MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

AN *hp*–ADAPTIVE NEWTON-DISCONTINUOUS-GALERKIN FINITE ELEMENT APPROACH FOR SEMILINEAR ELLIPTIC BOUNDARY VALUE PROBLEMS

PAUL HOUSTON AND THOMAS P. WIHLER

Abstract. In this paper we develop an *hp*–adaptive procedure for the numerical solution of general second-order semilinear elliptic boundary value problems, with possible singular perturbation. Our approach combines both adaptive Newton schemes and an *hp*–version adaptive discontinuous Galerkin finite element discretisation, which, in turn, is based on a robust *hp*–version *a posteriori* residual analysis. Numerical experiments underline the robustness and reliability of the proposed approach for various examples.

1. Introduction

The subject of this paper is the adaptive numerical approximation of second-order semilinear elliptic problems of the form

$$-\epsilon\Delta u + u = f(\cdot, u) \text{ in } \Omega, \qquad u = 0 \text{ on } \partial\Omega. \tag{1}$$

Here, $\Omega \subset \mathbb{R}^2$ is an open and bounded Lipschitz domain, $\epsilon \in (0, 1]$ represents a (possibly small singular perturbation) parameter, $f : \overline{\Omega} \times \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function, and $u : \Omega \to \mathbb{R}$ is an unknown solution; in the sequel, we will omit to explicitly express the dependence of f on the first argument, and simply write $f(u)$ instead. Problems of this type appear in a wide range of application areas of practical interest, such as, for example, nonlinear reaction-diffusion in ecology and chemical models [12,14,25,46,47], economy [8], or classical and quantum physics [9, 10, 32, 52].

Partial differential equations (PDEs) of the form (1) may admit a unique solution, no solution at all, or more typically a multitude of solutions, or indeed infinitely many such solutions. Moreover, in the singularly perturbed case, i.e., when $0 < \epsilon \ll 1$, solutions of (1), when they exist, may contain sharp layers in the form of interior/boundary layers, or isolated spike–like solutions, and their numerical approximation represents a challenging computational task. Indeed, to efficiently and reliably compute discrete approximations to the analytical solution u of (1),

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

Mathematics Institute, University of Bern, CH-3012 Bern, Switzerland

E-mail addresses: Paul.Houston@nottingham.ac.uk, wihler@math.unibe.ch.

2010 *Mathematics Subject Classification.* 65N30.

Key words and phrases. Newton method, semilinear elliptic problems, adaptive finite element methods, discontinuous Galerkin methods, *hp*–adaptivity.

TW acknowledges the support of the Swiss National Science Foundation (SNF), Grant No. 200021-162990.

it is essential to exploit *a posteriori* bounds which not only provide information regarding the size of the discretisation error, measured in some appropriate norm, but also yield local error indicators which may subsequently be employed to enrich the underlying approximation space in an adaptive manner. Of course, a key aspect of this general solution procedure is the design and implementation of a nonlinear solver which can efficiently compute the approximation u_h to u; we shall return to this issue below.

In general, the traditional approach exploited within the literature for the design of adaptive finite element methods, for example, is to first discretise the underlying PDE problem, in our case (1), and to derive an *a posteriori* error bound for the resulting (nonlinear) scheme; this is typically a very mathematically challenging task. However, once such a bound has been established, then given a suitable initial mesh and polynomial approximation order, the underlying nonlinear system of discrete equations arising from the underlying finite element discretisation may be solved based on employing, for example, a (damped) Newton iteration. Denoting this computed numerical approximation by u_h, the size of the error between u and u_h may then be estimated by exploiting this *a posteriori* error bound. If this bound is below a given user tolerance, then sufficient accuracy has been attained and the adaptive algorithm may be terminated. Otherwise, the computational mesh (h–refinement) or the polynomial degree (p–refinement), or both (hp–refinement) are locally enriched based on identifying regions in the domain where the elementwise error indicators, which stem from the *a posteriori* error bound, are locally large. On the basis of this new finite element space, a new approximation u_h to u may be computed, and the whole process repeated until either the desired accuracy has been attained, or a maximum number of refinement steps have been completed.

Stimulated by the work undertaken in the article [5], we consider an alternative approach based on the so-called adaptive Newton-Galerkin paradigm for the numerical approximation of nonlinear problems of the type (1). More precisely, this general technique is based on applying local Newton-type linearisations on the continuous level that allow for the approximation of the semilinear PDE (1) by a sequence of linearised problems. These resulting *linear* PDEs are then discretised by means of an adaptive finite element procedure, which, in turn, is based on a suitable *a posteriori* residual analysis. The adaptive Newton-Galerkin procedure provides an *interplay* between the (adaptive, or damped) Newton method and the adaptive finite element approach, whereby we either perform a Newton step (if the Newton linearisation effect dominates) or enrich the current finite element space based on the above *a posteriori* residual indicators (in the case that the finite element discretisation constitutes the main source of error); for related work we refer to [16, 27], or the articles [11, 21, 31] on (derivative-free) fixed-point iteration schemes. Finally, we point to the works [15, 33] dealing with modelling errors in linearised models.

In the current article, we extend the work undertaken in [5] to the framework of hp–version adaptive interior penalty discontinuous Galerkin (DG) schemes, thereby giving rise to hp–adaptive Newton-discontinuous Galerkin (NDG) methods. Here, the proof of the resulting *a posteriori* residual bound for the interior penalty DG discretisation of the underlying linearised PDE problem is based on two key steps: firstly, we introduce a suitable residual operator on a given enriched space, which, when measured in an appropriate norm, is equivalent to the error measured in terms

of the underlying DG energy norm. Secondly, an upper bound on the norm of the residual operator is derived based on exploiting the general techniques developed in the articles [34, 35, 55]; we also refer to [56] for the application to convection–diffusion problems, and to [19,38] for the treatment of strongly monotone quasilinear PDEs, cf., also, [18, 20] for hp–version two-grid DG methods. The proof of this upper bound crucially relies on the approximation of discontinuous finite element functions by conforming ones, cf., also, [40] for the h–version case. Moreover, in the current setting, following [53], particular care is devoted to the derivation of ϵ-robust approximation estimates. The resulting *a posteriori* bound consists of two key terms: one stemming from the Newton linearisation error, and the second which measures the approximation error in the underlying DG scheme. On the basis of this general hp–version bound, we devise a fully automatic hp–adaptive NDG scheme for the numerical approximation of PDEs of the form (1). Indeed, the performance of the resulting adaptive strategy is demonstrated on both the Bratu and Ginzburg Landau problems; moreover, the superiority of exploiting hp–enrichment of the DG finite element space, in comparison with standard mesh adaptation (h–refinement), will be highlighted.

The structure of this article is as follows. In Section 2 we briefly outline the adaptive (damped) Newton linearisation procedure employed within this article. The hp–version interior penalty DG discretisation of the resulting linearised PDE problem is then given in Section 3. Section 4 is devoted to the derivation of a residual-based *a posteriori* bound. On the basis of this bound in Section 5 we design a suitable adaptive refinement strategy, which controls both the error arising in the Newton linearisation, as well as the error in the hp–DG finite element scheme; in the latter case, we exploit automatic hp–refinement of the underlying finite element space. The performance of this proposed algorithm is demonstrated for a series of numerical examples presented in Section 6. Finally, in Section 7 we summarise the work presented in this article and discuss potential future extensions.

2. Newton Linearisation

2.1. **An Adaptive Newton Approach.** We will briefly revisit an adaptive 'black-box' prediction-type Newton algorithm from [5], and refer to [23] for more sophisticated approaches in more specific situations. Let us consider two Banach spaces X, Y, with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. Then, given an open subset $\Xi \subset X$, and a (possibly nonlinear) operator $\mathsf{F}_\epsilon : \Xi \to Y$, we are interested in solving the *nonlinear* operator equation

$$\mathsf{F}_\epsilon(u) = 0, \tag{2}$$

for some unknown zeros $u \in \Xi$. Supposing that the Fréchet derivative F'_ϵ of F_ϵ exists in Ξ (or in a suitable subset), the classical Newton method for solving (2) starts from an initial guess $u_0 \in \Xi$, and generates a sequence $\{u_n\}_{n\geq 1} \subset X$ that is defined iteratively by the *linear* equation

$$\mathsf{F}'_\epsilon(u_n)(u_{n+1} - u_n) = -\mathsf{F}_\epsilon(u_n), \qquad n \geq 0. \tag{3}$$

Naturally, for this iteration to be well-defined, we need to assume that $\mathsf{F}'_\epsilon(u_n)$ is invertible for all $n \geq 0$, and that $\{u_n\}_{n\geq 0} \subset \Xi$.

In order to improve the reliability of the Newton method (3) in the case that the initial guess u_0 is relatively far away from a root $u_\infty \in \Xi$ of F_ϵ, $\mathsf{F}_\epsilon(u_\infty) = 0$,

introducing some damping in the Newton method is a well-known remedy. In that case (3) is rewritten as

$$u_{n+1} = u_n - \Delta t_n \mathsf{F}'_\epsilon(u_n)^{-1}\mathsf{F}_\epsilon(u_n), \qquad n \geq 0, \tag{4}$$

where $\Delta t_n > 0$, $n \geq 0$, is a damping parameter that may be adjusted *adaptively* in each iteration step. The selection of the Newton parameter Δt_n is based on the following idea from [5]: provided that $\mathsf{F}'_\epsilon(u)$ is invertible on a suitable subset of $\Xi \subset X$, we define the *Newton-Raphson transform* by

$$u \mapsto \mathsf{NF}(u) := -\mathsf{F}'_\epsilon(u)^{-1}\mathsf{F}_\epsilon(u);$$

see, e.g., [48]. Then, rearranging terms in (4), we notice that

$$\frac{u_{n+1} - u_n}{\Delta t_n} = \mathsf{NF}(u_n), \qquad n \geq 0,$$

i.e., (4) can be seen as the discretisation of the dynamical system

$$\dot{u}(t) = \mathsf{NF}(u(t)), \quad t \geq 0, \qquad u(0) = u_0, \tag{5}$$

by the forward Euler scheme, with step size $\Delta t_n > 0$. For $t \in [0, \infty)$, the solution $u(t)$ of (5), if it exists, defines a trajectory in X that starts at u_0, and that will potentially converge to a zero of F_ϵ as $t \to \infty$. Indeed, this can be seen (formally) from the integral form of (5), that is,

$$\mathsf{F}_\epsilon(u(t)) = \mathsf{F}_\epsilon(u_0)e^{-t}, \qquad t \geq 0,$$

which implies that $\mathsf{F}_\epsilon(u(t)) \to 0$ as $t \to \infty$.

Now taking the view of dynamical systems, our goal is to compute an upper bound for the value of the step sizes $\Delta t_n > 0$ from (4), $n \geq 0$, so that the discrete forward Euler solution $\{u_n\}_{n\geq 0}$ from (4) stays reasonably close to the continuous solution of (5). Specifically, a Taylor expansion analysis (see [5, Section 2] for details) reveals that

$$u(t) = u_0 + t\mathsf{NF}(u_0) + \frac{t^2}{2h_0}\eta_{h_0} + \mathcal{O}(t^3) + \mathcal{O}(t^2 h_0 \|\mathsf{NF}(u_0)\|_X^2),$$

where, for any sufficiently small $h_0 > 0$, we let $\eta_{h_0} = \mathsf{NF}(u_0 + h_0\mathsf{NF}(u_0)) - \mathsf{NF}(u_0)$. Hence, after the first time step of length $\Delta t_0 > 0$ there holds

$$u(\Delta t_0) - u_1 = \frac{\Delta t_0^2}{2h_0}\eta_{h_0} + \mathcal{O}(\Delta t_0^3) + \mathcal{O}(\Delta t_0^2 h_0 \|\mathsf{NF}(u_0)\|_X^2), \tag{6}$$

where u_1 is the forward Euler solution from (4).

Given a prescribed tolerance $\tau > 0$, we proceed along the lines of [5, Section 2.3] by defining

$$\Delta t_0 := \min\left\{\sqrt{2\tau h_0 \|\eta_{h_0}\|_X^{-1}}, 1\right\},$$

as well as the update

$$h_1 := \gamma \Delta t_0 \|\mathsf{NF}(u_0)\|_X^{-2},$$

where $\gamma > 0$ is a parameter. Then, from (6) we infer that

$$\|u(\Delta t_0) - u_1\|_X \leq \tau + \mathcal{O}(\Delta t_0^2 h_1 \|\mathsf{NF}(u_0)\|_X^2). \tag{7}$$

This leads to the following *adaptive Newton algorithm*.

Algorithm 2.1. Fix a tolerance $\tau > 0$, a parameter $\gamma > 0$, and $h^{\max} > 0$. Set $n \leftarrow 0$.

1: Start the Newton iteration with an initial guess $u_0 \in \Xi$.
2: **if** $n = 0$ **then** choose

$$\Delta t_0 = \min\left\{\sqrt{2\tau\,\|\mathsf{NF}(u_0)\|_X^{-1}}, 1\right\},$$

based on [5, Algorithm 2.1] (cf. also [4]),
3: **else**let $\kappa_n = \Delta t_{n-1}$ and $h_n = \min\left\{\gamma\kappa_n\|\mathsf{NF}(u_n)\|_X^{-2}, h^{\max}\right\}$; define the Newton step size

$$\Delta t_n = \min\left\{\sqrt{2\tau h_n\,\|\mathsf{NF}(u_n + h_n\mathsf{NF}(u_n)) - \mathsf{NF}(u_n)\|_X^{-1}}, 1\right\}. \tag{8}$$

4: **end if**
5: Compute u_{n+1} based on the Newton iteration (4), and go to (3:) with $n \leftarrow n+1$.

We notice that the minimum in (8) ensures that the step size Δt_n is chosen to be 1 whenever possible. Indeed, this is required in order to guarantee quadratic convergence of the Newton iteration close to a root (provided that the root is simple).

Furthermore, we remark that the prescribed tolerance τ in the above adaptive strategy will typically be fixed *a priori*. Here, for highly nonlinear problems featuring numerous or even infinitely many solutions, it is typically mandatory to select $\tau \ll 1$ small in order to remain within the attractor of the given initial guess. This is particularly important if the starting value is relatively far away from a solution.

As final comment, we point out that within the definition of h_n in Algorithm 2.1, we impose a maximal value of $h^{\max}$; this is necessary to avoid roundoff issues as the Newton iteration approaches a root, whereby we expect $\|\mathsf{NF}(u_n)\|_X$ to be extremely small. Here, in view of the estimate (7), it would also be possible to select an individual upper bound, e.g., $h_n^{\max} = \mathcal{O}(\tau\kappa_n^{-2}\|\mathsf{NF}(u_n)\|_X^{-2})$, in each Newton iteration step.

2.2. **Application to Semilinear PDEs.** In this article, we suppose that a (not necessarily unique) solution $u \in X := H_0^1(\Omega)$ of (1) exists; here, we denote by $H_0^1(\Omega)$ the standard Sobolev space of functions in $H^1(\Omega) = W^{1,2}(\Omega)$ with zero trace on $\partial\Omega$. Furthermore, signifying by $X' = H^{-1}(\Omega)$ the dual space of X, and upon defining the map $\mathsf{F}_\epsilon : X \to X'$ through

$$\langle \mathsf{F}_\epsilon(u), v\rangle := \int_\Omega \{\epsilon\nabla u\cdot\nabla v + uv - f(u)v\}\,\mathsf{d}\boldsymbol{x} \qquad \forall v \in X, \tag{9}$$

where $\langle\cdot,\cdot\rangle$ is the dual product in $X' \times X$, the above problem (1) can be written as a nonlinear operator equation in X':

$$u \in X: \qquad \mathsf{F}_\epsilon(u) = 0. \tag{10}$$

For any subset $D \subseteq \Omega$, we denote by $\|\cdot\|_{0,D}$ the L^2-norm on D; in the case when $D = \Omega$, we simply write $\|\cdot\|_0$ in lieu of $\|\cdot\|_{0,\Omega}$. With this notation, we note that the space X is equipped with the norm

$$\|u\|_X^2 := \epsilon\|\nabla u\|_0^2 + \|u\|_0^2, \qquad u \in X.$$

The Fréchet-derivative of the operator F_ϵ from (10) at $u \in X$ is given by

$$\langle \mathsf{F}_\epsilon'(u)w, v\rangle = \int_\Omega \{\epsilon\nabla w\cdot\nabla v + wv - f'(u)wv\}\,\mathsf{d}\boldsymbol{x}, \qquad v, w \in X = H_0^1(\Omega),$$

where we write $f' \equiv \partial_u f$. We note that, if there is a constant $\omega > 1$ for which $f'(u) \in L^\omega(\Omega)$, then $\mathsf{F}'_\epsilon(u)$ is a well-defined linear and bounded mapping from X to X'; see [5, Lemma A.1].

Now given an initial guess $u_0 \in X$, the adaptive Newton method (4) for (10) is defined iteratively by finding $u_{n+1} \in X$ from $u_n \in X$, $n \geq 0$, such that

$$\mathsf{F}'_\epsilon(u_n)(u_{n+1} - u_n) = -\Delta t_n \mathsf{F}_\epsilon(u_n),$$

in X'. When applied to (9) and (10), this turns into

$$\begin{aligned}\int_\Omega \{\epsilon\nabla(u_{n+1} - u_n) \cdot \nabla v + (u_{n+1} - u_n)v - f'(u_n)(u_{n+1} - u_n)v\}\,\mathsf{d}\boldsymbol{x} \\ = -\Delta t_n \int_\Omega \{\epsilon\nabla u_n \cdot \nabla v + u_n v - f(u_n)v\}\,\mathsf{d}\boldsymbol{x} \qquad \forall v \in X.\end{aligned}$$

Hence, for $n \geq 0$, the updated Newton iterate u_{n+1} is defined through the *linear* weak formulation

$$\begin{aligned}\int_\Omega \{\epsilon\nabla\widehat{\mathfrak{u}}_{n+1} \cdot \nabla v + \widehat{\mathfrak{u}}_{n+1}v - f'(u_n)\widehat{\mathfrak{u}}_{n+1}v\}\,\mathsf{d}\boldsymbol{x} \\ = \Delta t_n \int_\Omega \{f(u_n) - f'(u_n)u_n\}v\,\mathsf{d}\boldsymbol{x} \qquad \forall v \in X,\end{aligned} \tag{11}$$

where $\widehat{\mathfrak{u}}_{n+1} = u_{n+1} - (1 - \Delta t_n)u_n$. Incidentally, if there exists a constant δ with $\epsilon^{-1}(f'(u_n) - 1) \leq \delta < C_{\mathsf{PF}}^{-2}$ on Ω, where $C_{\mathsf{PF}} = C_{\mathsf{PF}}(\Omega) > 0$ is the constant in the Poincaré-Friedrichs inequality on Ω,

$$\|w\|_0 \leq C_{\mathsf{PF}}\|\nabla w\|_0 \qquad \forall w \in X,$$

then (11) is a linear second-order diffusion-reaction problem that is coercive on X. In particular, (11) exhibits a unique solution $u_{n+1} \in X$ in this case.

