Communications on Applied Mathematics and Computation
https://doi.org/10.1007/s42967-021-00133-6

ORIGINAL PAPER

Discontinuous Galerkin Methods for a Class of Nonvariational Problems

Andreas Dedner[1] · **Tristan Pryer**[2]

Received: 22 September 2020 / Revised: 1 March 2021 / Accepted: 22 March 2021

Abstract
We extend the finite element method introduced by Lakkis and Pryer (SIAM J. Sci. Comput. 33(2): 786–801, 2011) to approximate the solution of second-order elliptic problems in nonvariational form to incorporate the discontinuous Galerkin (DG) framework. This is done by viewing the "finite element Hessian" as an auxiliary variable in the formulation. Representing the finite element Hessian in a discontinuous setting yields a linear system of the same size and having the same sparsity pattern of the compact DG methods for variational elliptic problems. Furthermore, the system matrix is very easy to assemble; thus, this approach greatly reduces the computational complexity of the discretisation compared to the continuous approach. We conduct a stability and consistency analysis making use of the unified frameworkset out in Arnold et al. (SIAM J. Numer. Anal. 39(5): 1749–1779, 2001/2002). We also give an a posteriori analysis of the method in the case where the problem has a strong solution. The analysis applies to any consistent representation of the finite element Hessian, and thus is applicable to the previous works making use of continuous Galerkin approximations. Numerical evidence is presented showing that the method works well also in a more general setting.

Keywords Nonvariational problems · Discontinuous Galerkin · Error estimates · Adaptivity

Mathematics Subject Classification 65N15 · 65N30 · 65Y20 · 65N12

1 Introduction

Linear, second-order, nonvariational partial differential equations (PDEs) are those which are given in the form

✉ Andreas Dedner
A.S.Dedner@warwick.ac.uk

Tristan Pryer
tmp38@bath.ac.uk

[1] Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

[2] Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

$$-\boldsymbol{A} : \mathrm{D}^2 u = f, \tag{1}$$

where $\boldsymbol{X} : \boldsymbol{Y} = \operatorname{trace}\left(\boldsymbol{X}^{\intercal}\boldsymbol{Y}\right)$ is the Frobenious inner product between matrices. If the matrix $\boldsymbol{A}$ is differentiable, then there is an equivalence between this problem and its variational sibling

$$-\operatorname{div}(\boldsymbol{A}\nabla u) + \mathrm{D}\boldsymbol{A}\nabla u = f, \tag{2}$$

where

$$\mathrm{D}\boldsymbol{A} = \left(\sum_{i=1}^{d} \partial_i a_{i,1}(\boldsymbol{x}), \cdots, \sum_{i=1}^{d} \partial_i a_{i,d}(\boldsymbol{x}) \right), \tag{3}$$

and d is the dimension under consideration. Rewriting in this form is sometimes undesirable. For example, if the coefficient matrix $\boldsymbol{A}$ has near singular derivatives, the problem will become advection dominated and possibly unstable for conforming finite element methods. There is a wealth of material on the treatment of advection-dominated problems [c.f., 18, 19]. If $\boldsymbol{A}$ is not differentiable, then the problem has no variational structure. In this case, standard finite element methods cannot be applied.

In a previous work [32], a finite element method for the approximation of the nonvariational problem (1) was introduced. This involved the introduction of a *finite element Hessian* represented in the same finite element space as the solution (modulo boundary conditions). The applications of the discrete representation of a Hessian of a piecewise function are becoming broader, for example, it can be used to drive anisotropic adaptive algorithms [1, 40], as a notion of discrete convexity [2] and in the design of finite element methods for nonlinear fourth-order problems [37].

The algebraic formulation of the C^0 Galerkin approximation of the nonvariational problem requires the solution of large sparse $(d+1)^2N^2$ linear system [32, Lem 3.3], where N is the number of degrees of freedom. Equivalently, using a Schur complement argument, this can be reduced to an N^2 full linear system. The reason that this system is full is due to the global nature of the $\mathrm{L}^2(\Omega)$ projection operator into a continuous finite element space. The motivation for extending the nonvariational finite element method into the discontinuous setting is the massive gain in computational efficiency over the continuous case. Indeed, due to the local representation of the projection operators in these discontinuous spaces, we are able to make massive computational savings, in that the system matrix becomes sparse and is the same size as that of a standard discontinuous Galerkin stiffness matrix for the Laplacian.

There has been a plethera of other finite element methods posed for linear nonvariational PDEs exist including [38] in which the authors pose a stable discontinuous Galerkin scheme for linear nonvariational problems using discrete analogs of the classical methods used to prove existence of strong solutions, see also [29] where similar techniques are used over curved domains. In [22–24] the author considers a least squares formulation, see also [31, 35] for related formulations. Many of these formulations share the same disadvantage in the conditioning of the system matrix—the condition number scales like a fourth-order problem. As already mentioned, one of the desirable features of the formulation presented here is that it scales in the same fashion as a discontinuous Galerkin method for a second-order variational problem, a property shared by the method in [25] under the Cordes condition.

We are particularly interested in nonvariational PDEs due to their relation to general fully nonlinear PDEs

$$\mathscr{F}\left(\mathrm{D}^2 u\right) = 0, \tag{4}$$

which are of significant current research. In the literature finite element methods have been presented to solve this general class of problem. For example in [8] the author presents a C^1 finite element method shows stability and consistency (hence convergence) of the scheme which requires a high degree of smoothness on the exact solution. In [20, 21] the authors give a method in which they approximate the general second-order fully nonlinear PDE by a sequence of fourth-order quasilinear PDEs. This is reminiscent of the vanishing viscosity method introduced for classically studying first-order fully nonlinear PDEs. Efficiency of any method used to approximate a problem such as this is key. Each of the methods is computationally costly due to their reliance on C^1 finite elements [8, 21] or mixed methods [20].

In [5], a generic framework was set up to prove convergence of numerical approximations to the viscosity solutions of degenerate elliptic fully nonlinear PDEs. This involved constructing monotone sequences of approximations which are typically applied to finite difference approximations of the nonlinear problem [c.f., 36]. The assumption of consistency made in the [5] framework is incompatible with finite element methods; however, an extremely important observation made in [28] is that the consistency condition may be weakened to incorporate the finite element case using a localisation argument (in the case of isotropic diffusion).

A posteriori analysis of linear nonvariational problems is less standard than for those of variational type. Typically, results are based on "closeness" conditions of Cordes type [12] which guarantees existence of a unique smooth, H^2 solution. Under these assumptions, it is relatively straightforward to derive a posteriori bounds in the natural norm for the problem, H^2, or a mesh-dependent equivalent. For other methods, this has been done in [24]. It has even been shown that adaptive methods for these problems converge optimally in terms of the number of degrees of freedom [30].

We show similar a posteriori bounds that make it straightforward to incorporate the method into well-developed software package for finite element methods, shown here using Dune [6, 7]. In this work, we are interested in the asymptotic behaviour of the discontinuous approximation and the computational gains using this method. In a subsequent work, we will study the computational gains using the discontinuous framework presented over the continuous one given in [32], as well as exploit the powerful parallelisation capabilities of the package.

The rest of the paper is set out as follows. In Sect. 2, we formally introduce the model problem and give a brief review of known classical facts about nonvariational PDEs. In Sect. 3, we examine the discretisation of the nonvariational method in the discontinuous Galerkin framework, making use of the unified framework set out in [4] to derive a very general formulation of the finite element Hessian represented as a discontinuous object. We present some examples and examine the natural question of what happens when we try to eliminate the finite element Hessian from the formulation. In Sect. 4, we conduct an a posteriori analysis of the method and show upper and lower bounds to the error in an H^2-like mesh-dependent norm. Finally, in Sect. 5, we detail a summary of extensive numerical experiments aimed at examining convergence and robustness of the method presented.

2 Problem Formulation

In this section, we formulate the model problem, fix notation and give some basic assumptions. In addition, we review the existence and uniqueness of the nonvariational problems. Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, be a connected domain with polygonal boundary. The Lebesgue spaces are defined as

$$\mathrm{L}^2(\Omega) = \left\{ \phi : \int_\Omega |\phi(\boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x} < \infty \right\} \quad \text{and} \quad \mathrm{L}^\infty(\Omega) = \left\{ \phi : \sup_{\boldsymbol{x}\in\Omega} |\phi(\boldsymbol{x})| < \infty \right\}, \tag{5}$$

and the Sobolev and Hilbert spaces

$$\mathrm{W}^{k,p}(\Omega) = \{\phi \in \mathrm{L}^p(\Omega) : \mathrm{D}^{\boldsymbol{\alpha}}\phi \in \mathrm{L}^p(\Omega), \quad \text{for } |\boldsymbol{\alpha}| \leqslant k\} \quad \text{and} \quad \mathrm{H}^k(\Omega) := \mathrm{W}^{k,2}(\Omega). \tag{6}$$

These are equipped with the norms

$$\|\phi\|^2_{\mathrm{L}^2(\Omega)} = \int_\Omega |\phi|^2 \, \mathrm{d}\boldsymbol{x}, \qquad \|\phi\|_{\mathrm{L}^\infty(\Omega)} = \sup_{\boldsymbol{x}\in\Omega} |\phi(\boldsymbol{x})|, \tag{7}$$

$$\|v\|^p_{\mathrm{W}^{k,p}(\Omega)} = \sum_{|\boldsymbol{\alpha}|\leqslant k} \|\mathrm{D}^{\boldsymbol{\alpha}} v\|^p_{\mathrm{L}^p(\Omega)} \quad \text{and} \quad |v|^p_{\mathrm{W}^{k,p}(\Omega)} = \sum_{|\boldsymbol{\alpha}|=k} \|\mathrm{D}^{\boldsymbol{\alpha}} v\|^p_{\mathrm{L}^p(\Omega)}, \tag{8}$$

where $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_d\}$ is a multi-index, $|\boldsymbol{\alpha}| = \sum_{i=1}^d \alpha_i$ and derivatives $\mathrm{D}^{\boldsymbol{\alpha}}$ are understood in a weak sense. We pay particular attention to the cases $k = 1, 2$ and

$$\mathrm{H}^1_0(\Omega) := \text{closure of } C^\infty_0(\Omega) \text{ in } \mathrm{H}^1(\Omega). \tag{9}$$

The model problem in strong form is: find $u \in \mathrm{H}^2(\Omega) \cap \mathrm{H}^1_0(\Omega)$ such that

$$\langle \mathscr{L}u, \phi \rangle = \langle f, \phi \rangle, \qquad \forall\, \phi \in \mathrm{H}^1_0(\Omega), \tag{10}$$

where the data $f \in \mathrm{L}^2(\Omega)$ are prescribed and $\mathscr{L}$ is a general linear, second-order, uniformly elliptic partial differential operator. Let $\boldsymbol{A} \in \mathrm{L}^\infty(\Omega)^{d\times d}$, we then define

$$\begin{aligned} \mathscr{L} : \ \mathrm{H}^2(\Omega) \cap \mathrm{H}^1_0(\Omega) &\to \mathrm{L}^2(\Omega), \\ u &\mapsto \mathscr{L}u := -\boldsymbol{A} : \mathrm{D}^2 u. \end{aligned} \tag{11}$$

We assume that $\boldsymbol{A}$ is uniformly positive definite, i.e., there exists a $\gamma > 0$ such that for all $\boldsymbol{x}$

$$\boldsymbol{y}^{\mathsf{T}} \boldsymbol{A}(\boldsymbol{x}) \boldsymbol{y} \geqslant \gamma |\boldsymbol{y}|^2, \quad \forall\, \boldsymbol{y} \in \mathbb{R}^d, \tag{12}$$

and we call γ the *ellipticity constant*.

