ARTICLE IN PRESS

Journal of Computational and Applied Mathematics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Adaptive C^0 interior penalty methods for Hamilton–Jacobi–Bellman equations with Cordes coefficients

Susanne C. Brenner [a,1], Ellya L. Kawecki [b,*]

[a] Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA
[b] Department of Mathematics, University College London, London, WC1H 0AY, UK

ARTICLE INFO

Article history:
Received 12 November 2019
Received in revised form 29 May 2020

Keywords:
Finite element methods
Interior penalty methods
Partial differential equations
A posteriori estimates
Mesh refinement

ABSTRACT

In this paper we conduct a priori and a posteriori error analysis of the C^0 interior penalty method for Hamilton–Jacobi–Bellman equations, with coefficients that satisfy the Cordes condition. These estimates show the quasi-optimality of the method, and provide one with an adaptive finite element method. In accordance with the proven regularity theory, we only assume that the solution of the Hamilton–Jacobi–Bellman equation belongs to H^2.

1. Introduction

The goal of this paper is to conduct a priori and a posteriori error analysis of the C^0 interior penalty finite element method (FEM) for the approximation of strong solutions of the following nondivergence form Hamilton–Jacobi–Bellman Dirichlet boundary-value problem. Find $u : \Omega \to \mathbb{R}$ such that

$$\sup_{\alpha\in\Lambda}\{A^\alpha : D^2u - f^\alpha\} = 0 \quad \text{a.e. in } \Omega, \tag{1.1}$$

$$u = g \quad \text{on } \partial\Omega, \tag{1.2}$$

where $\Omega \subset \mathbb{R}^d$, $d \geq 2$ is convex, and g is the restriction of a given $H^2(\Omega)$ function to $\partial\Omega$. We assume that

$$\Lambda \text{ is a compact metric space, and } A, f \in C(\overline{\Omega} \times \Lambda), \tag{1.3}$$

which in turn define the collection of functions $\{f^\alpha\}_{\alpha\in\Lambda}$, $\{A^\alpha\}_{\alpha\in\Lambda}$ as follows: for each $\alpha \in \Lambda$, $f^\alpha : x \mapsto f(x,\alpha)$, $A^\alpha : x \mapsto A(x,\alpha)$. We assume that the defined collection of coefficients is uniformly elliptic in the following sense: there exist constants $0 < \mu_1 \leq \mu_2 < \infty$ such that

$$\mu_1|\xi|^2 \leq \xi^T A^\alpha \xi \leq \mu_2|\xi|^2 \quad \text{a.e. in } \Omega,\ \forall \xi \in \mathbb{R}^d, \quad \forall \alpha \in \Lambda, \tag{1.4}$$

and satisfies the following Cordes condition [1] uniformly in α: there exists $\varepsilon \in (0, 1]$ such that

$$\frac{|A^\alpha|}{\mathrm{Tr}(A^\alpha)} \leq \frac{1}{\sqrt{d-1+\varepsilon}} \quad \text{a.e. in } \Omega \quad \forall \alpha \in \Lambda. \tag{1.5}$$

* Corresponding author.
E-mail addresses: brenner@math.lsu.edu (S.C. Brenner), e.kawecki@ucl.ac.uk (E.L. Kawecki).

[1] Susanne C. Brenner was supported in part by the National Science Foundation, USA under Grant Nos. DMS-16-20273 and DMS-19-13035.

https://doi.org/10.1016/j.cam.2020.113241
0377-0427/

Please cite this article as: S.C. Brenner and E.L. Kawecki, Adaptive C^0 interior penalty methods for Hamilton–Jacobi–Bellman equations with Cordes coefficients, Journal of Computational and Applied Mathematics (2020) 113241, https://doi.org/10.1016/j.cam.2020.113241.

In the case that Λ is a singleton set, we simply assume that $A \in L^\infty(\Omega)$ satisfies (1.4)–(1.5), and $f \in L^2(\Omega)$. In this case (1.1)–(1.2) become the following linear nondivergence form elliptic equation

$$A : D^2 u = f \quad \text{a.e. in } \Omega, \tag{1.6}$$

$$u = g \quad \text{on } \partial\Omega. \tag{1.7}$$

Remarkably, in two dimensions, uniform ellipticity implies the Cordes condition (1.5) (cf. [2]).

A solution u of (1.1)–(1.2) is called a strong solution if it belongs to $H^2(\Omega)$, i.e., the weak derivatives of u up to second order belong to $L^2(\Omega)$. This means that (1.1) holds a.e. with respect to the Lebesgue measure. The linear problem (1.6)–(1.7) is of interest, as it arises in the linearisation of (1.1)–(1.2), as well as other fully nonlinear elliptic partial differential equations, such as the Monge–Ampère (MA) equation. The MA equation, and (1.1) encompass a variety of modern applications, such as differential geometry, engineering, finance, economics, and stochastic optimal control problems.

Regularity: Since each $A^\alpha \in L^\infty(\Omega; \mathbb{R}^{d\times d}_{\text{Sym}})$, under the current hypotheses, in general a strong solution $u \in H^2(\Omega)$ may not belong to $H^s(\Omega)$ for any $s > 2$. As such, we shall only assume that the true solution $u \in H^2(\Omega)$.

One should note that under different hypotheses on the behaviour of the data A, f, and $\partial\Omega$, the solution of the linear problem (1.6)–(1.7) may possess higher Sobolev regularity, and integrability, and may even be classically differentiable.

- Calderon–Zygmund theory of strong solutions [3]: if $A \in C^0(\overline{\Omega}; \mathbb{R}^{d\times d})$, $f \in L^p(\Omega)$, $1 < p < \infty$, and $\partial\Omega \in C^{1,1}$, then $u \in W^{2,p}(\Omega)$.
- Classical solutions: if $A \in C^{0,\alpha}(\overline{\Omega}; \mathbb{R}^{d\times d})$, $\alpha \in (0, 1)$, $f \in C^{0,\alpha}(\overline{\Omega})$ and $\partial\Omega \in C^{2,\alpha}$, then $u \in C^{2,\alpha}(\overline{\Omega})$.

The fully nonlinear problem (1.1)–(1.2) may also admit classical solutions, again provided that A, f, and $\partial\Omega$ are sufficiently regular. In particular, if $A, f, \partial\Omega \in C^\infty$ and Λ is a finite set, then $u \in C^0(\overline{\Omega}) \cap C^{2,\alpha}(\Omega)$ for some $\alpha > 0$ (cf. [4], Theorem 1, and note that A is not required to satisfy (1.5)). See also [5]. We seek to avoid such assumptions, as polytopal domains do not possess such regularity, and in linearising (1.1)–(1.2), we cannot in general hope that the coefficients will have these properties either. See [6–9] for finite element methods approximating elliptic equations on curved domains.

The main challenge in designing a numerical method for (1.1)–(1.2) (aside from the nonlinearity) is the nondivergence form structure of the equation. Upon linearising (1.1)–(1.2), one arrives at a sequence of problems of the form (1.6)–(1.7). However, in general one cannot express $A : D^2 u = \nabla \cdot (A\nabla u) - (\nabla \cdot A) \cdot \nabla u$, as $A \in L^\infty(\Omega)$, and thus may not possess sufficient regularity. This means that (1.6)–(1.7) (and resultingly (1.1)–(1.2)) does not possess a weak formulation, and so, one cannot base a finite element method on that weak formulation. That said, this has not stopped the development of numerical methods for (1.1)–(1.2) and (1.6)–(1.7), often relying on the existence and uniqueness theory of the underlying equation, with methods dependent upon the different assumptions upon the coefficients and data, domain boundary, and resulting solution regularity outlined above. In particular, when $A \in L^\infty(\Omega; \mathbb{R}^{d\times d})$, $f \in L^2(\Omega)$, and Ω is convex, one has [7,10,11], and if $A \in C^0(\overline{\Omega}; \mathbb{R}^{d\times d})$, $f \in L^p(\Omega)$, $1 < p < \infty$, and $\partial\Omega \in C^{1,1}$ one has [12,13].

The papers [7,14] present and analyse discontinuous Galerkin FEMs that utilise a discrete analogue of the Miranda–Talenti estimate; the current paper utilises a similar approach. However, the method of this paper does not involve the inclusion of additional bilinear forms which numerically enforce a discrete Miranda–Talenti estimate (as in [7,14]), and thus is simpler to implement.

The approach of [10] is a mixed FEM, also relying on a variant of the Miranda–Talenti estimate, in this paper, the author was successful in proving a priori and a posteriori error estimates, as well as convergence of the adaptive method. This approach was further extended to the nonlinear setting of (1.1)–(1.2) in [15].

The papers [12,13] both employ a numerical analogue of the freezing of coefficients technique utilised in the Calderon–Zygmund theory of strong solutions to (1.1)–(1.2), however, the method of the present paper allows for more general coefficients and domains. For FEMs approximating (1.1) with oblique boundary conditions, see [6,9].

The fully nonlinear setting of (1.1)–(1.2) has seen several advancements in the literature, in the elliptic case [15–20], as well as the parabolic setting [21]. The most recent development (to the knowledge of the authors) [20] relies on a discrete Miranda–Talenti estimate for continuous finite element functions.

