DISCONTINUOUS GALERKIN FINITE ELEMENT APPROXIMATION OF QUASILINEAR ELLIPTIC BOUNDARY VALUE PROBLEMS I: THE SCALAR CASE

PAUL HOUSTON*, JANICE ROBSON†, AND ENDE RÜLS

Abstract. We develop a one-parameter family of \( h^p \)-version discontinuous Galerkin finite element methods, parameterised by \( \theta \in [-1, 1] \), for the numerical solution of quasilinear elliptic

\[
\begin{aligned}
-\nabla \cdot (\mu(x,|\nabla u|)\nabla u) &= f \quad \text{in } \Omega, \quad (1.1) \\
u &= g_D \quad \text{on } \Gamma_D, \quad (1.2) \\
\mu(x,|\nabla u|)\frac{\partial u}{\partial n} &= g_N \quad \text{on } \Gamma_N, \quad (1.3)
\end{aligned}
\]

where \( f \in L^2(\Omega) \), \( g_D \in H^{1/2}(\Gamma_D) \) and \( g_N \in L^2(\Gamma_N) \).

1. Introduction. Let \( \Omega \) be a bounded open set in \( \mathbb{R}^d \), \( d \geq 2 \), with Lipschitz continuous boundary \( \Gamma = \Gamma_D \cup \Gamma_N \), where \( \Gamma_D \) has positive \((d-1)\)-dimensional surface measure, and denote by \( \nu = (\nu_1, \ldots, \nu_d)^T \) the unit outward normal vector to \( \Gamma \), defined almost everywhere on \( \Gamma \). We consider the following elliptic boundary value problem:

where \( f \in L^2(\Omega) \), \( g_D \in H^{1/2}(\Gamma_D) \) and \( g_N \in L^2(\Gamma_N) \).

In the second part of this work, see [21], we shall focus on the following non-Newtonian flow problem: given \( f \in L^2(\Omega)^d \), \( g_D \in H^{1/2}(\Gamma_D)^d \) and \( g_N \in L^2(\Gamma_N)^d \), find \((u,p)\) such that

\[
\begin{aligned}
-\nabla \cdot (\mu(x,e(u))e(u)) + \nabla p &= f \quad \text{in } \Omega, \quad (1.4) \\
\text{div } u &= 0 \quad \text{in } \Omega, \quad (1.5) \\
u &= g_D \quad \text{on } \Gamma_D, \quad (1.6) \\
\{ \mu(x,e(u))e(u) - pI \} \cdot n &= g_N \quad \text{on } \Gamma_N, \quad (1.7)
\end{aligned}
\]

where \( u = (u_1, \ldots, u_d)^T \) is the velocity, \( p \) is the pressure, \( f = (f_1, \ldots, f_d)^T \) is the applied force, \( I \) is the \( d \times d \) identity matrix, \( e(u) \) is the symmetric \( d \times d \) strain tensor.

* Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK, email: Paul.Houston@le.ac.uk. Supported by the EPSRC (Grant GR/R76615).
†Computing Laboratory, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK, email: Janice.Robson@cs.ox.ac.uk. Computing Laboratory, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK, email: Ende.Rüls@cs.ox.ac.uk. Janice Robson and Endre Rüls acknowledge the support of the EPSRC, the Smith Institute, and Unilever UK.

1
whose entries are

\[ e_{ij}(u) = \frac{1}{2} \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \quad i, j = 1, \ldots, d, \]

and \( \| e(u) \| \) is the Frobenius norm of \( e(u) \) defined by

\[ \| e(u) \|^2 = e(u) : e(u) = \sum_{i,j=1}^{d} [e_{ij}(u)]^2. \]

In particular, if \( X = \text{diag}(x_1, \ldots, x_d) \) is a diagonal matrix, then the Frobenius norm \( \| X \| = (X^T X)^{1/2} \) of \( X \) is equal to the Euclidean norm \( (x^T x)^{1/2} \) of the vector \( x = (x_1, \ldots, x_d)^T \) consisting of the diagonal entries of \( X \).

When \( \Gamma_N \) is empty, the pressure \( p \) in problem (1.4)–(1.7) is only determined up to a constant; in that case, we supplement the problem with the condition

\[ \int_{\Omega} p \, dx = 0. \tag{1.8} \]

We shall assume throughout that the function \( \mu \) satisfies the following assumption.

\((\text{A})\) \( \mu \in C(\bar{\Omega} \times [0, \infty)) \) and there exist positive constants \( m_\mu \) and \( M_\mu \) such that

\[ m_\mu (t - s) \leq \mu(x, t) t - \mu(x, s) s \leq M_\mu (t - s), \quad t \geq s \geq 0, \quad x \in \bar{\Omega}. \tag{1.9} \]

When \( \mu \) satisfies (1.9), it follows from [4], Lemma 2.1 (for the case of \( d = 2 \), the case of \( d \geq 2 \) being analogous), that there exist positive constants \( C_1 \) and \( C_2 \), \( C_1 \geq C_2 \), such that for all \( d \times d \) real symmetric matrices \( Y \) and \( Z \), and all \( x \in \bar{\Omega} \),

\[ [\mu(x, Y) - \mu(x, Z)] \leq C_1 |Y - Z|, \tag{1.10} \]

\[ C_2 |Y - Z|^2 \leq (\mu(x, Y) - \mu(x, Z)) : (Y - Z). \tag{1.11} \]

By choosing \( Y = \text{diag}(y_1, \ldots, y_d) \) and \( Z = \text{diag}(z_1, \ldots, z_d) \), in particular, we deduce that (1.10) and (1.11) also hold when \( Y \) and \( Z \) are elements of \( \mathbb{R}^d \) where then \( | \cdot | \) signifies the Euclidean norm on \( \mathbb{R}^d \).

For the sake of notational simplicity we shall suppress the dependence of \( \mu \) on \( x \) and write \( \mu(t) \) instead of \( \mu(x, t) \). In fact, in many physical applications \( \mu \) is independent of \( x \). For example, the Carreau law

\[ \mu(t) = \mu_\infty + (\mu_0 - \mu_\infty)(1 + \lambda^2 t^2)^{-\frac{\tau}{2}}, \]

where \( \lambda > 0, 0 < r \leq 2 \) and \( 0 < \mu_\infty < \mu_0 \) satisfies (1.9) with \( m_\mu = \mu_\infty \) and \( M_\mu = \mu_0 \).

In recent years there has been considerable interest in discontinuous Galerkin finite element methods for the numerical solution of a wide range of partial differential equations. We shall not attempt to give an extensive survey of this area of research; the reader is referred to [11] for a detailed review. Discontinuous Galerkin Finite Element Methods (DGFEEMs) were introduced in the early 1970s for the numerical solution of first-order hyperbolic problems (see [30, 26, 23, 24, 12, 13, 14]). Simultaneously, but quite independently, they were proposed as nonstandard schemes for the approximation of second-order elliptic equations [27, 32, 1]. The recent upsurge of interest in this class of techniques has been stimulated by the computational convenience of DGFEEMs due to a high degree of locality, the need to approximate advection-dominated diffusion problems without excessive numerical stabilisation, the
necessity to accommodate high-order \( hp- \) and spectral element discretisations for first-order hyperbolic equations and advection-diffusion problems \([17, 25]\), and the desire to handle nonlinear hyperbolic problems in a locally conservative manner and without auxiliary numerical stabilisation \([10, 15]\); see also \([8, 9]\) for the error analysis of the local version of the DGFEM in the elliptic case, as well as \([2, 5]\) and \([28]\).

In the case of linear elliptic boundary value problems, two prominent techniques have emerged, referred to, respectively, as the symmetric and non-symmetric interior penalty DGFEM (cf. \([1, 29]\)). A common feature of the two methods is that the associated bilinear forms involve terms that penalise the jump \([u_h]_e\) in the numerical solution \(u_h\) over internal faces \(e\) in the subdivision of \(\Omega\). For example, in the case of Poisson’s equation \(\nabla \cdot \tau(u) = f\), with \(\tau(u) = \nabla u\), the bilinear form \(B_h(\cdot, \cdot)\) associated with the symmetric version of the interior penalty DGFEM includes the penalty term

\[
\sum_{e \in \mathcal{E}} \int_e \left( \sigma_e \left[ \frac{u_h}{\| \cdot \|_e} \right] - \left( \tau (u_h) \cdot \nu \right) \right) \, ds,
\]

where \(\nu\) denotes a unit normal vector assigned to \(e\) and \(\left( \tau (u_h) \cdot \nu \right)\) is the arithmetic average of the values of \(\tau (u_h) \cdot \nu\) on the two sides of the face \(e\). The bilinear form \(B_h(\cdot, \cdot)\) is then symmetric, and is also coercive if the positive penalty parameter \(\sigma_e\) is chosen to be sufficiently large, depending on the local mesh size and the local polynomial degree. In contrast, the bilinear form \(B_{NS}(\cdot, \cdot)\) corresponding to the non-symmetric version of the interior penalty DGFEM includes the penalty term

\[
\sum_{e \in \mathcal{E}} \int_e \left( \sigma_e \left[ \frac{u_h}{\| \cdot \|_e} \right] + \left( \tau (u_h) \cdot \nu \right) \right) \, ds,
\]

The plus sign in front of \(\left( \tau (u_h) \cdot \nu \right)\) ensures that \(B_{NS}(\cdot, \cdot)\) is coercive for any positive value of the penalty parameter \(\sigma_e\), although this desirable feature is achieved at the expense of rendering the bilinear form non-symmetric.

The penalty terms for the symmetric and the non-symmetric versions of the interior penalty DGFEMs are particular incarnations of the more general expression

\[
\sum_{e \in \mathcal{E}} \int_e \left( \sigma_e \left[ \frac{u_h}{\| \cdot \|_e} \right] + \theta \left( \tau (u_h) \cdot \nu \right) \right) \, ds,
\]

with \(\theta = -1\) corresponding to the symmetric and \(\theta = 1\) to the non-symmetric case; we remark that \(\theta = 0\) corresponds to the so-called incomplete interior penalty method studied by Sun and Wheeler, cf. \([31, 16]\). The relative merits of these methods have been widely discussed in the literature (see, for example, \([20]\) for a comparison in the context of duality-based \textit{a posteriori} error estimation).