3. *hp*–DG Discretisation

3.1. **Meshes, Spaces, and DG Flux Operators.**

We will employ a standard *hp*–DG setting; see, e.g., [35, 56].

3.1.1. *Meshes and DG Spaces.* Let $\mathcal{T}$ be a subdivision of Ω into disjoint open parallelograms κ such that $\overline{\Omega} = \bigcup_{\kappa\in\mathcal{T}} \overline{\kappa}$. We assume that $\mathcal{T}$ is shape-regular, and that each $\kappa \in \mathcal{T}$ is an affine image of the unit square $\widehat{\kappa} = (0,1)^2$; i.e., for each $\kappa \in \mathcal{T}$ there exists an affine element mapping $\Psi_\kappa : \widehat{\kappa} \to \kappa$ such that $\kappa = \Psi_\kappa(\widehat{\kappa})$. By h_κ we denote the element diameter of $\kappa \in \mathcal{T}$, $h = \max_{\kappa\in\mathcal{T}_h} h_\kappa$ is the mesh size, and $\boldsymbol{n}_\kappa$ signifies the unit outward normal vector to κ on $\partial\kappa$. Furthermore, we assume that $\mathcal{T}$ is of bounded local variation, i.e., there exists a constant $\rho_1 \geq 1$, independent of the element sizes, such that $\rho_1^{-1} \leq {}^{h_\kappa}\!/_{h_{\kappa'}} \leq \rho_1$, for any pair of elements $\kappa, \kappa' \in \mathcal{T}$ which share a common edge $e = (\partial\kappa \cap \partial\kappa')^\circ$. In this context, let us consider the set $\mathcal{E}$ of all one-dimensional open edges of all elements $\kappa \in \mathcal{T}$. Further, we denote by $\mathcal{E}_\mathcal{I}$ the set of all edges e in $\mathcal{E}$ that are contained in Ω (interior edges). Additionally, introduce $\mathcal{E}_\mathcal{B}$ to be the set of boundary edges consisting of all $e \in \mathcal{E}$ that are contained in $\partial\Omega$. In our analysis, we allow the meshes to be 1-irregular, i.e., each edge of an element $\kappa \in \mathcal{T}$ may contain (at most) one hanging node, which we assume to be located at the centre of e. Suppose that e is an edge of an element $\kappa \in \mathcal{T}$; then, by h_e, we denote the length of e. Due to our assumptions on the subdivision $\mathcal{T}$ we have that, if $e \subset \partial\kappa$, then h_e is commensurate with h_κ, the diameter of κ.

For a nonnegative integer k, we denote by $\mathcal{Q}_k(\widehat{\kappa})$ the set of all tensor-product polynomials on $\widehat{\kappa}$ of degree at most k in each co-ordinate direction. To each $\kappa \in \mathcal{T}$ we assign a polynomial degree p_κ (local approximation order). We store the quantities h_κ and p_κ in the vectors $\boldsymbol{h} = \{h_\kappa : \kappa \in \mathcal{T}\}$ and $\boldsymbol{p} = \{p_\kappa : \kappa \in \mathcal{T}\}$, respectively, and consider the DG finite element space

$$\mathcal{V}_{\mathsf{DG}} = \{v \in L^2(\Omega) : \ v|_\kappa \circ \Psi_\kappa \in \mathcal{Q}_{p_\kappa}(\widehat{\kappa}) \quad \forall \kappa \in \mathcal{T}\} \, . \tag{12}$$

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 $\rho_2 \geq 1$ independent of $\boldsymbol{h}$ and $\boldsymbol{p}$, such that, for any pair of neighbouring elements $\kappa, \kappa' \in \mathcal{T}$, we have $\rho_2^{-1} \leq {}^{p_\kappa}/_{p_{\kappa'}} \leq \rho_2$. Moreover, for an edge $e = (\partial\kappa \cap \partial\kappa')^\circ$ shared by two elements $\kappa, \kappa' \in \mathcal{T}$, we define $p_e := {}^1/_2(p_\kappa + p_{\kappa'})$, or $p_e = p_\kappa$ if $e = (\partial\kappa \cap \partial\Omega)^\circ$, for some $\kappa \in \mathcal{T}$, is a boundary edge.

3.1.2. *Jump and Average Operators.* Let κ and κ' be two adjacent elements of $\mathcal{T}$, and $\boldsymbol{x}$ an arbitrary point on the interior edge $e \in \mathcal{E}_\mathcal{I}$ given by $e = (\partial\kappa \cap \partial\kappa')^\circ$. Furthermore, let v and $\boldsymbol{q}$ be scalar- and vector-valued functions, respectively, that are sufficiently smooth inside each element κ, κ'. Then, the averages of v and $\boldsymbol{q}$ at $\boldsymbol{x} \in e$ are given by

$$\langle\!\langle v \rangle\!\rangle = \frac{1}{2}(v|_\kappa + v|_{\kappa'}), \qquad \langle\!\langle \boldsymbol{q} \rangle\!\rangle = \frac{1}{2}(\boldsymbol{q}|_\kappa + \boldsymbol{q}|_{\kappa'}),$$

respectively. Similarly, the jumps of v and $\boldsymbol{q}$ at $\boldsymbol{x} \in e$ are given by

$$[\![v]\!] = v|_\kappa \, \boldsymbol{n}_\kappa + v|_{\kappa'} \, \boldsymbol{n}_{\kappa'}, \qquad [\![\boldsymbol{q}]\!] = \boldsymbol{q}|_\kappa \cdot \boldsymbol{n}_\kappa + \boldsymbol{q}|_{\kappa'} \cdot \boldsymbol{n}_{\kappa'},$$

respectively. On a boundary edge $e \in \mathcal{E}_\mathcal{B}$, we set $\langle\!\langle v \rangle\!\rangle = v$, $\langle\!\langle \boldsymbol{q} \rangle\!\rangle = \boldsymbol{q}$ and $[\![v]\!] = v\boldsymbol{n}$, with $\boldsymbol{n}$ denoting the unit outward normal vector on the boundary $\partial\Omega$.

Furthermore, we introduce, for an edge $e \in \mathcal{E}$, the discontinuity penalisation parameter σ by

$$\sigma|_e = \frac{p_e^2}{h_e}. \tag{13}$$

We conclude this section by equipping the DG space $\mathcal{V}_{\mathsf{DG}}$ with the DG norm

$$\|v\|_{\mathsf{DG}}^2 := \epsilon \, \|\nabla_\mathcal{T} v\|_0^2 + \|v\|_0^2 + \int_\mathcal{E} (\epsilon\sigma + \sigma^{-1}) |[\![v]\!]|^2 \, \mathsf{d}s, \tag{14}$$

which is induced by the DG inner product

$$(v, w)_{\mathsf{DG}} = \int_\Omega \{\epsilon \nabla_\mathcal{T} v \cdot \nabla_\mathcal{T} w + vw\} \, \mathsf{d}\boldsymbol{x} + \int_\mathcal{E} (\epsilon\sigma + \sigma^{-1}) [\![w]\!] \cdot [\![v]\!] \, \mathsf{d}s. \tag{15}$$

Here, $\nabla_\mathcal{T}$ is the element-wise gradient operator. For an element $\kappa \in \mathcal{T}$ we shall also use the norm

$$\|v\|_{\epsilon,\kappa}^2 := \epsilon \, \|\nabla v\|_{0,\kappa}^2 + \|v\|_{0,\kappa}^2 \, ,$$

for $v \in H^1(\kappa)$.

3.1.3. *Conforming Subspaces.* For a given DG finite element space $\mathcal{V}_{\mathsf{DG}}$, cf. (12), we define the extended space

$$\mathcal{W}_{\mathsf{DG}} := H_0^1(\Omega) + \mathcal{V}_{\mathsf{DG}}.$$

With this notation, the following result holds.

Lemma 3.1. *There exists a linear operator* $\mathsf{A}_{\mathsf{DG}} : \mathcal{W}_{\mathsf{DG}} \to H_0^1(\Omega)$ *such that*

$$\begin{aligned} \|w - \mathsf{A}_{\mathsf{DG}} w\|_0^2 &\leq C_{(16)} \sum_{e\in\mathcal{E}} \int_e \sigma^{-1} |[\![w]\!]|^2 \,\mathsf{d}s, \\ \|\nabla_{\mathcal{T}}(w - \mathsf{A}_{\mathsf{DG}} w)\|_0^2 &\leq C_{(16)} \sum_{e\in\mathcal{E}} \int_e \sigma |[\![w]\!]|^2 \,\mathsf{d}s, \end{aligned} \tag{16}$$

for any $w \in \mathcal{W}_{\mathsf{DG}}$*, where* $C_{(16)} > 0$ *is a constant independent of* $\mathcal{T}$ *and of* $\boldsymbol{p}$*.*

Proof. Consider the space $\mathcal{V}_{\mathsf{DG}}^{\shortparallel} := \mathcal{V}_{\mathsf{DG}} \cap H_0^1(\Omega)$, and denote by $\mathsf{P}_{\mathsf{DG}}^{\shortparallel} : \mathcal{V}_{\mathsf{DG}} \to \mathcal{V}_{\mathsf{DG}}^{\shortparallel}$ the orthogonal projection with respect to the inner product defined in (15), i.e.,

$$w \in \mathcal{V}_{\mathsf{DG}} : \quad (w - \mathsf{P}_{\mathsf{DG}}^{\shortparallel} w, v)_{\mathsf{DG}} = 0 \qquad \forall v \in \mathcal{V}_{\mathsf{DG}}^{\shortparallel}.$$

Then, defining the subspace $\mathcal{V}_{\mathsf{DG}}^{\perp} := (\mathsf{id} - \mathsf{P}_{\mathsf{DG}}^{\shortparallel})\mathcal{V}_{\mathsf{DG}}$, we have the direct sum $\mathcal{V}_{\mathsf{DG}} = \mathcal{V}_{\mathsf{DG}}^{\shortparallel} \oplus \mathcal{V}_{\mathsf{DG}}^{\perp}$, as well as

$$\mathcal{W}_{\mathsf{DG}} = H_0^1(\Omega) \oplus \mathcal{V}_{\mathsf{DG}}^{\perp}. \tag{17}$$

Based on our assumptions on the mesh $\mathcal{T}$, and referring to [56, Theorem 4.4], there exists an operator $\mathsf{I}_{hp} : \mathcal{V}_{\mathsf{DG}} \to H_0^1(\Omega)$ that satisfies

$$\begin{aligned} \sum_{\kappa\in\mathcal{T}} \|v - \mathsf{I}_{hp} v\|_{L^2(\kappa)}^2 &\leq C \sum_{e\in\mathcal{E}} \int_e \sigma^{-1} |[\![v]\!]|^2 \,\mathsf{d}s, \\ \sum_{\kappa\in\mathcal{T}} \|\nabla(v - \mathsf{I}_{hp} v)\|_{L^2(\kappa)}^2 &\leq C \sum_{e\in\mathcal{E}} \int_e \sigma |[\![v]\!]|^2 \,\mathsf{d}s, \end{aligned}$$

for any $v \in \mathcal{V}_{\mathsf{DG}}$. By virtue of (17), we can now construct the operator A_{DG} as follows: for any $w \in \mathcal{W}_{\mathsf{DG}}$, there exist unique representatives $w_0 \in H_0^1(\Omega)$ and $w_{\mathsf{DG}}^{\perp} \in \mathcal{V}_{\mathsf{DG}}^{\perp}$ with $w = w_0 + w_{\mathsf{DG}}^{\perp}$. Hence, defining $\mathsf{A}_{\mathsf{DG}} w := w_0 + \mathsf{I}_{hp} w_{\mathsf{DG}}^{\perp} \in H_0^1(\Omega)$, and employing the previous estimates, we obtain

$$\|\nabla_{\mathcal{T}}(w - \mathsf{A}_{\mathsf{DG}} w)\|_0^2 = \sum_{\kappa\in\mathcal{T}} \|\nabla(w_{\mathsf{DG}}^{\perp} - \mathsf{I}_{hp} w_{\mathsf{DG}}^{\perp})\|_{L^2(\kappa)}^2 \leq C \sum_{e\in\mathcal{E}} \int_e \sigma |[\![w_{\mathsf{DG}}^{\perp}]\!]|^2 \,\mathsf{d}s.$$

Since $w_0 \in H_0^1(\Omega)$, we notice that $[\![w_0]\!]|_e = \mathbf{0}$ for all $e \in \mathcal{E}$; thereby,

$$\|\nabla_{\mathcal{T}}(w - \mathsf{A}_{\mathsf{DG}} w)\|_0^2 \leq C \sum_{e\in\mathcal{E}} \int_e \sigma |[\![w]\!]|^2 \,\mathsf{d}s,$$

which proves the second bound in (16). The first inequality results from an analogous argument. □

Remark 3.2. We note that any $v \in H_0^1(\Omega)$ satisfies $[\![v]\!] = \mathbf{0}$ on $\mathcal{E}$; thereby, in view of (16), it follows that $\mathsf{A}_{\mathsf{DG}} v = v$ for all $v \in H_0^1(\Omega)$. Furthermore, for $w \in \mathcal{W}_{\mathsf{DG}}$, upon application of the triangle inequality and Lemma 3.1, we deduce that

$$\begin{aligned} \|\mathsf{A}_{\mathsf{DG}} w\|_X^2 &= \epsilon \|\nabla \mathsf{A}_{\mathsf{DG}} w\|_0^2 + \|\mathsf{A}_{\mathsf{DG}} w\|_0^2 \\ &\leq 2\epsilon \|\nabla w\|_0^2 + 2\|w\|_0^2 + 2\epsilon \|\nabla(w - \mathsf{A}_{\mathsf{DG}} w)\|_0^2 + 2\|w - \mathsf{A}_{\mathsf{DG}} w\|_0^2 \\ &\leq 2\epsilon \|\nabla w\|_0^2 + 2\|w\|_0^2 + 2C_{(16)} \sum_{e\in\mathcal{E}} \int_e (\epsilon\sigma + \sigma^{-1}) |[\![w]\!]|^2 \,\mathsf{d}s. \end{aligned}$$

Thus the following stability estimate holds

$$\|\mathsf{A}_{\mathsf{DG}} w\|_X \leq C_{(18)} \|w\|_{\mathsf{DG}} \qquad \forall w \in \mathcal{W}_{\mathsf{DG}}, \tag{18}$$

where $C_{(18)} = \sqrt{2\max(1, C_{(16)})}$.

3.2. **Linear hp–DG Approximation.** The hp–version interior penalty DG discretisation of (11) is given by: find $u_{n+1}^{\mathsf{DG}} \in \mathcal{V}_{\mathsf{DG}}$ from u_n^{DG} such that

$$a_{\mathsf{DG}}(u_n^{\mathsf{DG}}; u_{n+1}^{\mathsf{DG}}, v) = \ell_{\mathsf{DG}}(u_n^{\mathsf{DG}}; v) \qquad \forall v \in \mathcal{V}_{\mathsf{DG}}. \tag{19}$$

Here, for a method parameter $\theta \in [-1, 1]$ and a penalty parameter $C_\sigma \geq 0$, we define the forms

$$\begin{aligned} a_{\mathsf{DG}}(u_n^{\mathsf{DG}}; u_{n+1}^{\mathsf{DG}}, v) := & \int_\Omega \left\{ \epsilon \nabla_{\mathcal{T}} \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} \cdot \nabla_{\mathcal{T}} v + \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} v - f'(u_n^{\mathsf{DG}}) \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} v \right\} \mathrm{d}\boldsymbol{x} \\ & - \int_{\mathcal{E}} \left\{ \langle\!\langle \epsilon \nabla_{\mathcal{T}} \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} \rangle\!\rangle \cdot [\![v]\!] + \theta [\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!] \cdot \langle\!\langle \epsilon \nabla_{\mathcal{T}} v \rangle\!\rangle \right\} \mathrm{d}s \\ & + C_\sigma \int_{\mathcal{E}} \epsilon \sigma [\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!] \cdot [\![v]\!] \, \mathrm{d}s, \end{aligned} \tag{20}$$

and

$$\ell_{\mathsf{DG}}(u_n^{\mathsf{DG}}; v) = \int_\Omega \widehat{\mathfrak{f}}(u_n^{\mathsf{DG}}) v \, \mathrm{d}\boldsymbol{x},$$

for $v \in \mathcal{V}_{\mathsf{DG}}$, where for $n \geq 0$, we set

$$\begin{aligned} \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} &:= u_{n+1}^{\mathsf{DG}} - (1 - \Delta t_n) u_n^{\mathsf{DG}}, \\ \widehat{\mathfrak{f}}(u_n^{\mathsf{DG}}) &:= \Delta t_n (f(u_n^{\mathsf{DG}}) - f'(u_n^{\mathsf{DG}}) u_n^{\mathsf{DG}}). \end{aligned} \tag{21}$$

The choices $\theta \in \{-1, 0, 1\}$ correspond, respectively, to the non-symmetric (NIPG), incomplete (IIPG), and symmetric (SIPG) interior penalty DG schemes; cf. [51]. For the IIPG and SIPG methods, the penalty parameter C_σ must be chosen sufficiently large to guarantee stability of the underlying DG scheme, cf. [54], for example. Furthermore, an additional constraint on the minimal value of C_σ will be introduced in Proposition 4.1 below.