Definition 1 (Strong solution) A *strong solution* of (1) is a function $u \in \mathrm{H}^2(\Omega) \cap \mathrm{H}^1_0(\Omega)$, that is a twice weakly differentiable function, which satisfies the problem almost everywhere.

Theorem 1 (Existence and regularity of a strong solution of (1) [12]) *Let $\Omega \subset \mathbb{R}^d$ be a convex polytope. Suppose $\boldsymbol{A} \in \mathrm{L}^\infty(\Omega)^{d\times d}$ is uniformly elliptic and $f \in \mathrm{L}^2(\Omega)$. Suppose further that $\boldsymbol{A}$ satisfies the Cordes condition, that there exists $0 < \alpha \in \mathrm{L}^\infty(\Omega)$ and $0 < \beta, \gamma \in \mathbb{R}$ with $\beta + \gamma < 1$ such that*

$$|\mathbf{I} : \boldsymbol{Y} - \alpha(x)\boldsymbol{A} : \boldsymbol{Y}| \leqslant \beta|\boldsymbol{Y}| + \gamma|\mathbf{I} : \boldsymbol{Y}|, \quad \forall\, \boldsymbol{Y} \in \mathrm{Sym}^+(\mathbb{R}^{d\times d}). \tag{13}$$

Then, (1) *admits a unique strong solution. There also exists a constant independent of u such that*

$$\|u\|_{\mathrm{H}^2(\Omega)} \leqslant C\|f\|_{\mathrm{L}^2(\Omega)}. \tag{14}$$

Remark 1 (Less regular solutions) Note that the theory of viscosity solutions has been developed for non-classical solutions of (10), if the problem data do not satisfy the regularity assumed above, see [27].

Assumption 1 (Inf-sup condition) From hereon in, we will assume that the problem satisfies an inf-sup condition, that is, with

$$a(w, v) := \int_\Omega -\boldsymbol{A} : \mathrm{D}^2 w v \,\mathrm{d}x, \tag{15}$$

then, for all $w \in \mathrm{H}^2(\Omega) \cap \mathrm{H}^1_0(\Omega)$

$$\sup_{v\in \mathrm{L}^2(\Omega)} \frac{a(w, v)}{\|v\|_{\mathrm{L}^2(\Omega)}} \geqslant C\|\Delta w\|_{\mathrm{L}^2(\Omega)}. \tag{16}$$

This is true under a variety of conditions. For example, those of Theorem 1. Note that for $d = 2$, this criterion is satisfied for all essentially bounded, symmetric positive definite $\boldsymbol{A}$.

3 Discretisation

Let $\mathscr{T}$ be a shape regular triangulation of Ω, namely, $\mathscr{T}$ is a finite family of sets such that

1) $K \in \mathscr{T}$ implies K is an open simplex (segment for $d = 1$, triangle for $d = 2$, tetrahedron for $d = 3$),
2) for any $K, J \in \mathscr{T}$ we have that $\overline{K} \cap \overline{J}$ is a full subsimplex (i.e., it is either $\emptyset$, a vertex, an edge, a face, or the whole of $\overline{K}$ and $\overline{J}$) of either $\overline{K}$ and $\overline{J}$ and
3) $\bigcup_{K\in\mathscr{T}} \overline{K} = \overline{\Omega}$.

We use the convention where $h : \Omega \to \mathbb{R}$ denotes the *mesh size function* of $\mathscr{T}$, i.e.,

$$h(\boldsymbol{x}) := \max_{\overline{K}\ni \boldsymbol{x}} h_K, \tag{17}$$

where h_K is the diameter of K. We let $\mathscr{E}$ be the skeleton (set of common interfaces) of the triangulation $\mathscr{T}$ and say $e \in \mathscr{E}$ if e is on the interior of Ω and $e \in \partial\Omega$ if e lies on the boundary $\partial\Omega$.

Let $\mathbb{P}^k(\mathscr{T})$ denote the space of piecewise polynomials of degree k over the triangulation $\mathscr{T}$, i.e.,

$$\mathbb{P}^k(\mathscr{T}) = \left\{\phi : \ \phi|_K \in \mathbb{P}^k(K)\right\}, \tag{18}$$

and introduce the *finite element spaces*

$$\mathbb{V} = \mathbb{V}(\mathscr{T}, k) := \mathbb{P}^k(\mathscr{T}), \tag{19}$$

$$\mathbb{V}_0 = \mathbb{V}_0(\mathscr{T}, k) := \left\{\phi \in \mathbb{P}^k(\mathscr{T}) : \ \phi|_{\partial\Omega} = 0\right\}, \tag{20}$$

to be the usual spaces of discontinuous piecewise polynomial functions.

Remark 2 (Generalised Hessian) Assume $v \in \mathrm{H}^2(\Omega)$, let $\boldsymbol{n} : \partial\Omega \to \mathbb{R}^d$ be the outward pointing normal of Ω, then the Hessian $\mathrm{D}^2 v$ of v, satisfies the following identity:

$$\int_\Omega \mathrm{D}^2 v \, \phi \, \mathrm{d}\boldsymbol{x} = -\int_\Omega \nabla v \otimes \nabla \phi \, \mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} \nabla v \otimes \boldsymbol{n} \, \phi \, \mathrm{d}s, \quad \forall \phi \in \mathrm{H}^1(\Omega). \tag{21}$$

If $v \in \mathrm{H}^1(\Omega)$, the right-hand side of (21) is still well defined in view of duality, in this case we set

$$\left\langle \mathrm{D}^2 v \,|\, \phi \right\rangle = -\int_\Omega \nabla v \otimes \nabla \phi \, \mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} \nabla v \otimes \boldsymbol{n} \, \phi \, \mathrm{d}s, \quad \forall \phi \in \mathrm{H}^1(\Omega), \tag{22}$$

where the last term is understood as a duality pairing.

Definition 2 (Broken Sobolev spaces, trace spaces) We introduce the broken Sobolev space

$$\mathrm{H}^k(\mathscr{T}) := \left\{\phi : \ \phi|_K \in \mathrm{H}^k(K), \text{ for each } K \in \mathscr{T}\right\}. \tag{22}$$

We also make use of functions defined in these broken spaces restricted to the skeleton of the triangulation. This requires an appropriate trace space

$$\mathcal{T}(\mathscr{E}) := \prod_{K\in\mathscr{T}} \mathrm{L}^2(\partial K) = \prod_{K\in\mathscr{T}} \mathrm{H}^{\frac{1}{2}}(K). \tag{24}$$

Definition 3 (Jumps, averages and tensor jumps) We define average, jump and tensor jump operators for arbitrary scalar functions $v \in \mathcal{T}(\mathscr{E})$, vectors $\boldsymbol{v} \in \mathcal{T}(\mathscr{E})^d$ and matrices $\boldsymbol{V} \in \mathcal{T}(\mathscr{E})^{d\times d}$ as

$$\{v\} = \frac{1}{2}\left(v|_{K_1} + v|_{K_2}\right), \qquad \{\boldsymbol{v}\} = \frac{1}{2}\left(\boldsymbol{v}|_{K_1} + \boldsymbol{v}|_{K_2}\right), \tag{25}$$

$$[\![v]\!] = v|_{K_1}\boldsymbol{n}_{K_1} + v|_{K_2}\boldsymbol{n}_{K_2}, \qquad [\![\boldsymbol{v}]\!] = \left(\boldsymbol{v}|_{K_1}\right)\cdot\boldsymbol{n}_{K_1} + \left(\boldsymbol{v}|_{K_2}\right)\cdot\boldsymbol{n}_{K_2}, \tag{26}$$

$$[\![\boldsymbol{V}]\!] = \boldsymbol{V}|_{K_1}\boldsymbol{n}_{K_1} + \boldsymbol{V}|_{K_2}\boldsymbol{n}_{K_2}, \qquad [\![\boldsymbol{v}]\!]_\otimes = \boldsymbol{v}|_{K_1}\otimes\boldsymbol{n}_{K_1} + \boldsymbol{v}|_{K_2}\otimes\boldsymbol{n}_{K_2}. \tag{27}$$

Note that on the boundary of the domain $\partial\Omega$, the jump and average operators are defined as

$$\{v\}\Big|_{\partial\Omega} := v, \qquad \{\boldsymbol{v}\}\Big|_{\partial\Omega} := \boldsymbol{v}, \tag{28}$$

$$[\![v]\!]\Big|_{\partial\Omega} := v\boldsymbol{n}, \qquad [\![\boldsymbol{v}]\!]\Big|_{\partial\Omega} := \boldsymbol{v}\cdot\boldsymbol{n}, \tag{29}$$

$$[\![\boldsymbol{V}]\!]\Big|_{\partial\Omega} := \boldsymbol{V}\boldsymbol{n}, \qquad [\![\boldsymbol{v}]\!]_{\otimes}\Big|_{\partial\Omega} := \boldsymbol{v}\otimes\boldsymbol{n}. \tag{30}$$

We will often use the following Proposition which we state in full for clarity but whose proof is merely using the identities in Definition 3.

Proposition 1 (Elementwise integration) *For a generic vector-valued function* $\boldsymbol{p}$ *and scalar-valued function* ϕ*, we have*

$$\sum_{K\in\mathcal{T}}\int_K \operatorname{div}(\boldsymbol{p})\phi\,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathcal{T}}\left(-\int_K \boldsymbol{p}\cdot\nabla_h\phi\,\mathrm{d}\boldsymbol{x} + \int_{\partial K}\phi\boldsymbol{p}\cdot\boldsymbol{n}_K\,\mathrm{d}s\right), \tag{31}$$

where $\nabla_h = \left(\mathrm{D}_h\right)^{\mathsf{T}}$ *is the elementwise spatial gradient. Furthermore, If we have* $\boldsymbol{p}\in\mathcal{T}(\mathcal{E}\cup\partial\Omega)^d$ *and* $\phi\in\mathcal{T}(\mathcal{E}\cup\partial\Omega)$*, the following identity holds*

$$\sum_{K\in\mathcal{T}}\int_{\partial K}\phi\boldsymbol{p}\cdot\boldsymbol{n}_K\,\mathrm{d}s = \int_{\mathcal{E}}[\![\boldsymbol{p}]\!]\{\phi\}\,\mathrm{d}s + \int_{\mathcal{E}\cup\partial\Omega}[\![\phi]\!]\cdot\{\boldsymbol{p}\}\,\mathrm{d}s = \int_{\mathcal{E}\cup\partial\Omega}[\![\boldsymbol{p}\phi]\!]\,\mathrm{d}s. \tag{32}$$

An equivalent tensor formulation of (31)–(32) *is*

$$\sum_{K\in\mathcal{T}}\int_K \mathrm{D}_h\boldsymbol{p}\phi\,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathcal{T}}\left(-\int_K \boldsymbol{p}\otimes\nabla_h\phi\,\mathrm{d}\boldsymbol{x} + \int_{\partial K}\phi\boldsymbol{p}\otimes\boldsymbol{n}_K\,\mathrm{d}s\right), \tag{33}$$

where

$$\sum_{K\in\mathcal{T}}\int_{\partial K}\phi\boldsymbol{p}\otimes\boldsymbol{n}_K\,\mathrm{d}s = \int_{\mathcal{E}}[\![\boldsymbol{p}]\!]_{\otimes}\{\phi\}\,\mathrm{d}s + \int_{\mathcal{E}\cup\partial\Omega}[\![\phi]\!]\otimes\{\boldsymbol{p}\}\,\mathrm{d}s = \int_{\mathcal{E}\cup\partial\Omega}[\![\boldsymbol{p}\phi]\!]_{\otimes}\,\mathrm{d}s. \tag{34}$$

