set $\mu = \min(r + 1, s)$, where r is the degree of used polynomials. Then*

$$|u - u_h|_{1,2,\Omega} \leq \begin{cases} C_8 |u|_{k+1,q,\Omega} h^{\mu - \frac{2}{q}}, & q \in [1, 2), \\ C_8 |u|_{k+1,q,\Omega} h^{\mu - 1}, & q \in [2, \infty). \end{cases} \quad (7.1)$$

Proof. Neglecting the last term on the right-hand side of (4.1) gives us $|u - u_h|_{1,2,\Omega}^2 \leq A(u, u - u_h) - A(u_h, u - u_h)$, using Galerkin orthogonality following from (2.2), (3.10) and $H_h^r \subset H^1(\Omega)$ for a piecewise Lagrange interpolation yields $A(u, u - u_h) - A(u_h, u - u_h) = A(u, u - \pi_h u) - A(u_h, u - \pi_h u)$. The fact that $\pi_h u$ is also zero

Table 6.3: 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 _{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
$\alpha = 1.5, r = 2$								
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
$\alpha = 1.5, r = 3$								
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
$\alpha = 1.5, r = 4$								
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
$\alpha = 0.5, r = 2$								
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 6.4: 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.

$\alpha = 1.5, r = 1$								
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

on $\partial\Omega$ and the Hölder inequality implies that $A(u, u - \pi_h u) - A(u_h, u - \pi_h u) = \int_{\Omega} \nabla(u - u_h) \cdot \nabla(u - \pi_h u) dx \leq |u - u_h|_{1,2,\Omega} |u - \pi_h u|_{1,2,\Omega}$. Dividing by $|u - u_h|_{1,2,\Omega}$ leads to the estimate $|u - u_h|_{1,2,\Omega} \leq |u - \pi_h u|_{1,2,\Omega}$. Now Theorem 5.2 for $H^1(T)$ -seminorm gives us the sought estimate. \square

Further, we can improve estimates in Theorem 5.3 in such a way that $\rho_1(t) = C_8 t$

for all $t \geq 0$ in the case that the exact solution satisfies the following condition:

$$G \subset \partial\Omega, \quad |G| > 0, \quad |u| \geq \varepsilon > 0 \quad \text{on } G. \quad (7.2)$$

Then the improved error estimate is a consequence of the strong monotonicity of the form A :

THEOREM 7.2. *Let $u \in H^1(\Omega)$ and let the conditions (7.2) hold. Then there exists a constant $C_9 = C_9(\Omega, G, \varepsilon) > 0$ such that*

$$A(u, u - v) - A(v, u - v) \geq C_9 \|u - v\|_{1,2,\Omega}^2 \quad \forall v \in H^1(\Omega). \quad (7.3)$$

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

$$\begin{aligned} 2(|u|^\alpha u - |v|^\alpha v)(u - v) &= |u|^\alpha(2u^2 - 2uv) + |v|^\alpha(2v^2 - 2uv) \\ &\geq |u|^\alpha(u^2 - 2uv + v^2) + |v|^\alpha(v^2 - 2uv + u^2) = (|u|^\alpha + |v|^\alpha)(u - v)^2. \end{aligned} \quad (7.4)$$

Now (7.4) and (2.3) imply that $A(u, u - v) - A(v, u - v) \geq \|u - v\|_{1,2,\Omega}^2 + \frac{1}{2}\kappa\varepsilon^\alpha \|u - v\|_{0,2,G}^2$. The existence of a constant C_9 from the statement of this theorem follows from Poincaré's inequality $\|u\|_{1,2,\Omega} \leq c_P(\|u\|_{1,2,\Omega} + \|u\|_{0,2,G})$. \square

For the DGM we get analogical results with the norm $\|\cdot\|$ replacing $\|\cdot\|_{1,2,\Omega}$ and the seminorm $|\cdot|_h$ replacing $|\cdot|_{1,2,\Omega}$. An interesting problem is the analysis of the FEM or DGM combined with the use of numerical integration.

REFERENCES

- [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] Babuška I.: Private communication, Austin 2017.
- [3] O. Bartoš: Discontinuous Galerkin method for the solution of boundary-value problems in non-smooth domains. Master Thesis, Faculty of Mathematics and Physics, Charles University, Praha 2017.
- [4] Ciarlet P.G.: *The Finite Element Method for Elliptic Problems*. North Holland, Amsterdam, 1978.
- [5] Feistauer M., Roskovec F., Sändig A.-M.: Discontinuous Galerkin method for an elliptic problem with nonlinear Newton boundary conditions in a polygon. *IMA J. Numer. Anal.* (to appear).
- [6] Ganesh M., Steinbach O.: Boundary element methods for potential problems with nonlinear boundary conditions. *Mathematics of Computation* 70 (2000), 1031–1042.
- [7] Ganesh M., Steinbach O.: Nonlinear boundary integral equations for harmonic problems. *Journal of Integral Equations and Applications* 11 (1999), 437–459.
- [8] 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.
- [9] Krížek, 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.
- [10] Liu L., Krížek, M.: Finite element analysis of a radiation heat transfer problem. *J. Comput. Math.* 16 (1998), 327–336.
- [11] Moreau R., Ewans J. W.: An analysis of the hydrodynamics of aluminium reduction cells. *J. Electrochem. Soc.* 31 (1984), 2251–2259.
- [12] Pick, L., Kufner, A., John, O., Fučík, S.: *Function Spaces*. De Gruyter Series in Nonlinear Analysis and Applications 14, Berlin, 2013.
- [13] Rudin W.: *Real and complex analysis*, McGraw-Hill, 1987