The purpose of this paper and its companion article \([21]\) is to formulate and analyse the natural extensions to quasilinear elliptic PDEs of interior penalty \(hp-\)DGFEM. To the best of our knowledge, our paper is the first attempt in this direction. For the \textit{a priori} error analysis of the \(h\)-version local discontinuous Galerkin finite element approximation of \((1.1) - (1.3)\) and \((1.4) - (1.7)\), we refer to the articles of Bastina and Gatica \([7]\) and Gatica, González and Meddahi \([18]\), respectively.

The paper is structured as follows. In Section 2 we formulate the \(hp\)-version discontinuous Galerkin finite element approximation to \((1.1) - (1.3)\). By using a corollary of Brouwer’s Fixed Point Theorem, we show that the discrete problem has a unique solution \(u_{DG}\) in the finite element space. Section 3 discusses the error analysis of
the method in the broken $H^1(\Omega)$-norm. For sufficiently large values of the positive penalty parameter $\sigma_r$ involved in the definition of the method, depending on the local mesh size and the local polynomial degree, the semilinear form associated with the method is uniformly monotone; together with the Lipschitz continuity of the semi-linear form with respect to its first argument, this then leads to precisely the same $h$-optimal and mildly $p$-suboptimal rate of convergence in the broken $H^1(\Omega)$-norm as in the case of a linear elliptic PDE approximated by the interior penalty DG-FEM, cf. [22]. More precisely, if $u \in C^1(\Omega) \cap H^k(\Omega)$, $k \geq 2$, then, using discontinuous piecewise polynomials of degree $p \geq 1$, the error between $u$ and $u_{DG}$, measured in the broken $H^1(\Omega)$-norm, is $O(h^{k+1}/p^{k-3/2})$. A similar result will be proved in the companion-paper [21] for (1.4)–(1.7). Section 4 is devoted to numerical experiments. We close with a brief discussion of some open problems in Section 5.

2. Finite element spaces. Let us suppose for simplicity that $\Omega$ is a bounded open polyhedral domain in $\mathbb{R}^d$, and let $\mathcal{T}$ be a subdivision of $\Omega$ into disjoint open element domains $\kappa$ such that $\Omega = \bigcup_{\kappa \in \mathcal{T}} \kappa$, where $\mathcal{T}$ is regular or 1-irregular, i.e., each face of $\kappa$ in $\mathcal{T}$ has at most one hanging node, the barycenter of the face. We assume that the family of subdivisions $\mathcal{T}$ is shape-regular (see, for example, pp. 61, 113, and Remark 2.2, p. 114, in [6]) and each $\kappa \in \mathcal{T}$ is an affine image of a fixed master element $\kappa^*$, i.e., $\kappa = F_\kappa(\kappa^*)$ for all $\kappa \in \mathcal{T}$, where $\kappa^*$ is either the open unit simplex or the open unit hypercube in $\mathbb{R}^d$. For a nonnegative integer $k$, we denote by $\mathcal{P}_k(\kappa)$ the set of polynomials of total degree $k$ on $\kappa$. When $\kappa^*$ is the unit hypercube, we also consider $Q_k(\kappa)$, the set of all tensor-product polynomials of degree $k$ in each coordinate direction. To each $\kappa \in \mathcal{T}$ we assign a nonnegative integer $p_\kappa$ (local polynomial degree) and a nonnegative integer $s_\kappa$ (local Sobolev index), collect the $p_\kappa$, $s_\kappa$ and $F_\kappa$ in the vectors $\mathbf{p} = \{p_\kappa : \kappa \in \mathcal{T}\}$, $\mathbf{s} = \{s_\kappa : \kappa \in \mathcal{T}\}$ and $\mathbf{F} = \{F_\kappa : \kappa \in \mathcal{T}\}$, respectively, and consider the finite element space

$$SP(\Omega, \mathcal{T}, \mathbf{F}) = \{v \in L^2(\Omega) : v|_\kappa \circ F_\kappa \in \mathcal{P}_{p_\kappa}(\kappa) \quad \forall \kappa \in \mathcal{T}\},$$

where $\mathcal{R}$ is either $\mathcal{P}$ or $\mathcal{Q}$.

We shall suppose that the polynomial degree vector $\mathbf{p}$, with $p_\kappa \geq 1$ for each $\kappa \in \mathcal{T}$, has bounded local variation, i.e., there exists a constant $\rho \geq 1$ such that, for any pair of elements $\kappa$ and $\kappa'$ which share a $(d-1)$-dimensional face,

$$\rho^{-1} \leq \frac{p_\kappa}{p_{\kappa'}} \leq \rho.$$  

We assign to the subdivision $\mathcal{T}$ the broken Sobolev space of composite index $s$,

$$H^s(\Omega, \mathcal{T}) = \{v \in L^2(\Omega) : v|_\kappa \in H^s(\kappa) \quad \forall \kappa \in \mathcal{T}\},$$

equipped with the broken Sobolev norm and corresponding seminorm, respectively,

$$||v||_{s, \mathcal{T}} = \left( \sum_{\kappa \in \mathcal{T}} ||v||^2_{H^s(\kappa)} \right)^{\frac{1}{2}}, \quad |v|_{s, \mathcal{T}} = \left( \sum_{\kappa \in \mathcal{T}} |v|^2_{H^s(\kappa)} \right)^{\frac{1}{2}}.$$  

When $s_\kappa = s$ for all $\kappa \in \mathcal{T}$, we shall write $H^s(\Omega, \mathcal{T})$, $||v||_{s, \mathcal{T}}$ and $|v|_{s, \mathcal{T}}$.

Let us consider the set $\mathcal{E}$ of all open $(d-1)$-dimensional faces (open edges when $d = 2$ or open faces when $d = 3$) of all elements $\kappa \in \mathcal{T}$. Given that $\mathcal{T}$ may be irregular, since hanging nodes are permitted in the DG-FEM, $\mathcal{E}$ will be understood to contain the smallest common $(d-1)$-dimensional faces of neighbouring elements (cf. Figure 2.1). Further, we denote by $\mathcal{E}_m$ the set of all $e \in \mathcal{E}$ that are contained
in $\Omega$, we let $\Gamma_{\text{int}} = \{ x \in \Omega : x \in e \text{ for some } e \in \mathcal{E}_{\text{int}} \}$ and we introduce the set $\mathcal{E}_{\Gamma}$ of $(d-1)$-dimensional boundary faces contained in the subset $\Gamma_{\Gamma}$ of $\Gamma$. Implicit in these definitions is the assumption that $\mathcal{T}$ respects the decomposition of $\Gamma$ in the sense that each $e \in \mathcal{E}$ that lies on $\Gamma$ belongs to the interior of exactly one of $\Gamma_{\Gamma}$ and $\Gamma_{\text{ext}}$.

Suppose that $e$ is a $(d-1)$-dimensional face of an element $\kappa \in \mathcal{T}$; then, the following inverse inequalities hold: there exists a positive constant $C_3$, independent of the discretisation parameters, such that

$$
\|w\|_{L^2(e)}^2 \leq C_3 \frac{h_e^2}{\kappa} \|w\|_{L^2(e)}^2 \quad \text{and} \quad \|\nabla w\|_{L^2(e)}^2 \leq C_3 \frac{h_e^2}{\kappa} \|\nabla w\|_{L^2(e)}^2
$$

for all $w \in \mathcal{P}(\Omega, \mathcal{T}, \mathcal{F})$. Here, $h_e$ is the diameter of the face $e$. Due to our assumption that the subdivision $\mathcal{T}$ is shape-regular, if $e \subset \partial \kappa$ then $h_e$ in these inequalities can be replaced by $h_{\kappa}$, the diameter of $\kappa$, at the expense of altering the constant $C_3$.

Given that $e \in \mathcal{E}_{\text{int}}$, there exist indices $i$ and $j$ such that $i > j$ and $\kappa_i$ and $\kappa_j$ share the face $e$; we define the (element-numbering-dependent) jump of $v \in H^1(\Omega, \mathcal{T})$ across $e$ and the mean value of $v$ on $e$ by

$$
[v]_e = v|_{\partial \kappa_i \cap e} - v|_{\partial \kappa_j \cap e} \quad \text{and} \quad \langle v \rangle_e = \frac{1}{2} (v|_{\partial \kappa_i \cap e} + v|_{\partial \kappa_j \cap e}),
$$

respectively. If there is no danger of confusion, the subscript $e$ will be suppressed. Additionally, we associate with the face $e$ the unit normal vector $\nu$ which points from $\kappa_i$ to $\kappa_j$.

With these notations and $\theta \in [-1, 1]$, we introduce the semilinear form

$$
B(w, v) = \sum_{\kappa \in \mathcal{T}} \int_{\kappa} \mu(\nabla w) \nabla w : \nabla v \, dx
- \int_{\Gamma_{\Gamma}} \mu(\nabla w) \frac{\partial w}{\partial n} \, d\sigma - \int_{\Gamma_{\text{int}}} \mu(\nabla w) \frac{\partial w}{\partial n} \, [w] \, d\sigma
+ \theta \int_{\Gamma_{\Gamma}} \mu(\nabla w) \frac{\partial w}{\partial n} \, d\sigma + \theta \int_{\Gamma_{\text{int}}} \mu(\nabla w) \frac{\partial w}{\partial n} \, [w] \, d\sigma
+ \int_{\Gamma_{\Gamma}} \sigma w v \, ds + \int_{\Gamma_{\text{int}}} \sigma [w] \, [v] \, d\sigma,
$$

(2.2)

and the linear functional

$$
\ell(v) = \sum_{\kappa \in \mathcal{T}} \int_{\kappa} f v \, dx + \int_{\Gamma_{\Gamma}} \sigma g_\Gamma v \, ds + \int_{\Gamma_N} g_{\text{Neumann}} v \, ds.
$$

(2.3)
Here, \( h^{-1}|_e = h^{-1}_e \) for all \( e \subset \Gamma_D \cup \Gamma_{\text{int}} \). Let \( \kappa \in \mathcal{T} \) and let \( e \) be a \((d-1)\)-dimensional face of \( \partial \kappa \). The discontinuity penalisation parameter \( \sigma \), featuring in \( B(\cdot, \cdot) \) and \( \ell(\cdot) \) above, is defined by

\[
\sigma|_e = \sigma_e = \frac{(p^2)_e}{h_e} \quad \text{for} \quad e \subset \Gamma_D \cup \Gamma_{\text{int}},
\]

with the convention that if \( e \subset \Gamma_D \), and thereby \( e \subset \partial \kappa \cap \Gamma_D \) for some \( \kappa \in \mathcal{T} \), then \((p^2)_e = p^2_\kappa\). Here \( \alpha \) is a positive constant whose size will be fixed later on. We shall see that, at least for the purposes of the analysis pursued here, \( \alpha \geq (1 + \|f\|C^2T^{-1}C_3^{-1}C_3C_d) \) with \( C_d = 2^d d + 2d \) and \( \theta \in [-1, 1] \) will suffice.