In addition for matrix-valued $\boldsymbol{V}$*, we have that*

$$\sum_{K\in\mathcal{T}}\int_K \left(\mathrm{D}_h\boldsymbol{p}\right):\boldsymbol{V}\,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathcal{T}}\left(-\int_K \boldsymbol{p}:\mathrm{D}_h\boldsymbol{V}\,\mathrm{d}\boldsymbol{x} + \int_{\partial\Omega}(\boldsymbol{V}\boldsymbol{p})\cdot\boldsymbol{n}\,\mathrm{d}s\right) \tag{35}$$

and

$$\sum_{K\in\mathcal{T}}\int_{\partial\Omega}(\boldsymbol{V}\boldsymbol{p})\cdot\boldsymbol{n}\,\mathrm{d}s = \int_{\mathcal{E}}[\![\boldsymbol{V}]\!]\cdot\{\boldsymbol{p}\}\,\mathrm{d}s + \int_{\mathcal{E}\cup\partial\Omega}[\![\boldsymbol{p}]\!]_{\otimes}:\{\boldsymbol{V}\}\,\mathrm{d}s = \int_{\mathcal{E}\cup\partial\Omega}[\![\boldsymbol{V}\boldsymbol{p}]\!]\,\mathrm{d}s. \tag{36}$$

3.1 Construction of an Appropriate Discrete Hessian

We now use the framework set out in [4] to construct a general notion of discrete Hessian. We first give a definition using a flux formulation.

Definition 4 (Generalised finite element Hessian: flux formulation) Let $u\in\mathrm{H}^2(\mathcal{T})$, $\widehat{U}:\mathrm{H}^1(\mathcal{T})\to\mathcal{T}(\mathcal{E}\cup\partial\Omega)$ be a linear form and $\widehat{\boldsymbol{p}}:\mathrm{H}^2(\mathcal{T})\times\mathrm{H}^1(\mathcal{T})^d\to\mathcal{T}(\mathcal{E}\cup\partial\Omega)^d$ a bilinear form representing approximations to u and ∇u over the skeleton of the triangulation. Then, we define the generalised finite element Hessian $\boldsymbol{H}[u]$ as the solution of

$$\int_K \boldsymbol{H}[u]\,\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} = -\int_K \boldsymbol{p}\otimes\nabla_h\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} + \int_{\partial K}\widehat{\boldsymbol{p}}_K\otimes\boldsymbol{n}\,\boldsymbol{\Phi}\,\mathrm{d}s,\quad \forall\,\boldsymbol{\Phi}\in\mathrm{H}^1(\mathscr{T}), \tag{37}$$

$$\int_K \boldsymbol{p}\otimes\boldsymbol{q}\,\mathrm{d}\boldsymbol{x} = -\int_K u\,\mathrm{D}_h\boldsymbol{q}\,\mathrm{d}\boldsymbol{x} + \int_{\partial K}\boldsymbol{q}\otimes\boldsymbol{n}\,\widehat{U}_K\,\mathrm{d}s \tag{38}$$

for all $\boldsymbol{\Phi}\in\mathbb{V}$.

We now present the primal formulation for the generalised finite element Hessian.

Theorem 2 (Generalised finite element Hessian: primal form) *Let $u\in\mathrm{H}^2(\mathscr{T})$ and let $\widehat{U}$ and $\widehat{\boldsymbol{p}}$ be defined as in Definition 4. Then, the generalised finite element Hessian $H[u]$ is given for each $\boldsymbol{\Phi}\in\mathbb{V}$ as*

$$\begin{aligned}\int_\Omega \boldsymbol{H}[u]\,\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} = &-\int_\Omega \nabla_h u\otimes\nabla_h\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} + \int_{\mathscr{E}\cup\partial\Omega}[\![\boldsymbol{\Phi}]\!]\otimes\{\widehat{\boldsymbol{p}}\}\,\mathrm{d}s + \int_{\mathscr{E}}\{\boldsymbol{\Phi}\}[\![\widehat{\boldsymbol{p}}]\!]_\otimes\,\mathrm{d}s\\ &-\int_{\mathscr{E}}\{\widehat{U}-u\}[\![\nabla_h\boldsymbol{\Phi}]\!]_\otimes\,\mathrm{d}s - \int_{\mathscr{E}\cup\partial\Omega}[\![\widehat{U}-u]\!]\otimes\{\nabla_h\boldsymbol{\Phi}\}\,\mathrm{d}s.\end{aligned} \tag{39}$$

Proof Note that in view of Definition 3 for generic vector fields $\boldsymbol{q}\in\mathbb{W}$ and $v\in\mathbb{V}$, we have the following identity:

$$\sum_{K\in\mathscr{T}}\int_{\partial K} v\boldsymbol{q}\otimes\boldsymbol{n}\,\mathrm{d}s = \int_{\mathscr{E}\cup\partial\Omega}[\![v]\!]\otimes\{\boldsymbol{q}\}\,\mathrm{d}s + \int_{\mathscr{E}}\{v\}[\![\boldsymbol{q}]\!]_\otimes\,\mathrm{d}s. \tag{40}$$

Then summing (37) over $K\in\mathscr{T}$ and making use of the identity (40) we see

$$\begin{aligned}\int_\Omega \boldsymbol{H}[u]\,\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} &= \sum_{K\in\mathscr{T}}\int_K \boldsymbol{H}[u]\,\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathscr{T}}\left(-\int_K \boldsymbol{p}\otimes\nabla_h\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} + \int_{\partial K}\widehat{\boldsymbol{p}}_K\otimes\boldsymbol{n}\,\boldsymbol{\Phi}\right)\mathrm{d}s\\ &= -\int_\Omega \boldsymbol{p}\otimes\nabla_h\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} + \int_{\mathscr{E}\cup\partial\Omega}[\![\boldsymbol{\Phi}]\!]\otimes\{\widehat{\boldsymbol{p}}_K\}\,\mathrm{d}s + \int_{\mathscr{E}}\{\boldsymbol{\Phi}\}[\![\widehat{\boldsymbol{p}}_K]\!]_\otimes\,\mathrm{d}s.\end{aligned} \tag{41}$$

Using the same argument for (38)

$$\begin{aligned}\int_\Omega \boldsymbol{p}\otimes\boldsymbol{q}\,\mathrm{d}\boldsymbol{x} &= \sum_{K\in\mathscr{T}}\int_K \boldsymbol{p}\otimes\boldsymbol{q}\,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathscr{T}}\left(-\int_K u\,\mathrm{D}_h\boldsymbol{q}\,\mathrm{d}\boldsymbol{x} + \int_{\partial K}\boldsymbol{q}\otimes\boldsymbol{n}\,\widehat{U}_K\,\mathrm{d}s\right)\\ &= -\int_\Omega u\,\mathrm{D}_h\boldsymbol{q}\,\mathrm{d}\boldsymbol{x} + \int_{\mathscr{E}\cup\partial\Omega}[\![\widehat{U}]\!]\otimes\{\boldsymbol{q}\}\,\mathrm{d}s + \int_{\mathscr{E}}\{\widehat{U}\}[\![\boldsymbol{q}]\!]_\otimes\,\mathrm{d}s.\end{aligned} \tag{42}$$

Note that, again making use of (40), we have for each $\boldsymbol{q}\in\mathrm{H}^1(\mathscr{T})^d$ and $v\in\mathrm{H}^1(\mathscr{T})$ that

$$\int_\Omega \boldsymbol{q}\otimes\nabla_h v\,\mathrm{d}\boldsymbol{x} = -\int_\Omega \mathrm{D}_h\boldsymbol{q}\,v\,\mathrm{d}\boldsymbol{x} + \int_{\mathscr{E}\cup\partial\Omega}\{\boldsymbol{q}\}\otimes[\![v]\!]\,\mathrm{d}s + \int_{\mathscr{E}}[\![\boldsymbol{q}]\!]_\otimes\{v\}\,\mathrm{d}s. \tag{43}$$

Taking $v=u$ in (43) and substituting into (38), we see

$$\int_\Omega \boldsymbol{p}\otimes\boldsymbol{q}\,\mathrm{d}\boldsymbol{x} = \int_\Omega \boldsymbol{q}\otimes\nabla_h u\,\mathrm{d}\boldsymbol{x} + \int_{\mathscr{E}\cup\partial\Omega}[\![\widehat{U}-u]\!]\otimes\{\boldsymbol{q}\}\,\mathrm{d}s + \int_{\mathscr{E}}\{\widehat{U}-u\}[\![\boldsymbol{q}]\!]_\otimes\,\mathrm{d}s. \tag{44}$$

Now choosing $\boldsymbol{q} = \nabla_h \boldsymbol{\Phi}$ and substituting (44) into (37), we arrive at the fully generalised finite element Hessian given by (39).

Remark 3 (Consistent representations of the gradient operator) If one were interested in consistent representations of other derivatives, for example the gradient operator, one would need to modify the proof of Theorem 2. Examples of consistent gradient representations can be found in [4]. See also [10, 11, 17]. Using this methodology, it should be possible to construct an entire hierarchy of derivatives.

Example 1 An example of a DG formulation for the approximation to the Hessian, $\mathrm{D}^2 u$, can be derived by taking the fluxes in the following way:

$$\widehat{U} = \begin{cases} \{u_h\} & \text{over } \mathscr{E}, \\ 0 & \text{on } \partial\Omega, \end{cases} \tag{45}$$

$$\widehat{\boldsymbol{p}} = \{\nabla_h u_h\} \text{ on } \mathscr{E} \cup \partial\Omega. \tag{46}$$

The result is a discrete representation of the Hessian $\boldsymbol{H}[u_h]$ as a unique element of $\mathbb{V}^{d\times d}$ such that

$$\int_\Omega \boldsymbol{H}[u_h]\, \boldsymbol{\Phi}\, \mathrm{d}\boldsymbol{x} = -\int_\Omega \nabla_h u_h \otimes \nabla_h \boldsymbol{\Phi}\, \mathrm{d}\boldsymbol{x} + \int_{\mathscr{E}\cup\partial\Omega} [\![u_h]\!] \otimes \{\nabla_h \boldsymbol{\Phi}\} + [\![\boldsymbol{\Phi}]\!] \otimes \{\nabla_h u_h\}\, \mathrm{d}s.$$

We can also derive unsymmetric approximations with fluxes similar to those used for the NIPG method. This is given by taking $\theta = -1$ in the following, while the symmetric version is given by $\theta = 1$:

$$\begin{aligned} \int_\Omega \boldsymbol{H}[u_h]\, \boldsymbol{\Phi}\, \mathrm{d}\boldsymbol{x} &= -\int_\Omega \nabla_h u_h \otimes \nabla_h \boldsymbol{\Phi}\, \mathrm{d}\boldsymbol{x} \\ &\quad + \int_{\mathscr{E}\cup\partial\Omega} \theta [\![u_h]\!] \otimes \{\nabla_h \boldsymbol{\Phi}\} + [\![\boldsymbol{\Phi}]\!] \otimes \{\nabla_h u_h\}\, \mathrm{d}s \\ &= \int_\Omega \mathrm{D}_h^2 u_h \boldsymbol{\Phi}\, \mathrm{d}\boldsymbol{x} - \int_{\mathscr{E}} [\![\nabla_h u_h]\!]_{\otimes} \{\boldsymbol{\Phi}\}\, \mathrm{d}s \\ &\quad + \int_{\mathscr{E}\cup\partial\Omega} \theta [\![u_h]\!] \otimes \{\nabla_h \boldsymbol{\Phi}\}\, \mathrm{d}s, \quad \forall\, \boldsymbol{\Phi} \in \mathbb{V}. \end{aligned} \tag{47}$$