The following estimates

$$\|D^2 v\|_{L^2(\Omega)} \leq \|\Delta v\|_{L^2(\Omega)}, \quad \forall v \in H^2(\Omega) \cap H^1_0(\Omega) \tag{1.8}$$

$$\|v\|_{H^2(\Omega)} \leq C\|\Delta v\|_{L^2(\Omega)}, \forall v \in H^2(\Omega) \cap H^1_0(\Omega) \tag{1.9}$$

are the so called Miranda–Talenti estimates, and hold when the domain Ω is convex. The approaches of [19,20] rely upon renormalising the nonlinear problem with the following parameter,

$$\gamma^\alpha := \frac{A^\alpha : I}{A^\alpha : A^\alpha} \in L^\infty(\Omega), \tag{1.10}$$

for each $\alpha \in \Lambda$.

Theorem 3 of [19] provides the existence and uniqueness of a function u belonging to the space

$$H := H^2(\Omega) \cap H^1_0(\Omega),$$

that satisfies (1.1)–(1.2), in the case that $g \equiv 0$. Treating the case of inhomogeneous boundary data follows in a manner similar to that of [19], Theorem 3. With the aim of invoking the Browder–Minty Theorem, we first define $F_\gamma : H^2(\Omega) \to L^2(\Omega)$ by

$$F_\gamma[u] := \sup_{\alpha\in\Lambda}\{\gamma^\alpha(A^\alpha : D^2u - f^\alpha)\}, \tag{1.11}$$

and proceed to define $a_g : H \to H'$ (where H' denotes the dual space of H) by

$$a_g(u; v) := (F_\gamma[u+g], \Delta v)_{L^2(\Omega)} \quad u, v \in H. \tag{1.12}$$

One can show that a_g is strictly monotone, and Lipschitz continuous on H, yielding the existence and uniqueness of a function $u_0 \in H$ such that

$$a_g(u_0; v) = 0 \quad \forall v \in H. \tag{1.13}$$

Finally, we uniquely define $u := u_0 + g$, which satisfies (1.1)–(1.2). This provides us with the following theorem.

Theorem 1.1. *Assume that $\Omega \subset \mathbb{R}^d$ is a convex domain, and that the collection $\{A^\alpha\}_{\alpha\in\Lambda}$ satisfies (1.4)–(1.5). Furthermore, assume that $g \in H^2(\Omega)$. Then, there exists a unique strong solution $u \in H^2(\Omega)$ of the following HJB equation:*

$$\begin{aligned} \sup_{\alpha\in\Lambda}\{A^\alpha : D^2u - f^\alpha\} &= 0 \quad \text{a.e. in } \Omega, \\ u &= g \quad \text{on } \partial\Omega. \end{aligned} \tag{1.14}$$

Contributions: In this paper we obtain a priori and a posteriori error estimates under the assumption that the true solution belongs to $H^2(\Omega)$. We note that the method we present has been considered in [20], in the homogeneous Dirichlet case, where the authors prove stability, and a priori error estimates for the problem (1.1)–(1.2), as well as the fully nonlinear Hamilton–Jacobi–Bellman equation. Our approach to the stability analysis is distinct from that of [20], as we also consider the case of inhomogeneous boundary data. Furthermore, the recent publication [22] provides the existence of an enriching operator when $p \geq 2$, and $d \in \{2, 3\}$, which removes the restriction upon the polynomial degree $p \in \{2, 3\}$, when $d = 3$ present in [20] (cf. [20] Remark 4). Furthermore, we also undertake a posteriori error analysis for this problem, and justify that one may utilise the scheme to approximate solutions to the fully nonlinear Monge–Ampère equation (see Section 5).

As mentioned, a motivation of this paper is to develop a numerical method for the Monge–Ampère (MA) equation. The (MA) equation is a prototypical fully nonlinear elliptic equation, arising in differential geometry, optimal transport, engineering and fluid dynamics: given a nonnegative $f : \Omega \to \mathbb{R}^+$, and $g : \partial\Omega \to \mathbb{R}$, find $u : \Omega \to \mathbb{R}$ such that

$$\begin{cases} \det D^2u = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{1.15}$$

In general solutions of (1.15) may not be unique; a simple example is given in the case $d = 2$, $g \equiv 0$, where it is clear that if u satisfies (1.15), then so does $-u$. The existence of multiple solutions to (1.15) poses a significant challenge in the design of numerical methods. In [23] (c.f. [23] Section 1.4), the authors implement a standard nine-point stencil finite difference method (FDM) for an example of (1.15) that has at most two solutions, with a smooth right-hand side, and with the choice of domain $\Omega = (0, 1)^2$. One would hope that the proposed FDM may have the same uniqueness property, that is, that there exist at most two solutions to the numerical method. However, upon implementing this method on a 4×4 grid, and solving the resulting nonlinear system by applying Newton's method, they obtain sixteen different numerical solutions by varying the initial guess of the Newton's method.

As mentioned in [23], one may conjecture that this phenomena extrapolates, causing Newton's method to potentially converge to $2^{(N-2)^2}$ different solutions on and $N \times N$ grid, by varying the initial guess. When designing a numerical scheme, it is important that one knows which solution the method is converging to, without needing too much prior knowledge of the true solution (Newton's method is well known to be conditionally convergent, often requiring that the initial guess is sufficiently close to the true solution). A variant of the FDM implemented and discussed in [23] was proposed in [24], with an additional *selection* criteria, which in essence singles out a particular numerical solution.

We overcome this difficulty, by using a long standing result due to N. Krylov [25], which allows one to characterise the MA equation (1.15) as a HJB equation, if and only if u is convex. In the case that $u \in W^{2,\infty}(\Omega)$ is *uniformly* convex, and $d = 2$, we are able to further show that the resulting HJB equation is equivalent to one with a control set Λ, and data A, f that satisfy (1.3)–(1.5). Moreover, the resulting numerical scheme is uniquely solvable. For other numerical methods for the approximation of solutions to the MA problem, see [16,26–28].

Domain assumptions: In the current section, we have assumed that $\Omega \subset \mathbb{R}^d$, $d \geq 2$, is convex, as this is a sufficient assumption of Theorem 1.1. However, in Section 2 we provide the numerical scheme, and from this point on, we further assume that $d \in \{2, 3\}$ and that Ω is polytopal.

This paper is laid out as follows. In Section 2, we introduce the discrete problem, and prove the stability of the associated bilinear form. Section 3 is devoted to convergence analysis; we prove quasi-optimal apriori error estimates and a posteriori error estimates in a H^2-type norm. In Section 4, we propose the linearisation scheme and adaptive

scheme. Section 5 is devoted to applications to the Monge–Ampère problem. In Section 6 we implement the proposed finite element method (as well as the adaptive version) in FEniCS [29], confirming the theoretical results of the paper. Finally, in Section 7, we provide concluding remarks on what has been achieved in this paper.

2. The discrete problem

As mentioned in the introduction, from this point on, we shall further assume that $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$ is convex and polytopal. Let $\mathscr{T}_h$ be a simplicial triangulation of Ω and $V_h \subset H^1(\Omega)$ be the continuous Lagrange finite element space of order $p \geq 2$ associated with $\mathscr{T}_h$, and denote $V_{h,0} := V_h \cap H^1_0(\Omega)$. We denote by D^2_h and Δ_h, the piecewise Hessian and Laplacian, respectively. Furthermore, we shall make use of the following mesh dependent (semi)norm for $u \in H^2(\Omega; \mathscr{T}_h) := \{v \in L^2(\Omega) : v|_K \in H^2(K) \forall K \in \mathscr{T}_h\}$

$$\|u\|_h^2 := \int_\Omega |D^2_h u|^2 + \sum_{e \in \mathscr{E}_h^i} \frac{\sigma}{h_e} \|[\![\partial u/\partial n]\!]_e\|^2_{L^2(e)}, \tag{2.1}$$

and we note that $\|\cdot\|_h$ is indeed a norm on $V_{h,0}$.

The discrete problem is posed as follows: we seek $u_h \in V_h$ satisfying

$$\begin{aligned} a_h(u_h; v) &:= \int_\Omega F_\gamma[u_h]\, \Delta_h v + \sum_{e \in \mathscr{E}_h^i} \frac{\sigma}{h_e} \int_e [\![\partial u_h/\partial n]\!]_e [\![\partial v/\partial n]\!]_e = 0 \quad \forall v \in V_{h,0}, \\ u_h|_{\partial\Omega} &:= g_h, \end{aligned} \tag{2.2}$$

where $g_h \in V_h$ is a suitable approximation of g (the derivatives in F_γ defined by (1.11) are considered piecewise), $\mathscr{E}_h^i$ is the set of internal edges of $\mathscr{T}_h$, $[\![\cdot]\!]_e$ denotes the jump across an edge e, and σ is a positive constant.

We remark that if $g \equiv 0$, and we instead seek $u_h \in V_{h,0}$, then (2.2) coincides with the method presented in [20]. We also note that the scheme is consistent in the following sense: if $u \in H^2(\Omega)$ satisfies (1.1)–(1.2), then

$$a_h(u; v) = 0 \quad \forall v \in V_{h,0}. \tag{2.3}$$

The above holds, since u satisfies (1.1)–(1.2), and $u \in H^2(\Omega)$ so $[\![\partial u/\partial n]\!]_e = 0$ for $e \in \mathscr{E}_h^i$. The following theorem and corollary are from [20]. As mentioned in the introduction, the results that follow, as presented in [20] hold for $d = 2$, for any $p \geq 2$, and for $d = 3$ if $p \in \{2, 3\}$. However, this occurs because the proofs rely on the existence of an operator $E_h : V_{h,0} \to H^2(\Omega) \cap H^1_0(\Omega)$ (called an enriching operator), that in particular satisfies the following estimate:

$$\|E_h v - v\|_h \leq C_* \sum_{e \in \mathscr{E}_h^i} \frac{1}{h_e} \|[\![\partial v/\partial n]\!]_e\|^2_{L^2(e)}, \quad \forall v \in V_{h,0} \tag{2.4}$$

where the constant C_* is (in principle) a computable, positive constant dependent only on the shape regularity of $\mathscr{T}_h$. A particular construction of such an operator is provided in [20] and uses the C^1 family of Clough–Tocher spaces, which leads to the aforementioned restriction when $d = 3$ (cf. [20], Remark 4). However, in the recent paper [22], the existence of an operator that satisfies (2.4) has been proven, only assuming $p \geq 2$, for $d \in \{2, 3\}$. Thus, the proceeding results hold for $d \in \{2, 3\}$ and $p \geq 2$.