4. hp–Version *A Posteriori* Analysis

4.1. **A DG Residual.** We introduce a residual operator

$$\mathsf{R}_\epsilon : \mathcal{W}_{\mathsf{DG}} \to \mathcal{W}_{\mathsf{DG}}',$$

where $\mathcal{W}_{\mathsf{DG}}'$ is the dual space of $\mathcal{W}_{\mathsf{DG}}$, as follows: given the operator A_{DG} constructed in Lemma 3.1, and $w \in \mathcal{W}_{\mathsf{DG}}$, let us define

$$\begin{aligned} \langle \mathsf{R}_\epsilon(w), v \rangle := & \int_\Omega \{ \epsilon \nabla_{\mathcal{T}} w \cdot \nabla \mathsf{A}_{\mathsf{DG}} v + w \mathsf{A}_{\mathsf{DG}} v - f(w) \mathsf{A}_{\mathsf{DG}} v \} \, \mathrm{d}\boldsymbol{x} \\ & + C_\sigma \int_{\mathcal{E}} (\epsilon \sigma + \sigma^{-1}) [\![w]\!] \cdot [\![v]\!] \, \mathrm{d}s \qquad \forall v \in \mathcal{W}_{\mathsf{DG}}, \end{aligned} \tag{22}$$

with σ from (13), and C_σ appearing in (20). Furthermore, for $w \in \mathcal{W}_{\mathsf{DG}}$, we introduce the norm

$$|\!|\!| \mathsf{R}_\epsilon(w) |\!|\!| := \sup_{\phi \in \mathcal{W}_{\mathsf{DG}}} \frac{\langle \mathsf{R}_\epsilon(w), \phi \rangle}{\|\phi\|_{\mathsf{DG}}}. \tag{23}$$

For a solution $u \in H_0^1(\Omega)$ of (1), we again note that $[\![u]\!] = \mathbf{0}$ on $\mathcal{E}$, and, hence, due to (9) and (10), we conclude that

$$\langle \mathsf{R}_\epsilon(u), v \rangle = 0 \qquad \forall v \in \mathcal{W}_{\mathsf{DG}}. \tag{24}$$

Moreover, the following result shows that, under suitable conditions on the nonlinearity f, the norm $|||\mathsf{R}_\epsilon(\cdot)|||$ defined in (23) is directly related to the DG-norm given in (14). In this sense, we may employ the norm $|||\mathsf{R}_\epsilon(\cdot)|||$ as a natural measure for the approximation in the Newton-DG formulation (19).

Proposition 4.1. *Suppose that there exist constants $\varrho_0 > -1$ and $L \geq 0$ such that f satisfies*

$$\varrho_0 \leq -f', \qquad \text{and} \qquad |f'| \leq L, \tag{25}$$

on $\overline{\Omega} \times \mathbb{R}$. Furthermore, assume that the penalty parameter C_σ is sufficiently large so that

$$C_\sigma \geq \frac{c_0}{2} + \frac{C_{(16)}(1+L)^2}{2c_0},$$

where $C_{(16)}$ is the constant arising in the bounds (16)*, and $c_0 = 1 + \min(0, \varrho_0) > 0$. Then, for any weak solution $u \in H^1_0(\Omega)$ of* (1)*, the following bounds hold*

$$\frac{c_0}{2}\|u - w\|_{\mathsf{DG}} \leq |||\mathsf{R}_\epsilon(w)||| \leq \sqrt{2}\max\left(C_{(18)}(1+L), C_\sigma\right)\|u - w\|_{\mathsf{DG}} \tag{26}$$

for all $w \in \mathcal{W}_{\mathsf{DG}}$, where $C_{(18)}$ is the constant arising in (18)*.*

Proof. The two bounds are proved separately. Let $w \in \mathcal{W}_{\mathsf{DG}}$, then employing (24), and noting that $\mathsf{A}_{\mathsf{DG}}u = u$, cf. Remark 3.2, we obtain

$$\begin{aligned}
\langle \mathsf{R}_\epsilon(w), w - u\rangle &= \langle \mathsf{R}_\epsilon(u) - \mathsf{R}_\epsilon(w), u - w\rangle \\
&= \epsilon\int_\Omega \nabla_{\mathcal{T}}(u-w)\cdot\nabla(u - \mathsf{A}_{\mathsf{DG}}w)\,\mathrm{d}\boldsymbol{x} + \int_\Omega (u-w)(u-\mathsf{A}_{\mathsf{DG}}w)\,\mathrm{d}\boldsymbol{x} \\
&\quad - \int_\Omega (f(u)-f(w))(u-\mathsf{A}_{\mathsf{DG}}w)\,\mathrm{d}\boldsymbol{x} + C_\sigma\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![u-w]\!]|^2\,\mathrm{d}s \\
&= \|u-w\|^2_{\mathsf{DG}} + \epsilon\int_\Omega \nabla_{\mathcal{T}}(u-w)\cdot\nabla_{\mathcal{T}}(w-\mathsf{A}_{\mathsf{DG}}w)\,\mathrm{d}\boldsymbol{x} \\
&\quad + \int_\Omega (u-w)(w-\mathsf{A}_{\mathsf{DG}}w)\,\mathrm{d}\boldsymbol{x} - \int_\Omega (f(u)-f(w))(u-w)\,\mathrm{d}\boldsymbol{x} \\
&\quad - \int_\Omega (f(u)-f(w))(w-\mathsf{A}_{\mathsf{DG}}w)\,\mathrm{d}\boldsymbol{x} \\
&\quad + (C_\sigma - 1)\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![u-w]\!]|^2\,\mathrm{d}s.
\end{aligned}$$

Given the assumptions on f stated in (25) hold, we conclude that

$$-(f(u)-f(w))(u-w) \geq \varrho_0|u-w|^2, \qquad |f(u)-f(w)| \leq L|u-w|,$$

on $\overline{\Omega}\times\mathbb{R}$. Thus, applying the Cauchy-Schwarz inequality, we arrive at

$$\begin{aligned}
\langle \mathsf{R}_\epsilon(w), w-u\rangle &\geq (1+\min(0,\varrho_0))\|u-w\|^2_{\mathsf{DG}} - \epsilon\|\nabla_{\mathcal{T}}(u-w)\|_0\|\nabla_{\mathcal{T}}(w-\mathsf{A}_{\mathsf{DG}}w)\|_0 \\
&\quad - (1+L)\|u-w\|_0\|w-\mathsf{A}_{\mathsf{DG}}w\|_0 \\
&\quad + (C_\sigma - 1 - \min(0,\varrho_0))\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![u-w]\!]|^2\,\mathrm{d}s.
\end{aligned}$$

Setting $c_0 = 1 + \min(0, \varrho_0)$, we deduce that

$$\begin{aligned}\langle \mathsf{R}_\epsilon(w), w-u\rangle \geq c_0\|u-w\|_{\mathsf{DG}}^2 - \frac{c_0\epsilon}{2}\|\nabla_{\mathcal{T}}(u-w)\|_0^2 - \frac{\epsilon}{2c_0}\|\nabla_{\mathcal{T}}(w-\mathsf{A}_{\mathsf{DG}}w)\|_0^2 \\ - \frac{c_0}{2}\|u-w\|_0^2 - \frac{(1+L)^2}{2c_0}\|w-\mathsf{A}_{\mathsf{DG}}w\|_0^2 \\ + (C_\sigma - c_0)\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![u-w]\!]|^2\,\mathsf{d}s.\end{aligned}$$

By virtue of Lemma 3.1, and noting that $[\![u]\!] = \mathbf{0}$ on $\mathcal{E}$, we get

$$\begin{aligned}\langle \mathsf{R}_\epsilon(w), w-u\rangle &\geq \frac{c_0}{2}\|u-w\|_{\mathsf{DG}}^2 \\ &\quad + \left(C_\sigma - \frac{c_0}{2} - \frac{C_{(16)}(1+L)^2}{2c_0}\right)\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![u-w]\!]|^2\,\mathsf{d}s \\ &\geq \frac{c_0}{2}\|u-w\|_{\mathsf{DG}}^2.\end{aligned}$$

This gives the first bound in (26). In order to show the second estimate, we employ (25) and the Cauchy-Schwarz inequality, for any $v \in \mathcal{W}_{\mathsf{DG}}$, to infer that

$$\begin{aligned}&\langle \mathsf{R}_\epsilon(w), v\rangle \\ &= \langle \mathsf{R}_\epsilon(w) - \mathsf{R}_\epsilon(u), v\rangle \\ &= \int_\Omega \{\epsilon\nabla_{\mathcal{T}}(w-u)\cdot\nabla\mathsf{A}_{\mathsf{DG}}v + (w-u)\mathsf{A}_{\mathsf{DG}}v - (f(w)-f(u))\mathsf{A}_{\mathsf{DG}}v\}\,\mathsf{d}\boldsymbol{x} \\ &\quad + C_\sigma\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})[\![w-u]\!]\cdot[\![v]\!]\,\mathsf{d}s \\ &\leq \epsilon\|\nabla_{\mathcal{T}}(w-u)\|_0\|\nabla\mathsf{A}_{\mathsf{DG}}v\|_0 + (1+L)\|w-u\|_0\|\mathsf{A}_{\mathsf{DG}}v\|_0 \\ &\quad + \left(C_\sigma^2 C_{(18)}^{-2}\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![w-u]\!]|^2\,\mathsf{d}s\right)^{1/2}\left(C_{(18)}^2\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![v]\!]|^2\,\mathsf{d}s\right)^{1/2} \\ &\leq \max\left(1+L, C_\sigma C_{(18)}^{-1}\right)\|u-w\|_{\mathsf{DG}}\left(\|\mathsf{A}_{\mathsf{DG}}v\|_X^2 + C_{(18)}^2\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![v]\!]|^2\,\mathsf{d}s\right)^{1/2}.\end{aligned}$$

Recalling the stability of A_{DG} from (18) yields

$$\langle \mathsf{R}_\epsilon(w), v\rangle \leq \sqrt{2}C_{(18)}\max\left(1+L, C_\sigma C_{(18)}^{-1}\right)\|u-w\|_{\mathsf{DG}}\|v\|_{\mathsf{DG}}.$$

This implies the second bound in (26), and, thus, completes the proof. □

4.2. ***A Posteriori*** **Residual Analysis.** In this section we develop a residual–based *a posteriori* numerical analysis for the hp–NDG method (19).

4.2.1. *hp–Approximation Estimates.* Let $v \in \mathcal{W}_{\mathsf{DG}}$ be arbitrary, and consider $\mathsf{A}_{\mathsf{DG}}v \in H_0^1(\Omega)$ as in Lemma 3.1. Then, we may choose $\phi^{\mathsf{DG}} \in \mathcal{V}_{\mathsf{DG}}$ such that, for all $\kappa \in \mathcal{T}$, the stability bound

$$\|\mathsf{A}_{\mathsf{DG}}v - \phi^{\mathsf{DG}}\|_{0,\kappa} \leq \|\mathsf{A}_{\mathsf{DG}}v\|_{0,\kappa},$$

as well as the approximation estimate

(27)

$$\|\nabla(\mathsf{A}_{\mathsf{DG}}v - \phi^{\mathsf{DG}})\|_{0,\kappa}^2 + \frac{p_\kappa^2}{h_\kappa^2}\|\mathsf{A}_{\mathsf{DG}}v - \phi^{\mathsf{DG}}\|_{0,\kappa}^2 \leq C_{(27)}\left(\|\nabla\mathsf{A}_{\mathsf{DG}}v\|_{0,\kappa}^2 + \|\mathsf{A}_{\mathsf{DG}}v\|_{0,\kappa}^2\right)$$

hold *simultaneously*, where $C_{(27)}$ is a positive constant, independent of $\boldsymbol{h}, \boldsymbol{p}$, and of $\mathsf{A}_{\mathsf{DG}}v$; see [41, § 3.1]. Since $\epsilon \in (0,1]$, we infer the bounds

$$\epsilon\|\nabla(\mathsf{A}_{\mathsf{DG}}v - \phi^{\mathsf{DG}})\|_{0,\kappa}^2 \leq C_{(27)}\|\mathsf{A}_{\mathsf{DG}}v\|_{\epsilon,\kappa}^2,$$

and

$$\epsilon^{1/2}\|\nabla\phi^{\mathsf{DG}}\|_{0,\kappa} \leq \epsilon^{1/2}\|\nabla(\mathsf{A}_{\mathsf{DG}}v - \phi^{\mathsf{DG}})\|_{0,\kappa} + \epsilon^{1/2}\|\nabla\mathsf{A}_{\mathsf{DG}}v\|_{0,\kappa} \leq C_{(28)}\|\mathsf{A}_{\mathsf{DG}}v\|_{\epsilon,\kappa}. \tag{28}$$

Moreover, following the approach outlined in [53] (see also [5]), we deduce from the above estimates that

(29)

$$\|\mathsf{A}_{\mathsf{DG}}v - \phi^{\mathsf{DG}}\|_{0,\kappa}^2 \leq \min\left(1, C_{(27)}\epsilon^{-1}h_\kappa^2 p_\kappa^{-2}\right)\|\mathsf{A}_{\mathsf{DG}}v\|_{\epsilon,\kappa}^2 \leq \max\left(1, C_{(27)}\right)\alpha_\kappa^2\|\mathsf{A}_{\mathsf{DG}}v\|_{\epsilon,\kappa}^2,$$

where, for $\kappa \in \mathcal{T}$,

$$\alpha_\kappa := \min\left(1, \epsilon^{-1/2}h_\kappa p_\kappa^{-1}\right). \tag{30}$$

Furthermore, applying a multiplicative trace inequality, that is,

$$\|\psi\|_{0,\partial\kappa}^2 \leq C_{(31)}\left(h_\kappa^{-1}\|\psi\|_{0,\kappa}^2 + \|\psi\|_{0,\kappa}\|\nabla\psi\|_{0,\kappa}\right), \qquad \psi \in H^1(\kappa), \tag{31}$$

we obtain

$$\|\mathsf{A}_{\mathsf{DG}}v - \phi^{\mathsf{DG}}\|_{0,\partial\kappa}^2 \leq C_{(31)}\max\left(1, C_{(27)}\right)\widetilde{\beta}_\kappa^2\|\mathsf{A}_{\mathsf{DG}}v\|_{\epsilon,\kappa}^2,$$

where, for $\kappa \in \mathcal{T}$, we define

$$\widetilde{\beta}_\kappa := \sqrt{h_\kappa^{-1}\alpha_\kappa^2 + \epsilon^{-1/2}\alpha_\kappa}.$$

Noting the bound

$$\widetilde{\beta}_\kappa^2 = \epsilon^{-1/2}\alpha_\kappa\left(\epsilon^{1/2}h_\kappa^{-1}\alpha_\kappa + 1\right) \leq \epsilon^{-1/2}\alpha_\kappa(p_\kappa^{-1} + 1) \leq 2\epsilon^{-1/2}\alpha_\kappa,$$

we deduce that

$$\|\mathsf{A}_{\mathsf{DG}}v - \phi^{\mathsf{DG}}\|_{0,\partial\kappa} \leq C_{(32)}\beta_\kappa\|\mathsf{A}_{\mathsf{DG}}v\|_{\epsilon,\kappa}, \tag{32}$$

where

$$\beta_\kappa := \epsilon^{-1/4}\alpha_\kappa^{1/2}. \tag{33}$$

4.2.2. *Upper* A Posteriori *Residual Bound.* In order to derive an *a posteriori* residual estimate for the *hp*–NDG discretisation (19), we recall the residual

$$\begin{aligned}\left\langle \mathsf{R}_\epsilon(u_{n+1}^{\mathsf{DG}}), v\right\rangle &= \int_\Omega \left\{\epsilon\nabla_{\mathcal{T}}u_{n+1}^{\mathsf{DG}}\cdot\nabla\mathsf{A}_{\mathsf{DG}}v + u_{n+1}^{\mathsf{DG}}\mathsf{A}_{\mathsf{DG}}v - f(u_{n+1}^{\mathsf{DG}})\mathsf{A}_{\mathsf{DG}}v\right\}\,\mathrm{d}\boldsymbol{x}\\ &\quad + C_\sigma\int_{\mathcal{E}}(\epsilon\sigma + \sigma^{-1})[\![u_{n+1}^{\mathsf{DG}}]\!]\cdot[\![v]\!]\,\mathrm{d}s \equiv T_1 + T_2,\end{aligned}$$

cf. (22), where we define

$$\begin{aligned}T_1 &:= \int_\Omega\left\{\epsilon\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}\cdot\nabla\mathsf{A}_{\mathsf{DG}}v + \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}\mathsf{A}_{\mathsf{DG}}v - (f'(u_n^{\mathsf{DG}})\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + \widehat{\mathfrak{f}}(u_n^{\mathsf{DG}}))\mathsf{A}_{\mathsf{DG}}v\right\}\,\mathrm{d}\boldsymbol{x}\\ &\quad + C_\sigma\int_{\mathcal{E}}(\epsilon\sigma + \sigma^{-1})[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\cdot[\![v]\!]\,\mathrm{d}s,\\ T_2 &:= (1 - \Delta t_n)\left\langle \mathsf{R}_\epsilon(u_n^{\mathsf{DG}}), v\right\rangle\\ &\quad - \int_\Omega\left\{f(u_{n+1}^{\mathsf{DG}}) - f(u_n^{\mathsf{DG}}) - f'(u_n^{\mathsf{DG}})(u_{n+1}^{\mathsf{DG}} - u_n^{\mathsf{DG}})\right\}\mathsf{A}_{\mathsf{DG}}v.\end{aligned}$$