The \( hp \)-DGFM approximation of problem (1.1)-(1.3) is: find \( u_{DG} \in \mathcal{S}P(\Omega, \mathcal{T}, \mathbf{F}) \) such that

\[
B(u_{DG}, v) = \ell(v) \quad \forall v \in \mathcal{S}P(\Omega, \mathcal{T}, \mathbf{F}).
\]

**Remark 2.1.** In (2.2), the role of the fourth and fifth integral (namely those that are multiplied by \( \theta \)) is, respectively, to weakly and approximately enforce the Dirichlet boundary condition \( u = g_D \) on \( \Gamma_D \) and the continuity condition \( [u] = 0 \) on \( \Gamma_{\text{int}} \) satisfied by the analytical solution \( u \). The choice of the factors \( \mu(h^{-1}|[w - g_D]|) \) and \( \mu(h^{-1}|[w]|) \) appearing in the corresponding integrands is, in principle, arbitrary. However, in the present context our choice has been guided by the following three requirements:

(a) when the problem is linear we would like our scheme to collapse to a standard \( hp \)-DGFM scheme (cf. [1], [2] or [29], for example);

(b) in our analysis, we wish to make use of the monotonicity condition (A). This, in turn dictates that the arguments of \( \mu \) in the two relevant terms should be multiples of \([w - g_D]\) and \([w]\), respectively;

(c) while any multiple of \([w - g_D]\) and \([w]\) would have been appropriate, for reasons of scaling we have chosen to use \( h^{-1}|w - g_D| \) and \( h^{-1}\|w\| \), so that the terms \( \mu(h^{-1}|w - g_D|) \) and \( \mu(h^{-1}\|w\|) \) resemble \( \mu(\|\nabla w\|) \).

Before embarking on the proof of the existence and uniqueness of solutions to (2.5), we shall make some preparatory observations.

Let us consider \( \mathcal{S}P(\Omega, \mathcal{T}, \mathbf{F}) \) equipped with the norm \( \|\cdot\|_{1,h} \) defined by

\[
\|v\|_{1,h} = \left( \sum_{e \in \mathcal{T}_h} \int_E |\nabla v|^2 \, dx + \int_{\Gamma_D} \sigma v^2 \, ds + \int_{\Gamma_{\text{int}}} \sigma |v|^2 \, ds \right)^{1/2},
\]

induced by the inner product \( \langle \cdot, \cdot \rangle_{1,h} \), where

\[
(w, v)_{1,h} = \sum_{e \in \mathcal{T}_h} \int_E \nabla w \cdot \nabla v \, dx + \int_{\Gamma_D} \sigma w v \, ds + \int_{\Gamma_{\text{int}}} \sigma |w| |v| \, ds.
\]

The next two lemmas stem, respectively, from (1.10) and (1.11).

**Lemma 2.2.** The semilinear form \( B(\cdot, \cdot) \) is Lipschitz-continuous in its first argument in the sense that

\[
|B(w_1, v) - B(w_2, v)| \leq C_4\|w_1 - w_2\|_{1,h}\|v\|_{1,h} \quad \forall w_1, w_2, v \in \mathcal{S}P(\Omega, \mathcal{T}, \mathbf{F}),
\]

where \( C_4 = \max\{C_1, 1\} + C_1(3C_3C\theta^{-1})^{1/2}, \theta \in [-1, 1] \) and \( \alpha > 0 \).
Proof. Using the fact that \( \frac{\partial v}{\partial n} = |\nabla v \cdot n| \leq |\nabla v| \), we have that

\[
\begin{align*}
|B(w_1, v) - B(w_2, v)| & \leq \sum_{\kappa \in T} \int_{\partial \kappa} |\mu| \left( |\nabla w_1| \nabla w_1 - \mu |\nabla w_2| \nabla w_2 \right) |\nabla v| ds \nonumber \\
& \quad + \int_{\Gamma_D} |\mu| |\nabla w_1| \nabla w_1 - \mu |\nabla w_2| \nabla w_2| |v| ds \nonumber \\
& \quad + \int_{\Gamma_{int}} \langle |\mu| |\nabla w_1| \nabla w_1 - \mu |\nabla w_2| \nabla w_2| \rangle |v| ds \nonumber \\
& \quad + \| \mu \|_{L^2(\Gamma_{int})} \left( |w_1 - w_2| |v| + \int_{\Gamma_{int}} |\mu| |w_1 - w_2| |v| ds \right) \nonumber \\
& \quad + \| \mu \|_{L^2(\Gamma_{int})} \left( |h^{-1}| |w_1| |v| + \int_{\Gamma_{int}} |h^{-1}| |w_1| |v| ds \right) \nonumber \\
& \quad \equiv T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_7. \tag{2.7}
\end{align*}
\]

Let \( w = w_1 - w_2 \). Applying (1.10) and the Cauchy–Schwarz inequality, it follows that

\[
T_1 \leq C_1 \left( \sum_{\kappa \in T} \int_{\partial \kappa} |\nabla w|^2 ds \right)^{1/2} \left( \sum_{\kappa \in T} \int_{\partial \kappa} |\nabla v|^2 ds \right)^{1/2}. \tag{2.8}
\]

For \( T_6 \) and \( T_7 \), we have

\[
T_6 \leq \left( \int_{\Gamma_D} |\mu| |w|^2 ds \right)^{1/2} \left( \int_{\Gamma_D} |\mu| |v|^2 ds \right)^{1/2}, \tag{2.9}
\]

\[
T_7 \leq \left( \int_{\Gamma_{int}} |\mu| |w|^2 ds \right)^{1/2} \left( \int_{\Gamma_{int}} |\mu| |v|^2 ds \right)^{1/2}. \tag{2.10}
\]

For \( T_2 \), (1.10) and the Cauchy–Schwarz inequality yield the bound

\[
T_2 \leq C_1 \int_{\Gamma_D} |\nabla w| |v| ds \leq C_1 \left( \int_{\Gamma_D} |\nabla w|^2 ds \right)^{1/2} \left( \int_{\Gamma_D} |\mu| |v|^2 ds \right)^{1/2}. \tag{2.11}
\]

Hence, using the second of the inverse inequalities (2.1) and recalling the definition of the penalty parameter \( \sigma \) on \( e \subset \Gamma_D \), we have that

\[
T_2 \leq C_1 (C_3 \sigma^{-1} 2d)^{1/2} \left( \sum_{\kappa \in T} \int_{\partial \kappa} |\nabla w|^2 ds \right)^{1/2} \left( \int_{\Gamma_D} |\mu| |v|^2 ds \right)^{1/2}, \tag{2.12}
\]

where, \( 2d \) denotes the maximum number of faces an element may possess which lie on the boundary of the computational domain \( \Omega \). Analogously,