3.2 The Discontinuous Nonvariational Finite Element Method

We are now in a position to state the numerical method for the approximation of (1). We look to find $u_h \in \mathbb{V}_0$ together with $\boldsymbol{H}[u_h] \in \mathbb{V}^{d\times d}$ such that

$$\mathscr{A}_h\big(u_h, \boldsymbol{\Psi}\big) = l(\boldsymbol{\Psi}), \quad \forall\, \boldsymbol{\Psi} \in \mathbb{V}_0 \tag{48}$$

with

$$\mathscr{A}_h\big(u_h, \boldsymbol{\Psi}\big) := \int_\Omega -\boldsymbol{A} : \boldsymbol{H}[u_h] \boldsymbol{\Psi}\, \mathrm{d}\boldsymbol{x} + \int_{\mathscr{E}\cup\partial\Omega} \sigma h^{-1} [\![u_h]\!] \cdot [\![\boldsymbol{\Psi}]\!]\, \mathrm{d}s, \tag{49}$$

$$l(\Psi) := \int_\Omega f\Psi \,\mathrm{d}\boldsymbol{x}, \tag{50}$$

where the *penalisation parameter* $\sigma > 0$ is to be chosen sufficiently large.

Using the L^2 projection operator $\mathrm{P}_k : \mathrm{L}^2(\Omega) \to \mathbb{V}$ defined for $v \in \mathrm{L}^2(\Omega)$ through

$$\int_\Omega \mathrm{P}_k(v)\Psi \,\mathrm{d}\boldsymbol{x} = \int_\Omega v\Psi \,\mathrm{d}\boldsymbol{x}, \quad \forall \Psi \in \mathbb{V}, \tag{51}$$

it is possible to elliminate the finite element Hessian from the bilinear form for sufficiently smooth $\boldsymbol{A}$.

Lemma 1 (Elimination of the finite element Hessian in a general setting) *If the fluxes are chosen as in Example 1, then*

$$\begin{aligned}\mathscr{A}_h\big(u_h, \Psi\big) = &\int_\Omega D_h\big(\mathrm{P}_k(\Psi\boldsymbol{A})\big)\nabla_h u_h \,\mathrm{d}\boldsymbol{x} - \int_{\mathscr{E}\cup\partial\Omega} \theta[\![u_h]\!] \cdot \{D_h\big(\mathrm{P}_k(\Psi\boldsymbol{A})\big)\} \,\mathrm{d}s \\ &- \int_{\mathscr{E}\cup\partial\Omega} [\![\mathrm{P}_k(\Psi\boldsymbol{A})]\!] \cdot \{\nabla_h u_h\} \,\mathrm{d}s + \int_{\mathscr{E}\cup\partial\Omega} \sigma h^{-1}[\![u_h]\!] \cdot [\![\Psi]\!] \,\mathrm{d}s.\end{aligned} \tag{52}$$

Proof This follows from the following identity:

$$\begin{aligned}\int_\Omega -\boldsymbol{A} : \boldsymbol{H}[u_h]\Psi \,\mathrm{d}\boldsymbol{x} &= \int_\Omega -\boldsymbol{H}[u_h] : (\Psi\boldsymbol{A}) \,\mathrm{d}\boldsymbol{x} = \int_\Omega -\boldsymbol{H}[u_h] : \mathrm{P}_k(\Psi\boldsymbol{A}) \,\mathrm{d}\boldsymbol{x} \\ &= \int_\Omega D_h\big(\mathrm{P}_k(\Psi\boldsymbol{A})\big)\nabla_h u_h \,\mathrm{d}\boldsymbol{x} - \int_{\mathscr{E}\cup\partial\Omega} \theta[\![u_h]\!] \cdot \{D_h\big(\mathrm{P}_k(\Psi\boldsymbol{A})\big)\} \,\mathrm{d}s \\ &\quad - \int_{\mathscr{E}\cup\partial\Omega} [\![\mathrm{P}_k(\Psi\boldsymbol{A})]\!] \cdot \{\nabla_h u_h\} \,\mathrm{d}s.\end{aligned} \tag{53}$$

Remark 4 The solution of the problem in this form is nontrivial due to the global $\mathrm{L}^2(\Omega)$ projection appearing in the formulation. However, in the discontinuous setting, the global $\mathrm{L}^2(\Omega)$ projection is in fact computable locally. We may actually exploit this fact to optimise our schemes efficiency. We will discuss this further in the sequel.

Example 2 (Laplacian formulation) Note that if in (1) we have that $\boldsymbol{A} = \boldsymbol{I}$, then we have that

$$f = -A : \mathrm{D}^2 u = -\Delta u \tag{54}$$

and our bilinear form reduces to

$$\begin{aligned}\mathscr{A}_h\big(u_h, \Psi\big) = &\int_\Omega \big(\nabla_h \Psi\big) \cdot \nabla_h u_h \,\mathrm{d}\boldsymbol{x} - \int_{\mathscr{E}\cup\partial\Omega} \theta[\![u_h]\!] \cdot \{\nabla_h \Psi\} \,\mathrm{d}s \\ &- \int_{\mathscr{E}\cup\partial\Omega} ([\![\Psi]\!] \cdot \{\nabla_h u_h\} - \sigma h^{-1}[\![u_h]\!] \cdot [\![\Psi]\!]) \,\mathrm{d}s,\end{aligned} \tag{55}$$

since $\mathrm{P}_k(\Psi\boldsymbol{A}) = \Psi\boldsymbol{I}$.

The nonvariational finite element method, thus, coincides with the classical (symmetric) interior penalty method for the Laplacian [16].

Remark 5 (Relation to standard DG methods) It is not difficult to prove that choosing to numerical fluxes in the same way as presented in [4, Table 3.2] results in the same correlation to the DG methods summarised in the aforementioned paper for the case that $\boldsymbol{A}$ is constant. For brevity, we will not prove this here.

Note that when $\boldsymbol{A}$ is not constant, we have that the nonvariational finite element method does *not* coincide with its standard variational finite element counterpart. Indeed, the method is able to successfully cope with classes of advection dominated problems not falling under the variational framework without special treatment [32, §4.2] which is illustrated by the result of Lemma 1.

We conclude this section with a proof of consistency of the method and then show that Galerkin orthogonality holds.

Lemma 2 (Consistency) *Let $u \in \mathrm{H}^2(\Omega)$ and assume that the numerical fluxes are chosen in a consistent fashion in the sense of* [4, §3.1], *that is*,

$$\widehat{U} = u|_{\mathscr{E}\cup\partial\Omega}, \tag{56}$$

$$\widehat{\boldsymbol{p}} = \nabla u|_{\mathscr{E}\cup\partial\Omega}. \tag{57}$$

Then for $\boldsymbol{\Phi} \in \mathbb{V}$

$$\int_\Omega \boldsymbol{H}[u]\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} = \int_\Omega \mathrm{D}^2 u\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x}. \tag{58}$$

Therefore, we have that $\boldsymbol{H}[u] = \mathrm{P}_k\big(\mathrm{D}^2 u\big)$.

Proof Applying Proposition 1 to the first term in the definition of $\boldsymbol{H}[u]$ yields

$$\begin{aligned}\int_\Omega \boldsymbol{H}[u]\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} &= \int_\Omega \mathrm{D}^2 u\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x} + \int_{\mathscr{E}} [\![\widehat{\boldsymbol{p}} - \nabla u]\!]_{\otimes}\{\boldsymbol{\Phi}\}\,\mathrm{d}s + \int_{\mathscr{E}\cup\partial\Omega} \{\widehat{\boldsymbol{p}} - \nabla u\} \otimes [\![\boldsymbol{\Phi}]\!]\,\mathrm{d}s \\ &\quad - \int_{\mathscr{E}} \{\widehat{U} - u\}[\![\nabla_h\boldsymbol{\Phi}]\!]_{\otimes}\,\mathrm{d}s - \int_{\mathscr{E}\cup\partial\Omega} [\![\widehat{U} - u]\!] \otimes \{\nabla_h\boldsymbol{\Phi}\}\,\mathrm{d}s \\ &= \int_\Omega \mathrm{D}^2 u\boldsymbol{\Phi}\,\mathrm{d}\boldsymbol{x}, \quad \forall\,\boldsymbol{\Phi} \in \mathbb{V},\end{aligned} \tag{59}$$

which proves the results under the consistency conditions on the fluxes.

Lemma 3 (Galerkin orthogonality) *Let $u \in \mathrm{H}^2(\Omega) \cap \mathrm{H}^1_0(\Omega)$ be a strong solution to the problem* (1) *and let $u_h \in \mathbb{V}_0$ be its nonvariational finite element approximation. Assume that the numerical fluxes $\widehat{U}$ and $\widehat{\boldsymbol{p}}$ are consistent, then we have the following orthogonality result*:

$$\mathscr{A}_h\big(u_h - u, \boldsymbol{\Psi}\big) = J(\boldsymbol{\Psi}), \quad \forall\,\boldsymbol{\Psi} \in \mathbb{V}_0, \tag{60}$$

with the error functional given by

$$J(\boldsymbol{\Psi}) = \int_\Omega \big(\mathrm{D}^2 u - \boldsymbol{H}[u]\big) : (\boldsymbol{A}\boldsymbol{\Psi})\,\mathrm{d}\boldsymbol{x}. \tag{61}$$

Proof Using the consistency result and that $[\![u]\!] = 0$, we conclude

$$\begin{aligned}\mathscr{A}_h\big(u_h - u, \Psi\big) &= \mathscr{A}_h\big(u_h, \Psi\big) + \int_\Omega \boldsymbol{A} : \boldsymbol{H}[u]\Psi \,\mathrm{d}\boldsymbol{x} = l(\Psi) + \int_\Omega \boldsymbol{H}[u] : (\boldsymbol{A}\Psi)\,\mathrm{d}\boldsymbol{x} \\ &= -\int_\Omega \boldsymbol{A} : \mathrm{D}^2 u\Psi - \boldsymbol{H}[u] : (\boldsymbol{A}\Psi)\,\mathrm{d}\boldsymbol{x} = J(\Psi),\end{aligned}$$

concluding the proof.

Remark 6 If $\boldsymbol{A}$ is piecewise constant, then since $\boldsymbol{H}[u] = \mathrm{P}_k\big(\mathrm{D}^2 u\big)$ we have $J(\Psi) = 0$ and we recover the usual Galerkin orthogonality $\mathscr{A}_h\big(u_h - u, \Psi\big) = 0$.