Theorem 2.1. *One has that for any $v_h \in V_{h,0}$,*

$$\|D^2_h v_h\|_{L^2(\Omega)} \leq \|\Delta_h v_h\|_{L^2(\Omega)} + C_{\mathbf{MT}} \left(\sum_{e \in \mathscr{E}_h^i} \frac{1}{h_e} \|[\![\partial v_h/\partial n]\!]\|^2_{L^2(e)} \right)^{1/2}, \tag{2.5}$$

where the constant $C_{\mathbf{MT}}$ is independent of h.

Corollary 2.2. *One has that for any $v \in V_{h,0}$, and all $t \in (0, 1)$,*

$$\|\Delta_h v\|^2_{L^2(\Omega)} \geq (1 - t)\|D^2_h v_h\|^2_{L^2(\Omega)} - \frac{C^2_{\mathbf{MT}}}{t} \left(\sum_{e \in \mathscr{E}_h^i} \frac{1}{h_e} \|[\![\partial v/\partial n]\!]\|^2_{L^2(e)} \right). \tag{2.6}$$

We now prove a strict monotonicity result for a_h (a variant of [20], Lemma 7), provided that σ is sufficiently large.

Lemma 2.3. *One has that for any $u, v \in V_h$, such that $u - v \in V_{h,0}$*

$$a_h(u; u - v) - a_h(v; u - v) \geq \delta(1 - \sqrt{1 - \varepsilon})\|u - v\|^2_h,$$

for any $\delta \in (0, 1)$, independent of h, u and v, provided σ is sufficiently large (dependent on δ).

Proof. Take u, v, as in the hypotheses of the lemma, and denote $w := u - v \in V_{h,0}$. Denoting I to be the $d \times d$ identity matrix, by (1.5), we have that

$$\begin{aligned} |\gamma^\alpha A^\alpha - I|^2 &= (\gamma^\alpha A^\alpha - I) : (\gamma^\alpha A^\alpha - I) = (\gamma^\alpha)^2 (A^\alpha : A^\alpha) - 2\gamma^\alpha (A^\alpha : I) + I : I \\ &= -\frac{(A^\alpha : I)^2}{(A^\alpha : A^\alpha)} + d \\ &\leq -(d - 1 + \varepsilon) + d = 1 - \varepsilon \quad \text{a.e. in } \Omega, \quad \forall \alpha \in \Lambda. \end{aligned} \tag{2.7}$$

Inequality (2.7), Theorem 2.1, and Corollary 2.2 imply that for any $t \in (0, 1)$

$$\begin{aligned} \int_\Omega (F_\gamma[u] - F_\gamma[v])\, \Delta_h w &\geq \|\Delta_h w\|^2_{L^2(\Omega)} - \int_\Omega \sup_{\alpha \in \Lambda} \{|(\gamma^\alpha A^\alpha - I) : D_h^2 w|\} |\Delta_h w| \\ &\geq \|\Delta_h w\|^2_{L^2(\Omega)} - \sqrt{1-\varepsilon} \|D_h^2 w\|_{L^2(\Omega)} \|\Delta_h w\|_{L^2(\Omega)} \\ &\geq (1 - \sqrt{1-\varepsilon}/2) \|\Delta_h w\|^2_{L^2(\Omega)} - (\sqrt{1-\varepsilon}/2) \|D_h^2 w\|^2_{L^2(\Omega)} \\ &\geq \left[(1-t)(1 - \sqrt{1-\varepsilon}/2) - \sqrt{1-\varepsilon}/2 \right] \|D_h^2 w\|^2_{L^2(\Omega)} - \frac{C_{\mathbf{MT}}^2}{t} \sum_{e \in \mathscr{E}_h^i} \frac{1}{h_e} \|[\![\partial w / \partial n]\!]\|^2_{L^2(e)}. \end{aligned}$$

Now, for a given $\delta \in (0, 1)$, we set $t = t(\delta, \varepsilon) := (1-\delta)(1 - \sqrt{1-\varepsilon})/(1 - \sqrt{1-\varepsilon}/2) \in (0, 1)$. This gives us

$$\begin{aligned} a_h(u; u - v) - a_h(v; u - v) &\geq \delta(1 - \sqrt{1-\varepsilon}) \|w\|_h^2 \\ &\quad + \left(\sigma(1 - \delta(1 - \sqrt{1-\varepsilon})) - C_{\mathbf{MT}}^2 \frac{(1 - \sqrt{1-\varepsilon}/2)^2}{(1-\delta)(1 - \sqrt{1-\varepsilon})} \right) \sum_{e \in \mathscr{E}_h^i} \frac{1}{h_e} \|[\![\partial w / \partial n]\!]\|^2_{L^2(e)} \\ &\geq \delta(1 - \sqrt{1-\varepsilon}) \|w\|_h^2 \end{aligned}$$

provided that σ satisfies

$$\sigma \geq \frac{C_{\mathbf{MT}}^2 (1 - \sqrt{1-\varepsilon}/2)^2}{(1-\delta)(1 - \sqrt{1-\varepsilon})(1 - \delta(1 - \sqrt{1-\varepsilon}))} =: C(\delta, \varepsilon). \quad \square \tag{2.8}$$

Remark 2.4 (*Dependence of σ on ε*)**.** From (2.8), for a fixed value of $\delta \in (0, 1)$, we can see that the monotonicity of a_h requires that σ is sufficiently large, dependent on ε. From the identity $1 - \sqrt{1-\varepsilon} = \varepsilon/(1 + \sqrt{1-\varepsilon})$, we find that $\varepsilon/2 \leq 1 - \sqrt{1-\varepsilon} \leq \varepsilon$. We also have that $1 - \delta \leq 1 - \delta(1 - \sqrt{1-\varepsilon}) \leq 1$ and $1/4 \leq (1 - \sqrt{1-\varepsilon}/2)^2 \leq 1$. It follows that

$$\frac{C_{\mathbf{MT}}^2}{4(1-\delta)} \varepsilon^{-1} \leq C(\delta, \varepsilon) \leq \frac{2 C_{\mathbf{MT}}^2}{(1-\delta)^2} \varepsilon^{-1}.$$

Therefore, (2.8) holds if and only if $\sigma \geq C_\delta C_{\mathbf{MT}}^2 \varepsilon^{-1}$, for some positive constant C_δ dependent only on δ.

The following lemma is a simple consequence of the Lipschitz continuity of F_γ (with Lipschitz constant $\sqrt{d}+1$, see (3.8) below), and the definition of the norm $\|\cdot\|_h$.

Lemma 2.5. *One has that for any u, v, $w \in V_h$,*

$$|a_h(u; w) - a_h(v; w)| \leq (\sqrt{d} + 1) \|u - v\|_h \|w\|_h.$$

The following proof is motivated by the proof of Theorem 1.1.

Theorem 2.6. *Under the hypotheses of Lemma 2.3, there exists a unique $u_h \in V_h$ satisfying (2.2).*

Proof. Let us define $a_{g_h} : V_{h,0} \times V_{h,0} \to \mathbb{R}$, by $a_{g_h}(u_h; v) := a_h(u + g_h; v)$ for all u_h, $v \in V_{h,0}$. Lemmas 2.3 and 2.5 then imply that for all u_h, v, $w \in V_{h,0}$, and for any $\delta \in (0, 1)$ (so long as σ is sufficiently large, dependent on δ)

$$\begin{aligned} a_{g_h}(u_h; u_h - v) - a_{g_h}(v; u_h - v) &\geq \delta(1 - \sqrt{1-\varepsilon}) \|u_h - v\|_h^2, \\ |a_{g_h}(u_h; w) - a_{g_h}(v; w)| &\leq C \|u_h - v\|_h \|w\|_h, \end{aligned}$$

where the constant C is independent of u_h, v, w. Thus a_{g_h} is strongly monotone and Lipschitz continuous, and so the Browder–Minty Theorem implies the existence and uniqueness of $u_{h,0} \in V_{h,0}$ such that

$$a_{g_h}(u_{h,0}, v) = a_h(u_{h,0} + g_h; v) = 0 \quad \forall v \in V_{h,0}.$$

Thus, we may uniquely define $u_h := u_{h,0} + g_h$, which satisfies (2.2). $\square$

3. Convergence analysis

3.1. A priori error analysis

For the remainder of the paper, we assume that the parameter σ is chosen such that there exists a unique $u_h \in V_h$ satisfying (2.2).