Here, $\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}$ and $\widehat{\mathfrak{f}}(u_n^{\mathsf{DG}})$ are given in (21), and $v \in \mathcal{W}_{\mathsf{DG}}$ is again arbitrary. Recalling (19), we note that

$$\begin{aligned}\int_\Omega \Big\{\epsilon\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}\cdot\nabla_{\mathcal{T}}\phi^{\mathsf{DG}} + \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}\phi^{\mathsf{DG}} - (f'(u_n^{\mathsf{DG}})\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + \widehat{\mathfrak{f}}(u_n^{\mathsf{DG}}))\phi^{\mathsf{DG}}\Big\}\,\mathsf{d}\boldsymbol{x}\\ = \int_{\mathcal{E}}\big\{\langle\!\langle\epsilon\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}\rangle\!\rangle\cdot[\![\phi^{\mathsf{DG}}]\!] + \theta[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\cdot\langle\!\langle\epsilon\nabla_{\mathcal{T}}\phi^{\mathsf{DG}}\rangle\!\rangle\big\}\,\mathsf{d}s\\ - C_\sigma\int_{\mathcal{E}}\epsilon\sigma[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\cdot[\![\phi^{\mathsf{DG}}]\!]\,\mathsf{d}s,\end{aligned}$$

with $\phi^{\mathsf{DG}} \in \mathcal{V}_{\mathsf{DG}}$ as in Section 4.2.1 above. Therefore,

$$\begin{aligned}T_1 = \int_\Omega \big\{\epsilon\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}\cdot\nabla_{\mathcal{T}}(\mathsf{A}_{\mathsf{DG}}v-\phi^{\mathsf{DG}}) + \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}(\mathsf{A}_{\mathsf{DG}}v-\phi^{\mathsf{DG}})\big\}\,\mathsf{d}\boldsymbol{x}\\ - \int_\Omega (f'(u_n^{\mathsf{DG}})\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + \widehat{\mathfrak{f}}(u_n^{\mathsf{DG}}))(\mathsf{A}_{\mathsf{DG}}v-\phi^{\mathsf{DG}})\,\mathsf{d}\boldsymbol{x}\\ + \int_{\mathcal{E}}\big\{\langle\!\langle\epsilon\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}\rangle\!\rangle\cdot[\![\phi^{\mathsf{DG}}]\!] + \theta[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\cdot\langle\!\langle\epsilon\nabla_{\mathcal{T}}\phi^{\mathsf{DG}}\rangle\!\rangle\big\}\,\mathsf{d}s\\ + C_\sigma\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\cdot[\![v]\!]\,\mathsf{d}s - C_\sigma\int_{\mathcal{E}}\epsilon\sigma[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\cdot[\![\phi^{\mathsf{DG}}]\!]\,\mathsf{d}s.\end{aligned}$$

Performing elementwise integration by parts in the first integral, and proceeding as in the proof of [38, Theorem 3.2], the following estimate can be established:

$$\begin{aligned}C|T_1| \le \sum_{\kappa\in\mathcal{T}}\|\epsilon\Delta\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} - \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + f'(u_n^{\mathsf{DG}})\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + \widehat{\mathfrak{f}}(u_n^{\mathsf{DG}})\|_{0,\kappa}\|\mathsf{A}_{\mathsf{DG}}v-\phi^{\mathsf{DG}}\|_{0,\kappa}\\ + \sum_{\kappa\in\mathcal{T}}\|\epsilon[\![\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa\setminus\partial\Omega}\|\mathsf{A}_{\mathsf{DG}}v-\phi^{\mathsf{DG}}\|_{0,\partial\kappa}\\ + \left(\sum_{\kappa\in\mathcal{T}}\frac{\epsilon p_\kappa^2}{h_\kappa}\|[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2\right)^{1/2}\epsilon^{1/2}\|\nabla_{\mathcal{T}}\phi^{\mathsf{DG}}\|_0\\ + \left(C_\sigma^2\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]|^2\,\mathsf{d}s\right)^{1/2}\left(\int_{\mathcal{E}}(\epsilon\sigma+\sigma^{-1})|[\![v]\!]|^2\,\mathsf{d}s\right)^{1/2}\\ + \left(C_\sigma^2\sum_{\kappa\in\mathcal{T}}\frac{\epsilon^2\beta_\kappa^2p_\kappa^4}{h_\kappa^2}\|[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2\right)^{1/2}\left(\sum_{\kappa\in\mathcal{T}}\beta_\kappa^{-2}\|[\![\phi^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2\right)^{1/2}.\end{aligned}$$

Here, C is a positive constant independent of $\boldsymbol{h}$, $\boldsymbol{p}$, and ϵ, and β_κ is defined in (33). Observing that $[\![\mathsf{A}_{\mathsf{DG}}v]\!] = \mathbf{0}$ on $\mathcal{E}$, and recalling (32), we infer the bound

$$\begin{aligned}\sum_{\kappa\in\mathcal{T}}\beta_\kappa^{-2}\|[\![\phi^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2 = \sum_{\kappa\in\mathcal{T}}\beta_\kappa^{-2}\|[\![\phi^{\mathsf{DG}}-\mathsf{A}_{\mathsf{DG}}v]\!]\|_{0,\partial\kappa}^2 \le C\sum_{\kappa\in\mathcal{T}}\beta_\kappa^{-2}\|\phi^{\mathsf{DG}}-\mathsf{A}_{\mathsf{DG}}v\|_{0,\partial\kappa}^2\\ \le C\|\mathsf{A}_{\mathsf{DG}}v\|_X^2.\end{aligned}$$

Additionally, exploiting (28), (29), and (32), yields

$$
\begin{aligned}
C|T_1| \leq & \sum_{\kappa\in\mathcal{T}} \|\epsilon\Delta\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} - \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + f'(u_n^{\mathsf{DG}})\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + \widehat{\mathfrak{f}}(u_n^{\mathsf{DG}})\|_{0,\kappa}\alpha_\kappa\|\mathsf{A}_{\mathsf{DG}}v\|_{\epsilon,\kappa} \\
& + \sum_{\kappa\in\mathcal{T}} \|\epsilon[\![\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa\setminus\partial\Omega}\beta_\kappa\|\mathsf{A}_{\mathsf{DG}}v\|_{\epsilon,\kappa} \\
& + \left(\sum_{\kappa\in\mathcal{T}} \frac{\epsilon p_\kappa^2}{h_\kappa}\|[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2\right)^{1/2}\|\mathsf{A}_{\mathsf{DG}}v\|_X \\
& + \left(C_\sigma^2\sum_{\kappa\in\mathcal{T}}\left(\frac{\epsilon p_\kappa^2}{h_\kappa}+\frac{h_\kappa}{p_\kappa^2}\right)\|[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2\,\mathsf{d}s\right)^{1/2}\|v\|_{\mathsf{DG}} \\
& + \left(C_\sigma^2\sum_{\kappa\in\mathcal{T}}\frac{\epsilon^2\beta_\kappa^2 p_\kappa^4}{h_\kappa^2}\|[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2\right)^{1/2}\|\mathsf{A}_{\mathsf{DG}}v\|_X,
\end{aligned}
$$

with α_κ defined in (30). Observing that $\alpha_\kappa \leq \epsilon^{-1/2}h_\kappa p_\kappa^{-1}$ yields

$$
\max\left(\frac{\epsilon p_\kappa^2}{h_\kappa}+\frac{h_\kappa}{p_\kappa^2}, \frac{\epsilon^2\beta_\kappa^2 p_\kappa^4}{h_\kappa^2}\right) \leq \frac{\epsilon p_\kappa^3}{h_\kappa}+\frac{h_\kappa}{p_\kappa^2}.
$$

Hence, applying the Cauchy-Schwarz inequality, and making use of (18), we arrive at

$$
|T_1| \leq C\left(\sum_{\kappa\in\mathcal{T}}\eta_{n,\kappa}^2\right)^{1/2}\|v\|_{\mathsf{DG}},
$$

where, for any $\kappa\in\mathcal{T}$, we define the local residual indicators

$$
\begin{aligned}
\eta_{n,\kappa}^2 := & \,\alpha_\kappa^2\|\epsilon\Delta\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} - \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + f'(u_n^{\mathsf{DG}})\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + \widehat{\mathfrak{f}}(u_n^{\mathsf{DG}})\|_{0,\kappa}^2 \\
& + \beta_\kappa^2\epsilon^2\|[\![\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa\setminus\partial\Omega}^2 + \max\left(1,C_\sigma^2\right)\left(\frac{\epsilon p_\kappa^3}{h_\kappa}+\frac{h_\kappa}{p_\kappa^2}\right)\|[\![\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2.
\end{aligned} \tag{34}
$$

In order to deal with the term T_2, we apply elementwise integration by parts to obtain

$$
\begin{aligned}
& \int_\Omega\{\epsilon\nabla_{\mathcal{T}}u_n^{\mathsf{DG}}\cdot\nabla\mathsf{A}_{\mathsf{DG}}v + u_n^{\mathsf{DG}}\mathsf{A}_{\mathsf{DG}}v - f(u_n^{\mathsf{DG}})\mathsf{A}_{\mathsf{DG}}v\}\,\mathsf{d}\boldsymbol{x} \\
& \quad = -\sum_{\kappa\in\mathcal{T}}\int_\kappa\{\epsilon\Delta u_n^{\mathsf{DG}} - u_n^{\mathsf{DG}} + f(u_n^{\mathsf{DG}})\}\mathsf{A}_{\mathsf{DG}}v\,\mathsf{d}\boldsymbol{x} + \int_{\mathcal{E}_{\mathcal{I}}}[\![\epsilon\nabla_{\mathcal{T}}u_n^{\mathsf{DG}}]\!]\mathsf{A}_{\mathsf{DG}}v\,\mathsf{d}s.
\end{aligned}
$$

Furthermore, we define the lifting operator $\mathsf{L}:\mathcal{V}_{\mathsf{DG}}\to\mathcal{V}_{\mathsf{DG}}$ by

$$
w\mapsto\mathsf{L}(w):\qquad \int_\Omega\mathsf{L}(w)\phi^{\mathsf{DG}}\,\mathsf{d}\boldsymbol{x} = \int_{\mathcal{E}_{\mathcal{I}}}[\![\nabla_{\mathcal{T}}w]\!]\phi^{\mathsf{DG}}\,\mathsf{d}s\qquad\forall\phi^{\mathsf{DG}}\in\mathcal{V}_{\mathsf{DG}};
$$

cf., e.g., [6, 49]. Thereby, we note that

$$
\begin{aligned}
& \int_\Omega\{\epsilon\nabla_{\mathcal{T}}u_n^{\mathsf{DG}}\cdot\nabla\mathsf{A}_{\mathsf{DG}}v + u_n^{\mathsf{DG}}\mathsf{A}_{\mathsf{DG}}v - f(u_n^{\mathsf{DG}})\mathsf{A}_{\mathsf{DG}}v\}\,\mathsf{d}\boldsymbol{x} \\
& \quad = -\int_\Omega\{\epsilon\Delta_{\mathcal{T}}u_n^{\mathsf{DG}} - u_n^{\mathsf{DG}} + f(u_n^{\mathsf{DG}}) - \epsilon\mathsf{L}(u_n^{\mathsf{DG}})\}\mathsf{A}_{\mathsf{DG}}v\,\mathsf{d}\boldsymbol{x} \\
& \qquad + \int_{\mathcal{E}_{\mathcal{I}}}[\![\epsilon\nabla_{\mathcal{T}}u_n^{\mathsf{DG}}]\!](\mathsf{A}_{\mathsf{DG}}v-\phi^{\mathsf{DG}})\,\mathsf{d}s - \int_\Omega\epsilon\mathsf{L}(u_n^{\mathsf{DG}})(\mathsf{A}_{\mathsf{DG}}v-\phi^{\mathsf{DG}})\,\mathsf{d}\boldsymbol{x},
\end{aligned}
$$

where $\Delta_{\mathcal{T}}$ is the elementwise Laplacian operator. Applying the Cauchy-Schwarz inequality, and incorporating the bounds from Section 4.2.1, we deduce that

$$\begin{aligned}
&\left| \int_\Omega \{\epsilon \nabla_{\mathcal{T}} u_n^{\mathsf{DG}} \cdot \nabla \mathsf{A}_{\mathsf{DG}} v + u_n^{\mathsf{DG}} \mathsf{A}_{\mathsf{DG}} v - f(u_n^{\mathsf{DG}}) \mathsf{A}_{\mathsf{DG}} v\} \, \mathrm{d}\boldsymbol{x} \right| \\
&\quad \leq \|\epsilon \Delta_{\mathcal{T}} u_n^{\mathsf{DG}} - u_n^{\mathsf{DG}} + f(u_n^{\mathsf{DG}}) - \epsilon \mathsf{L}(u_n^{\mathsf{DG}})\|_0 \|\mathsf{A}_{\mathsf{DG}} v\|_0 \\
&\qquad + \sum_{\kappa \in \mathcal{T}} \epsilon \|[\![\nabla_{\mathcal{T}} u_n^{\mathsf{DG}}]\!]\|_{0,\partial\kappa \setminus \partial\Omega} \|\mathsf{A}_{\mathsf{DG}} v - \phi^{\mathsf{DG}}\|_{0,\partial\kappa} \\
&\qquad + \sum_{\kappa \in \mathcal{T}} \epsilon \|\mathsf{L}(u_n^{\mathsf{DG}})\|_{0,\kappa} \|\mathsf{A}_{\mathsf{DG}} v - \phi^{\mathsf{DG}}\|_{0,\kappa} \\
&\quad \leq \|\epsilon \Delta_{\mathcal{T}} u_n^{\mathsf{DG}} - u_n^{\mathsf{DG}} + f(u_n^{\mathsf{DG}}) - \epsilon \mathsf{L}(u_n^{\mathsf{DG}})\|_0 \|\mathsf{A}_{\mathsf{DG}} v\|_X \\
&\qquad + C \sum_{\kappa \in \mathcal{T}} \beta_\kappa \epsilon \|[\![\nabla_{\mathcal{T}} u_n^{\mathsf{DG}}]\!]\|_{0,\partial\kappa \setminus \partial\Omega} \|\mathsf{A}_{\mathsf{DG}} v\|_{\epsilon,\kappa} \\
&\qquad + C \sum_{\kappa \in \mathcal{T}} \alpha_\kappa \epsilon \|\mathsf{L}(u_n^{\mathsf{DG}})\|_{0,\kappa} \|\mathsf{A}_{\mathsf{DG}} v\|_{\epsilon,\kappa} \\
&\quad \leq \|\epsilon \Delta_{\mathcal{T}} u_n^{\mathsf{DG}} - u_n^{\mathsf{DG}} + f(u_n^{\mathsf{DG}}) - \epsilon \mathsf{L}(u_n^{\mathsf{DG}})\|_0 \|\mathsf{A}_{\mathsf{DG}} v\|_X \\
&\qquad + C \left(\sum_{\kappa \in \mathcal{T}} \left(\beta_\kappa^2 \epsilon^2 \|[\![\nabla_{\mathcal{T}} u_n^{\mathsf{DG}}]\!]\|_{0,\partial\kappa \setminus \partial\Omega}^2 + \alpha_\kappa^2 \epsilon^2 \|\mathsf{L}(u_n^{\mathsf{DG}})\|_{0,\kappa}^2 \right) \right)^{1/2} \|\mathsf{A}_{\mathsf{DG}} v\|_X .
\end{aligned}$$

Recalling (18), we get

$$\begin{aligned}
&\left| \int_\Omega \{\epsilon \nabla_{\mathcal{T}} u_n^{\mathsf{DG}} \cdot \nabla \mathsf{A}_{\mathsf{DG}} v + u_n^{\mathsf{DG}} \mathsf{A}_{\mathsf{DG}} v - f(u_n^{\mathsf{DG}}) \mathsf{A}_{\mathsf{DG}} v\} \, \mathrm{d}\boldsymbol{x} \right| \\
&\quad \leq \|\epsilon \Delta_{\mathcal{T}} u_n^{\mathsf{DG}} - u_n^{\mathsf{DG}} + f(u_n^{\mathsf{DG}}) - \epsilon \mathsf{L}(u_n^{\mathsf{DG}})\|_0 \|v\|_{\mathsf{DG}} \\
&\qquad + C \left(\sum_{\kappa \in \mathcal{T}} \left(\beta_\kappa^2 \epsilon^2 \|[\![\nabla_{\mathcal{T}} u_n^{\mathsf{DG}}]\!]\|_{0,\partial\kappa \setminus \partial\Omega}^2 + \alpha_\kappa^2 \epsilon^2 \|\mathsf{L}(u_n^{\mathsf{DG}})\|_{0,\kappa}^2 \right) \right)^{1/2} \|v\|_{\mathsf{DG}}.
\end{aligned}$$

Furthermore, we have

$$\left| C_\sigma \int_{\mathcal{E}} (\epsilon\sigma + \sigma^{-1}) [\![u_n^{\mathsf{DG}}]\!] \cdot [\![v]\!] \, \mathrm{d}s \right| \leq C \left(\sum_{\kappa \in \mathcal{T}} C_\sigma^2 \left(\frac{\epsilon p_\kappa^2}{h_\kappa} + \frac{h_\kappa}{p_\kappa^2} \right) \|[\![u_n^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2 \right)^{1/2} \|v\|_{\mathsf{DG}}.$$

Finally, using (18) once more, we note that

$$\begin{aligned}
&\int_\Omega \{f(u_{n+1}^{\mathsf{DG}}) - f(u_n^{\mathsf{DG}}) - f'(u_n^{\mathsf{DG}})(u_{n+1}^{\mathsf{DG}} - u_n^{\mathsf{DG}})\} \mathsf{A}_{\mathsf{DG}} v \\
&\qquad\qquad \leq C_{(18)} \|f(u_{n+1}^{\mathsf{DG}}) - f(u_n^{\mathsf{DG}}) - f'(u_n^{\mathsf{DG}})(u_{n+1}^{\mathsf{DG}} - u_n^{\mathsf{DG}})\|_0 \|v\|_{\mathsf{DG}}.
\end{aligned}$$

In summary, we obtain the bound $|T_2| \leq C \delta_{n,\Omega} \|v\|_{\mathsf{DG}}$, where

$$\delta_{n,\Omega} := \left(\sum_{\kappa \in \mathcal{T}} (\delta_{n,\kappa}^{(1)})^2 \right)^{1/2} + \left(\sum_{\kappa \in \mathcal{T}} (\delta_{n,\kappa}^{(2)})^2 \right)^{1/2}, \tag{35}$$

with

$$\delta_{n,\kappa}^{(1)} := (1-\Delta t_n)\Bigg[\|\epsilon\Delta_{\mathcal{T}} u_n^{\mathsf{DG}} - u_n^{\mathsf{DG}} + f(u_n^{\mathsf{DG}}) - \epsilon \mathsf{L}(u_n^{\mathsf{DG}})\|_{0,\kappa}^2 + \epsilon^2\left(\beta_\kappa^2\|[\![\nabla_{\mathcal{T}} u_n^{\mathsf{DG}}]\!]\|_{0,\partial\kappa\setminus\partial\Omega}^2 + \alpha_\kappa^2\|\mathsf{L}(u_n^{\mathsf{DG}})\|_{0,\kappa}^2\right) + C_\sigma^2\left(\frac{\epsilon p_\kappa^2}{h_\kappa} + \frac{h_\kappa}{p_\kappa^2}\right)\|[\![u_n^{\mathsf{DG}}]\!]\|_{0,\partial\kappa}^2\Bigg]^{1/2}, \tag{36}$$

and

$$\delta_{n,\kappa}^{(2)} := \|f(u_{n+1}^{\mathsf{DG}}) - f(u_n^{\mathsf{DG}}) - f'(u_n^{\mathsf{DG}})(u_{n+1}^{\mathsf{DG}} - u_n^{\mathsf{DG}})\|_{0,\kappa}. \tag{37}$$

Thus we have proved the following key result.