Definition 5 ($\mathrm{H}^1(\mathscr{T})$ and $\mathrm{H}^2(\mathscr{T})$ norms) We introduce the broken $\mathrm{H}^1(\mathscr{T})$ and $\mathrm{H}^2(\mathscr{T})$ norms as

$$\|u_h\|_{\mathrm{DG},1}^2 := \|\nabla_h u_h\|_{\mathrm{L}^2(\Omega)}^2 + h^{-1}\|[\![u_h]\!]\|_{\mathrm{L}^2(\mathscr{E})}^2, \tag{62}$$

$$\|u_h\|_{\mathrm{DG},2}^2 := \left\|\mathrm{D}_h^2 u_h\right\|_{\mathrm{L}^2(\Omega)}^2 + h^{-1}\|[\![\nabla_h u_h]\!]\|_{\mathrm{L}^2(\mathscr{E})}^2 + h^{-3}\|[\![u_h]\!]\|_{\mathrm{L}^2(\mathscr{E})}^2. \tag{63}$$

These are equivalent to their continuous equivalent norms for functions in $\mathbb{V}$.

Proposition 2 (Projection approximation in $\mathbb{V}$) *Let* $\mathrm{P}_k : \mathrm{L}^2(\Omega) \to \mathbb{V}$ *be the* $\mathrm{L}^2(\Omega)$ *orthogonal projection operator defined by* (51). *Using standard approximation arguments, we have that*

$$\begin{cases} \|v - \mathrm{P}_k v\|_{\mathrm{DG},1} \leqslant Ch^k |v|_{\mathrm{H}^{k+1}(\Omega)}, \\ \|v - \mathrm{P}_k v\|_{\mathrm{L}^2(\Omega)} \leqslant Ch^{k+1} |v|_{\mathrm{H}^{k+1}(\Omega)}. \end{cases} \tag{64}$$

Lemma 4 (Stability of $\boldsymbol{H}$ [37, Theorem 4.10]) *Let* $\boldsymbol{H}$ *be defined as in Example* 1. *Then the DG Hessian is stable in the sense that*

$$\left\|\mathrm{D}_h^2 v_h - \boldsymbol{H}[v_h]\right\|_{\mathrm{L}^2(\Omega)}^2 \leqslant C\left(\int_{\mathscr{E}} h^{-1}\left|[\![\nabla_h v_h]\!]\right|^2 + h^{-3}\left|[\![v_h]\!]\right|^2 \mathrm{d}s\right). \tag{65}$$

Consequently, we have

$$\|\boldsymbol{H}[v_h]\|_{\mathrm{L}^2(\Omega)}^2 \leqslant C\|v_h\|_{\mathrm{DG},2}^2. \tag{66}$$

4 A Posteriori Analysis

The discrete bilinear form (49) only makes sense over $\mathrm{H}^2(\mathscr{T}) \times \mathrm{H}^2(\mathscr{T})$. To allow for an a posteriori bound we require an extension to ensure the appropriate stability arguments can be applied. We require the bilinear form to be extended to $\mathrm{H}^2(\mathscr{T}) \times \mathrm{L}^2(\Omega)$. To do this for $(u, v) \in \mathrm{H}^2(\mathscr{T}) \times \mathrm{L}^2(\Omega)$, we define

$$\mathscr{A}_h(u,v) := \int_\Omega \boldsymbol{A} : \boldsymbol{H}(u)v + \int_{\mathscr{E}} \sigma h_e^{-1} [\![\nabla u]\!] \cdot [\![\mathrm{P}_k v]\!]. \tag{67}$$

Remark 7 Notice that the modified bilinear form (67) coincides with (49) over $\mathbb{V} \times \mathbb{V}$ and that it satisfies

$$\mathscr{A}_h(u,v) \leqslant C \, |\!|\!| u |\!|\!|_{\mathrm{DG},2} \|v\|_{\mathrm{L}^2(\Omega)} \text{ for} (u,v) \in \left(\mathrm{H}^2(\mathscr{T}) \times \mathrm{L}^2(\Omega)\right). \tag{68}$$

Assumption 2 (Existence of an H^2 reconstruction) We will assume there exists an operator $\mathscr{R} : \mathbb{V} \to \mathrm{H}^2(\Omega)$ such that

$$|\!|\!| \mathscr{R}(u_h) - u_h |\!|\!|^2_{\mathrm{DG},2} \leqslant C \left(\left\| h^{-1/2} [\![\nabla u_h]\!] \right\|^2_{\mathrm{L}^2(\mathscr{E})} + \left\| h^{-3/2} [\![u_h]\!] \right\|^2_{\mathrm{L}^2(\mathscr{E})} \right). \tag{69}$$

Such an example is given, for $d = 2$, in [26, Lem 3.1] and consists of averaging techniques onto macro-elements on the Hsieh-Clough-Tocher space or in [9], for $d = 3$, using virtual element spaces.

Proposition 3 (Abstract upper bound) *Let $u \in \mathrm{H}^2(\Omega)$ solve (1), $u_h \in \mathbb{V}$ be the finite element approximation given by (48) and $\mathscr{R}(u_h) \in \mathrm{H}^2(\Omega)$ be a post-processor satisfying (69). Then,*

$$\begin{aligned} &\left\| u - \mathscr{R}(u_h) \right\|_{\mathrm{H}^2(\Omega)} \\ &\leqslant \frac{1}{C} \sup_{v \in \mathrm{L}^2(\Omega)} \frac{\left(l(v) - \mathscr{A}_h(u_h, v) + \mathscr{A}_h\big(u_h - \mathscr{R}(u_h), v\big) + \mathscr{A}_h\big(\mathscr{R}(u_h), v\big) - \mathscr{A}\big(\mathscr{R}(u_h), v\big) \right)}{\|v\|_{\mathrm{L}^2(\Omega)}}. \end{aligned} \tag{70}$$

Proof Making use of the Miranda-Talenti inequality [34, 39] and the inf-sup condition (16) we see that

$$C \left\| u - \mathscr{R}(u_h) \right\|_{\mathrm{H}^2(\Omega)} \leqslant \tilde{C} \left\| \Delta\big(u - \mathscr{R}(u_h)\big) \right\|_{\mathrm{L}^2(\Omega)} \leqslant \sup_{v \in \mathrm{L}^2(\Omega)} \frac{\mathscr{A}\big(u - \mathscr{R}(u_h), v\big)}{\|v\|_{\mathrm{L}^2(\Omega)}}. \tag{71}$$

Now, adding and subtracting appropriate terms, we have

$$\begin{aligned} \mathscr{A}\big(u - \mathscr{R}(u_h), v\big) &= l(v - v_h) - \mathscr{A}_h\big(u_h, v - v_h\big) \\ &\quad + \mathscr{A}_h\big(u_h - \mathscr{R}(u_h), v\big) + \mathscr{A}_h\big(\mathscr{R}(u_h), v\big) - \mathscr{A}\big(\mathscr{R}(u_h), v\big), \end{aligned} \tag{72}$$

and the result follows.

Theorem 3 (A posteriori bound) *Let $u \in \mathrm{H}^2(\Omega)$ solve (1) and $u_h \in \mathbb{V}$ be the finite element approximation given by (48). Then*

$$|\!|\!| u - u_h |\!|\!|_{\mathrm{DG},2} \leqslant C \left(\sum_{K \in \mathscr{T}} \eta_K^2 + \sum_{e \in \mathscr{E}} \eta_e^2 \right)^{1/2}, \tag{73}$$

where

$$\eta_K^2 := \left\| f + \boldsymbol{A} : \boldsymbol{H}(u_h) \right\|_{\mathrm{L}^2(K)}^2, \tag{74}$$

$$\eta_e^2 := h_e^{-1} \left\| [\![\nabla u_h]\!] \right\|_{\mathrm{L}^2(e)}^2 + h_e^{-3} \left\| [\![u_h]\!] \right\|_{\mathrm{L}^2(e)}^2. \tag{75}$$

Proof Beginning from Proposition 3 we note that the bound can be split into a residual component $\mathscr{I}_1$, a nonconformity component $\mathscr{I}_2$ and an inconsistency component $\mathscr{I}_3$ as follows:

$$\begin{aligned} \left\| u - \mathscr{R}(u_h) \right\|_{\mathrm{H}^2(\Omega)} \leqslant \frac{1}{C} \Bigg(& \sup_{v \in \mathrm{L}^2(\Omega), \|v\|_{\mathrm{L}^2(\Omega)} \leqslant 1} \left(l(v) - \mathscr{A}_h\big(u_h, v\big) \right) \\ & + \sup_{v \in \mathrm{L}^2(\Omega), \|v\|_{\mathrm{L}^2(\Omega)} \leqslant 1} \mathscr{A}_h\big(u_h - \mathscr{R}(u_h), v\big) \\ & + \sup_{v \in \mathrm{L}^2(\Omega), \|v\|_{\mathrm{L}^2(\Omega)} \leqslant 1} \mathscr{A}_h\big(\mathscr{R}(u_h), v\big) - \mathscr{A}\big(\mathscr{R}(u_h), v\big) \Bigg) \\ =: \ & \mathscr{I}_1 + \mathscr{I}_2 + \mathscr{I}_3. \end{aligned} \tag{76}$$

Now, we proceed to bound these term by term. In view of Cauchy-Schwartz, we have

$$\begin{aligned} \mathscr{I}_1 & \leqslant \sup_{v \in \mathrm{L}^2(\Omega), \|v\|_{\mathrm{L}^2(\Omega)} \leqslant 1} \int_\Omega \big(f - \boldsymbol{A} : \boldsymbol{H}(u_h) \big) v \\ & \leqslant \left(\sum_{K \in \mathscr{T}} \left\| f - \boldsymbol{A} : \boldsymbol{H}(u_h) \right\|_{\mathrm{L}^2(K)}^2 \right)^{1/2}. \end{aligned} \tag{77}$$

For the nonconformity term, we note that

$$\mathscr{I}_2 \leqslant C \left| \! \left| \! \left| \mathscr{R}(u_h) - u_h \right| \! \right| \! \right|_{\mathrm{DG},2} \leqslant C \left(\sum_{e \in K} h_e^{-1} \left\| [\![\nabla_h v_h]\!] \right\|_{\mathrm{L}^2(e)}^2 + h_e^{-3} \left\| [\![v_h]\!] \right\|_{\mathrm{L}^2(e)}^2 \right)^{1/2}, \tag{78}$$

by the properties of the reconstruction given in (69).