Remark 3.1 (*Choice of g_h*). In practice, one may use a variety of numerical approximations g_h of g, for example, the L^2 projection, or some suitable interpolant. However, for the density argument of Remark 3.3 it is useful to define g_h to be the unique element of V_h that satisfies

$$\int_\Omega D_h^2 g_h : D_h^2 v + \int_{\partial\Omega} g_h v + \sum_{e\in\mathscr{E}_h^i} \frac{\sigma}{h_e}\int_e [\![\partial g_h/\partial n]\!]_e [\![\partial v/\partial n]\!]_e = \int_\Omega D_h^2 g : D_h^2 v + \int_{\partial\Omega} g v \quad \forall v \in V_h. \tag{3.1}$$

We first prove a quasi-optimal error estimate for the error $\|u-u_h\|_h$, where $u \in H^2(\Omega)$ satisfies (1.1)–(1.2). Let $v \in V_h$ satisfy $v|_{\partial\Omega} = g_h$, where g_h satisfies (3.1). The triangle inequality gives us

$$\|u-u_h\|_h \le \|u-v\|_h + \|v-u_h\|_h. \tag{3.2}$$

Lemma 2.3, (2.3), and Lemma 2.5 imply that for any $\delta \in (0,1)$ (denoting $c_{\delta,\varepsilon} := \delta(1-\sqrt{1-\varepsilon})$)

$$\begin{aligned} c_{\delta,\varepsilon}\|v-u_h\|_h^2 &\le a_h(u_h; u_h - v) - a_h(v; u_h - v) \\ &= a_h(u; u_h - v) - a_h(v; u_h - v) \\ &\le (\sqrt{d}+1)\|u-v\|_h \|u_h - v\|_h. \end{aligned}$$

Thus,

$$\|v-u_h\|_h \le c_{\delta,\varepsilon}^{-1}(\sqrt{d}+1)\|u-v\|_h. \tag{3.3}$$

Combining (3.2) with (3.3), we arrive at the following quasi-optimal error estimate.

Theorem 3.2. *If $u_h \in V_h$ satisfies (2.2), then*

$$\|u-u_h\|_h \le C_\sharp \Big(\inf_{v\in V_h : v|_{\partial\Omega} = g_h} \|u-v\|_h\Big), \tag{3.4}$$

where $C_\sharp := 1 + \delta^{-1}(1-\sqrt{1-\varepsilon})^{-1}(\sqrt{d}+1)$.

Remark 3.3. Estimate (3.4) in combination with a density argument shows that

$$\lim_{h\to 0} \|u-u_h\|_h = 0.$$

Moreover, the Poincaré–Friedrichs inequality for piecewise H^2 functions (cf. [30,31]), implies that there exists a positive constant C independent of h such that

$$\|u-u_h\|_{H^1(\Omega)} + \|u-u_h\|_{L^\infty(\Omega)} \le C\|u-u_h\|_h, \tag{3.5}$$

and so

$$\lim_{h\to 0}(\|u-u_h\|_{H^1(\Omega)} + \|u-u_h\|_{L^\infty(\Omega)}) = 0.$$

3.2. A posteriori error analysis

The *a posteriori* error analysis is based on an enriching operator $E_h : V_{h,0} \to H^2(\Omega)\cap H_0^1(\Omega)$ that satisfies (2.4). We first consider the homogeneous case, $g \equiv 0$. In this case, we have that

$$\|u-u_h\|_h \le \|u - E_h u_h\|_h + \|u_h - E_h u_h\|_h, \tag{3.6}$$

and note that the monotonicity of a_g on H implies that

$$\|u-E_h u_h\|_h^2 = \|D^2(u-E_h u_h)\|_{L^2(\Omega)}^2 \le \frac{a_g(u; u-E_h u_h) - a_g(E_h u_h; u - E_h u_h)}{1-(1-\varepsilon)^{\frac{1}{2}}}. \tag{3.7}$$

Furthermore, it follows from (1.8), (1.12), and (1.13), that

$$\begin{aligned} &a_g(u; u-E_h u_h) - a_g(E_h u_h; u - E_h u_h) \\ &\quad = -(F_\gamma[E_h u_h], \Delta(u-E_h u_h))_{L^2(\Omega)} \\ &\quad = (-F_\gamma[u_h] + (F_\gamma[u_h] - F_\gamma[E_h u_h]), \Delta(u-E_h u_h))_{L^2(\Omega)} \\ &\quad \le (\|F_\gamma[u_h]\|_{L^2(\Omega)} + (\sqrt{d}+1)\|u_h - Eu_h\|_h)\|u - Eu_h\|_h, \end{aligned} \tag{3.8}$$

where we used the inequality

$$\sup_{\alpha\in\Lambda}|\gamma^\alpha A^\alpha| \le \sup_{\alpha\in\Lambda}|\gamma^\alpha A^\alpha - I| + |I| \le \sqrt{1-\varepsilon} + \sqrt{d} < \sqrt{d}+1 \tag{3.9}$$

that follows from (2.7). Combining (3.7) and (3.8), we find

$$\|u-E_hu_h\|_h \le \frac{1}{1-(1-\varepsilon)^{\frac12}}\left(\|F_\gamma[u_h]\|_{L^2(\Omega)} + (\sqrt{d}+1)\|u_h-E_hu_h\|_h\right), \tag{3.10}$$

which, together with (3.6) implies

$$\|u-u_h\|_h \le \frac{1}{1-(1-\varepsilon)^{\frac12}}\left(\|F_\gamma[u_h]\|_{L^2(\Omega)} + (\sqrt{d}+2)\|u_h-E_hu_h\|_h\right). \tag{3.11}$$

In view of (2.4) and (3.11), we arrive at the following *a posteriori* error estimate.

Theorem 3.4. *If $g\equiv 0$, then we have that*

$$\|u-u_h\|_h \le \frac{1}{1-(1-\varepsilon)^{\frac12}}\left(\|F_\gamma[u_h]\|_{L^2(\Omega)} + (\sqrt{d}+2)\sqrt{C_*}\left(\sum_{e\in\mathscr{E}_h^i}\frac{1}{h_e}\int_e[\![\partial u_h/\partial n]\!]_e^2\,ds\right)^{\frac12}\right). \tag{3.12}$$

We utilise Theorem 3.4 to prove the analogous result in the inhomogeneous setting.

Theorem 3.5. *We have that*

$$\begin{aligned}\|u-u_h\|_h \le{}& \frac{\|F_\gamma[u_h]\|_{L^2(\Omega)} + (\sqrt{d}+2)(1+\sqrt{C_*/\sigma})\|g-g_h\|_h}{1-(1-\varepsilon)^{\frac12}}\\ &+ \frac{(\sqrt{d}+2)\sqrt{C_*}\left(\sum_{e\in\mathscr{E}_h^i}\frac{1}{h_e}\int_e[\![\partial u_h/\partial n]\!]_e^2\,ds\right)^{\frac12}}{1-(1-\varepsilon)^{\frac12}}.\end{aligned} \tag{3.13}$$

Proof. Define $u_0 := u-g\in H^2(\Omega)\cap H_0^1(\Omega)$. We see that u_0 satisfies

$$\sup_{\alpha\in\Lambda}\{A^\alpha : D^2u_0 - g^\alpha\} = \sup_{\alpha\in\Lambda}\{A^\alpha : D^2u - f^\alpha\} = 0 \quad \text{a.e. in } \Omega,$$
$$u_0 = 0 \quad \text{on } \partial\Omega,$$

where $g^\alpha := f^\alpha - A^\alpha : D^2g$. Defining $u_{h,0} = u_h - g_h \in V_{h,0}$, by Theorem 3.4, we have that

$$\|u_0-u_{h,0}\|_h \le \frac{\|\sup_{\alpha\in\Lambda}\{\gamma^\alpha(A^\alpha : D_h^2u_{h,0}-g^\alpha)\}\|_{L^2(\Omega)} + (\sqrt{d}+2)\sqrt{C_*}\left(\sum_{e\in\mathscr{E}_h^i}\frac{1}{h_e}\int_e[\![\partial u_{h,0}/\partial n]\!]_e^2\,ds\right)^{\frac12}}{1-(1-\varepsilon)^{\frac12}}. \tag{3.14}$$

Let us denote $g_h^\alpha := f^\alpha - A^\alpha : D_h^2g_h$. The triangle inequality and (3.9) imply that

$$\begin{aligned}\|\sup_{\alpha\in\Lambda}\{\gamma^\alpha(A^\alpha : D_h^2u_{h,0}-g^\alpha)\}\|_{L^2(\Omega)} &\le \|\sup_{\alpha\in\Lambda}\{\gamma^\alpha(A^\alpha : D_h^2u_{h,0}-g_h^\alpha)\}\|_{L^2(\Omega)}\\ &\quad + \|\sup_{\alpha\in\Lambda}\{\gamma^\alpha(A^\alpha : D_h^2u_{h,0}-g^\alpha)\} - \sup_{\alpha\in\Lambda}\{\gamma^\alpha(A^\alpha : D_h^2u_{h,0}-g_h^\alpha)\}\|_{L^2(\Omega)}\\ &\le \|F_\gamma[u_h]\|_{L^2(\Omega)} + (\sqrt{d}+1)\|g-g_h\|_h,\end{aligned}$$

as well as

$$\left(\sum_{e\in\mathscr{E}_h^i}\frac{1}{h_e}\int_e[\![\partial u_{h,0}/\partial n]\!]_e^2\,ds\right)^{\frac12} \le \left(\sum_{e\in\mathscr{E}_h^i}\frac{1}{h_e}\int_e[\![\partial u_h/\partial n]\!]_e^2\,ds\right)^{\frac12} + \|g-g_h\|_h/\sqrt{\sigma}.$$