Theorem 4.2. *For the hp–NDG method* (19)*, the following upper* a posteriori *residual bound holds*

$$|||\mathsf{R}_\epsilon(u_{n+1}^{\mathsf{DG}})||| \le \mathsf{E}(u_n^{\mathsf{DG}}, u_{n+1}^{\mathsf{DG}}, \boldsymbol{h}, \boldsymbol{p}) \equiv C\left(\delta_{n,\Omega}^2 + \eta_{n,\Omega}^2\right)^{1/2},$$

where C is a positive constant, independent of $\boldsymbol{h}$, $\boldsymbol{p}$, the penalty parameter C_σ, and ϵ. Moreover, $\eta_{n,\Omega}^2 = \sum_{\kappa\in\mathcal{T}}\eta_{n,\kappa}^2$, where $\eta_{n,\kappa}$, $\kappa\in\mathcal{T}$, and $\delta_{n,\Omega}$ are given in (34) *and* (35)–(37)*, respectively.*

4.2.3. *Lower* A Posteriori *Residual Bounds.* In this section, we briefly discuss the derivation of local lower residual bounds in terms of the error indicators $\eta_{n,\kappa}$, $\kappa\in\mathcal{T}$, and some data oscillation terms; for further details, we refer to [5, §4.4.2] and [35, 38, 56]. To this end, given an edge $e\in\mathcal{E}$, we write $\omega_e = \cup\{\kappa\in\mathcal{T} : e\subset\partial\kappa\}$, where, for simplicity of exposition, we make the assumption that the mesh is regular; cf. [38, Remark 3.9] for more general situations. Thereby, for any $v\in H_0^1(\omega_e)$, following [5], we note that

$$\begin{aligned}
&\int_e \epsilon[\![\nabla_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}]\!] v\,\mathsf{d}s \\
&\quad= \int_{\omega_e}\left\{\epsilon\nabla_{\mathcal{T}} u_{n+1}^{\mathsf{DG}}\cdot\nabla v + u_{n+1}^{\mathsf{DG}} v - f(u_{n+1}^{\mathsf{DG}})v\right\}\mathsf{d}\boldsymbol{x} \\
&\qquad+ \int_{\omega_e}\left\{\epsilon\Delta_{\mathcal{T}}\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} - \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + f_h'(u_n^{\mathsf{DG}})\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + \widehat{\mathfrak{f}}_h(u_n^{\mathsf{DG}})\right\} v\,\mathsf{d}\boldsymbol{x} \\
&\qquad+ (1-\Delta t_n)\int_{\omega_e}\left\{\epsilon\Delta_{\mathcal{T}} u_n^{\mathsf{DG}} - u_n^{\mathsf{DG}} + f(u_n^{\mathsf{DG}}) - \epsilon\mathsf{L}(u_n^{\mathsf{DG}})\right\} v\,\mathsf{d}\boldsymbol{x} \\
&\qquad+ (1-\Delta t_n)\left[\int_{\omega_e}\epsilon\mathsf{L}(u_n^{\mathsf{DG}})v\,\mathsf{d}\boldsymbol{x} - \int_e \epsilon[\![\nabla_{\mathcal{T}} u_n^{\mathsf{DG}}]\!] v\,\mathsf{d}s\right], \\
&\qquad+ \int_{\omega_e}\left\{\Delta t_n(f(u_n^{\mathsf{DG}}) - f_h(u_n^{\mathsf{DG}})) + (f'(u_n^{\mathsf{DG}}) - f_h'(u_n^{\mathsf{DG}}))(u_{n+1}^{\mathsf{DG}} - u_n^{\mathsf{DG}})\right\} v\,\mathsf{d}\boldsymbol{x} \\
&\qquad+ \int_{\omega_e}\left\{f(u_{n+1}^{\mathsf{DG}}) - f(u_n^{\mathsf{DG}}) - f'(u_n^{\mathsf{DG}})(u_{n+1}^{\mathsf{DG}} - u_n^{\mathsf{DG}})\right\} v\,\mathsf{d}\boldsymbol{x},
\end{aligned}$$

where we use the subindex notation $\cdot_h$ to denote the L^2–projection of a corresponding function onto $\mathcal{V}_{\mathsf{DG}}$. For $v\in H_0^1(\kappa)$, $\kappa\in\mathcal{T}$, an analogous identity to (38) holds, whereby the integrals over ω_e are replaced by integrals over κ, with the left–hand side identically equal to zero. On the basis of these equalities, proceeding as

in [35, 38], we deduce the following local lower bounds for the first and third terms present in the *a posteriori* error indicator $\eta_{n,\kappa}$, $\kappa \in \mathcal{T}$, cf. (34).

Theorem 4.3. *Given a solution $u \in H_0^1(\Omega)$ of (1), then for the hp–NDG method defined in (19), the following local lower bounds hold.*

(a) For each element $\kappa \in \mathcal{T}$:

$$\begin{aligned} &\alpha_\kappa \left\| \epsilon \Delta \widehat{\mathsf{u}}_{n+1}^{\mathsf{DG}} - \widehat{\mathsf{u}}_{n+1}^{\mathsf{DG}} + f_h'(u_n^{\mathsf{DG}}) \widehat{\mathsf{u}}_{n+1}^{\mathsf{DG}} + \widehat{\mathsf{f}}_h(u_n^{\mathsf{DG}}) \right\|_{0,\kappa} \\ &\qquad \leq C p_\kappa \left(\left| \! \left| \! \left| \mathsf{R}_\epsilon(u_{n+1}^{\mathsf{DG}}) \right| \! \right| \! \right|_\kappa + \alpha_\kappa p_\kappa^{\lambda-1} \left[\delta_{n,\kappa}^{(1)} + \delta_{n,\kappa}^{(2)} + \delta_{n,\kappa}^{h} \right] \right). \end{aligned} \tag{38}$$

(b) For any edge $e \in \mathcal{E}$:

$$\left(\frac{\epsilon p_e^3}{h_e} + \frac{h_e}{p_e^2} \right)^{1/2} \left\| [\![\widehat{\mathsf{u}}_{n+1}^{\mathsf{DG}}]\!] \right\|_{0,e} \leq C p_e^{1/2} \sum_{\kappa \subset \omega_e} \left[(\epsilon\sigma + \sigma^{-1})^{1/2} \left\| [\![u - u_{n+1}^{\mathsf{DG}}]\!] \right\|_{0,\partial\kappa} + \delta_{n,\kappa}^{(1)} \right]. \tag{39}$$

Here, $\lambda \in (1/2, 1]$, α_e and β_e denote the restriction of α_κ and β_κ, to an edge $e \subset \partial\kappa$, $\kappa \in \mathcal{T}$, and $C > 0$ is a constant, independent of the discretisation parameters and ϵ. Moreover, for any $\Lambda \subset \Omega$, which is formed from the union of a subset of elements belonging to the finite element mesh $\mathcal{T}$, we signify by $\left| \! \left| \! \left| \mathsf{R}_\epsilon(u_n^{\mathsf{DG}}) \right| \! \right| \! \right|_\Lambda$ the localised variant of $\left| \! \left| \! \left| \mathsf{R}_\epsilon(u_n^{\mathsf{DG}}) \right| \! \right| \! \right|$ defined over Λ. Finally, for $\kappa \in \mathcal{T}$, the data oscillation term $\delta_{n,\kappa}^h$ is defined by

$$\delta_{n,\kappa}^h = \|\Delta t_n (f(u_n^{\mathsf{DG}}) - f_h(u_n^{\mathsf{DG}})) + (f'(u_n^{\mathsf{DG}}) - f_h'(u_n^{\mathsf{DG}}))(u_{n+1}^{\mathsf{DG}} - u_n^{\mathsf{DG}})\|_{0,\kappa}.$$

Remark 4.4. Deriving suitable local lower bounds on the second term arising within the *a posteriori* error estimator $\eta_{n,\kappa}$, $\kappa \in \mathcal{T}$, involving the gradient jump of the computed numerical solution, is technically more demanding in the current setting. In the case when $\epsilon = \mathcal{O}(1)$, then an analogous bound to the one derived in [35] may be established, subject to the addition of the corresponding linearisation terms; in the *hp*–setting, we note that this is suboptimal by a factor of $p_e^{\lambda+1/2}$. On the other hand, assuming that the polynomial degree is kept fixed, then employing cut-off functions on ω_e, of sufficiently small support, cf. [53], ϵ-robust h-version lower bounds may be deduced, cf. [56]. However, in the current *hp*–setting, we note that the inverse estimates required to establish *hp*–version ϵ-robust lower bounds for the gradient flux term are currently unavailable within the literature. Notwithstanding this issue, we shall demonstrate numerically in Section 6 that the upper *a posteriori* bound derived within this article is indeed sharp.

Furthermore, in contrast to the h–version approach in [5], we remark that the local efficiency bounds above are slightly suboptimally scaled with respect to the local polynomial degrees due to the need of applying p–dependent norm equivalence results (involving cut-off functions), cf. [42]. We remark that *hp*–version *a posteriori* error indicators, which are based on equilibrated flux reconstruction, may be shown to be robust with respect to the polynomial degree; see, for example, [24, 29]. In this latter approach, however, the resulting *a posteriori* error indicators are implicit in the sense that local problems posed on patches of elements must be numerically approximated in order to compute the elementwise error indicators. In the context of controlling both discretisation and linearisation error within this setting, we refer to the article [28].

5. *hp*–Adaptive NDG Scheme

In this section, we will discuss how the *a posteriori* bound from Theorem 4.2 can be exploited in the design of an *hp*–adaptive NDG algorithm for the numerical approximation of (1).

5.1. ***hp*–Adaptive Refinement Procedure.** In order to enrich the finite element space $\mathcal{V}_{\mathsf{DG}}$, we shall apply an *hp*–adaptive refinement algorithm which is based on the following two ingredients:

(a) Element marking: Each element κ in the computational mesh $\mathcal{T}$ may be marked for refinement on the basis of the size of the local residual indicators $\eta_{n,\kappa}$, cf. (34), $n \geq 0$. To this end, several strategies, such as equidistribution, fixed fraction, Dörfler marking, optimized mesh criterion, and so on, cf. [36], for example, have been proposed within the literature. For the purposes of this article, we employ the *maximal strategy*: here, we refine the set of elements $\kappa \in \mathcal{T}$ which satisfy the condition

$$\eta_{n,\kappa} > \Upsilon \max_{\kappa \in \mathcal{T}} \eta_{n,\kappa},$$

where $0 < \Upsilon < 1$ is a given parameter. On the basis of [22, 39, 50], throughout this article, we set $\Upsilon = 1/3$.

(b) hp–Refinement criterion: Once an element $\kappa \in \mathcal{T}$ has been marked for refinement, a decision must be made regarding whether to subdivide the element (*h*–refinement) or to increase the local degree of the polynomial approximation on element κ (*p*–refinement). Several strategies have been proposed within the literature; for a review of *hp*–refinement algorithms, we refer to [43]. Here we employ the *hp*–refinement strategy developed in [37] where the local regularity of the analytical solution is estimated on the basis of truncated local Legendre expansions of the computed numerical solution, cf., also, [26, 30].

5.2. **Fully Adaptive Newton-Galerkin Method.** We now propose a procedure that provides an *interplay* of the Newton linearisation and automatic *hp*–finite element mesh refinements based on the *a posteriori* residual estimate from Theorem 4.2 (as outlined in the previous Section 5.1). To this end, we make the assumption that the NDG sequence $\left\{u_{n+1}^{\mathsf{DG}}\right\}_{n\geq 0}$ given by (19) is well-defined as long as the iterations are being performed.

Algorithm 5.1. Given a (coarse) starting mesh $\mathcal{T}$ in Ω, with an associated (low-order) polynomial degree distribution $\boldsymbol{p}$, and an initial guess $u_0^{\mathsf{DG}} \in \mathcal{V}_{\mathsf{DG}}$. Set $n \leftarrow 0$.

1: Determine the Newton step size parameter Δt_n based on u_n^{DG} by the adaptive procedure from Algorithm 2.1; the Newton-Raphson transform $\mathsf{NF}(u_n^{\mathsf{DG}})$ required for the computation of the step size parameter Δt_n is approximated using the *hp*–DG method on the current mesh.

2: Compute the DG solution $\widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}}$ from (19), and $u_{n+1}^{\mathsf{DG}} = \widehat{\mathfrak{u}}_{n+1}^{\mathsf{DG}} + (1 - \Delta t_n) u_n^{\mathsf{DG}}$. Furthermore, evaluate the corresponding residual indicators $\{\eta_{n,\kappa}\}_{\kappa\in\mathcal{T}}$, and $\delta_{n,\Omega}$ from (34) and (35)–(36), respectively.

3: **if**

$$\delta_{n,\Omega}^2 \leq \Lambda\, \eta_{n,\Omega}^2 \tag{40}$$

holds, for some given parameter $\Lambda > 0$, **then** *hp*–refine the space $\mathcal{V}_{\mathsf{DG}}$ adaptively based on the marking criterion and the *hp*–strategy outlined in Section 5.1; go

back to step (1:) with the new mesh $\mathcal{T}$ (and based on the previously computed solution u_{n+1}^{DG} interpolated on the refined mesh).

4: **else**, i.e., if (40) is not fulfilled, then set $n \leftarrow n+1$, and perform another Newton step by going back to (1:).

5: **end if**

Remark 5.2. We note that our computational experience suggests that the choice of the element marking strategy can directly affect the robustness of the NDG scheme, particularly, when the numerical solution is far away from a given solution. Indeed, it is essential to employ a marking scheme which adaptively adjusts the number of elements marked for refinement at each step of the adaptive process; algorithms such as the fixed fraction method which only mark a fixed percentage of elements at each refinement level can lead to slow convergence of the combined adaptive Newton-Galerkin approach.

Remark 5.3. The balancing criterion (40) ensures that the linearization error is bounded uniformly by the (global) discretization error. In view of the (local) lower residual bounds presented in Section 4.2.3, this allows us to estimate the linearization indicator $\delta_{n,\Omega}$ defined in (35) in terms of the residual bounds from Theorem 4.3 and the normal jumps of the discrete solution, cf. Remark 4.4. Furthermore, for Λ sufficiently small, the linearization errors arising in (38) and (39) could, at least in parts, be absorbed into the right-hand sides of these bounds; cf. [27, Theorem 4.8].

6. Numerical Experiments

In this section we present a series of numerical experiments to demonstrate the practical performance of the proposed hp–adaptive refinement strategy outlined in Algorithm 5.1. To this end, throughout this section we select $\tau = 0.1$, $\gamma = 0.5$, and $h^{\max} = 10^6$ in Algorithm 2.1, the penalty parameter $C_\sigma = 10$ and $\theta = 1$ (SIPG) in the interior penalty DG scheme (19), cf. (20), and $\Lambda = 0.5$ in Algorithm 5.1, cf. [5]. Throughout this section we shall compare the performance of the proposed hp–adaptive refinement strategy with the corresponding algorithm based on exploiting only local mesh subdivision, i.e., h–refinement. Furthermore, within each inner linear iteration, we employ the *direct* MUltifrontal Massively Parallel Solver (MUMPS) [1–3]; in particular, in Theorem 4.2, we do not take into account any linear algebra errors resulting from iterative solvers (cf., e.g., [27]).

Example 6.1. In this first example, we consider the Bratu problem

$$\epsilon\Delta u + \mathrm{e}^u = 0 \quad \text{in } (0,1)^2,$$

i.e., $f(u) = \mathrm{e}^u + u$, subject to homogeneous Dirichlet boundary conditions on $\partial\Omega$. Writing $\lambda = 1/\epsilon$, we recall that there exists a critical parameter value λ_c $(= 1/\epsilon_c)$, such that for $\lambda > \lambda_c$ $(\epsilon < \epsilon_c)$ the problem has no solution, for $\lambda = \lambda_c$ $(\epsilon = \epsilon_c)$ there exists exactly one solution, and for $\lambda < \lambda_c$ $(\epsilon > \epsilon_c)$ there are two solutions. In the one–dimensional setting, an analytical expression for λ_c is available, cf. [7, 13, 17]; for the two–dimensional case, calculations have revealed that $\lambda_c = 6.808124423$ $(\epsilon_c = 0.146883332)$ to 9 decimal places, see [17, 44, 45], and the references cited therein.