For the inconsistency term, we have that

$$\begin{aligned} \mathscr{I}_3 = & \sup_{v \in \mathrm{L}^2(\Omega), \|v\|_{\mathrm{L}^2(\Omega)} \leqslant 1} \frac{1}{C} \int_\Omega \boldsymbol{A} : \big(\boldsymbol{H}(\mathscr{R}(u_h)) - \mathrm{D}^2 \mathscr{R}(u_h) \big) v \\ \leqslant & \, C \|\boldsymbol{A}\|_{\mathrm{L}^\infty(\Omega)} \left\| \boldsymbol{H}(\mathscr{R}(u_h)) - \mathrm{D}^2 \mathscr{R}(u_h) \right\|_{\mathrm{L}^2(\Omega)} \\ \leqslant & \, C \|\boldsymbol{A}\|_{\mathrm{L}^\infty(\Omega)} \Big(\left\| \boldsymbol{H}(\mathscr{R}(u_h)) - \boldsymbol{H}(u_h) \right\|_{\mathrm{L}^2(\Omega)} + \left\| \boldsymbol{H}(u_h) - \mathrm{D}_h^2 u_h \right\|_{\mathrm{L}^2(\Omega)} \\ & + \left\| \mathrm{D}_h^2 u_h - \mathrm{D}^2 \mathscr{R}(u_h) \right\|_{\mathrm{L}^2(\Omega)} \Big). \end{aligned} \tag{79}$$

Making use of Lemma 4 and the properties of the reconstruction (69), we see that

$$\begin{aligned}\left\|\boldsymbol{H}(\mathscr{R}(u_h)) - \boldsymbol{H}(u_h)\right\|_{\mathrm{L}^2(\Omega)} &\leqslant C\,|||\mathscr{R}(u_h) - u_h|||_{\mathrm{DG},2} \\ &\leqslant C\left(\sum_{e\in\mathscr{E}} h_e^{-1}\left\|[\![\nabla_h v_h]\!]\right\|^2_{\mathrm{L}^2(e)} + h_e^{-3}\left\|[\![v_h]\!]\right\|^2_{\mathrm{L}^2(e)}\right)^{1/2}.\end{aligned} \tag{80}$$

Again from Lemma 4 we have that

$$\left\|\boldsymbol{H}(u_h) - \mathrm{D}_h^2 u_h\right\|_{\mathrm{L}^2(\Omega)} \leqslant C\left(\sum_{e\in\mathscr{E}} h_e^{-1}\left\|[\![\nabla_h v_h]\!]\right\|^2_{\mathrm{L}^2(e)} + h_e^{-3}\left\|[\![v_h]\!]\right\|^2_{\mathrm{L}^2(e)}\right)^{1/2}, \tag{81}$$

and finally (69) gives that

$$\left\|\mathrm{D}_h^2 u_h - \mathrm{D}^2\mathscr{R}(u_h)\right\|_{\mathrm{L}^2(\Omega)} \leqslant C\left(\sum_{e\in\mathscr{E}} h_e^{-1}\left\|[\![\nabla_h v_h]\!]\right\|^2_{\mathrm{L}^2(e)} + h_e^{-3}\left\|[\![v_h]\!]\right\|^2_{\mathrm{L}^2(e)}\right)^{1/2}. \tag{82}$$

Hence, we have that

$$\mathscr{I}_3 \leqslant C\left(\sum_{e\in\mathscr{E}} h_e^{-1}\left\|[\![\nabla_h v_h]\!]\right\|^2_{\mathrm{L}^2(e)} + h_e^{-3}\left\|[\![v_h]\!]\right\|^2_{\mathrm{L}^2(e)}\right)^{1/2}. \tag{83}$$

Collecting (77), (78) and (83) yields the desired result.

Proposition 4 (A posteriori lower bound) *Using the notation of Theorem 3, through standard a posteriori techniques, we have a lower bound of the form*

$$\eta_K + \sum_{e\in\partial K}\eta_e \leqslant C\left(\left\|\mathrm{D}^2 u - \boldsymbol{H}[u_h]\right\|_{\mathrm{L}^2(\hat{K})} + \sum_{K\in\hat{K}}\eta_{I,K}\right), \tag{84}$$

where

$$\eta_{I,K}^2 := \left\|\left(P_{k-2}\boldsymbol{A} - \boldsymbol{A}\right) : \boldsymbol{H}(u_h)\right\|^2_{\mathrm{L}^2(K)} + \left\|f - P_{k-2}f\right\|^2_{\mathrm{L}^2(K)}. \tag{85}$$

5 Numerical Experiments

In this section, we detail numerical experiments carried out in the finite element package Dune-Fem [15] which is based on the Dune software framework [6, 7]. The code makes use of the newly developed Python frontend [15] and the unified form language [3] is used to provide the problem data. The code will be made freely available within the Dune-Fem-tutorial in a future release [13].

We present some benchmark problems designed such that the exact solution is known. In each of the experiments the domain $\Omega = [0,1]^2$ and we consider the coefficient matrix to be

$$\boldsymbol{A}(\boldsymbol{x}) = \begin{bmatrix} 1 & b(\boldsymbol{x}) \\ b(\boldsymbol{x}) & a(\boldsymbol{x}) \end{bmatrix} \tag{86}$$

varying $a(\boldsymbol{x})$ and $b(\boldsymbol{x})$. We study three different choices using $\boldsymbol{x} = (x_1, x_2)$ as follows.

1) (Coercive) In this test, we take the components of $\boldsymbol{A}$ such that the differential operator can be written in variational form and is coercive, fitting into a standard analytical framework:

$$a(\boldsymbol{x}) = 1 - \ln\Big(\big(x_1 - 1/2\big)^2 + 10^{-4}\Big), \tag{87}$$

$$b(\boldsymbol{x}) = 0. \tag{88}$$

2) (Continuous not H^2) In this test, we take $\boldsymbol{A}$ such that it is comparable to [32, §4.4]

$$a(\boldsymbol{x}) = 2, \tag{89}$$

$$b(\boldsymbol{x}) = \big(x_1^2 x_2^2\big)^{1/3}. \tag{90}$$

3) (Discontinuous not $W^{1,\infty}$) In our third test, a is discontinuous and not in $W^{1,\infty}$. We take $b \equiv 0$ and choose

$$a(\boldsymbol{x}) = \alpha(|\boldsymbol{x}|_\infty) \tag{91}$$

with

$$\alpha(s) = \frac{1}{100} + 1\,000 \begin{cases} \sqrt{\frac{1}{4} - s}, & \text{if } s < \frac{1}{4}, \\ \cos \pi s, & \text{otherwise.} \end{cases} \tag{92}$$

Note that the initial grid is chosen so that it aligns with the discontinuity.

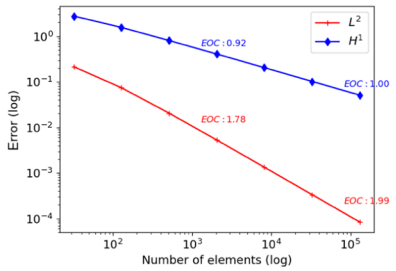
We study the behaviour of our method for polynomial degrees $k = 1, 2, 3$, choosing the forcing such that the exact solution is given by the following two choices.

1) (Smooth solution): $u(\boldsymbol{x}) = \sin 2\pi x_1 \sin 2\pi x_2$.
2) ($H^2 \backslash H^3$ solution): $u(\boldsymbol{x}) = \begin{cases} \frac{1}{4}\Big(\cos 8\pi \Big|\boldsymbol{x} - \frac{1}{2}\Big|^2 + 1\Big), & \text{if } \Big|\boldsymbol{x} - \frac{1}{2}\Big|^2 \leqslant \frac{1}{8}, \\ 0, & \text{otherwise.} \end{cases}$

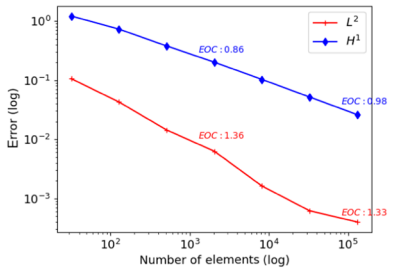
The penalty parameter is chosen as $\sigma = \lambda_{\boldsymbol{A},\max} 5k(k+1)$ where $\lambda_{\boldsymbol{A},\max}$ is the maximum eigenvalue of $\boldsymbol{A}$.

Remark 8 (Compatibility of spaces for u_h and $\boldsymbol{H}(u_h)$) Note that it is not actually required that the approximations u_h and $\boldsymbol{H}(u_h)$ are represented in the same finite element space. Computationally we observe similar results when $\boldsymbol{H}(u_h)$ is one degree lower than u_h and suboptimal convergence rates when $\boldsymbol{H}(u_h)$ is two orders lower than u_h. Other choices do not appear to be stable.

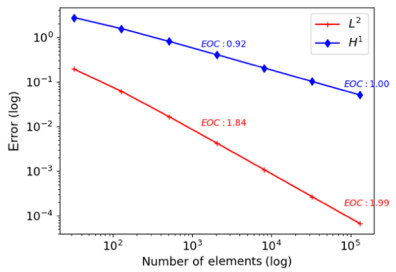
Benchmarking for these tests is shown in Figs. 1, 2 and 3. For linear polynomials, the $\mathrm{H}^2(\mathscr{T})$-error does not converge since the piecewise Hessian of u_h vanishes. So we only show the errors in the $\mathrm{L}^2(\Omega)$ and $\mathrm{H}^1(\mathscr{T})$ norms. For the smooth solution (left column), we clearly see that the method converges optimally in both norms. While for the less smooth solution convergence is slowed as would be expected when approximating an $\mathrm{H}^2(\Omega)$ solution. For polynomial orders 2 and 3 convergence is optimal for all norms studied here when approximating a smooth solution and between 1 and 1.5 for



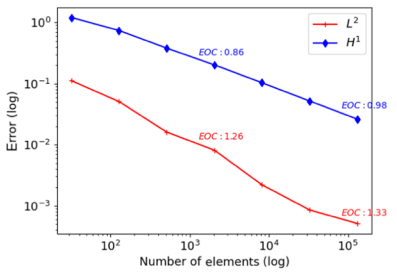
(a) $u \in C^\infty(\Omega)$ and $\boldsymbol{A}$ is coercive



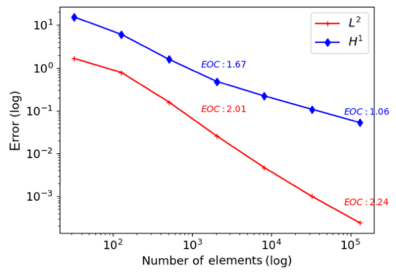
(b) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is coercive



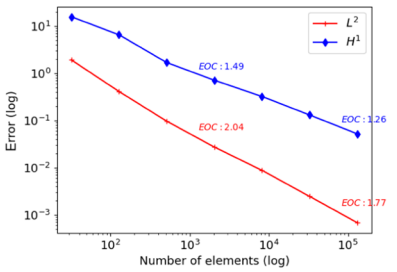
(c) $u \in C^\infty(\Omega)$ and $\boldsymbol{A}$ is continuous but not $\mathrm{H}^2(\Omega)$



(d) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is continuous but not $\mathrm{H}^2(\Omega)$



(e) $u \in C^\infty(\Omega)$ and $\boldsymbol{A}$ is discontinuous

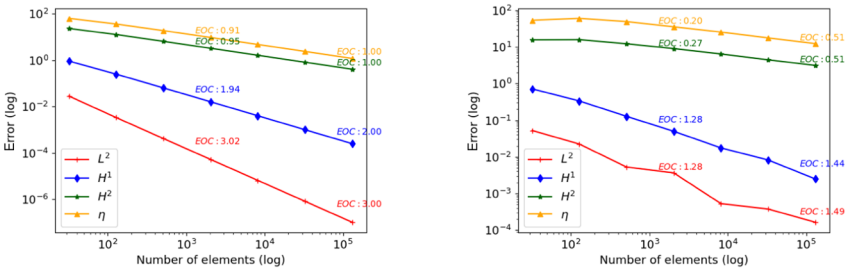


(f) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is discontinuous

Fig. 1 Convergence rates for the error measured in $\mathrm{H}^1(\Omega)$ and $\mathrm{L}^2(\Omega)$ for $k = 1$ testing the case when u is either a prescribed smooth solution or $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$. We also test three different choices for diffusion coefficient $\boldsymbol{A}$ that are coercive, continuous but not H^2 and discontinuous. The rates are optimal in $\mathrm{H}^1(\Omega)$ in all cases and in $\mathrm{L}^2(\Omega)$ in most cases. Note we did not present the $\mathrm{H}^2(\Omega)$ rates nor the estimate since neither will converge for $k = 1$

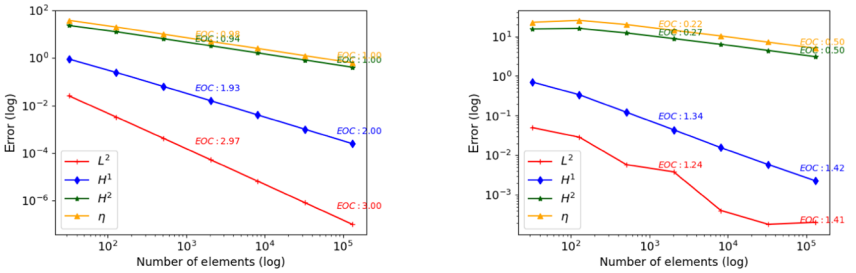
the $\mathrm{H}^2(\Omega)$ solution. The convergence rate hardly depends on the smoothness of A. The only case where some clear dependency is visible is in the $\mathrm{L}^2(\Omega)$ errors for quadratic polynomials with the discontinuous $\boldsymbol{A}$. In this case, the order of the $\mathrm{L}^2(\Omega)$ error seems to equal the convergence in the $\mathrm{H}^1(\mathscr{T})$ norm, i.e., is not optimal (see Fig. 2).