Applying the above two estimates to (3.14), and using the triangle inequality once more provides

$$\begin{aligned}\|u-u_h\|_h &\le \|u_0-u_{h,0}\|_h + \|g-g_h\|_h\\ &\le \frac{\|F_\gamma[u_h]\|_{L^2(\Omega)} + (\sqrt{d}+2)(1+\sqrt{C_*/\sigma})\|g-g_h\|_h + (\sqrt{d}+2)\sqrt{C_*}\left(\sum_{e\in\mathscr{E}_h^i}\frac{1}{h_e}\int_e[\![\partial u_h/\partial n]\!]_e^2\,ds\right)^{\frac12}}{1-(1-\varepsilon)^{\frac12}},\end{aligned}$$

as desired. □

According to Theorem 3.5, the error estimator

$$\eta_h := \|F_\gamma[u_h]\|_{L^2(\Omega)} + \|D_h^2(g-g_h)\|_{L^2(\Omega)} + \left(\sum_{e\in\mathscr{E}_h^i}\frac{1}{h_e}\int_e [\![\partial g_h/\partial n]\!]_e^2\,ds\right)^{\frac{1}{2}} + \left(\sum_{e\in\mathscr{E}_h^i}\frac{1}{h_e}\int_e [\![\partial u_h/\partial n]\!]_e^2\,ds\right)^{\frac{1}{2}}, \tag{3.15}$$

is reliable. On the other hand, the local efficiency of η_h (modulo data approximation terms) is obvious because

$$\|F_\gamma[u_h]\|_{L^2(\Omega)} \le \|\sup_{\alpha\in\Lambda}\{|\gamma^\alpha A^\alpha : D_h^2(u_h-u)|\}\|_{L^2(\Omega)} \le (\sqrt{d}+1)\|D_h^2(u_h-u)\|_{L^2(\Omega)}, \tag{3.16}$$

$$\frac{1}{h_e}\int_e [\![\partial u_h/\partial n]\!]_e^2\,ds = \frac{1}{h_e}\int_e [\![\partial(u-u_h)/\partial n]\!]_e^2\,ds \quad \forall e\in\mathscr{E}_h^i. \tag{3.17}$$

We denote the local indicators as follows for $e\in\mathscr{E}_h^i$, and $K\in\mathscr{T}_h$:

$$\eta_K(u_h) := \|F_\gamma[u_h]\|_{L^2(K)}, \qquad \eta_K^{g_h} := \|D^2(g-g_h)\|_{L^2(K)},$$
$$\eta_e^2(u_h) := \frac{1}{h_e}\|[\![\partial u_h/\partial n]\!]\|_{L^2(e)}^2, \qquad \eta_e^2(g_h) := \frac{1}{h_e}\|[\![\partial g_h/\partial n]\!]\|_{L^2(e)}^2.$$

4. Iterative scheme

Since the form a_h is nonlinear in the first argument, we shall employ an iterative scheme, in order to approximate the solution of (1.1)–(1.2). The method itself is referred to as a semismooth Newton's method (described in Algorithm 1), we cannot apply classical Newton's method, since a_h is not classically differentiable in the first argument, due to the presence of the supremum. The semismooth Newton's method presented is also provided in [20], and superlinear convergence results for a similar (discontinuous Galerkin) finite element method are proven in [19]. This particular semismooth Newton's method is also known as Howard's Algorithm [32,33].

In order to apply the semismooth Newton's method, we iteratively solve discrete problems that correspond to problems of the form (1.6)–(1.7). To this end, given a measurable function $\alpha:\Omega\to\Lambda$, let us define $a_\alpha: V_h\times V_{h,0}\to\mathbb{R}$, $\ell_\alpha: V_{h,0}\to\mathbb{R}$ by

$$a_\alpha(u,v) := (\gamma^\alpha A^\alpha : D_h^2 u, \Delta_h v)_{L^2(\Omega)} + \sum_{e\in\mathscr{E}_h^i}\frac{\sigma}{h_e}\int_e [\![\partial u_h/\partial n]\!]_e [\![\partial v/\partial n]\!]_e\,ds$$
$$\ell_\alpha(v) := (\gamma^\alpha f^\alpha, \Delta_h v)_{L^2(\Omega)}.$$

One can see that the discrete problems of finding $u\in V_h$ such that $u|_{\partial\Omega}=g_h$, and

$$a_\alpha(u,v) = \ell_\alpha(v) \quad \forall v\in V_{h,0},$$

is equivalent to (2.2), in the case that Λ is a singleton set.

Algorithm 1 Semismooth Newton's method

Require: $\Omega\subset\mathbb{R}^d$, tol $\in\mathbb{R}^+$, itermax $\in\mathbb{N}$, $\mathscr{T}_h$ a mesh on $\overline{\Omega}$, V_h, $V_{h,0}$, Λ, $\{A^\alpha,\gamma^\alpha,f^\alpha\}_{\alpha\in\Lambda}$, $u_h^0, g_h\in V_h$

```
k ← 0
r ← 1
u_h^0 ← u_h^0
while k < itermax and r > tol do
    Select an arbitrary α_k ∈ argmax F_γ[u_h^k]
    u_h^{k+1} ← the solution of
        a_{α_k}(u,v) = ℓ_{α_k}(v)  ∀v ∈ V_{h,0},
        u|_{∂Ω} = g_h                                   (4.1)
    r ← ‖u_h^{k+1} − u_h^k‖_{L^∞(Ω)}
    u_h^k ← u_h^{k+1}
    k ← k + 1
end while
```

Remark 4.1 (*Choice of* α_k)**.** The maximiser α_k in Algorithm 1 is a function $\alpha_k : \Omega \to \Lambda$. In practice, α_k may be represented as a vector with Λ-valued entries (similar to the representation of a finite element function as a vector of degrees of freedom). The dimension of this vector is typically dependent on the dimension of the finite element space (c.f. [34] Algorithm 5.1).

The following algorithm (Algorithm 2) describes the adaptive scheme. A general adaptive scheme is defined by iterating the following procedure:

$$\text{Solve} \mapsto \text{Estimate} \mapsto \text{Mark} \mapsto \text{Refine}.$$

There are several potential marking schemes that one could consider (for example Dörfler marking [35]); for the experiments of this section, we implement the maximum marking strategy (described in Algorithm 2) with newest vertex bisection (that is, a marked simplex is bisected, and then the generated node is joined to the closest vertex, so that the refinement procedure does not result in hanging nodes).

Algorithm 2 Adaptive finite element method

Require: $\Omega \subset \mathbb{R}^d$, tol $\in (0, 1)$, itermax $\in \mathbb{N}$, $\theta \in (0, 1]$, $\mathscr{T}_0$ an initial mesh on $\overline{\Omega}$
$k \leftarrow 0$, $\eta_0 \leftarrow 1$
while $k <$ itermax **and** $r >$ tol **do**
  Solve: $u_k \leftarrow$ the solution of Algorithm 1
  Estimate: For all $K \in \mathscr{T}_k$, $e \in \mathscr{E}_h^i(\mathscr{T}_k)$, calculate $\eta_K(u_k)$, $\eta_K^{g_k}$, $\eta_e(u_k)$, $\eta_e(g_k)$
  $\eta_k \leftarrow \max\{\max_{K\in\mathscr{T}_k} \eta_K(u_k), \max_{K\in\mathscr{T}_k} \eta_K^{g_k}, \max_{e\in\mathscr{E}_h^i(\mathscr{T}_k)} \eta_e(u_k), \max_{e\in\mathscr{E}_h^i(\mathscr{T}_k)} \eta_e(g_k)\}$
  Mark:
  for $e \in \mathscr{E}_h^i(\mathscr{T}_k)$ **do**
    if $\eta_e > \theta\eta_k$ **then**
      Mark e
    end if
  end for
  for $K \in \mathscr{T}_k$ **do**
    if $\eta_K > \theta\eta_k$ **then**
      Mark K
    end if
  end for
  Refine: Define $\mathscr{T}_{k+1}$ by bisecting all marked simplices, all simplices whose boundary contains a marked edge, and joining created hanging nodes to closest vertices.
  $k \leftarrow k + 1$
end while

4.1. Solving the linear problem in FEniCS

At each step of Algorithm 1, we are required to solve a linear problem of the form Eq. 4.1. This is equivalent to solving a linear system. The following code snippet details how we define the bilinear form $a(\cdot,\cdot)$ and linear form $\ell(\cdot)$ in 4.1 in FEniCS (for simplicity we drop the α_k subscript). For simplicity of exposition, we assume that A, f, g, σ, h and $\mathscr{T}_h$ are given.

```
# defining finite element
fes = FiniteElement("CG", mesh.ufl_cell(), degree)

# defining finite element space
FES = FunctionSpace(mesh, fes)

# defining trial and test functions
uh = Function(FES)
v = TestFunction(FES)

# defining boundary data as L^2 projection
gd = project(g,FES)

# defining unit normal
n = FacetNormal(mesh)

# defining penalty parameter
```

```
gamma = (A00+A11)/(pow(A00,2)+2.0*pow(A01,2)+pow(A11,2))

# defining mesh penalty parameter
sig = sigma*pow(h, 1)

# defining jump stabilisation operator
def J_h(u,v,mesh):
    J1 = sig*(n[0]('+')*(u.dx(0)('+') u.dx(0)(' '))\\
        + n[1]('+')*(u.dx(1)('+') u.dx(1)(' ')))\\
        *(n[0]('+')*(v.dx(0)('+') v.dx(0)(' '))\\
        +n[1]('+')*(v.dx(1)('+') v.dx(1)(' ')))\\
        *dS(mesh,metadata={'quadrature_degree': quad_deg})
    return J1

# defining nondivergence part of the bilinear form
def ah(u,v,mesh):
    a = gamma*(A00*u.dx(0).dx(0)+A11*u.dx(1).dx(1)+A01*u.dx(1).dx(0)\\
        +A01*u.dx(0).dx(1))*(v.dx(0).dx(0)+v.dx(1).dx(1))\\
        *dx(mesh,metadata={'quadrature_degree': quad_deg})
    return a

# defining bilinear form a
    a = ah(u,v,mesh)+J_h(u,v,mesh)

# defining linear form l
    l = gamma*(f)*(v.dx(0).dx(0)+v.dx(1).dx(1))\\
        *dx(mesh,metadata={'quadrature_degree': quad_deg})
```

Remark 4.2 (*Boundary Data*)**.** We apply the Dirichlet boundary condition using the DirichletBC function in FEniCS. In the code snippet, and in our numerical examples, we take g_h to be the L^2 projection of g onto V_h. However, one could take g_h to be the unique element of V_h satisfying (3.1).