Following [44], we select the initial guess $u_0^{\mathsf{DG}} \in \mathcal{V}_{\mathsf{DG}}$ to be the L^2–projection of the function u_0 onto $\mathcal{V}_{\mathsf{DG}}$, where $u_0(x,y) = a\sin(\pi x)\sin(\pi y)$, and a is a given amplitude. Noting that the maximum amplitude of the critical solution computed

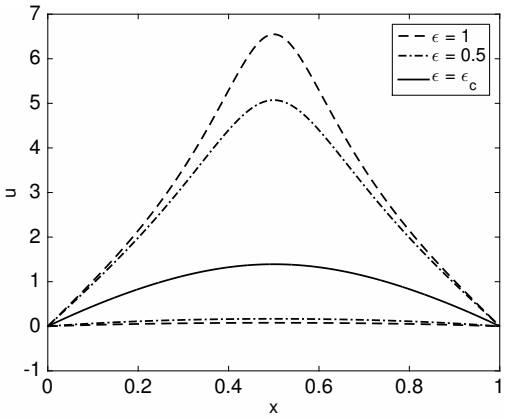
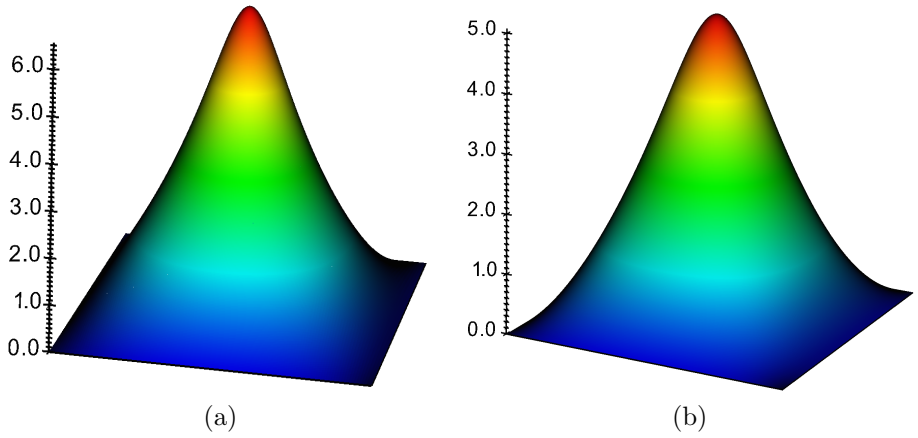


FIGURE 1. Bratu Problem. Slice at $y = 0.5$, $0 \leq x \leq 1$, of the upper and lower solutions computed with $\epsilon = 1$ and $\epsilon = 0.5$, together with the critical solution ($\epsilon = \epsilon_c$).



(a) (b)

FIGURE 2. Bratu Problem. Upper solution computed with: (a) $\epsilon = 1$; (b) $\epsilon = 0.5$.

with $\epsilon = \epsilon_c$ is approximately 1.39, selecting a to be smaller/larger than this value leads to convergence to the so–called lower/upper solution, respectively. With this in mind we select $a = 2$ when $\epsilon = \epsilon_c$, $a \in \{1/10, 6\}$ for $\epsilon = 1$, and $a \in \{1, 4\}$ for $\epsilon = 1/2$; in the latter two cases the smaller value of a is employed for the computation of the lower solution, while the larger value ensures convergence to the upper solution. In Figure 1 we plot a slice of each of the computed numerical solutions at $y = 0.5$, $0 \leq x \leq 1$. Here, we observe that the lower solutions tend to be rather flat in profile, while the upper solutions have a stronger peak in the middle of the computational domain, cf., also, Figure 2.

In Figure 3 we demonstrate the performance of the proposed *hp*–adaptive NDG algorithm, cf. Algorithm 5.1, for the computation of the lower and upper solutions when $\epsilon = 1$ and $\epsilon = 1/2$, as well as for the numerical approximation of

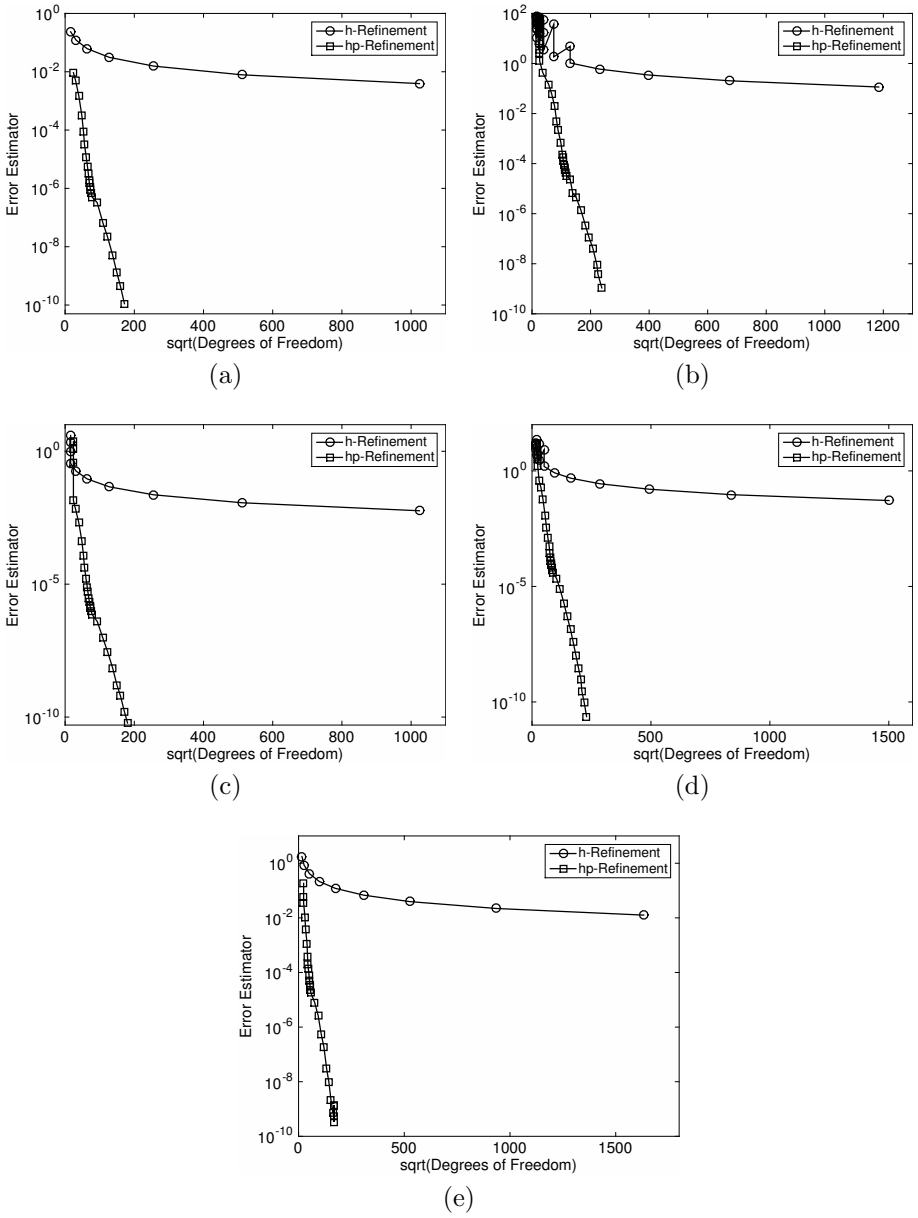


FIGURE 3. Bratu Problem. Comparison between h– and hp–refinement. (a) $\epsilon = 1$ (lower solution); (b) $\epsilon = 1$ (upper solution); (c) $\epsilon = 1/2$ (lower solution); (d) $\epsilon = 1/2$ (upper solution); (e) $\epsilon = \epsilon_c$ (critical solution);

the critical solution when $\epsilon = \epsilon_c$. In each case we plot the residual estimator $\mathsf{E} = \mathsf{E}(u_n^{\mathsf{DG}}, u_{n+1}^{\mathsf{DG}}, \boldsymbol{h}, \boldsymbol{p})$ from Theorem 4.2 (with the constant C set to 1) versus the square root of the number of degrees of freedom in the finite element space $\mathcal{V}_{\mathsf{DG}}$, based on employing both h– and hp–refinement. For each parameter value

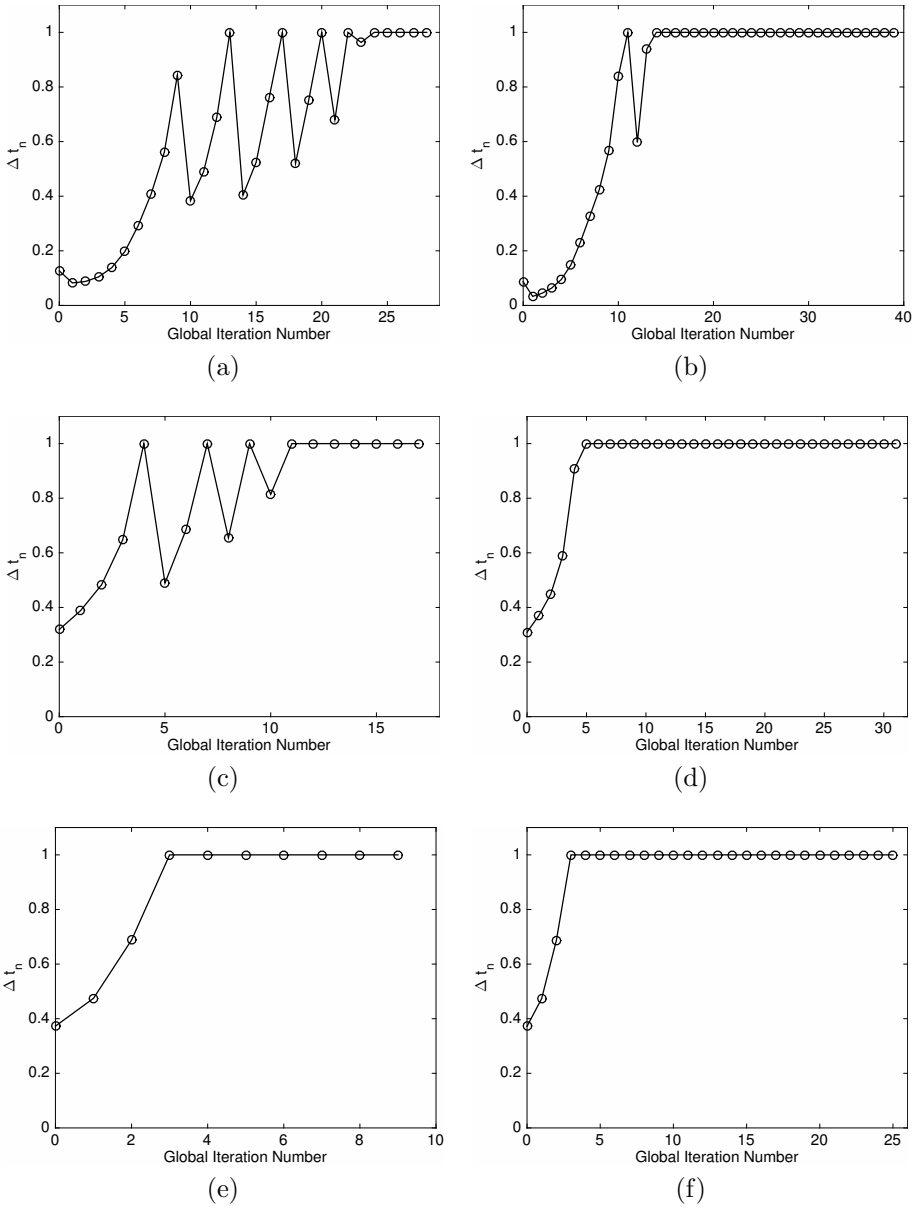


FIGURE 4. Bratu Problem. Damping parameter Δt_n. Left: h–refinement; right: hp–refinement. (a) & (b) $\epsilon = 1$ (upper solution); (c) & (d) $\epsilon = 1/2$ (upper solution); (e) & (f) $\epsilon = 1/2$ (lower solution).

we observe that the hp–refinement algorithm leads to an exponential decay of the residual estimator E as the finite element space $\mathcal{V}_{\mathsf{DG}}$ is adaptively enriched: on a linear-log plot, the convergence lines are roughly straight. Moreover, we observe the superiority of hp–refinement in comparison with a standard h–refinement algorithm, in the sense that the former refinement strategy leads to several orders

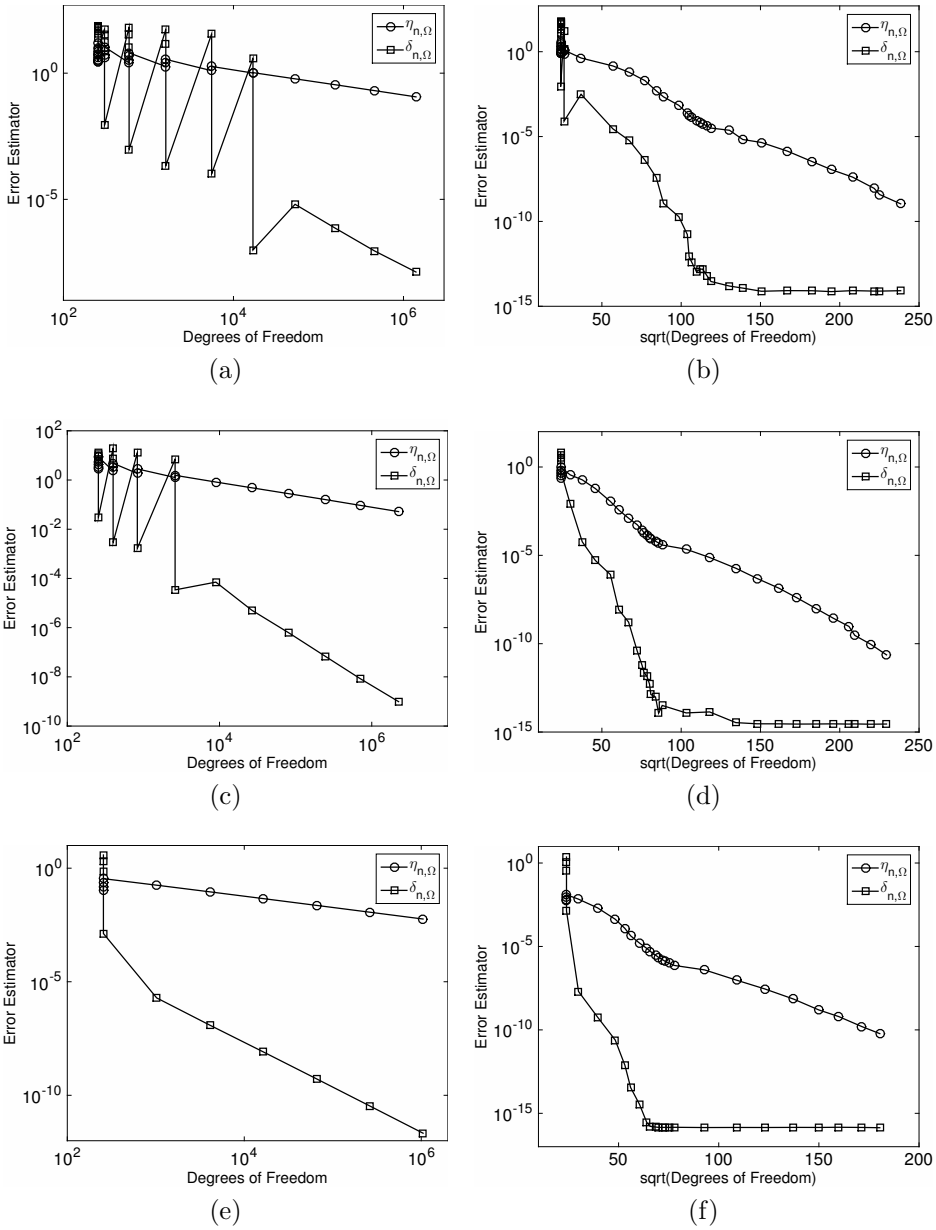


FIGURE 5. Bratu Problem. Individual error indicators $\eta_{n,\Omega}$ and $\delta_{n,\Omega}$. Left: h–refinement; right: hp–refinement. (a) & (b) $\epsilon = 1$ (upper solution); (c) & (d) $\epsilon = 1/2$ (upper solution); (e) & (f) $\epsilon = 1/2$ (lower solution).

of magnitude reduction in E, for a given number of degrees of freedom, than the corresponding quantity computed exploiting mesh subdivision only.