\[
T_3 \leq C_1 \int_{\Gamma_{int}} \langle |\mu| |w| |v| \rangle ds \leq C_1 \left( \int_{\Gamma_{int}} \sigma^{-1} |\nabla w|^2 ds \right)^{1/2} \left( \int_{\Gamma_{int}} |\mu| |v|^2 ds \right)^{1/2}. \tag{2.13}
\]
Let us write
\[ \int_{\Gamma_{\text{int}}} \sigma^{-1} \|\nabla w\|^2 \, ds = \sum_{e \in \mathcal{E}_{\text{int}}} \sigma_e^{-1} \int_{e} \|\nabla w\|^2 \, ds, \]
and, for \( e \in \mathcal{E}_{\text{int}} \), let \( \kappa \) and \( \kappa' \) be the two elements that share \( e \). Then,
\[ \int_{e} \|\nabla w\|^2 \, ds \leq \frac{1}{2} \int_{\kappa} \|\nabla w\|_{\kappa}^2 \, ds + \frac{1}{2} \int_{\kappa'} \|\nabla w\|_{\kappa'}^2 \, ds \]
\[ \leq C_3 \frac{\rho_e^2}{2h_e} \int_{\kappa} \|\nabla w\|^2 \, dx + C_1 \frac{\rho_e^2}{2h_e} \int_{\kappa'} \|\nabla w\|^2 \, dx \]
\[ \leq C_3 \frac{\rho^2_e}{h_e} \max \left\{ \int_{\kappa} \|\nabla w\|^2 \, dx, \int_{\kappa'} \|\nabla w\|^2 \, dx \right\}. \]
On recalling from the definition of \( \sigma \) that
\[ \sigma_e = \frac{\rho^2_e}{h_e} \quad \text{for } e \in \mathcal{E}_{\text{int}}, \]
we have that
\[ \sum_{e \in \mathcal{E}_{\text{int}}} \sigma_e^{-1} \int_{e} \|\nabla w\|^2 \, ds \leq C_3 \alpha^{-1} \sum_{e \in \mathcal{E}_{\text{int}}} \max \left\{ \kappa : e \in \partial \kappa \right\} \int_{e} \|\nabla w\|^2 \, dx. \]
Thanks to our assumption that no face \( e \) of any element \( \kappa \in \mathcal{T} \) contains more than one hanging node, it follows that no element \( \kappa \) can have more than \( 2d \cdot 2^{d-1} = 2^d d \) faces if \( \kappa \) is the \( d \)-dimensional hypercube, or more than \( (d + 1)d \) faces if \( \kappa \) is the \( d \)-dimensional simplex. On writing \( c_d = \max\{2^d d, (d + 1)d\} = 2^d d \), we then have that
\[ \sum_{e \in \mathcal{E}_{\text{int}}} \sigma_e^{-1} \int_{e} \|\nabla w\|^2 \, ds \leq C_3 \alpha^{-1} c_d \sum_{e \in \mathcal{T}} \int_{e} \|\nabla w\|^2 \, dx, \]
and hence
\[ T_3 \leq C_1 (C_3 \alpha^{-1} c_d)^{1/2} \left( \sum_{\kappa \in \mathcal{T}} \int_{\kappa} \|\nabla w\|^2 \, dx \right)^{1/2} \left( \int_{\Gamma_{\text{int}}} \sigma\|\nabla w\|^2 \, ds \right)^{1/2}. \tag{2.13} \]
For \( T_4 \), we have, in exactly the same way as for \( T_2 \) (only, exchanging \( v \) and \( w \)),
\[ T_4 \leq \theta |C_1 (C_3 \alpha^{-1} 2d|^{1/2} \left( \sum_{\kappa \in \mathcal{T}} \int_{\kappa} \|\nabla v\|^2 \, dx \right)^{1/2} \left( \int_{\Gamma_{\text{int}}} \sigma\|\nabla v\|^2 \, ds \right)^{1/2}. \tag{2.14} \]
For \( T_5 \), in the same way as for \( T_3 \) (only, exchanging \( v \) and \( w \)),
\[ T_5 \leq \theta |C_1 (C_3 \alpha^{-1} c_d)^{1/2} \left( \sum_{\kappa \in \mathcal{T}} \int_{\kappa} \|\nabla v\|^2 \, dx \right)^{1/2} \left( \int_{\Gamma_{\text{int}}} \sigma\|\nabla v\|^2 \, ds \right)^{1/2}. \tag{2.15} \]
Substituting the bounds on \( T_1, \ldots, T_7 \) into (2.7), recalling that \( w = w_1 - w_2 \), and collecting the constants, we deduce that
\[ |B(w_1, v) - B(w_2, v)| \leq C_4 \|w_1 - w_2\|_{1,h} \|v\|_{1,h} \quad \forall w_1, w_2, v \in S^p(\Omega, \mathcal{T}, \mathbf{F}), \tag{2.16} \]
where $C_4 = \max\{C_1, 1\} + C_4 (C_3 C_d \alpha^{-1})^{1/2}$ and $C_d = c_d + 2d = 2d + 2d$. \qed

We note, in particular, that if $\alpha \geq (1 + |\theta|)^2 C_1 C_2^{-1} C_3 C_d$, as we shall assume from now on, then $C_4 \leq \max\{C_1, 1\} + C_2^{1/2}$. Hence we may set $C_4 = \max\{C_1, 1\} + C_2^{1/2}$.

**Lemma 2.3.** Suppose that $\theta \in [-1, 1]$ and $\alpha \geq (1 + |\theta|)^2 C_1 C_2^{-1} C_3 C_d$; then, the semilinear form $B(\cdot, \cdot)$ is uniformly monotone in the sense that

$$B(w_1, w_1 - w_2) - B(w_2, w_1 - w_2) \geq \frac{1}{2} \min\{C_2, 1\} \|w_1 - w_2\|_{1, h}^2 \quad \forall w_1, w_2 \in \mathcal{S} \Omega(T, F).$$

**(2.17)**

**Proof.** Using (1.10) and (1.11) and writing $w = w_1 - w_2$, we have that

$$B(w_1, w_1 - w_2) - B(w_2, w_1 - w_2)$$

$$\geq C_2 \sum_{k \in T_k} \int_{\Omega} |\nabla w|^2 dx - C_1 (1 + |\theta|) \left( \int_{\Gamma_T} \sigma^{-1} |\nabla w|^2 ds \right)^{1/2} \left( \int_{\Gamma_D} \sigma |w|^2 ds \right)^{1/2}$$

$$- C_1 (1 + |\theta|) \left( \int_{\partial \Omega} \sigma^{-1} |\nabla w|^2 ds \right)^{1/2} \left( \int_{\Gamma_D} \sigma |w|^2 ds \right)^{1/2}$$

$$+ \int_{\Gamma_D} \sigma |w|^2 ds + \int_{\partial \Omega} \sigma |w|^2 ds.$$

In exactly the same way as in the case of terms $T_2$ and $T_3$ in the proof of Lemma 2.2, we have

$$\left( \int_{\Gamma_D} \sigma^{-1} |\nabla w|^2 ds \right)^{1/2} \leq (C_3 \alpha^{-1} 2d)^{1/2} \left( \sum_{k \in T_k} \int_{\Omega} |\nabla w|^2 dx \right)^{1/2}$$

and

$$\left( \int_{\partial \Omega} \sigma^{-1} |\nabla w|^2 ds \right)^{1/2} \leq (C_3 \alpha^{-1} c_d)^{1/2} \left( \sum_{k \in T_k} \int_{\Omega} |\nabla w|^2 dx \right)^{1/2}.$$

Therefore,

$$B(w_1, w_1 - w_2) - B(w_2, w_1 - w_2) \geq C_2 \sum_{k \in T_k} \int_{\Omega} |\nabla w|^2 dx$$

$$- (1 + |\theta|)^2 C_1 C_2 \alpha^{-1} 2d)^{1/2} \left( \sum_{k \in T_k} \int_{\Omega} |\nabla w|^2 dx \right)^{1/2} \left( \int_{\Gamma_D} \sigma |w|^2 ds \right)^{1/2}$$

$$- (1 + |\theta|)^2 C_1 C_2 \alpha^{-1} c_d)^{1/2} \left( \sum_{k \in T_k} \int_{\Omega} |\nabla w|^2 dx \right)^{1/2} \left( \int_{\partial \Omega} \sigma |w|^2 ds \right)^{1/2}$$

$$+ \int_{\Gamma_D} \sigma |w|^2 ds + \int_{\partial \Omega} \sigma |w|^2 ds.$$

Applying Cauchy’s inequality $ab \leq \frac{1}{2} a^2 + \frac{1}{2} b^2$ to the second and third terms on the right-hand side and recalling that $C_d = c_d + 2d$, we have

$$B(w_1, w_1 - w_2) - B(w_2, w_1 - w_2) \geq C_2 \left( 1 - \frac{(1 + |\theta|)^2 C_1 C_2 C_d}{2 C_2 \alpha} \right) \sum_{k \in T_k} \int_{\Omega} |\nabla w|^2 dx$$

$$+ \frac{1}{2} \int_{\Gamma_D} \sigma |w|^2 ds + \frac{1}{2} \int_{\partial \Omega} \sigma |w|^2 ds.$$
Thus, on selecting $\alpha$ such that $\alpha \geq (1 + \theta)^2 C_2^2 C_3^{-1} C_4$, we deduce (2.17). $\square$

Now we are ready to show the existence of a unique solution to (2.5). We shall make use of the following corollary to Brouwer’s Fixed Point Theorem (see, [19], p. 105).

**Proposition 2.4.** Let $H$ be a finite-dimensional Hilbert space whose scalar product and norm are denoted, respectively, by $(\cdot, \cdot)$ and $|\cdot|$. Let $P$ be a continuous mapping from $H$ into $H$ with the following property: there exists $\xi > 0$ such that

$$
(P(w), w) > 0 \quad \forall w \in H \quad \text{with } |w| = \xi.
$$

Then, there exists an element $u$ in $H$ such that

$$
|u| \leq \xi, \quad P(u) = 0.
$$

**Theorem 2.5.** Suppose that $\theta \in [-1, 1]$ and $\alpha \geq (1 + \theta)^2 C_2^2 C_3^{-1} C_4$; then, there exists a unique element $u_{1,h}$ in $SP(\Omega, \mathcal{T}, \mathcal{F})$ such that (2.5) holds.

Proof. As a first step in our argument, we shall rewrite the numerical method (2.5) as a nonlinear operator equation $P(u) = 0$ on $H \equiv SP(\Omega, \mathcal{T}, \mathcal{F})$. We shall do so by exploiting the Riesz Representation Theorem from Hilbert space theory.

It is a straightforward matter to show that $SP(\Omega, \mathcal{T}, \mathcal{F})$ is a finite-dimensional Hilbert space with the norm $|| \cdot ||_{1,h}$ induced by the inner product $(\cdot, \cdot)_{1,h}$. Let us consider a second norm, $|| \cdot ||_{1,h}^*$, on $SP(\Omega, \mathcal{T}, \mathcal{F})$ defined by

$$
||v||_{1,h}^* = \left( \sum_{v \in \mathcal{T}} \int_{\Omega} |\nabla v|^2 + v^2 \, dx + \int_{\Gamma_{\text{int}}} \sigma v^2 \, ds + \int_{\Gamma_D} \sigma v^2 \, ds + \int_{\Gamma_N} v^2 \, ds \right)^{1/2}.
$$

Since $SP(\Omega, \mathcal{T}, \mathcal{F})$ has finite dimension, the norms $|| \cdot ||_{1,h}$ and $|| \cdot ||_{1,h}^*$ are equivalent on $SP(\Omega, \mathcal{T}, \mathcal{F})$: that is, there exists a positive constant $c_6$, dependent on the dimension $\delta$ of $SP(\Omega, \mathcal{T}, \mathcal{F})$, such that

$$
||v||_{1,h} \leq ||v||_{1,h}^* \leq c_6 ||v||_{1,h} \quad \forall v \in SP(\Omega, \mathcal{T}, \mathcal{F}).
$$

Given any $w$ in $SP(\Omega, \mathcal{T}, \mathcal{F})$, consider the linear functional

$$
\psi_w : v \in SP(\Omega, \mathcal{T}, \mathcal{F}) \mapsto \psi_w(v) = B(w, v) - \ell(v) \in \mathbb{R}.
$$