5. Applications to the fully nonlinear Monge–Ampère equation

Let us consider the fully nonlinear Monge–Ampère (MA) equation:

$$\det D^2 u = f, \quad \text{in} \quad \Omega, \tag{5.1}$$

$$u = g, \quad \text{on} \quad \partial\Omega, \tag{5.2}$$

$$u \quad \text{is convex}, \tag{5.3}$$

where f and g are given functions, and f is assumed to be uniformly positive. Thanks to [25] we may characterise equation (5.1)–(5.3) as the following HJB problem:

$$\max_{W\in X}\{-W : D^2 u + 2f^{1/2}(\det W)^{1/2}\} = 0, \quad \text{in} \quad \Omega, \tag{5.4}$$

$$u = g, \quad \text{on} \quad \partial\Omega, \tag{5.5}$$

where $X := \{W \in \mathbb{R}^{2\times 2} : W \geq 0, W = W^T, \text{Trace}(W) = 1\}$.

However, the control set, X, contains degenerate matrices, which do not satisfy (1.5). This is remedied by the results (in particular Theorem 5.2) below, which prove that we may consider a restricted control set of matrices that satisfy (1.5) uniformly. The material that follows is present in [36], under the assumption of classical differentiability of the solution to the MA problem (5.4)–(5.5). Furthermore, similar results are also present in [25].

Theorem 5.1. *Let Ω be a bounded convex open subset of $\mathbb{R}^2$, and assume that $g \in H^2(\Omega)$, and that $f \in C^0(\overline{\Omega})$ is nonnegative. Let $X_\xi := \{W \in X : \det W \geq \xi\}$. Then, for any constant $\xi \in (0, 1/4]$, there exists a unique solution $u \in H^2(\Omega)$ of the following HJB equation*

$$\begin{aligned} \sup_{W\in X_\xi}\{-W : D^2 u + 2(\det W)^{1/2} f^{1/2}\}(x) = 0, \quad &\text{a.e. in } \Omega, \\ u(x) = g, \quad &\text{on } \partial\Omega. \end{aligned} \tag{5.6}$$

Proof. First note that as $\xi \leq 1/4$, one has that $(1/2)I \in X_\xi$, and so $X_\xi \neq \emptyset$. The set X_ξ also contains only positive definite matrices (since all elements of X_ξ are 2×2 matrices with positive trace and determinant), and in two dimensions uniform ellipticity implies the Cordes condition. Then, setting $\Lambda = X_\xi$, we can see that X_ξ is a compact metric space; using the Euclidean distance as a metric, and noting that $X_\xi = \mathscr{D}^{-1}([\xi, 1/4])$, where $\mathscr{D} : \Lambda \to \mathbb{R}$ given by

$$\mathscr{D}(W) := \det(W), \quad W \in X_\xi,$$

is a continuous function, we deduce that X_ξ is closed. Since each member of X_ξ is of unit trace, denoting the eigenvalues of $W \in X_\xi$ by λ_1, λ_2, we have that $|W|^2 = \lambda_1^2 + \lambda_2^2 = (\lambda_1 + \lambda_2)^2 - 2\lambda_1\lambda_2 = 1 - 2\det W \leq 1 - 2\xi < \infty$. Thus X_ξ is bounded. It then follows that X_ξ is compact.

We can apply Theorem 1.1, yielding existence of a unique $v \in H^2(\Omega)$ satisfying

$$\begin{cases} \sup\limits_{W \in X_\xi} \{W : D^2 v + 2(\det W)^{1/2} f^{1/2}\} = 0 \text{ in } \Omega, \\ u = -g \text{ on } \partial\Omega. \end{cases} \tag{5.7}$$

We then (uniquely) define $u := -v$. □

Theorem 5.2. *Let $d = 2$, assume that Ω is convex, that $g \in W^{2,\infty}(\Omega)$, and $f \in C^0(\overline{\Omega})$ is uniformly positive. Furthermore, assume that $u \in W^{2,\infty}(\Omega)$ is uniformly convex, and satisfies (5.1)–(5.2). Then, there exists $\xi \in (0, 1/4]$ dependent upon $|u|_{W^{2,\infty}(\Omega)}$, such that u is also the unique solution to*

$$\begin{cases} \sup\limits_{W \in X_\xi} \{-W : D^2 u + 2(\det W)^{1/2} f^{1/2}\} = 0 \quad \text{a.e. in } \Omega, \\ u = g \quad \text{on } \partial\Omega. \end{cases} \tag{5.8}$$

Proof. Let us define the map $A_u : \overline{\Omega} \to \mathbb{R}^{2\times 2}$ by:

$$A_u(x) := \frac{\operatorname{Cof}(D^2 u)}{\Delta u}, \tag{5.9}$$

note that this map is well defined, since u is uniformly convex, and so, its Laplacian is uniformly positive. Also, since $u \in W^{2,\infty}(\Omega)$, we have that $A_u \in L^\infty(\Omega)$. Furthermore, $\operatorname{Cof}(D^2 u)$ is symmetric, and

$$\operatorname{Tr}(A_u) = \frac{1}{\Delta u} \operatorname{Tr}(\operatorname{Cof}(D^2 u)) = \frac{\Delta u}{\Delta u} = 1,$$

and so $A_u : \overline{\Omega} \to X$. We see that A_u satisfies

$$\begin{aligned} &- A_u(x) : D^2 u(x) + 2\det(A_u(x))^{1/2} f^{1/2} \\ &= \frac{1}{\Delta u(x)}(-\operatorname{Cof}(D^2 u(x)) : D^2 u(x) + 2(\det(\operatorname{Cof} D^2 u(x)))^{1/2} f(x)^{1/2}) \\ &= \frac{2}{\Delta u(x)}(-\det D^2 u(x) + \det(D^2 u(x))^{1/2} f(x)^{1/2}) \\ &= \frac{2}{\Delta u(x)}(-\det D^2 u(x) + f(x)) = 0. \end{aligned} \tag{5.10}$$

We also obtain a lower bound on the determinant of A_u:

$$\begin{aligned} \det(A_u) &= \det\left(\frac{\operatorname{Cof}(D^2 u)}{\Delta u}\right) \\ &= \frac{\det(D^2 u)}{(\Delta u)^2} \\ &= \frac{f}{(\Delta u)^2} \geq \frac{\delta}{2|u|^2_{W^{2,\infty}(\Omega)}} =: \xi, \end{aligned}$$

where $\delta = \inf_{x \in \overline{\Omega}} f(x) > 0$, and so, $\xi > 0$.

Let us consider the following HJB equation: find $v \in H^2(\Omega)$ such that

$$\begin{cases} \sup\limits_{W \in X_\xi} \{-W : D^2 v + 2(\det W)^{1/2} f^{1/2}\} = 0, \ x \in \Omega, \\ v = g, \ x \in \partial\Omega. \end{cases} \tag{5.11}$$

There is an important difference between the set X and the set $X_\xi := \{W \in X : \det W \geq \xi\}$, which is that the latter set consists entirely of positive definite matrices. It then follows from Theorem 5.1 that there exists a unique $v \in H^2(\Omega)$ that satisfies (5.11).

We then see that the solution u of the MA equation satisfies (noting that $X_\xi \subseteq X$)

$$\sup_{W\in X_\xi}\{-W : D^2u + 2(\det W)^{1/2}f^{1/2}\} \le \sup_{W\in X}\{-W : D^2u + 2(\det W)^{1/2}f^{1/2}\} = 0 \quad \text{a.e. in } \Omega.$$

Since $A_u(x) \in X_\xi$ for a.e. $x \in \Omega$, from (5.10), we obtain

$$\sup_{W\in X_\xi}\{-W : D^2u + 2(\det W)^{1/2}f^{1/2}\} \ge -A_u(x) : D^2u + 2(\det A_u(x))^{1/2}f^{1/2} = 0 \quad \text{a.e. in } \Omega.$$

By combining these results, we obtain

$$\sup_{W\in X_\xi}\{-W : D^2u(x) + 2(\det W)^{1/2}f^{1/2}\} = 0 \quad \text{a.e. in } \Omega.$$

Since $u = g$ on $\partial\Omega$, and $u \in H^2(\Omega)$, by uniqueness $u = v$. □

6. Numerical results

Remark 6.1 (*PDE Coefficients*)**.** In Experiment 6.1, we consider the coefficient matrix given by $A_{ij} := (1+\delta_{ij})\frac{x_ix_j}{|x_i||x_j|}$ composed with an affine map. This example was considered in [14]. Furthermore, we multiply the coefficient matrix by an interface function χ_Ω (defined below), so that the coefficients have large jumps.