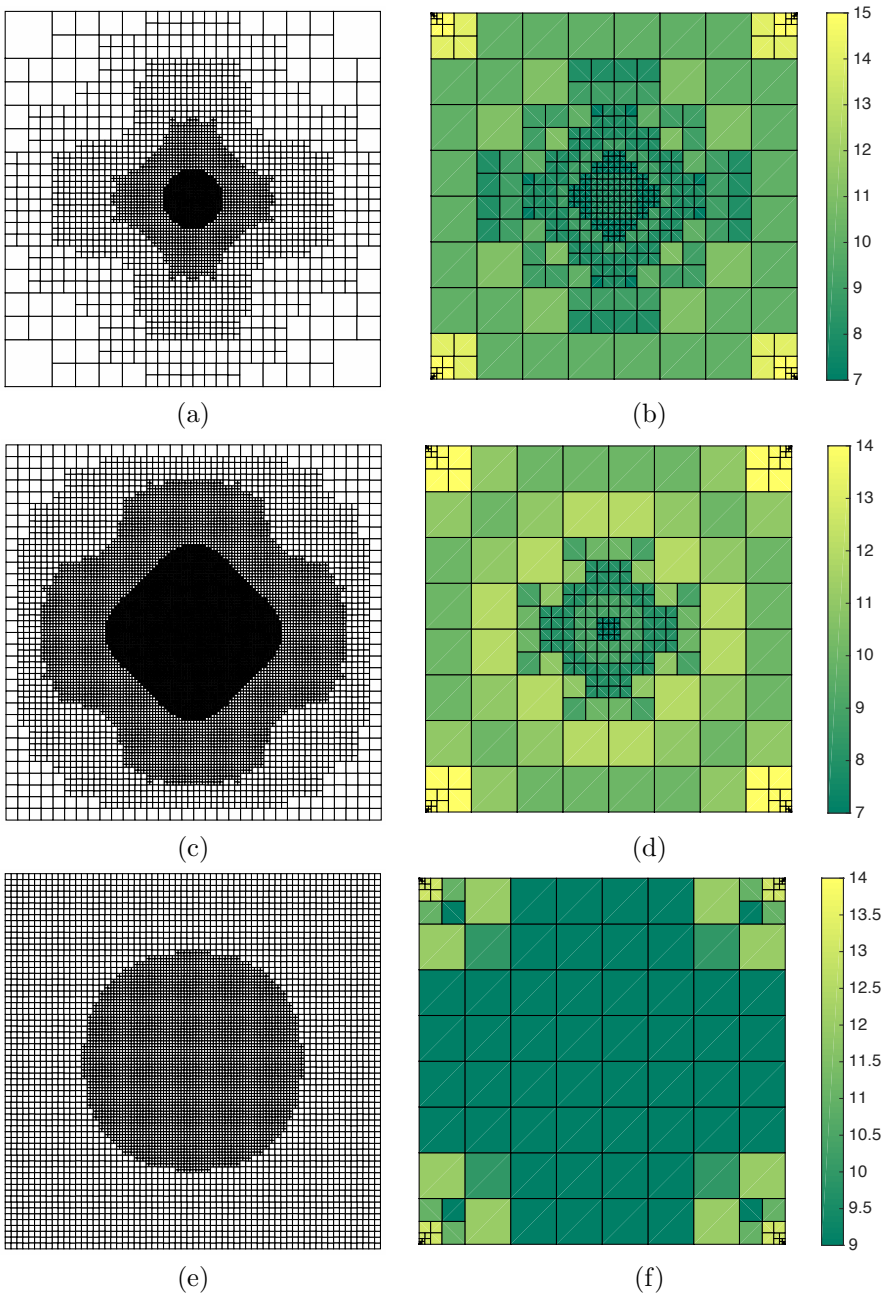


FIGURE 6. Bratu Problem. Computational meshes. Left: h–refinement; right: hp–refinement. (a) & (b) Upper solution computed with $\epsilon = 1$; (c) & (d) Upper solution computed with $\epsilon = 0.5$. (e) & (f) Critical solution.

In Figure 4 we plot the size of the Newton damping Δt_n versus the global iteration number. In many of the cases considered here $\Delta t_n = 1$ at all steps; for

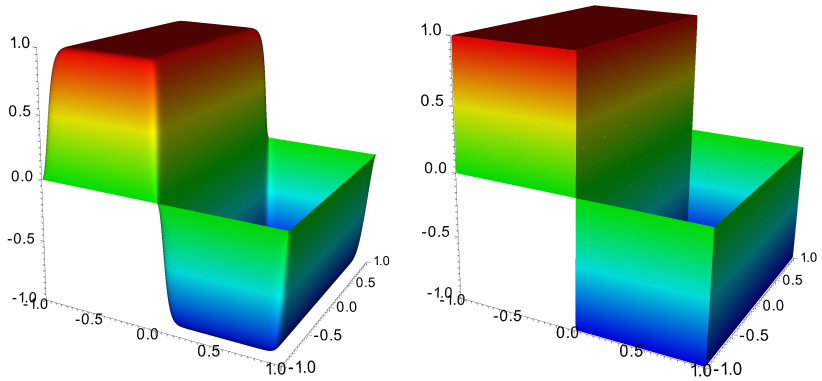


FIGURE 7. Ginzburg-Landau equation. Solution computed with: (a) $\epsilon = 10^{-3}$; (b) $\epsilon = 10^{-6}$.

brevity, these results have been omitted. For the cases presented in Figure 4, we observe that initially the damping parameter slowly increases when we are far away from the solution; once the damping parameter is close to unity, the condition

$$\delta_{n,\Omega}^2 \leq \Lambda\, \eta_{n,\Omega}^2$$

in Algorithm 5.1 becomes fulfilled in which case the finite element space $\mathcal{V}_{\mathsf{DG}}$ is adaptively enriched. In some cases, particularly at the early stages of the algorithm, refinement of $\mathcal{V}_{\mathsf{DG}}$ may then lead to a reduction in Δt_n, in which case further Newton steps are required before the next refinement can be undertaken. As the iterates approach the solution more closely, the size of the damping parameter typically remains approximately 1. To provide further detail concerning the performance of the proposed adaptive refinement strategy, for the cases depicted in Figure 4, in Figure 5 we plot the individual error indicators $\eta_{n,\Omega}$ and $\delta_{n,\Omega}$ in both the h– and hp–cases. Here, we clearly observe that, initially, the linearisation estimator $\delta_{n,\Omega}$ dominates the discretization error estimator $\eta_{n,\Omega}$, in which case further Newton steps are required; however, as the adaptive algorithm proceeds, the *a posteriori* error estimator E is dominated by the size of $\eta_{n,\Omega}$.

Finally, in Figure 6 we show the h– and hp–refined meshes generated for the numerical approximation of the upper solutions when $\epsilon = 1$ and $\epsilon = 1/2$, as well as for the critical solution. Here we observe that when h–refinement is employed, the mesh is concentrated in the vicinity of the peak in the solution located at the centre of the computational domain, cf. Figures 1 & 2. In the hp–setting, we observe that while some mesh refinement has been undertaken in the centre of the domain Ω, the corners of Ω have been significantly refined in order to resolve corner singularities typical for elliptic problems. Moreover, p–enrichement has been employed both in these corner regions, as well as in the vicinity of the peak in the computed solution. The corresponding meshes for the lower solutions are largely uniformly refined, due to the flat nature of the solution; for brevity, these have been omitted.

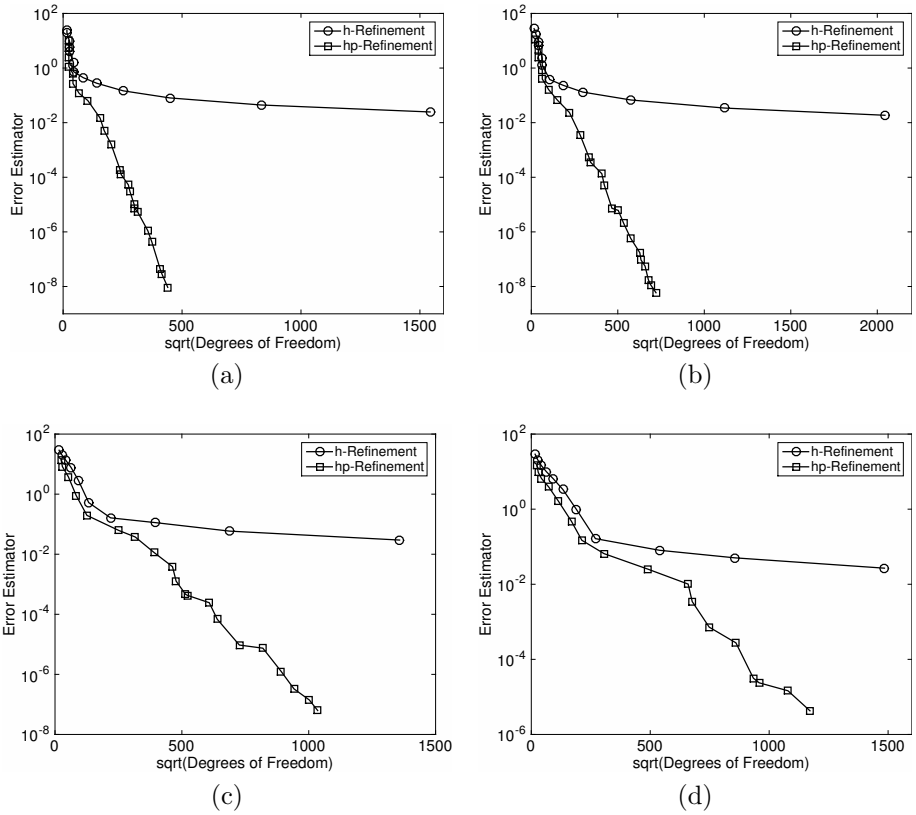


FIGURE 8. Ginzburg-Landau equation. Comparison between h– and hp–refinement. (a) $\epsilon = 10^{-3}$; (b) $\epsilon = 10^{-4}$; (c) $\epsilon = 10^{-5}$; (d) $\epsilon = 10^{-6}$.

Example 6.2. In this example, we consider the Ginzburg-Landau equation given by

$$-\epsilon\Delta u + u = u(2 - u^2) \quad \text{in } (-1,1)^2,$$

subject to homogeneous Dirichlet boundary conditions on $\partial\Omega$. Following [5], we first note that $u \equiv 0$ is a solution; moreover, any solution u appears in a pairwise fashion as $-u$. In the absence of boundary conditions, it is clear that $u = \pm 1$ are solutions of the Ginzburg-Landau equation. Thereby, in the presence of homogeneous Dirichlet boundary conditions, boundary layers will arise in the vicinity of $\partial\Omega$, whose width will be governed by the size of the diffusion coefficient ϵ. Here, we select the initial guess $u_0^{\mathsf{DG}} \in \mathcal{V}_{\mathsf{DG}}$ to be the L^2–projection of the function $u_0(x,y) = -\text{sgn}(x)$ onto $\mathcal{V}_{\mathsf{DG}}$, subject to the enforcement of the boundary conditions. In this case the solution to the Ginzburg-Landau equation will possess not only boundary layers, but also an internal layer along $x = 0$; in Figure 7 we plot the solution computed with both $\epsilon = 10^{-3}$ and $\epsilon = 10^{-6}$.

In Figure 8 we demonstrate the performance of the proposed hp–adaptive NDG algorithm, cf. Algorithm 5.1, for the computation of the solution to the Ginzburg-Landau equation when $\epsilon = 10^{-3},\ 10^{-4},\ 10^{-5},\ 10^{-6}$. In each case we plot the residual estimator E versus the square root of the number of degrees of freedom

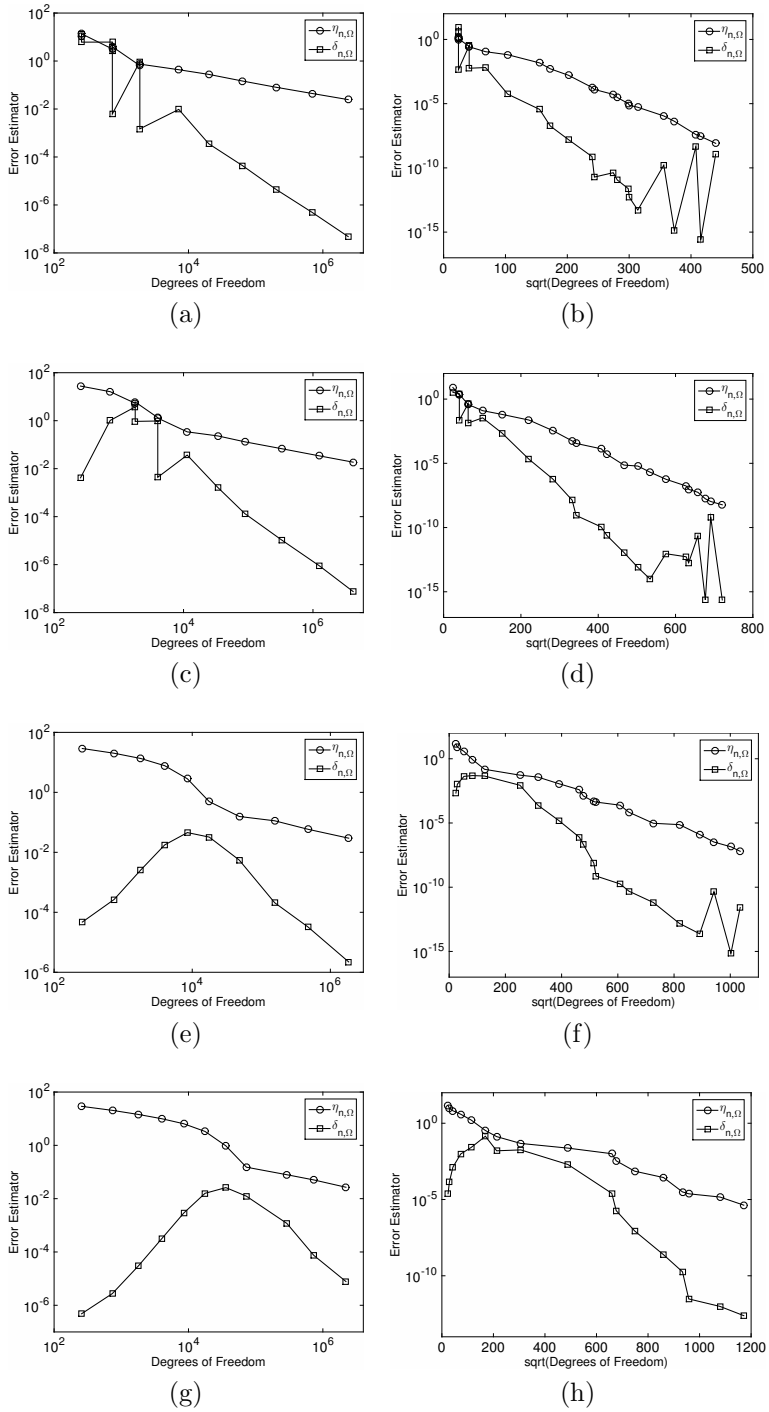
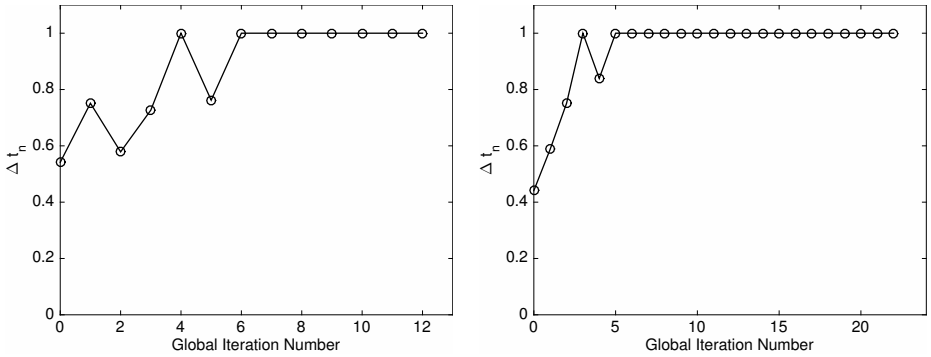


FIGURE 9. Ginzburg-Landau equation. Individual error indicators $\eta_{n,\Omega}$ and $\delta_{n,\Omega}$. Left: h–refinement; right: hp–refinement. (a) & (b) $\epsilon = 10^{-3}$; (c) & (d) $\epsilon = 10^{-4}$; (e) & (f) $\epsilon = 10^{-5}$; (g) & (h) $\epsilon = 10^{-6}$.



(a) (b)

FIGURE 10. Ginzburg-Landau equation. Damping parameter Δt_n for $\epsilon = 10^{-3}$. (a) h–refinement; (b) hp–refinement.

in the finite element space $\mathcal{V}_{\mathsf{DG}}$, based on employing both h– and hp–refinement. For each value of ϵ we again observe that the hp–refinement algorithm leads to an exponential decay of the residual estimator E as the finite element space $\mathcal{V}_{\mathsf{DG}}$ is adaptively enriched. Moreover, we again observe the superiority of exploiting hp–refinement in comparison with a standard h–refinement algorithm, in the sense that the former refinement strategy leads to several orders of magnitude reduction in E, for a given number of degrees of freedom, than the corresponding quantity computed using h–refinement only. Furthermore, we note that as ϵ is reduced, additional h–enrichment of the computational mesh is required before p–refinement is employed. Indeed, for $\epsilon = 10^{-6}$ we observe that there is an initial transient, before the hp–version convergence line becomes straight and exponential convergence is observed. The magnitude of the individual error indicators $\eta_{n,\Omega}$ and $\delta_{n,\Omega}$ are shown in Figure 9; as in the previous example, we observe that the discretisation error indicator $\eta_{n,\Omega}$ typically dominates the linearisation error indicator $\delta_{n,\Omega}$ as the adaptive Newton iteration proceeds. However, we observe that for $\epsilon = 10^{-3}$, 10^{-4}, and 10^{-5}, in the case when hp–refinement is employed, $\delta_{n,\Omega}$ is not very monotonic in the latter stages of the adaptive algorithm.

In Figure 10 we plot Δt_n versus the global iteration number for $\epsilon = 10^{-3}$; for the other values of ϵ considered here, the damping parameter was close to one on all of the meshes considered. As in the previous example, we again see an initial increase in Δt_n as the adaptive Newton algorithm proceeds, before the underlying mesh is adaptively refined. Again, in the early stages of the algorithm, enrichment of $\mathcal{V}_{\mathsf{DG}}$ may lead to some additional damping, before Δt_n tends to one.

Finally, in Figure 11 we plot the corresponding h– and hp–meshes generated for $\epsilon = 10^{-3}$ and $\epsilon = 10^{-6}$. Here, we clearly observe that the boundary and internal layers present in the analytical solution are refined by our adaptive mesh adaptation strategy; in particular, we emphasise that the NDG iterates converge to a solution which features the same topology as the initial guess, and, hence, does not switch between various attractors (corresponding to different solutions; see, e.g, [5]). In the hp–setting, we see that once the h–mesh has been sufficiently refined, then p–enrichment is employed.