The fact that $SP(\Omega, \mathcal{T}, \mathcal{F})$ is finite-dimensional implies that $\psi_w$ is a bounded linear functional on $SP(\Omega, \mathcal{T}, \mathcal{F})$. The actual bound of $\psi_w(v)$ is easily established: by the Cauchy–Schwarz inequality,

$$
|\ell(v)| \leq \left( \int_{\Omega} f^2 \, dx \right)^{1/2} \left( \int_{\Omega} v^2 \, dx \right)^{1/2} + \left( \int_{\Gamma_D} \sigma g_D^2 \, ds \right)^{1/2} \left( \int_{\Gamma_D} \sigma v^2 \, ds \right)^{1/2}
$$

$$
+ \left( \int_{\Gamma_N} g_N^2 \, ds \right)^{1/2} \left( \int_{\Gamma_N} v^2 \, ds \right)^{1/2},
$$

and therefore

$$
|\ell(v)| \leq \left( \int_{\Omega} f^2 \, dx + \int_{\Gamma_D} \sigma g_D^2 \, ds + \int_{\Gamma_N} g_N^2 \, ds \right)^{1/2}
$$

$$
\times \left( \int_{\Omega} v^2 \, dx + \int_{\Gamma_D} \sigma v^2 \, ds + \int_{\Gamma_N} v^2 \, ds \right)^{1/2}.
$$
which yields that
\[ |\ell(v)| \leq \left( \int_{\Omega} f^2 \, dx + \int_{\Gamma_D} \sigma g_N^2 \, ds + \int_{\Gamma_N} g_N^2 \, ds \right)^{1/2} \|v\|_{1,h}^2. \]
By the norm-equivalence (2.20), we then have that
\[ |\ell(v)| \leq C_5 \|v\|_{1,h} \quad \forall v \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}), \tag{2.21} \]
where
\[ C_5 = c_5 \left( \int_{\Omega} f^2 \, dx + \int_{\Gamma_D} \sigma g_N^2 \, ds + \int_{\Gamma_N} g_N^2 \, ds \right)^{1/2}. \]
On the other hand, by (2.6) with \( w_1 = w \) and \( w_2 = 0 \), we have that
\[ |B(w, v)| \leq C_4 \|w\|_{1,h} \|v\|_{1,h} \quad \forall v \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}), \]
and therefore,
\[ |\psi_w(v)| \leq (C_4 \|w\|_{1,h} + C_5) \|v\|_{1,h} \quad \forall v \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}). \]
Since the linear functional \( \psi_w \) is bounded (and therefore continuous) on \( \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \), by virtue of the Riesz Representation Theorem, there exists \( P(w) \) in \( \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \) such that
\[ \psi_w(v) = (P(w), v)_{1,h} \quad \forall v \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}). \tag{2.22} \]
As \( w \) passes through \( \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \), (2.22) defines the mapping
\[ w \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \mapsto P(w) \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \]

of \( \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \) into itself.

Next, we show that the mapping \( w \mapsto P(w) \) is Lipschitz continuous (and, thereby, continuous) in the norm \( \|\cdot\|_{1,h} \), uniformly in the mesh size \( h = \max h_e \) and the polynomial degree vector \( \mathbf{p} \). Clearly,
\[ (P(w_1) - P(w_2), v)_{1,h} = B(w_1, v) - B(w_2, v). \]
Hence, by Lemma 2.2,
\[ |(P(w_1) - P(w_2), v)_{1,h}| = |B(w_1, v) - B(w_2, v)| \leq C_4 \|w_1 - w_2\|_{1,h} \|v\|_{1,h} \tag{2.23} \]
for all \( w_1, w_2, v \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \), and therefore,
\[ \|P(w_1) - P(w_2)\|_{1,h} \leq \sup_{v \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F})} \frac{|(P(w_1) - P(w_2), v)_{1,h}|}{\|v\|_{1,h}} \leq C_4 \|w_1 - w_2\|_{1,h} \]
for all \( w_1, w_2 \in \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \). Thus, \( P \) is a Lipschitz continuous mapping of (the Hilbert space) \( \mathcal{SP}(\Omega, \mathcal{T}, \mathbf{F}) \) into itself, with Lipschitz constant \( C_4 \) independent of the discretisation parameters.
In order to apply Proposition 2.4, it remains to show that there exists $\xi > 0$ such that $(P(w), w)_{1,h} > 0$ for all $w$ in $\text{SP}(\Omega, \mathcal{T}, \mathbf{F})$ with $\|w\|_{1,h} = \xi$. Clearly,

$$(P(w), w)_{1,h} = \psi_w(w) = B(w, w) - \ell(w).$$

Now, Lemma 2.3, with $w_1 = w$, $w_2 = 0$ and $\alpha \geq (1 + \|\|)^2 C_2^2 C_3^{-1} C_3 C_d$, implies that

$$B(w, w) \geq \frac{1}{2} \min \{ C_2, 1 \} \|w\|^2_{1,h}. \quad (2.24)$$

Combining (2.24) and (2.21) gives

$$(P(w), w) = B(w, w) - \ell(w) \geq \frac{1}{2} \min \{ C_2, 1 \} \|w\|^2_{1,h} - C_5 \|w\|_{1,h} \quad \forall w \in \text{SP}(\Omega, \mathcal{T}, \mathbf{F}).$$

In particular, if $\xi \in \mathbb{R}$ and $\xi > 2 C_5 \min \{ C_2, 1 \}$, then, for any $w \in \text{SP}(\Omega, \mathcal{T}, \mathbf{F})$ such that $\|w\|_{1,h} = \xi$, we have that $(P(w), w)_{1,h} > 0$. By applying Proposition 2.4 we deduce the existence of a solution $u_{DG}$ to (2.5) in $\text{SP}(\Omega, \mathcal{T}, \mathbf{F})$.

To prove the uniqueness of the solution to (2.5), suppose that $u_{DG}$ and $u_{DG}'$ are two solutions to (2.5) in $\text{SP}(\Omega, \mathcal{T}, \mathbf{F})$. Then,

$$B(u_{DG}, v) - B(u_{DG}', v) = 0 \quad \forall v \in \text{SP}(\Omega, \mathcal{T}, \mathbf{F}),$$

and thereby also

$$B(u_{DG} - u_{DG}', u_{DG}) = B(u_{DG} - u_{DG}', u_{DG}') = 0.$$

On applying (2.17) with $w_1 = u_{DG}$, $w_2 = u_{DG}'$ and $\alpha \geq (1 + \|\|)^2 C_2^2 C_3^{-1} C_3 C_d$, we get

$$B(u_{DG}, u_{DG} - u_{DG}') \geq \frac{1}{2} \min \{ C_2, 1 \} \|u_{DG} - u_{DG}'\|^2_{1,h}. \quad (2.25)$$

Since the left-hand side of this inequality is equal to 0 and $\| \cdot \|_{1,h}$ is a norm on $\text{SP}(\Omega, \mathcal{T}, \mathbf{F})$, we deduce that $u_{DG} - u_{DG}' = 0$, which establishes the uniqueness of the solution to (2.5). $\square$

3. Error analysis of the method. We recall the following approximation result for the finite element space $\text{SP}(\Omega, \mathcal{T}, \mathbf{F})$.

**Lemma 3.1.** Suppose that $\kappa \in \mathcal{T}$ is a $d$-simplex or $d$-parallelepiped of diameter $h_{\kappa}$. Suppose further that $u_{|\kappa} \in H^k(\kappa)$, $h_{\kappa} \geq 0$, for $\kappa \in \mathcal{T}$. Then, there exists a sequence $z_{p_{\kappa}}^{h_{\kappa}}$ in $K_{p_{\kappa}}(\kappa)$, $p_{\kappa} = 1, 2, \ldots$, such that for $0 \leq q \leq k_{\kappa}$,

$$\|u - z_{p_{\kappa}}^{h_{\kappa}}\|_{H^q(\kappa)} \leq C \frac{h_{\kappa}^{k_{\kappa} - q}}{p_{\kappa}^{k_{\kappa} - q}} \|u\|_{H^k(\kappa)},$$

where $1 \leq s_{\kappa} \leq \min \{ p_{\kappa} + 1, k_{\kappa} \}$, $p_{\kappa} \geq 1$, for $\kappa \in \mathcal{T}$, and $C$ is a positive constant, independent of $u$ and the discretisation parameters.

**Proof.** For the proof, see Lemma 4.5 in [3] for $d = 2$; when $d > 2$ the argument is completely analogous. $\square$

Given $u \in H^2(\Omega, \mathcal{T})$, we now define $\Pi_{p}^{h_{\kappa}} u \in \text{SP}(\Omega, \mathcal{T}, \mathbf{F})$ by

$$(\Pi_{p}^{h_{\kappa}} u)_{|\kappa} = z_{p_{\kappa}}^{h_{\kappa}}(u_{|\kappa}).$$
Then, assuming that \( u_{|\kappa} \in H^{k_\kappa}(\kappa), \ k_\kappa \geq 2, \) for \( \kappa \in \mathcal{T}, \) and writing

\[
\eta = u - \Pi^b_h u,
\]

by virtue of Lemma 3.1 we have that

\[
\|\eta\|_{L^2(\Omega)}^2 \leq C \frac{h^{2k_{\kappa} + 1}}{p_{\kappa}^2} \|u\|_{H^{k_{\kappa}}(\kappa)}^2 \quad \text{and} \quad \|\nabla \eta\|_{L^2(\Omega)}^2 \leq C \frac{h^{2k_{\kappa} - 2}}{p_{\kappa}^2} \|u\|_{H^{k_{\kappa}}(\kappa)}^2,
\]

where \( 1 \leq s_{\kappa} \leq \min\{p_{\kappa} + 1, k_\kappa\}, \ p_\kappa \geq 1, \) for \( \kappa \in \mathcal{T}, \) and \( C \) is a positive constant, independent of \( u \) and the discretisation parameters.