Remark 6.2 (*Monge–Ampère*)**.** In Experiment 6.3, we consider a family of Monge–Ampère type problems with true solutions that have been slightly modified from an example that is present in [28] (cf. [28], Test 4). The modifications ensure that the true solutions are uniformly convex and belong to $W^{2,\infty}(\Omega)\setminus V_h$.

6.1. Experiment 1

In this experiment, we consider the following problems

$$\begin{cases} \displaystyle\sum_{i,j=1}^{2}(1+\delta_{ij})\frac{(x_i-0.5)}{|x_i-0.5|}\frac{(x_j-0.5)}{|x_j-0.5|}\chi_\Omega^N(x_1,x_2)D^2_{ij}u_s = f_s, \quad \text{in} \quad \Omega, \\ \qquad\qquad u_s = g_s, \quad \text{on} \quad \partial\Omega, \end{cases} \tag{6.1}$$

where $\Omega = (0,1)^2$. Furthermore, the interface function χ_Ω^N satisfies $\chi_\Omega^N = 1$ on $\Omega_1 := \cup_{i,j=0}^{N/2-1}\{2i/N < x_1 < (2i+1)/N, 2j/N < x_2 < (2j+1)/N\}$, and $\chi_\Omega^N = 1000$ on $\Omega\setminus\Omega_1$. In this case we take $N = 20$. In this case f_s and g_s are chosen so that the solution of (6.1) is given by $u(x) = |x|^{1+s}$. We consider the exponent $s \in \{0.01, 0.1, 0.2, \dots, 0.5\}$. It holds that $u_s \in H^{2+\delta}(\Omega)$, for arbitrary $\delta \in [0, s]$. Furthermore, u_s lacks regularity at the origin, and one can see in Fig. 1, the error estimator prioritises refinement towards the origin, in addition to further refinement in the areas of the domain where χ_Ω is the largest. We apply both Algorithm 2 with $\theta = 0.2$, and a uniform refinement procedure, so that we may compare the two approaches. For clarity, we denote the numerical solution by $u_{h,\text{adapt}}$, and $u_{h,\text{unif}}$ for the adaptive and uniform approach, respectively. We consider a variety of values of s, and polynomial degree, p, and calculate the error in the following (semi) norms: $\|\cdot\|_{L^2(\Omega)}$, $|\cdot|_{H^1(\Omega)}$, $\|\cdot\|_h$, and also calculate the error estimator η_h.

Case 1: $p = 4$, and $s = 0.01$. We observe that

$$\begin{aligned} \|u_{0.01} - u_{h,\text{adapt}}\|_{L^2(\Omega)} &= \mathcal{O}(\text{ndofs}^{-2}), \quad \|u_{0.01} - u_{h,\text{unif}}\|_{L^2(\Omega)} = \mathcal{O}(\text{ndofs}^{-1.01}), \\ |u_{0.01} - u_{h,\text{adapt}}|_{H^1(\Omega)} &= \mathcal{O}(\text{ndofs}^{-1}), \quad |u_{0.01} - u_{h,\text{unif}}|_{H^1(\Omega)} = \mathcal{O}(\text{ndofs}^{-0.51}), \\ \eta_{\text{adapt}}, \|u_{0.01} - u_{h,\text{adapt}}\|_h &= \mathcal{O}(\text{ndofs}^{-0.01}), \quad \eta_{\text{unif}}, \|u_{0.01} - u_{h,\text{unif}}\|_h = \mathcal{O}(\text{ndofs}^{-0.005}), \end{aligned}$$

and so, the adaptive method outperforms the uniform scheme. We also plot the effectivity index in Fig. 2, verifying (3.16)–(3.17), for the adaptive scheme.

Case 2: $p = 3$, and $s \in \{0.1, \dots, 0.5\}$. We observe that

$$\begin{aligned} \|u_s - u_{h,\text{adapt},s}\|_{L^2(\Omega)} &= \mathcal{O}(\text{ndofs}^{-(2+s)}), & \|u_{0.01} - u_{h,\text{unif},s}\|_{L^2(\Omega)} &= \mathcal{O}(\text{ndofs}^{-(1+s/2)}), \\ |u_s - u_{h,\text{adapt},s}|_{H^1(\Omega)} &= \mathcal{O}(\text{ndofs}^{-(1+s)}), & |u_{0.01} - u_{h,\text{unif},s}|_{H^1(\Omega)} &= \mathcal{O}(\text{ndofs}^{-(0.5+s/2)}), \\ \|u_s - u_{h,\text{adapt},s}\|_h &= \mathcal{O}(\text{ndofs}^{-s}), & \|u_{0.01} - u_{h,\text{unif},s}\|_h &= \mathcal{O}(\text{ndofs}^{-s/2}), \\ \eta_{\text{adapt},s} &= \mathcal{O}(\text{ndofs}^{-1.11}), & \eta_{\text{unif},s} &= \mathcal{O}(\text{ndofs}^{-s/2}). \end{aligned}$$

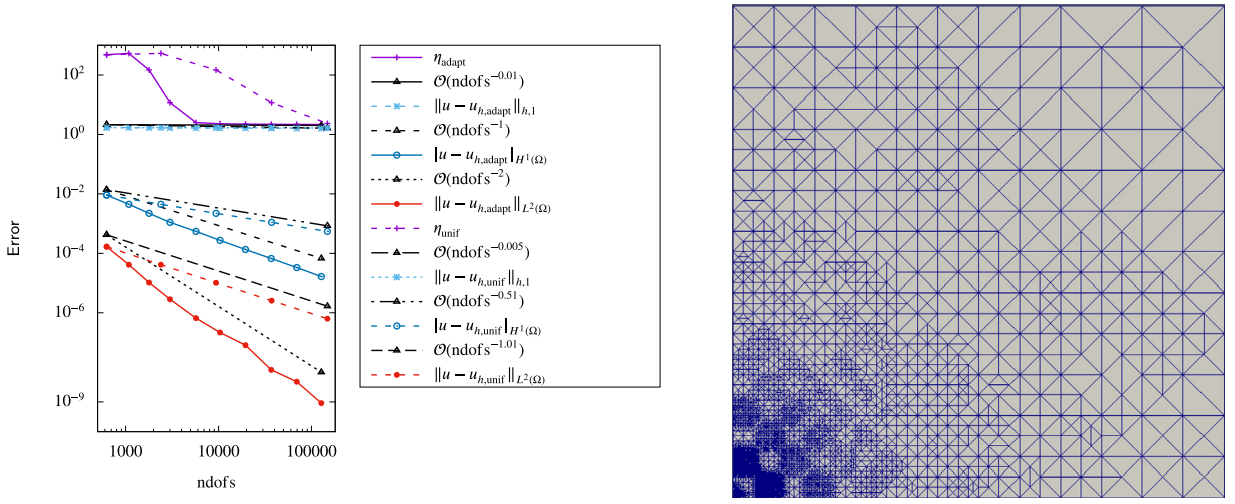

Fig. 1. On the left are the convergence rates for Experiment 6.1, with $s = 0.01$, and on the right is the final adapted mesh.

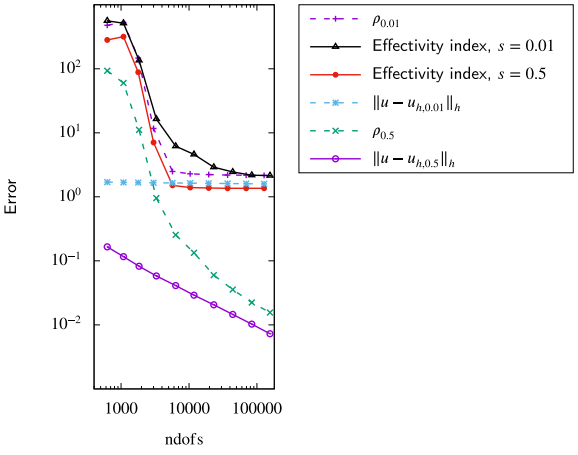

Fig. 2. Plot of the effectivity index, with error indicators and true errors for Experiment 6.1, with polynomial degree $p = 4$.

6.2. Experiment 2

In the previous experiment, we observed the advantage of applying the adaptive scheme, when compared with uniform refinement (see Fig. 1). However, the solution is known to possess H^s-regularity, with $s > 2$, and is known to lack regularity at the origin. We propose a second experiment, in which the solution is unknown, the right-hand is smooth, and the coefficients are indeed discontinuous (we choose a smooth right-hand side in order to surmise that any bad behaviour of the solution is due to the coefficients and regularity of $\partial\Omega$). In particular we consider the boundary value problem:

$$\begin{cases} \displaystyle\sum_{i,j=1}^{2}(1+\delta_{ij})\frac{(x_i-0.5)}{|x_i-0.5|}\frac{(x_j-0.5)}{|x_j-0.5|}\chi_{\Omega}^{N}(x_1,x_2)D_{ij}^2u_N = 1, \quad \text{in} \quad \Omega, \\ u_N = 0, \quad \text{on} \quad \partial\Omega, \end{cases} \tag{6.2}$$

where $\Omega = (0,1)^2$. We consider the case $N = 10$, and in this case the PDE theory implies that $u_N \in H^2(\Omega) \cap H_0^1(\Omega)$ (see (1.13)). We consider the polynomial degree $p = 2$, and an initial triangulation with a resolution that matches the indicator function (i.e., N squares in each coordinate direction, with each square further bisected into two triangles), and apply uniform mesh refinement, as well as adapted refinement (applying Algorithm 2), and compare the results.