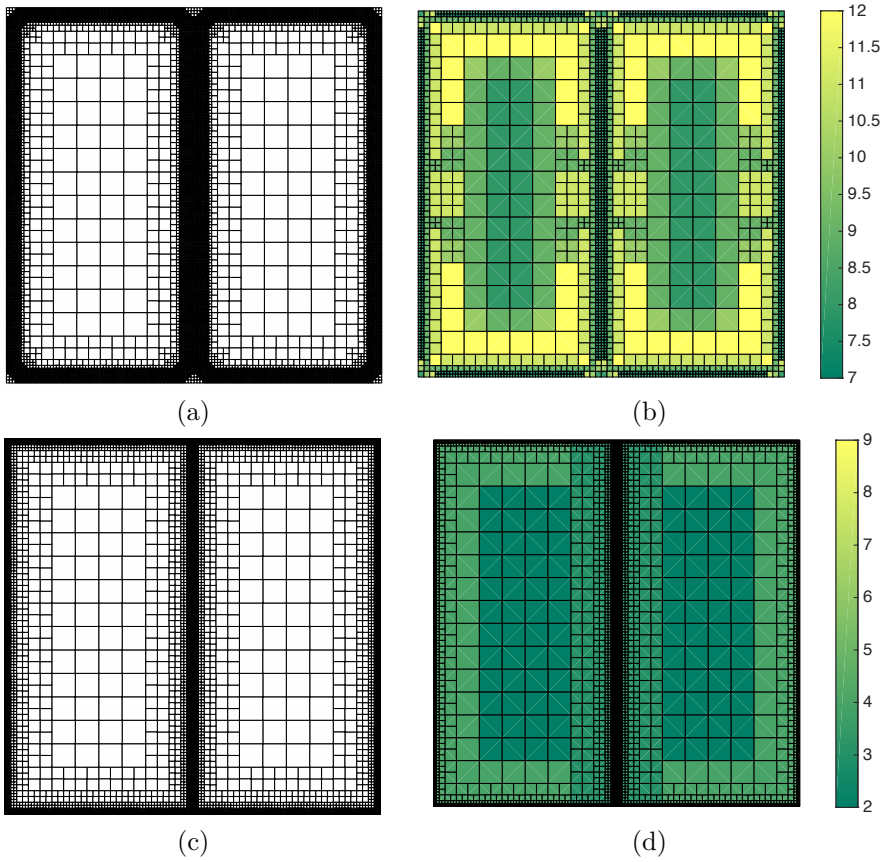


FIGURE 11. Ginzburg-Landau equation. Computational meshes. Left: *h*–refinement; right: *hp*–refinement. (a) & (b) $\epsilon = 10^{-3}$; (c) & (d) $\epsilon = 10^{-6}$.

Example 6.3. In order to test the computational efficiency and ϵ-robustness of the proposed *a posteriori* error bound stated in Theorem 4.2, in this final example we consider a variant of the Ginzburg-Landau equation which possesses a known analytical solution. More precisely, we consider the numerical approximation of

$$-\epsilon\Delta u + u = u(2-u^2) + \hat{f} \quad \text{in } (-1,1)^2, \tag{41}$$

subject to homogeneous Dirichlet boundary conditions on $\partial\Omega$, where, given the form of the computed solution depicted in Figure 7, we select $\hat{f}$ such that the analytical solution to (41) is given by

$$u = -16\left(\frac{\mathrm{e}^{-\frac{1}{\sqrt{\epsilon}(x+1)}} - \mathrm{e}^{\frac{1}{\sqrt{\epsilon}(x-1)}}}{\mathrm{e}^{-\frac{1}{\sqrt{\epsilon}(x+1)}} + \mathrm{e}^{\frac{1}{\sqrt{\epsilon}(x-1)}}}\right)\frac{\mathrm{e}^{\frac{2\sqrt{\epsilon}}{x^2-1}}}{\left(\mathrm{e}^{-\frac{\sqrt{\epsilon}}{x+1}} + \mathrm{e}^{\frac{\sqrt{\epsilon}}{x-1}}\right)^2}\frac{\mathrm{e}^{\frac{2\sqrt{\epsilon}}{y^2-1}}}{\left(\mathrm{e}^{-\frac{\sqrt{\epsilon}}{y+1}} + \mathrm{e}^{\frac{\sqrt{\epsilon}}{y-1}}\right)^2}.$$

Note that, for this choice, the source term $\hat{f}$ in (41) is negligible in most of the domain.

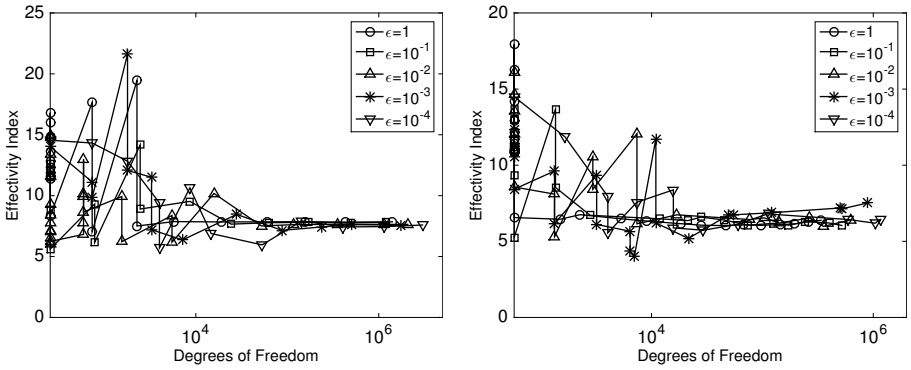


FIGURE 12. Approximate Ginzburg-Landau equation. Effectivity indices for $\epsilon = 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$. (a) h–refinement; (b) hp–refinement.

In Figure 12 we plot the effectivity indices (ratio of the *a posteriori* error bound and the true error, measured in the DG norm), for $\epsilon = 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$, based on employing both h– and hp–mesh refinement. Here, we observe that, in the initial stages of the adaptive process, the effectivity indices tend to oscillate between a value of around 5 and 20 as the damping parameter is continuously adjusted while the mesh is refined. However, as the adaptive strategy proceeds, and the damping parameter tends to unity, the effectivity indices tend to a value of around 7, uniformly with respect to ϵ. The performance of the underlying h– and hp–refinement strategies are quantitatively similar to the results presented in the previous example; thereby, for the sake of brevity these results are omitted.

7. Concluding remarks

In this article we have introduced the hp–version of the NDG scheme for the numerical approximation of second-order, singularly perturbed, semilinear elliptic boundary value problems. Here, the general approach is based on first linearising the underlying PDE problem on a continuous level, followed by subsequent discretisation of the resulting sequence of linear PDEs. For this latter task, in the current article we have exploited the hp–version of the interior penalty DG method. Furthermore, we have derived an ϵ-robust *a posteriori* bound which takes into account both the linearisation and discretisation errors. On the basis of this residual estimate, we have designed and implemented an hp–adaptive refinement algorithm which automatically controls both of these sources of error; the practical performance of this strategy has been studied for a series of numerical test problems. Future work will be devoted to the extension of this technique to more general nonlinear PDE problems, as well as to problems in three dimensions.

Acknowledgment. The authors would like to thank the anonymous referee for bringing Example 6.3 to their attention.

References

1. P.R. Amestoy, I.S. Duff, J. Koster, and J.-Y. L'Excellent, *A fully asynchronous multifrontal solver using distributed dynamic scheduling*, SIAM J. Mat. Anal. Appl. **23** (2001), no. 1, 15–41.
2. P.R. Amestoy, I.S. Duff, and J.-Y. L'Excellent, *Multifrontal parallel distributed symmetricand unsymmetric solvers*, Comput. Methods Appl. Mech. Eng. **184** (2000), 501–520.
3. P.R. Amestoy, A. Guermouche, J.-Y. L'Excellent, and S. Pralet, *Hybrid scheduling for the parallel solution of linear systems*, Parallel Computing **32** (2006), no. 2, 136–156.
4. M. Amrein and T.P. Wihler, *An adaptive Newton-method based on a dynamical systems approach*, Commun. Nonlinear Sci. Numer. Simul. **19** (2014), no. 9, 2958–2973.
5. ———, *Fully adaptive Newton-Galerkin methods for semilinear elliptic partial differential equations*, SIAM J. Sci. Comput. **37** (2015), no. 4, A1637–A1657.
6. D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal. **39** (2001), 1749–1779.
7. U.M. Ascher, M.M. Mattheij, and R.D. Russell, *Numerical solution of boundary value problems for ordinary differential equation*, SIAM, Philadelphia, PA, 1995.
8. G. Barles and J. Burdeau, *The Dirichlet problem for semilinear second-order degenerate elliptic equations and applications to stochastic exit time control problems*, Comm. Partial Differential Equations **20** (1995), no. 1-2, 129–178.
9. A. Barone, F. Esposito, C.J. Magee, and A.C. Scott, *Theory and applications of the Sine-Gordon equation*, Riv. Nuovo Cim. **1** (1971), 227–267.
10. H. Berestycki and P.-L. Lions, *Nonlinear scalar field equations. I. Existence of a ground state*, Arch. Rational Mech. Anal. **82** (1983), no. 4, 313–345.
11. C. Bernardi, J. Dakroub, G. Mansour, and T. Sayah, *A posteriori analysis of iterative algorithms for a nonlinear problem*, J. Sci. Comput. **65** (2015), no. 2, 672–697.
12. A. Borisyuk, B. Ermentrout, A. Friedman, and D. Terman, *Tutorials in mathematical biosciences. I*, Lecture Notes in Mathematics, vol. 1860, Springer-Verlag, Berlin, 2005, Mathematical neuroscience, Mathematical Biosciences Subseries.
13. D. Calvetti and L. Reichel, *Iterative methods for large continuation problems*, J. Comput. Appl. Math. **123** (2000), 217–240.
14. R.S. Cantrell and C. Cosner, *Spatial ecology via reaction-diffusion equations*, Wiley Series in Mathematical and Computational Biology, John Wiley & Sons, Ltd., Chichester, 2003.
15. A.L. Chaillou and M. Suri, *Computable error estimators for the approximation of nonlinear problems by linearized models*, Comput. Methods Appl. Mech. Engrg. **196** (2006), no. 1-3, 210–224.
16. ———, *A posteriori estimation of the linearization error for strongly monotone nonlinear operators*, J. Comput. Appl. Math. **205** (2007), no. 1, 72–87.
17. K.A. Cliffe, E. Hall, P. Houston, E.T. Phipps, and A.G. Salinger, *Adaptivity and a posteriori error control for bifurcation problems I: The Bratu problem*, Commun. Comput. Phys. **8** (2010), 845–865.
18. S. Congreve and P. Houston, *Two-grid hp-version discontinuous Galerkin finite element methods for quasi-Newtonian fluid flows*, Int. J. Numer. Anal. Model. **11** (2014), no. 3, 496–524.
19. S. Congreve, P. Houston, E. Süli, and T.P. Wihler, *Discontinuous Galerkin finite element approximation of quasilinear elliptic boundary value problems II: strongly monotone quasi-Newtonian flows*, IMA J. Numer. Anal. **33** (2013), no. 4, 1386–1415.
20. S. Congreve, P. Houston, and T.P. Wihler, *Two-grid hp-version discontinuous Galerkin finite element methods for second-order quasilinear elliptic PDEs*, J. Sci. Comput. **55** (2013), no. 2, 471–497.
21. S. Congreve and T.P. Wihler, *An iterative finite element method for strongly monotone quasilinear diffusion-reaction problems*, in preparation, 2015.
22. L. Demkowicz, *Computing with hp-adaptive finite elements. Vol. 1*, Chapman & Hall/CRC Applied Mathematics and Nonlinear Science Series, Chapman & Hall/CRC, Boca Raton, FL, 2007, One and two dimensional elliptic and Maxwell problems.
23. P. Deuflhard, *Newton methods for nonlinear problems*, Springer Series in Computational Mathematics, vol. 35, Springer-Verlag, Berlin, 2004, Affine invariance and adaptive algorithms.

24. V. Dolejší, Alexandre Ern, and Martin Vohralík, *hp-adaptation driven by polynomial-degree-robust a posteriori error estimates for elliptic problems*, SIAM J. Sci. Comput. **38** (2016), no. 5, A3220–A3246.
25. L. Edelstein-Keshet, *Mathematical models in biology*, Classics in Applied Mathematics, vol. 46, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005, Reprint of the 1988 original.
26. T. Eibner and J. M. Melenk, *An adaptive strategy for hp-FEM based on testing for analyticity*, Comput. Mech. **39** (2007), no. 5, 575–595.
27. L. El Alaoui, A. Ern, and M. Vohralík, *Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems*, Comput. Methods Appl. Mech. Engrg. **200** (2011), no. 37-40, 2782–2795.
28. A. Ern and M. Vohralík, *Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs*, SIAM J. Sci. Comput. **35** (2013), no. 4, A1761–A1791.
29. A. Ern and M. Vohralík, *Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations*, SIAM J. Numer. Anal. **53** (2015), no. 2, 1058–1081.
30. T. Fankhauser, T.P. Wihler, and M. Wirz, *The hp-adaptive FEM based on continuous Sobolev embeddings: isotropic refinements*, Comp. Math. Appl. **67** (2014), no. 4, 854–868.
31. E.M. Garau, P. Morin, and C. Zuppa, *Convergence of an adaptive Kačanov FEM for quasi-linear problems*, Appl. Numer. Math. **61** (2011), no. 4, 512–529.
32. J.D. Gibbon, I.N. James, and I.M. Moroz, *The Sine-Gordon equation as a model for a rapidly rotating baroclinic fluid*, Phys. Script. **20** (1979), 402–408.
33. W. Han, *A posteriori error analysis for linearization of nonlinear elliptic problems and their discretizations*, Math. Method Appl. Sci. **17** (1994), no. 7, 487–508.
34. P. Houston, D. Schötzau, and T.P. Wihler, *An hp-adaptive mixed discontinuous Galerkin FEM for nearly incompressible linear elasticity*, Comput. Methods Appl. Mech. Engrg. **195** (2006), no. 25-28, 3224–3246.
35. ———, *Energy norm a posteriori error estimation of hp-adaptive discontinuous Galerkin methods for elliptic problems*, Math. Models Methods Appl. Sci. **17** (2007), no. 1, 33–62.
36. P. Houston and E. Süli, *Adaptive finite element approximation of hyperbolic problems*, Error Estimation and Adaptive Discretization Methods in Computational Fluid Dynamics. Lect. Notes Comput. Sci. Engrg. (T. Barth and H. Deconinck, eds.), vol. 25, Springer, 2002, pp. 269–344.
37. ———, *A note on the design of hp-adaptive finite element methods for elliptic partial differential equations*, Comput. Methods Appl. Mech. Engrg. **194** (2005), no. 2-5, 229–243.
38. P. Houston, E. Süli, and T.P. Wihler, *A posteriori error analysis of hp-version discontinuous Galerkin finite-element methods for second-order quasi-linear elliptic PDEs*, IMA J. Numer. Anal. **28** (2008), no. 2, 245–273.
39. P. Houston and T.P. Wihler, *Adaptive energy minimisation for hp-finite element methods*, Comput. Math. Appl. **71** (2016), no. 4, 977 – 990.
40. O.A. Karakashian and F. Pascal, *A posteriori error estimation for a discontinuous Galerkin approximation of second order elliptic problems*, SIAM J. Numer. Anal. **41** (2003), 2374–2399.
41. M. Karkulik and J.M. Melenk, *Local high-order regularization and applications to hp-methods*, Comp. Math. Appl. **70** (2015), no. 7, 1606–1639.
42. J. M. Melenk and B. I. Wohlmuth, *On residual-based a posteriori error estimation in hp-FEM*, Adv. Comput. Math. **15** (2001), no. 1-4, 311–331, A posteriori error estimation and adaptive computational methods.
43. W.F. Mitchell and M.A. McClain, *A comparison of hp-adaptive strategies for elliptic partial differential equations*, ACM. Transactions on Mathematical Software **41** (2014), no. 1, 2:1–39.
44. A. Mohsen, *A simple solution of the Bratu problem*, Comput. Math. Appl. **67** (2014), 26–33.
45. A. Mohsen, L.F. Sedeek, and S.A. Mohamed, *New smoother to enhance multigrid–based methods for Bratu problem*, Appl. Math. Comput. **204** (2008), 325–339.
46. W.-M. Ni, *The mathematics of diffusion*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 82, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.
47. A. Okubo and S.A. Levin, *Diffusion and ecological problems: modern perspectives*, second ed., Interdisciplinary Applied Mathematics, vol. 14, Springer-Verlag, New York, 2001.

48. H.R. Schneebeli and T.P. Wihler, *The Newton-Raphson method and adaptive ODE solvers*, Fractals. Complex Geometry, Patterns, and Scaling in Nature and Society **19** (2011), no. 1, 87–99.
49. D. Schötzau, C. Schwab, and A. Toselli, *Mixed hp-DGFEM for incompressible flows*, SIAM J. Numer. Anal. **40** (2002), no. 6, 2171–2194 (electronic) (2003).
50. P. Solin, K. Segeth, and I. Dolezel, *Higher-order finite element methods*, Studies in advanced mathematics, Chapman & Hall/CRC, Boca Raton, London, 2004.
51. B. Stamm and T.P. Wihler, *hp-Optimal discontinuous Galerkin methods for linear elliptic problems*, Math. Comp. **79** (2010), no. 272, 2117–2133.
52. W.A. Strauss, *Existence of solitary waves in higher dimensions*, Comm. Math. Phys. **55** (1977), no. 2, 149–162.
53. R. Verfürth, *Robust a posteriori error estimators for a singularly perturbed reaction-diffusion equation*, Numer. Math. **78** (1998), no. 3, 479–493.
54. T.P. Wihler, P. Frauenfelder, and C. Schwab, *Exponential convergence of the hp-DGFEM for diffusion problems*, Comp. Math. Appl. **46** (2003), no. 1, 183–205.
55. L. Zhu, S. Giani, P. Houston, and D. Schötzau, *Energy norm a posteriori error estimation for hp-adaptive discontinuous Galerkin methods for elliptic problems in three dimensions*, Math. Models Methods Appl. Sci. **21** (2011), no. 2, 267–306.
56. L. Zhu and D. Schötzau, *A robust* a posteriori *error estimate for hp-adaptive DG methods for convection-diffusion equations*, IMA J. Numer. Anal. **31** (2011), 971–1005.