The multiplicative trace inequality asserts the existence of a positive constant \( C = C(d) \) such that

\[
\|\eta\|_{L^2(\partial \Omega)}^2 \leq C(d) \left( \|\eta\|_{L^2(\Omega)} \|\nabla \eta\|_{L^2(\Omega)} + h^{-1}_\kappa \|\eta\|_{H^{k_{\kappa}}(\kappa)}^2 \right).
\]

Hence,

\[
\|\eta\|_{L^2(\partial \Omega)}^2 \leq C \frac{h^{2k_{\kappa} - 1}}{p_{\kappa}^2} \|u\|_{H^{k_{\kappa}}(\kappa)}^2 \quad \text{and} \quad \|\nabla \eta\|_{L^2(\partial \Omega)}^2 \leq C \frac{h^{2k_{\kappa} - 3}}{p_{\kappa}^2} \|u\|_{H^{k_{\kappa}}(\kappa)}^2,
\]

where \( 1 \leq s_{\kappa} \leq \min\{p_{\kappa} + 1, k_\kappa\}, \ p_\kappa \geq 1, \) for \( \kappa \in \mathcal{T}, \) and \( C \) is a positive constant, independent of \( u \) and the discretisation parameters. Since \( \mathcal{T} \) is shape-regular and the polynomial degree vector \( \mathbf{p} \) has bounded local variation, it then follows, using (2.4), that

\[
\sum_{\kappa \in \mathcal{T}} \int_{\kappa} \|\nabla \eta\|^2 \, dx + \int_{\Gamma_D} \sigma^{-1} \|\nabla \eta\|^2 \, ds + \int_{\Gamma_D} \sigma^{-1} \|\nabla \eta\|^2 \, ds + \int_{\Gamma_{int}} \sigma \eta^2 \, ds + \int_{\Gamma_{int}} \sigma \|\eta\|^2 \, ds 
\leq C \sum_{\kappa \in \mathcal{T}} \frac{h^{2k_{\kappa} - 2}}{p_{\kappa}^2} \|u\|_{H^{k_{\kappa}}(\kappa)}^2,
\]

and therefore, also,

\[
|\eta|_{1,h} = \left( \sum_{\kappa \in \mathcal{T}} \int_{\kappa} |\nabla \eta|^2 \, dx + \int_{\Gamma_D} \sigma \eta^2 \, ds + \int_{\Gamma_{int}} \sigma \|\eta\|^2 \, ds \right)^{1/2} 
\leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{h^{2k_{\kappa} - 2}}{p_{\kappa}^2} \|u\|_{H^{k_{\kappa}}(\kappa)}^2 \right)^{1/2},
\]

where \( 1 \leq s_{\kappa} \leq \min\{p_{\kappa} + 1, k_\kappa\}, \ p_\kappa \geq 1, \) for \( \kappa \in \mathcal{T}, \) and \( C \) is a positive constant, independent of \( u \) and the discretisation parameters.

Our next result is concerned with the consistency of the semilinear form \( B(\cdot, \cdot). \)

**Lemma 3.2.** Suppose that \( u_{|\kappa} \in H^{k_\kappa}(\kappa), \ k_\kappa \geq 2, \) for \( \kappa \in \mathcal{T}. \) Assume, further, that \( \theta \in [-1, 1] \) and \( \alpha \geq (1 + \theta)^2 C^2 \gamma C_1 C_3 C_\theta; \) then, the semilinear form \( B(\cdot, \cdot), \) satisfies the following bound:

\[
|B(u, v) - B(\Pi^b_h u, v)| \leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{h^{2s_{\kappa} - 2}}{p_{\kappa}^2} \|u\|_{H^{s_{\kappa}}(\kappa)}^2 \right)^{1/2} \|v\|_{1,h} \quad \forall v \in \mathbf{S}(\Omega, \mathcal{T}, \mathbf{F}),
\]

(3.3)
where $1 \leq s_\kappa \leq \min\{p_\kappa + 1, k_\kappa\}$, $p_\kappa \geq 1$, for $\kappa \in \mathcal{T}$, and $C$ is a positive constant, independent of $u$, $v$, and the discretisation parameters.

Proof. We proceed in much the same way as in the proof of Lemma 2.2:

$$
|B(u, v) - B(\Pi_p^h u, v)| \leq \sum_{\kappa \in \mathcal{T}} \int_D \left| \mu |\nabla u| - \mu(|\nabla \Pi_p^h u|) |\nabla v| \right| dx
$$

$$
+ \int_D \left| \mu |\nabla u| - \mu(|\nabla \Pi_p^h u|) |\nabla v| \right| ds
$$

$$
+ \int_{\Gamma_D} \left| \mu |\nabla u| - \mu(|\nabla \Pi_p^h u|) |\nabla v| \right| ds
$$

$$
+ \left[ \int_{\Gamma_D} |\mu(h^{-1}|u - g_D|)(u - g_D) - \mu(h^{-1}|\Pi_p^h u - g_D|)(\Pi_p^h u - g_D)| |\nabla v| \right| ds
$$

$$
+ \left[ \int_{\Gamma_D} |\mu(h^{-1}|u|)|u| - \mu(h^{-1}|\Pi_p^h u|)|\Pi_p^h u| \right| |\nabla v| \right| ds
$$

$$
+ \int_{\Gamma_D} |\mu |\nabla u - \Pi_p^h u| |v| |ds + \int_{\Gamma_D} |\mu |\nabla v| |v| |ds
$$

$$
\equiv T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_7. \quad (3.4)
$$

Using (1.10), the Cauchy–Schwarz inequality and (3.2) with $\eta = u - \Pi_p^h u$, we get

$$
T_1 \leq C \left( \sum_{\kappa \in \mathcal{T}} \int_D |\nabla (u - \Pi_p^h u)|^2 dx \right)^{1/2} \left( \sum_{\kappa \in \mathcal{T}} \int_D |\nabla v|^2 dx \right)^{1/2}
$$

$$
\leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{\nu^2_{\kappa, s - 2}}{2k_\kappa - s} |u|_{H^{s, (\kappa)}}^2 \right)^{1/2} \left( \sum_{\kappa \in \mathcal{T}} \int_D |\nabla v|^2 dx \right)^{1/2} \quad \forall v \in \mathbf{P}^p(\Omega, \mathcal{T}, \mathbf{F}). \quad (3.5)
$$

where $1 \leq s_\kappa \leq \min\{p_\kappa + 1, k_\kappa\}$, $p_\kappa \geq 1$, for $\kappa \in \mathcal{T}$, and $C$ is a positive constant, independent of $u$, $v$, and the discretisation parameters.

For $T_2$, (1.10) and the Cauchy–Schwarz inequality yield the bound

$$
T_2 \leq C \left( \int_{\Gamma_D} |\nabla (u - \Pi_p^h u)| |v| ds \right)^{1/2} \left( \int_{\Gamma_D} |\nabla v|^2 ds \right)^{1/2}
$$

$$
\leq C \left( \int_{\Gamma_D} \sigma^{-1} |\nabla (u - \Pi_p^h u)|^2 ds \right)^{1/2} \left( \int_{\Gamma_D} \sigma |v|^2 ds \right)^{1/2}. \quad (3.6)
$$

Hence, using the bound on the second term on the left-hand side of (3.1), we have that

$$
T_2 \leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{\nu^2_{\kappa, s - 2}}{2k_\kappa - s} |u|_{H^{s, (\kappa)}}^2 \right)^{1/2} \left( \int_{\Gamma_D} \sigma |v|^2 ds \right)^{1/2} \quad \forall v \in \mathbf{P}^p(\Omega, \mathcal{T}, \mathbf{F}). \quad (3.7)
$$

where $1 \leq s_\kappa \leq \min\{p_\kappa + 1, k_\kappa\}$, $p_\kappa \geq 1$, for $\kappa \in \mathcal{T}$, and $C$ is a positive constant, independent of $u$, $v$ and the discretisation parameters.
 Analogously,

\[
T_3 \leq C_1 \int_{\Gamma_{\text{int}}} \left( \left\{ \nabla (u - \Pi_{p_{k_{e}}} u) \right\} \right) \|v\| \, ds
\leq C_1 \left( \int_{\Gamma_{\text{int}}} \sigma^{-1} \left\{ \nabla (u - \Pi_{p_{k_{e}}} u) \right\}^2 \, ds \right) \left( \int_{\Gamma_{\text{int}}} \sigma \|v\|^2 \, ds \right)^{1/2}
\leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{\lambda^{s_\kappa}}{p_{k_{e}}^2} \|u\|_{H^s(x)}^2 \right)^{1/2} \left( \int_{\Gamma_{\text{int}}} \sigma \|v\|^2 \, ds \right)^{1/2} \forall v \in \mathcal{S}^p(\Omega, \mathcal{T}, \mathbf{F}), \tag{3.8}
\]

where \(1 \leq s_\kappa \leq \min\{p_{k_{e}} + 1, k_{e}\}, p_{k_{e}} \geq 1\), for \(\kappa \in \mathcal{T}\), and \(C\) is a positive constant, independent of \(u, v\) and the discretisation parameters.

For \(T_4\), (1.10) and the Cauchy–Schwarz inequality yield the bound

\[
T_4 \leq \|h\| \left( \int_{\Gamma_{\text{D}}} |u - \Pi_{p_{k_{e}}} u| \|v\| \, ds \right)
\leq \|h\| \left( \int_{\Gamma_{\text{D}}} \sigma^{-1} \|u - \Pi_{p_{k_{e}}} u\|^2 \, ds \right)^{1/2} \left( \int_{\Gamma_{\text{D}}} \sigma \|v\|^2 \, ds \right)^{1/2}. \tag{3.9}
\]

Hence, using the second of the two inverse inequalities in (2.1) in the second factor on the right-hand side of (3.9), the bound on the fourth term on the left-hand side of (3.1) and recalling the definition of the penalty parameter \(\sigma_\kappa\) on \(e \subset \Gamma_{\text{D}}\), we have that

\[
T_4 \leq \|h\| \left( \sum_{\kappa \in \mathcal{T}} \frac{\lambda^{s_\kappa}}{p_{k_{e}}^2} \|u\|_{H^s(x)}^2 \right)^{1/2} \left( \sum_{\kappa \in \mathcal{T}} \int_{\kappa} |\nabla v|^2 \, ds \right)^{1/2} \forall v \in \mathcal{S}^p(\Omega, \mathcal{T}, \mathbf{F}), \tag{3.10}
\]

where \(1 \leq s_\kappa \leq \min\{p_{k_{e}} + 1, k_{e}\}, p_{k_{e}} \geq 1\), for \(\kappa \in \mathcal{T}\), and \(C\) is a positive constant, independent of \(u, v\) and the discretisation parameters. Analogously,

\[
T_5 \leq \|h\| \left( \sum_{\kappa \in \mathcal{T}} \frac{\lambda^{s_\kappa}}{p_{k_{e}}^2} \|u\|_{H^s(x)}^2 \right)^{1/2} \left( \sum_{\kappa \in \mathcal{T}} \int_{\kappa} |\nabla v|^2 \, ds \right)^{1/2} \forall v \in \mathcal{S}^p(\Omega, \mathcal{T}, \mathbf{F}), \tag{3.11}
\]

where \(1 \leq s_\kappa \leq \min\{p_{k_{e}} + 1, k_{e}\}, p_{k_{e}} \geq 1\), for \(\kappa \in \mathcal{T}\), and \(C\) is a positive constant, independent of \(u, v\) and the discretisation parameters.