We also study the condition number of the system matrix generated when assembling (49). The results are shown in Fig. 4. In contrast to many other methods, which rewrite the



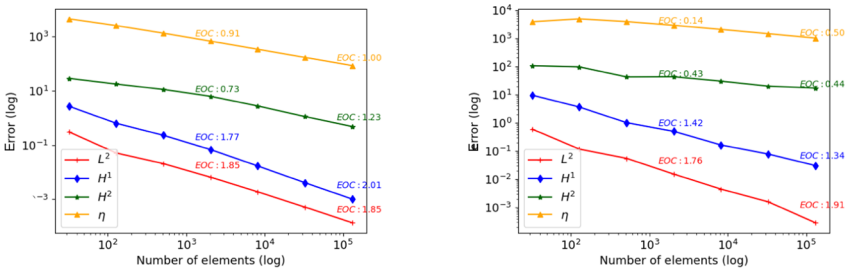
(a) $u \in C^{\infty}(\Omega)$ and $\boldsymbol{A}$ is coercive

(b) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is coercive



(c) $u \in C^{\infty}(\Omega)$ and $\boldsymbol{A}$ is continuous but not $\mathrm{H}^2(\Omega)$

(d) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is continuous but not $\mathrm{H}^2(\Omega)$



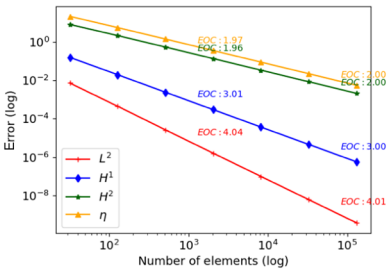
(e) $u \in C^{\infty}(\Omega)$ and $\boldsymbol{A}$ is discontinuous

(f) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is discontinuous

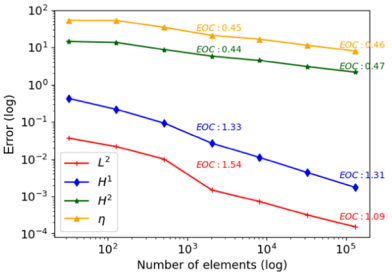
Fig. 2 Convergence rates for the error measured in $\mathrm{H}^2(\Omega)$, $\mathrm{H}^1(\Omega)$ and $\mathrm{L}^2(\Omega)$ for $k = 2$ and the estimator given in Theorem 3. We test the case when u is either a prescribed smooth solution or $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$. We also test three different choices for diffusion coefficient $\boldsymbol{A}$ that are coercive, continuous but not H^2 and discontinuous. The rates are optimal in $\mathrm{H}^2(\Omega)$ and $\mathrm{H}^1(\Omega)$ in all cases and in $\mathrm{L}^2(\Omega)$ in most cases. When the solution is not smooth the rates are slowed to an expected rate in-line with the regularity of u. In all cases, the estimate is efficient and robust

nonvariational model as a fourth-order problem, the condition number depends on the grid spacing in the same way as it does for second-order variational problem, i.e., it is $O(h^{-2})$.

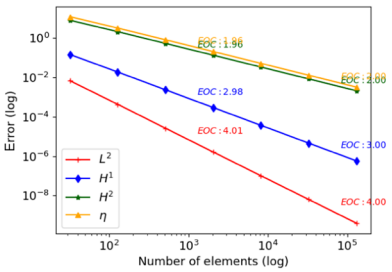
Finally, we study the behaviour of the residual error indicator and an adaptive scheme for the approximation of the solutions to these problems. The indicator is included in Figs. 2, 3 where its reliability is clearly visible. A comparison of these globally refined simulations with an adaptive simulation using an equal distribution strategy for locally refining the grid are given in Figs. 5 and 6. As to be expected, no advantage can be gained when approximating a smooth solution.



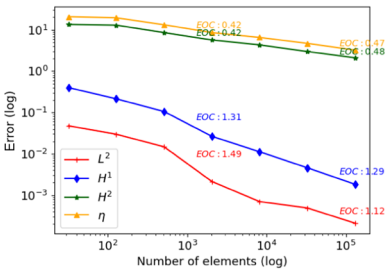
(a) $u \in C^\infty(\Omega)$ and $\boldsymbol{A}$ is coercive



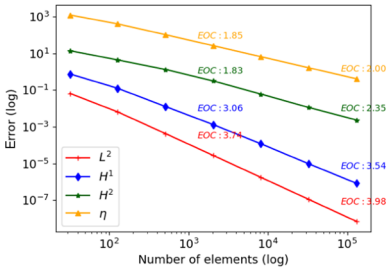
(b) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is coercive



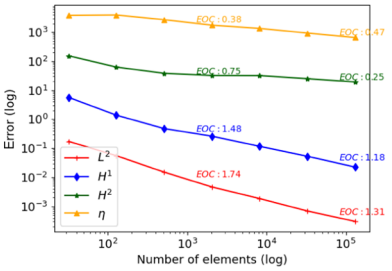
(c) $u \in C^\infty(\Omega)$ and $\boldsymbol{A}$ is continuous but not $\mathrm{H}^2(\Omega)$



(d) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is continuous but not $\mathrm{H}^2(\Omega)$



(e) $u \in C^\infty(\Omega)$ and $\boldsymbol{A}$ is discontinuous

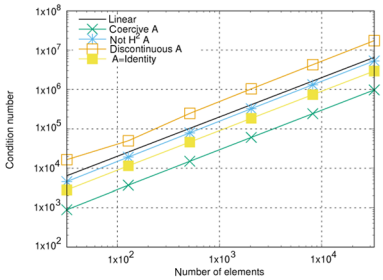


(f) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $\boldsymbol{A}$ is discontinuous

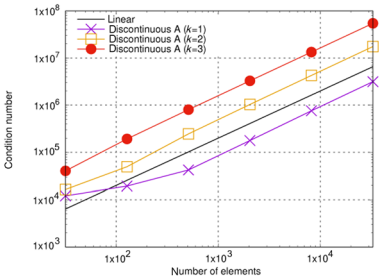
Fig. 3 Convergence rates for the error measured in $\mathrm{H}^2(\Omega)$, $\mathrm{H}^1(\Omega)$ and $\mathrm{L}^2(\Omega)$ for $k = 3$ and the estimator given in Theorem 3. We test the case when u is either a prescribed smooth solution or $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$. We also test three different choices for diffusion coefficient $\boldsymbol{A}$ that are coercive, continuous but not H^2 and discontinuous. The rates are optimal in $\mathrm{H}^2(\Omega)$ and $\mathrm{H}^1(\Omega)$ in all cases and in $\mathrm{L}^2(\Omega)$ in most cases. When the solution is not smooth the rates are slowed to an expected rate in-line with the regularity of u. In all cases, the estimate is efficient and robust.

For the $\mathrm{H}^2(\Omega)$ solution, the adaptive simulation approximately doubles the convergence rate of the scheme.

We conclude with a simulation for which we do not have an exact solution. We use the discontinuous $\boldsymbol{A}$ and choose a constant forcing $f \equiv 1\,000$. As before boundary conditions are equal to zero. Results are shown in Fig. 6 where we show the convergence of the residual indicator under global and local refinement. A visualisation of the function a, the resulting discrete solution, and the adaptive grid are given in Fig. 7.

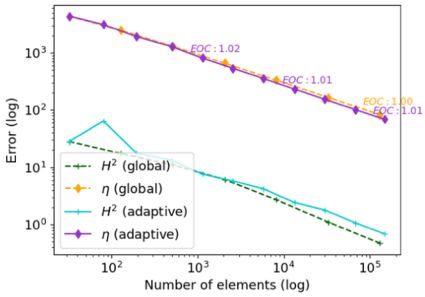


(a) Condition numbers for each of the operators (5.1). For comparison purposes we included the case of the Laplace problem ($\boldsymbol{A} = I$) where our method corresponds to the IP-DG method

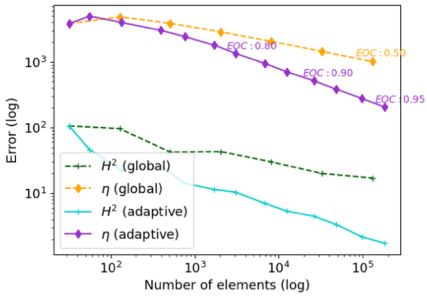


(b) Condition numbers for discontinuous coefficient $\boldsymbol{A}$ (91) with various polynomial degrees $k = 1, 2, 3$

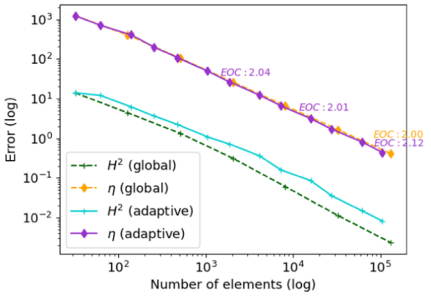
Fig. 4 Condition number estimates of the system matrix given by (49). Notice that the condition number grows in the same asymptotic fashion as the IP-DG method for the Laplacian ($\boldsymbol{A}$ =Identity). Also, the complexity does not change asymptotically as k is increased



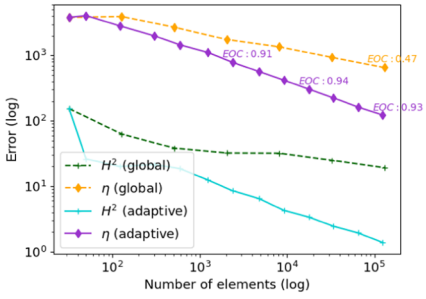
(a) $u \in C^{\infty}(\Omega)$ and $k = 2$



(b) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $k = 2$

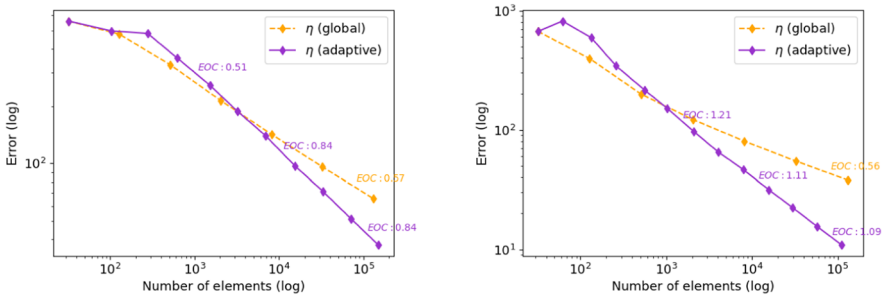


(c) $u \in C^{\infty}(\Omega)$ and $k = 3$



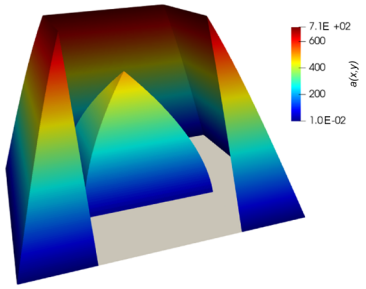
(d) $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and $k = 3$