Proceeding in the same manner, we obtain the following bounds on \(T_6\) and \(T_7\):

\[
T_6 \leq \left( \int_{\Gamma_{\text{D}}} \sigma \|u - \Pi_{p_{k_{e}}} u\|^2 \, ds \right)^{1/2} \left( \int_{\Gamma_{\text{D}}} \sigma v^2 \, ds \right)^{1/2}
\leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{\lambda^{s_\kappa}}{p_{k_{e}}^2} \|u\|_{H^s(x)}^2 \right)^{1/2} \left( \int_{\Gamma_{\text{D}}} \sigma \|v\|^2 \, ds \right)^{1/2} \forall v \in \mathcal{S}^p(\Omega, \mathcal{T}, \mathbf{F}), \tag{3.12}
\]

\[
T_7 \leq \left( \int_{\Gamma_{\text{int}}} \sigma \|u - \Pi_{p_{k_{e}}} u\|^2 \, ds \right)^{1/2} \left( \int_{\Gamma_{\text{int}}} \sigma \|v\|^2 \, ds \right)^{1/2}
\leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{\lambda^{s_\kappa}}{p_{k_{e}}^2} \|u\|_{H^s(x)}^2 \right)^{1/2} \left( \int_{\Gamma_{\text{int}}} \sigma \|v\|^2 \, ds \right)^{1/2} \forall v \in \mathcal{S}^p(\Omega, \mathcal{T}, \mathbf{F}), \tag{3.13}
\]
where \(1 \leq s_\kappa \leq \min\{p_\kappa + 1, k_\kappa\},\) \(p_\kappa \geq 1,\) for \(\kappa \in \mathcal{T},\) and \(C\) is a positive constant, independent of \(u, v\) and the discretisation parameters. Substituting the bounds on \(T_1, \ldots, T_7\) into (3.4) we arrive at (3.3). \(\square\)

Now we are ready to prove the main result of the paper.

**Theorem 3.3.** Let \(\Omega \subset \mathbb{R}^d\) be a bounded polyhedral domain, \(\mathcal{T} = \{\kappa\}\) a shape-
regular subdivision of \(\Omega\) into \(d\)-simplexes or \(d\)-parallelipipeds, and \(\mathbf{p}\) a polynomial degree vector of bounded local variation. Suppose, further, that \(\theta \in [-1, 1],\) \(\alpha \geq (1 + |\theta|)^2 C^0 C^{-1} C_3 C_4\) and assign to each face \(e \subset \mathcal{E}_m \cup \mathcal{E}_d\) the positive real number \(\sigma_e\) defined by (2.4) where \(h_e\) is the diameter of \(e.\) Then, assuming that \(u \in C^1(\Omega)\) and \(u|_e \in H^{k_e}(\kappa),\) \(k_e \geq 2,\) for \(\kappa \in \mathcal{T},\) the solution \(u_{\text{DG}} \in SP(\Omega, \mathcal{T}, \mathbf{F})\) of (2.5) satisfies the error bound

\[
\|u - u_{\text{DG}}\|_{1,h}^2 \leq C \sum_{\kappa \in \mathcal{T}} \frac{h^{2s_\kappa - 2}_{\kappa}}{p^{2s_\kappa - 2}_{\kappa}} \|u\|_{H^{s_\kappa}(\kappa)}^2,
\]

with \(1 \leq s_\kappa \leq \min\{p_\kappa + 1, k_\kappa\},\) \(p_\kappa \geq 1,\) for \(\kappa \in \mathcal{T},\) and \(C\) is a positive constant independent of \(u\) and the discretisation parameters.

Proof. Let us write

\[
u - u_{\text{DG}} = (u - \Pi_P^h u) + (\Pi_P^h u - u_{\text{DG}}) \equiv \eta + \xi.
\]

Note that since \(u \in C^1(\Omega) \cap H^2(\Omega, \mathcal{T}),\) we have that \(B(u, v) = \ell(v)\) for all \(v \in SP(\Omega, \mathcal{T}, \mathbf{F});\) in particular, \(B(u, \xi) = \ell(\xi)\). We begin by estimating \(\xi.\) Recalling (2.17), (2.5) and (3.3), we have that

\[
\frac{1}{2} \min\{C_2, 1\} \|\xi\|_{1,h}^2 = \frac{1}{2} \min\{C_2, 1\} \|\Pi_P^h u - u_{\text{DG}}\|_{1,h}^2
\]

\[
\leq B(\Pi_P^h u, \Pi_P^h u - u_{\text{DG}}) - B(u_{\text{DG}}, \Pi_P^h u - u_{\text{DG}})
\]

\[
= B(\Pi_P^h u, \xi) - \ell(\xi) = B(u, \xi) - B(u, \xi)
\]

\[
\leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{h^{2s_\kappa - 2}_{\kappa}}{p^{2s_\kappa - 2}_{\kappa}} \|u\|_{H^{s_\kappa}(\kappa)}^2 \right)^{1/2} \|\xi\|_{1,h}.
\]

Therefore,

\[
\|\xi\|_{1,h} \leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{h^{2s_\kappa - 2}_{\kappa}}{p^{2s_\kappa - 2}_{\kappa}} \|u\|_{H^{s_\kappa}(\kappa)}^2 \right)^{1/2},
\]

which, by the triangle inequality and (3.2), gives

\[
\|u - u_{\text{DG}}\|_{1,h} \leq \|\xi\|_{1,h} + \|\eta\|_{1,h} \leq C \left( \sum_{\kappa \in \mathcal{T}} \frac{h^{2s_\kappa - 2}_{\kappa}}{p^{2s_\kappa - 2}_{\kappa}} \|u\|_{H^{s_\kappa}(\kappa)}^2 \right)^{1/2},
\]

(3.14)

where \(1 \leq s_\kappa \leq \min\{p_\kappa + 1, k_\kappa\},\) \(p_\kappa \geq 1,\) for \(\kappa \in \mathcal{T},\) and \(C\) is a positive constant, independent of \(u\) and the discretisation parameters. \(\square\)

4. **Numerical experiments.** In this section we present a series of numerical experiments to highlight the practical performance of the interior penalty DG method introduced and analysed in this article for the numerical approximation of the quasi-linear elliptic boundary value problem (1.1)–(1.3). For simplicity, we restrict ourselves to two-dimensional model problems, i.e. \(d = 2;\) additionally, we note that throughout this section we select the constant appearing in the discontinuity penalisation parameter \(\sigma\) defined in (2.4) as follows: \(\alpha = 10.\)
4.1. Example 1. In this first example we take $\Omega \subset \mathbb{R}^2$ to be the square domain $(-1, 1)^2$ with $\Gamma_D = [-1, 1] \times (-1) \cup \{1\} \times [-1, 1]$ and $\Gamma_N = [-1, 1] \times \{1\} \cup (-1) \times [-1, 1]$. Furthermore, we set the nonlinear diffusion coefficient as follows:

$$
\mu(x, |\nabla u|) = 2 + \frac{1}{1 + |\nabla u|},
$$

the Dirichlet and Neumann boundary conditions, $g_D$ and $g_N$, respectively, and the forcing function $f$ are then chosen so that the analytical solution to (1.1)--(1.3) is given by

$$
u(x, y) = \cos(\pi x /2) \cos(\pi y /2),$$

cf. [7]. We remark that a simple calculation verifies that the coefficient $\mu$ defined in (4.1) satisfies our assumption (A) with $m_\mu = 2$ and $M_\mu = 3$.

We investigate the asymptotic convergence of the $hp$-DGFEM (2.5) on a sequence of successively finer uniform square meshes for $p \equiv p = 1, 2, 3, 4$ for different choices of $\theta$. Here, we consider the three most popular choices for $\theta$: $\theta = -1$ corresponding to the symmetric interior penalty method (SIPG), $\theta = 0$ corresponding to the incomplete interior penalty method (IIPG), and $\theta = 1$ corresponding to the nonsymmetric interior penalty method (NIPG). To this end, in Figure 4.1, we present a comparison of the DG–norm $\| \cdot \|_{1,h}$ of the error with the mesh function $h$ for each polynomial degree and each value of $\theta$. Here, we observe that $\| u - u_{DG} \|_{1,h}$ converges to zero, for each fixed $p$, at the rate $O(h^p)$ as the mesh is refined, in agreement with Theorem 3.3. In particular, we note that the error in the DG–norm is fairly insensitive to the choice of $\theta$; indeed, while an increase in the size of $\theta$ leads to a decrease in $\| u - u_{DG} \|_{1,h}$, the convergence lines are almost indistinguishable as the mesh is refined, especially in the case of odd-order polynomial degrees. Secondly, we investigate the convergence of the $hp$-DGFEM with $p$-enrichment on a fixed mesh. Since the analytical solution (4.2) is a real analytic function, we expect to see exponential rates of convergence as $p$ increases. In Figure 4.2, we plot the DG–norm of the error against $p$ on three different
square meshes for each value of $\theta$. In each case, we observe that on a linear–log scale, the convergence plots become straight lines as the spectral order $p$ is increased, thereby indicating exponential convergence in $p$.