Fig. 5 Comparison of simulations carried out on adaptive and globally refined grids driven by the estimate from Theorem 3. We test the case when u is either a prescribed smooth solution or $u \in \mathrm{H}^2(\Omega)/\mathrm{H}^3(\Omega)$ and when $\boldsymbol{A}$ is discontinuous. For the smooth solution, we find the uniform and adaptive schemes behave similarly, for the solution that is not $\mathrm{H}^3(\Omega)$, the adaptive scheme clearly outperforms the uniform one, although there is no gain from using $k = 3$

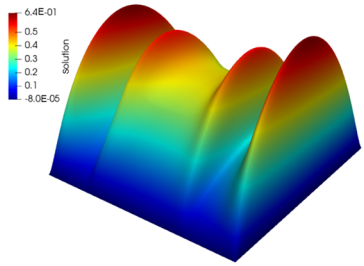


(a) Indicator with constant forcing and $k = 2$ **(b)** Indicator with constant forcing and $k = 3$

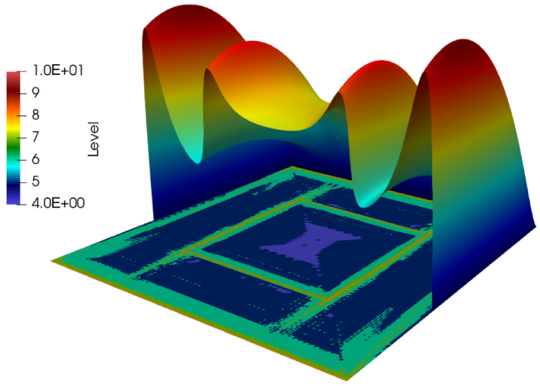
Fig. 6 Convergence of estimator under adaptive and globally refined grids. Here, we consider the case when $\boldsymbol{A}$ is discontinuous, shown in Fig. 7a, and the forcing is chosen $f = 100$. No exact solution is known for this case. There is a clear indication that the adaptive scheme outperforms the uniform one



(a) A visualisation of $a(\boldsymbol{x})$ where a slice has been cut out to show the discontinuity. The figure is scaled by 10^{-3}



(b) The adaptive approximation to the solution. Note there are boundary and interior layers to resolve



(c) The adaptive approximation has been sliced and overlaid on the underlying adaptive mesh. Colour indicates refinement level from the initial macro triangulation

Fig. 7 Visualisation of the problem coefficient, the solution and the underlying adaptive mesh refinement level. Notice the algorithm refines where the problem data are discontinuous and well as near boundary layers caused by the anisotropy of the diffusion tensor

6 Conclusions and Outlook

In this work, we have extended the framework from [32] for linear nonvariational problems to incorporate discontinuous approximations. We have derived a posteriori bounds for this problem and shown they are useful to drive adaptive algorithms. We would like to point out that the approach presented here can be directly applied to the case where the approximation u_h is chosen in a continuous finite element space but the finite element Hessian is defined in the discontinuous fashion described here.

In the numerical experiments, we note the method is well posed and converges optimally even for $\boldsymbol{A}$ that are discontinuous. The method is well suited to solve nonlinear problems [33] and this will be the topic of ongoing research.

Compliance with Ethical Standards

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Agouzal, A., Vassilevski, Y.: On a discrete Hessian recovery for P_1 finite elements. J. Numer. Math. **10**(1), 1–12 (2002)
2. Aguilera, N.E., Morin, P.: On convex functions and the finite element method. SIAM J. Numer. Anal. **47**(4), 3139–3157 (2009)
3. Alnæs, M.S., Logg, A., Ølgaard, K.B., Rognes, M.E., Wells, G.N.: Unified form language: a domain-specific language for weak formulations of partial differential equations. CoRR (2012). arXiv:1211.4047
4. Arnold, D.N., Brezzi, F., Cockburn, B., Marini, L.D.: Unified analysis of discontinuous Galerkin methods for elliptic problems. SIAM J. Numer. Anal. **39**(5), 1749–1779 (2002)
5. Barles, G., Souganidis, P.E.: Convergence of approximation schemes for fully nonlinear second order equations. Asymptot. Anal. **4**(3), 271–283 (1991)
6. Bastian, P., Blatt, M., Dedner, A., Engwer, C., Klöfkorn, R., Ohlberger, M., Sander, O.: A generic grid interface for parallel and adaptive scientific computing. I. Abstract framework. Computing **82**(2/3), 103–119 (2008)
7. Bastian, P., Blatt, M., Dedner, A., Engwer, C., Klöfkorn, R., Kornhuber, R., Ohlberger, M., Sander, O.: A generic grid interface for parallel and adaptive scientific computing. II. Implementation and tests in DUNE. Computing **82**(2/3), 121–138 (2008)
8. Böhmer, K.: On finite element methods for fully nonlinear elliptic equations of second order. SIAM J. Numer. Anal. **46**(3), 1212–1249 (2008)
9. Brenner, S.C., Sung, L.-Y.: Virtual enriching operators. Calcolo **56**(4), 1–25 (2019)
10. Buffa, A., Ortner, C.: Compact embeddings of broken Sobolev spaces and applications. IMA J. Numer. Anal. **29**(4), 827–855 (2009)
11. Burman, E., Ern, A.: Discontinuous Galerkin approximation with discrete variational principle for the nonlinear Laplacian. C. R. Math. Acad. Sci. Paris **346**(17/18), 1013–1016 (2008)
12. Cordes, H.O.: Über die erste randwertaufgabe bei quasilinearen differentialgleichungen zweiter ordnung in mehr als zwei variablen. Math. Ann. **131**(3), 278–312 (1956)

13. Dedner, A., Klöfkorn, R., Nolte, M.: Python bindings for the DUNE-FEM module. Zenodo (2020). https://doi.org/10.5281/zenodo.3706993
14. Dedner, A., Klöfkorn, R., Nolte, M., Ohlberger, M.: A generic interface for parallel and adaptive scientific computing: abstraction principles and the DUNE-FEM module. Computing **90**, 165–196 (2010)
15. Dedner, A., Nolte, M.: The Dune-Python Module. CoRR (2018). arXiv:1807.05252
16. Douglas, J. Jr., Dupont, T.: Interior penalty procedures for elliptic and parabolic Galerkin methods. In: Computing Methods in Applied Sciences (Second Internat. Sympos., Versailles, 1975). Lecture Notes in Phys., vol. 58, pp. 207–216. Springer, Berlin (1976)
17. Di Pietro, D.A., Ern, A.: Discrete functional analysis tools for discontinuous Galerkin methods with application to the incompressible Navier-Stokes equations. Math. Comput. **79**(271), 1303–1330 (2010)
18. Elman, H.C., Silvester, D.J., Wathen, A.J.: Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics. In: Numerical Mathematics and Scientific Computation. Oxford University Press, New York (2005)
19. Ern, A., Guermond, J.-L.: Theory and practice of finite elements. In: Antman, S.S., Marsden, J.E., Sirovich, L.(eds) Applied Mathematical Sciences, vol. 159. Springer, New York (2004)
20. Feng, X., Neilan, M.: Mixed finite element methods for the fully nonlinear Monge-Ampère equation based on the vanishing moment method. SIAM J. Numer. Anal. **47**(2), 1226–1250 (2009)
21. Feng, X., Neilan, M.: Vanishing moment method and moment solutions for fully nonlinear second order partial differential equations. J. Sci. Comput. **38**(1), 74–98 (2009)
22. Feng, X., Hennings, L., Neilan, M.: Finite element methods for second order linear elliptic partial differential equations in non-divergence form. Math. Comput. **86**(307), 2025–2051 (2017)
23. Feng, X., Neilan, M., Schnake, S.: Interior penalty discontinuous Galerkin methods for second order linear non-divergence form elliptic PDES. J. Sci. Comput. **74**(3), 1651–1676 (2018)
24. Gallistl, D.: Variational formulation and numerical analysis of linear elliptic equations in nondivergence form with Cordes coefficients. SIAM J. Numer. Anal. **55**(2), 737–757 (2017)
25. Gallistl, D.: Numerical approximation of planar oblique derivative problems in nondivergence form. Math. Comput. **88**(317), 1091–1119 (2019)
26. Georgoulis, E.H., Houston, P., Virtanen, J.: An a posteriori error indicator for discontinuous Galerkin approximations of fourth-order elliptic problems. IMA J. Numer. Anal. **31**(1), 281–298 (2011)
27. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order, 2nd edn. Springer, Berlin (1983)
28. Jensen, M., Smears, I.: On the convergence of finite element methods for Hamilton-Jacobi-Bellman equations. Technical report, 01 (2011)
29. Kawecki, E.L.: A DGFEM for nondivergence form elliptic equations with Cordes coefficients on curved domains. Numer. Methods Partial Differ. Equations **35**(5), 1717–1744 (2019)
30. Kawecki, E.L., Smears, I.: Convergence of adaptive discontinuous Galerkin and c^0-interior penalty finite element methods for Hamilton-Jacobi-Bellman and Isaacs equations (2020). arXiv:2006.07215
31. Lakkis, O., Mousavi, A.: A least-squares Galerkin approach to gradient and Hessian recovery for non-divergence-form elliptic equations (2019). arXiv:1909.00491
32. Lakkis, O., Pryer, T.: A finite element method for second order nonvariational elliptic problems. SIAM J. Sci. Comput. **33**(2), 786–801 (2011)
33. Lakkis, O., Pryer, T.: A nonvariational finite element method for fully nonlinear elliptic problems. Submitted–Tech report (2012). arXiv:1103.2970
34. Miranda, C.: Sulle equazioni ellittiche del secondo ordine di tipo non variazionale, a coefficienti discontinui. Ann. Mat. **63**(1), 353–386 (1963)
35. Mu, L., Ye, X.: A simple finite element method for non-divergence form elliptic equations. Int. J. Numer. Anal. Model. **14**(2), 306–311 (2017)
36. Oberman, A.M.: Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton-Jacobi equations and free boundary problems. SIAM J. Numer. Anal. **44**(2), 879–895 (2006) (electronic)
37. Pryer, T.: Discontinuous Galerkin methods for the p-biharmonic equation from a discrete variational perspective. Electron. Trans. Numer. Anal. **41**, 328–349 (2014)
38. Smears, I., Süli, E.: Discontinuous Galerkin finite element approximation of nondivergence form elliptic equations with Cordès coefficients. SIAM J. Numer. Anal. **51**(4), 2088–2106 (2013)
39. Talenti, G.: Sopra una classe di equazioni ellittiche a coefficienti misurabili. Ann. Mat. **69**(1), 285–304 (1965)
40. Vallet, M.-G., Manole, C.-M., Dompierre, J., Dufour, S., Guibault, F.: Numerical comparison of some Hessian recovery techniques. Int. J. Numer. Methods Eng. **72**(8), 987–1007 (2007)