Finally, in Figure 4.3 we plot the $L^2(\Omega)$–norm of the error against $h$ for each $p$ and each $\theta$. Here, we observe that, for each of the three choices of $\theta$, the error in the $L^2(\Omega)$–norm behaves like $O(h^{p+1})$ for odd $p$ and like $O(h^p)$ for even $p$. We remark that, in the case of a linear elliptic partial differential equation, the SIPG scheme ($\theta = -1$) is optimally accurate for both odd and even order polynomial degrees, cf. [20], for example; though both the IIPG and NIPG still exhibit the same lack of optimality when $p$ is even in this case. This loss of optimality of the SIPG
scheme for the numerical approximation of the quasilinear elliptic partial differential equation (1.1)–(1.3) when \( p \) is even is attributed to the loss of adjoint consistency of the interior penalty method (2.5). By this we mean that the integral-mean-value linearisation of the semilinear form \( B(\cdot, \cdot) \) in its first argument is a bilinear form that fails to be adjoint consistent with the bilinear form which arises from the integral-mean-value linearisation of the semilinear form in the (standard) weak formulation of the boundary value problem.

However, we remark that numerical experiments indicate that the SIPG method does not suffer from this sub-optimality for even \( p \) when the method is employed for the numerical approximation of semilinear elliptic partial differential equations.

4.2. Example 2. In this second example, we investigate the performance of the \( h p \)-DGFM (2.5) for a problem with a non-smooth solution. To this end, let \( \Omega \) be the L-shaped domain \((-1,1)^2 \setminus [0,1) \times (-1,0]\), with \( \Gamma_D = \partial \Omega \),

\[
\mu(x, |\nabla u|) = 1 + e^{-|\nabla u|^2},
\]

and select \( \gamma_D \) and \( f \) so that the analytical solution to (1.1)–(1.3) is given by

\[
u(x, y) = \cos(\pi y/2) \chi(x) x^{2.5},
\]

where \( \chi : \mathbb{R} \to \mathbb{R} \) denotes the characteristic function of the interval \((0,1) \subset \mathbb{R}\), cf. [7]. Again, as in the above example, the coefficient \( \mu \) defined in (4.3) satisfies our assumption (A) with \( m_\mu = 1 - \sqrt{2/e} \) and \( M_\mu = 1 \). The analytical solution given by (4.4) contains a singularity along the line \( x = 0 \); in particular, we note that \( u \) lies in the Sobolev space \( H^{2.5-\varepsilon}(\Omega) \), for any \( \varepsilon > 0 \).

In this example we again consider the convergence of the \( h p \)-DGFM (2.5) on a sequence of successively finer uniform square meshes for \( p = 1, 2, 3, 4 \) and \( \theta = -1, 0, 1 \). To this end, in Figure 4.4 we plot the DG–norm of the error against \( h \) for each \( p \) and each \( \theta \). Here, we observe that for each of the three methods considered the error, \( ||u - u_{DG}||_{1,h} \), converges to zero at the optimal rate \( O(h^{\min(p+1,k)-1}) \) predicted by

\[
||u - u_{DG}||_{1,h} \leq C h^{\min(p+1,k)-1}
\]
<table>
<thead>
<tr>
<th>$p$</th>
<th>$|u - u_{DG}|_{1,h}$</th>
<th>$k$</th>
<th>$|u - u_{DG}|_{1,h}$</th>
<th>$k$</th>
<th>$|u - u_{DG}|_{1,h}$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.745e-1</td>
<td>-</td>
<td>6.927e-1</td>
<td>-</td>
<td>6.737e-1</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>5.163e-3</td>
<td>6.34</td>
<td>5.033e-3</td>
<td>6.31</td>
<td>5.017e-3</td>
<td>6.30</td>
</tr>
<tr>
<td>4</td>
<td>1.021e-3</td>
<td>5.63</td>
<td>9.994e-4</td>
<td>5.62</td>
<td>9.949e-4</td>
<td>5.62</td>
</tr>
<tr>
<td>5</td>
<td>3.813e-4</td>
<td>4.41</td>
<td>3.731e-4</td>
<td>4.41</td>
<td>3.715e-4</td>
<td>4.41</td>
</tr>
<tr>
<td>6</td>
<td>1.759e-4</td>
<td>4.24</td>
<td>1.722e-4</td>
<td>4.24</td>
<td>1.718e-4</td>
<td>4.24</td>
</tr>
<tr>
<td>7</td>
<td>9.242e-5</td>
<td>4.18</td>
<td>9.044e-5</td>
<td>4.18</td>
<td>9.005e-5</td>
<td>4.18</td>
</tr>
<tr>
<td>8</td>
<td>5.327e-5</td>
<td>4.13</td>
<td>5.218e-5</td>
<td>4.12</td>
<td>5.198e-5</td>
<td>4.11</td>
</tr>
<tr>
<td>9</td>
<td>3.304e-5</td>
<td>4.06</td>
<td>3.237e-5</td>
<td>4.05</td>
<td>3.225e-5</td>
<td>4.05</td>
</tr>
<tr>
<td>10</td>
<td>2.170e-5</td>
<td>3.99</td>
<td>2.130e-5</td>
<td>3.97</td>
<td>2.123e-5</td>
<td>3.96</td>
</tr>
</tbody>
</table>

Table 4.1

**Example 2. Convergence of the DGFEM with $p$-refinement.**

Theorem 3.3 as $h$ tends to zero for each fixed $p$. Again, as in the previous example we see that the size of the error is insensitive to the choice of $\theta$, though an increase in $\theta$ does again lead to a marginal decrease in $\|u - u_{DG}\|_{1,h}$.

Finally, we investigate the asymptotic behaviour of the proposed methods with $p$-enrichment. In Table 4.1 we show the the DG-norm of the error and the computed rate of convergence $k$ for the SIPG, IPG and NIPG schemes on a uniform square mesh consisting of twelve elements. Here, we observe that the DG-norm of the error converges to zero at (approximately) the rate $O(p^{-4})$ for all three methods considered; this is twice the (optimal) rate predicted by Theorem 3.3. However, this behaviour is due to the presence of the singularity in $u$ arising on an inter-element boundary, rather than in the interior of an element $k$ in the mesh. Indeed, by employing approximation results in terms of weighted Sobolev norms, a *priori* error bounds which reflect this order-doubling in the rate of convergence of the DGFEM with $p$-refinement may be established, cf. [22].

### 4.3. Example 3

In this final example, we consider a problem for which the structural hypothesis (1.9) on the coefficient $\mu$ is violated. To this end, we set $\Omega = (0,1)^2$, with $\Gamma_D = \partial \Omega$ and

$$
\mu(x, |\nabla u|) = |\nabla u|^2, \quad 1 < r < \infty.
$$

Choosing $r = 3$, we select $g_D$ and $f$ so that the analytical solution to (1.1)-(1.3) is given by

$$
u(x, y) = e^{xy}.
$$

In Figure 4.5 we plot the DG-norm of the error against the mesh function $h$ for $p = 1, 2, 3, 4$ and $\theta = -1, 0, 1$. As in Example 1, we again observe that $\|u - u_{DG}\|_{1,h}$ converges to zero, for each fixed $p$, at the rate $O(h^p)$ as the mesh is refined, for each choice of $\theta$; this is in agreement with the optimal rate predicted by Theorem 3.3, even though the underlying hypotheses on $\mu$ no longer hold. As in the previous examples, we note that the error is relatively insensitive to changes in $\theta$. Furthermore, we note that the $L^2(\Omega)$-norm of the error behaves in an analogous manner as in Example 1, for a fixed polynomial degree as the mesh is refined; i.e. the $L^2(\Omega)$-norm of the error in $hp$-DGFEM converges to zero at the optimal rate $O(h^{p+1})$ for odd $p$ as $h$ tends to
zero, but at only the rate $O(h^p)$ when $p$ is even. For brevity, these numerical results are omitted.

Finally, in Figure 4.6 we plot $||u - u_{DG}||_{1,h}$ against the polynomial degree $p$ for each value of $\theta$ on a linear-log scale. Given that the analytical solution (4.6) is a real analytic function, we again observe exponential convergence of the error in the $hp$-DGFEM as $p$ is enriched for each fixed $h$ and each fixed $\theta$, cf. Example 1.

5. Concluding remarks. In this article we have developed a one-parameter family of $hp$-version discontinuous Galerkin finite element methods for the numerical solution of quasilinear elliptic equations in divergence-form on a bounded open
polyhedral domain \( \Omega \subset \mathbb{R}^d, d \geq 2 \). We then considered the analysis of the methods for the equation \(-\nabla \cdot (\mu(x, |\nabla u|)\nabla u) = f(x)\) subject to mixed Dirichlet–Neumann boundary conditions on \( \partial \Omega \), under the assumption that \( \mu \) is a real–valued function, \( \mu \in C(\bar{\Omega} \times [0, \infty)) \), and there exist positive constants \( m_{\mu} \) and \( M_{\mu} \) such that \( m_{\mu}(t-s) \leq \mu(t) - \mu(s) \leq M_{\mu}(t-s) \) for \( t \geq s \geq 0 \).

The discrete problem was shown to have a unique solution \( u_{DG} \) in the finite element space for any value of the parameter \( \theta \in [-1, 1] \). If \( u \in C^1(\bar{\Omega}) \cap H^2(\Omega) \), \( k \geq 2 \), then, with discontinuous piecewise polynomials of degree \( p \geq 1 \), the error between \( u \) and \( u_{DG} \), measured in the broken \( H^1 \)-norm, was proved to be \( O(h^{k-1}/p^{k-3/2}) \), where \( 1 \leq s \leq \min\{p+1, k\} \). The theoretical results were illustrated by numerical experiments. Provided that the structural hypothesis (1.9) on the nonlinearity is retained, the theoretical results of the paper are easily extended to quasilinear elliptic and parabolic problems containing lower order terms. In the absence of hypothesis (1.9), however, as would be the case with nonlinearities such as \( \mu(t) = (1+|t|^{r})^{1/r} \) or \( \mu(t) = \|t\|^{r-2}, \ 1 < r < \infty \), either one or both of the uniform Lipschitz continuity and the uniform monotonicity of the semilinear form \( B(\cdot, \cdot) \) is violated. The analysis is then much more complicated and is the subject of our current research.

REFERENCES