The solution is unknown, and so we plot the error estimator η_h in each case. Due to discrete Poincaré–Friedrichs' inequalities and the reliability and efficiency of the estimator, η_h may be used to as a predictor for the (semi)norms $\|\cdot\|_{L^2(\Omega)}$, $|\cdot|_{H^1(\Omega)}$, and $\|\cdot\|_h$. Since the convergence rates in $\|\cdot\|_{L^2(\Omega)}$, $|\cdot|_{H^1(\Omega)}$, as predicted by η_h are likely to be pessimistic, we calculate the error arising between successive meshes, and appeal to this to guide the convergence. In particular, we

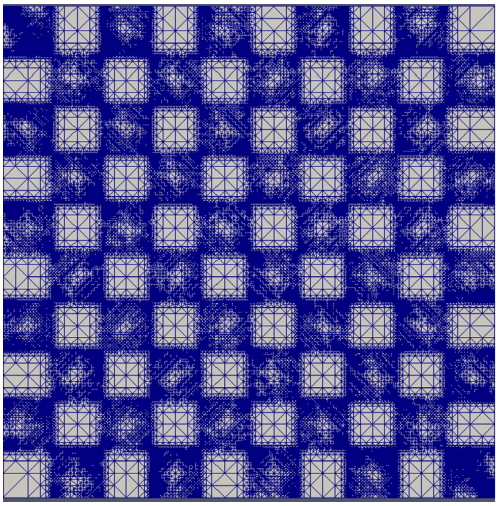

Fig. 3. Final adapted mesh for Experiment (6.2), with $N = 10$.

Table 1
$p = 2$, uniform refinement.

ndofs	$\|\theta_h\|_h$	EOC	$\|\theta_h\|_{H^1(\Omega)}$	EOC	$\|\theta_h\|_{L^2(\Omega)}$	EOC	η_h	EOC
1,681	0.212	0.000	$3.548 \cdot 10^{-3}$	0.000	$4.531 \cdot 10^{-4}$	0.000	169.505	0.000
6,561	0.252	0.126	$3.846 \cdot 10^{-3}$	$5.906 \cdot 10^{-2}$	$1.549 \cdot 10^{-4}$	−0.788	119.413	−0.257
25,921	0.161	−0.324	$1.389 \cdot 10^{-3}$	−0.741	$4.709 \cdot 10^{-5}$	−0.867	77.433	−0.315
103,041	$9.639 \cdot 10^{-2}$	−0.373	$4.21 \cdot 10^{-4}$	−0.865	$1.27 \cdot 10^{-5}$	−0.950	45.459	−0.386

Table 2
$p = 2$, adaptive refinement.

ndofs	$\|\theta_h\|_h$	EOC	$\|\theta_h\|_{H^1(\Omega)}$	EOC	$\|\theta_h\|_{L^2(\Omega)}$	EOC	η_h	EOC
1,093	0.186	0.000	$3.126 \cdot 10^{-3}$	0.000	$4.041 \cdot 10^{-4}$	0.000	234.229	0.000
2,911	0.188	$8.685 \cdot 10^{-3}$	$2.924 \cdot 10^{-3}$	$-6.823 \cdot 10^{-2}$	$2.729 \cdot 10^{-4}$	−0.401	172.288	−0.314
7,501	0.189	$4.851 \cdot 10^{-3}$	$2.291 \cdot 10^{-3}$	−0.258	$5.139 \cdot 10^{-5}$	−1.764	127.046	−0.322
20,159	0.137	−0.328	$1.285 \cdot 10^{-3}$	−0.584	$5.097 \cdot 10^{-5}$	$-8.282 \cdot 10^{-3}$	82.034	−0.442
52,503	$8.345 \cdot 10^{-2}$	−0.515	$4.466 \cdot 10^{-4}$	−1.104	$2.724 \cdot 10^{-5}$	−0.655	50.968	−0.497
132,973	$5.011 \cdot 10^{-2}$	−0.549	$1.47 \cdot 10^{-4}$	−1.196	$8.607 \cdot 10^{-6}$	−1.240	31.666	−0.512

define $\theta_k := u_k - u_{k-1}$, where the subscript k denotes the current refinement level, and appeal to the fact that for the norms under consideration $\|u - u_k\| \leq \|u - u_{k-1}\| + \|\theta_k\|$, and that the contribution $\|\theta_k\|$ should be the dominating term. We plot the final adapted mesh generated by the adaptive scheme in Fig. 3. The predictions show that the adaptive scheme outperforms uniform refinement, however, not to the same degree as is observed in Experiment 6.1, in the L^2- and H^1-norms. The exact values are provided in Tables 1–2.

6.3. Experiment 3

In this experiment, we consider the following Monge–Ampère problems

$$\det D^2 u_a = f_a, \quad \text{in} \quad \Omega, \tag{6.3}$$

$$u_a = g_a, \quad \text{on} \quad \partial\Omega, \tag{6.4}$$

on $\Omega = (0, 1)^2$.

Case 1: The functions f_a and g_a are chosen so that the true solution of (6.3)–(6.4) is given by

$$u_a(x_1, x_2) = |x_1 - a| \sin(x_1 - a) + 50.0(x_1^2 + x_2^2),$$

for $a \in \{0.4, 0.5\}$. Our initial mesh is a uniform triangulation on $\overline{\Omega}$ consisting of two squares (each further subdivided into two right-angled triangles) in the x_1 and x_2 direction, as such, we have that $u_{0.5}$ is piecewise smooth on the initial mesh (and all subsequent meshes, since each marked triangle is bisected), however, $u_{0.4}$ does not enjoy this piecewise smoothness property, and so, its approximation, $u_{h,a=0.4}$, does not converge as fast, as observed in Fig. 4. In both cases,

Table 3
$p = 4$, $a = 0.4$, uniform refinement.

ndofs	$\|e_h\|_h$	EOC	$\|e_h\|_{H^1(\Omega)}$	EOC	$\|e_h\|_{L^2(\Omega)}$	EOC	η_h	EOC
81	0.832	0.000	$2.22 \cdot 10^{-2}$	0.000	$1.477 \cdot 10^{-3}$	0.000	0.880	0.000
289	0.703	−0.133	$9.745 \cdot 10^{-3}$	−0.647	$3.105 \cdot 10^{-4}$	−1.226	0.797	$-7.714 \cdot 10^{-2}$
1,089	0.323	−0.586	$2.366 \cdot 10^{-3}$	−1.067	$3.586 \cdot 10^{-5}$	−1.627	0.359	−0.602
4,225	0.268	−0.138	$1.211 \cdot 10^{-3}$	−0.494	$1.231 \cdot 10^{-5}$	−0.789	0.333	$-5.501 \cdot 10^{-2}$
16,641	0.142	−0.461	$2.989 \cdot 10^{-4}$	−1.021	$9.441 \cdot 10^{-6}$	−0.194	0.164	−0.519

Table 4
$p = 4$, $a = 0.4$, adaptive refinement.

ndofs	$\|e_h\|_h$	EOC	$\|e_h\|_{H^1(\Omega)}$	EOC	$\|e_h\|_{L^2(\Omega)}$	EOC	η_h	EOC
81	0.832	0.000	$2.22 \cdot 10^{-2}$	0.000	$1.477 \cdot 10^{-3}$	0.000	0.880	0.000
139	0.888	0.120	$1.738 \cdot 10^{-2}$	−0.453	$9.446 \cdot 10^{-4}$	−0.827	0.953	0.148
341	0.610	−0.418	$9.273 \cdot 10^{-3}$	−0.700	$3.333 \cdot 10^{-4}$	−1.161	0.712	−0.324
839	0.468	−0.293	$3.159 \cdot 10^{-3}$	−1.196	$7.299 \cdot 10^{-5}$	−1.687	0.515	−0.359
1,961	0.260	−0.693	$1.426 \cdot 10^{-3}$	−0.937	$2.891 \cdot 10^{-5}$	−1.091	0.317	−0.571
4,385	0.184	−0.427	$5.369 \cdot 10^{-4}$	−1.214	$2.488 \cdot 10^{-5}$	−0.187	0.217	−0.470
9,341	0.117	−0.600	$1.815 \cdot 10^{-4}$	−1.434	$3.954 \cdot 10^{-6}$	−2.432	0.148	−0.506
19,609	$8.074 \cdot 10^{-2}$	−0.502	$7.139 \cdot 10^{-5}$	−1.258	$3.293 \cdot 10^{-6}$	−0.247	$9.937 \cdot 10^{-2}$	−0.540

Table 5
$p = 4$, $a = 0.5$, uniform refinement.

ndofs	$\|e_h\|_h$	EOC	$\|e_h\|_{H^1(\Omega)}$	EOC	$\|e_h\|_{L^2(\Omega)}$	EOC	η_h	EOC
81	$8.886 \cdot 10^{-4}$	0.000	$3.136 \cdot 10^{-5}$	0.000	$2.558 \cdot 10^{-6}$	0.000	$9.902 \cdot 10^{-4}$	0.000
289	$9.265 \cdot 10^{-5}$	−1.777	$2.092 \cdot 10^{-6}$	−2.128	$8.781 \cdot 10^{-8}$	−2.651	$1.082 \cdot 10^{-4}$	−1.740
1,089	$1.114 \cdot 10^{-5}$	−1.597	$1.308 \cdot 10^{-7}$	−2.090	$2.808 \cdot 10^{-9}$	−2.595	$1.332 \cdot 10^{-5}$	−1.579
4,225	$1.32 \cdot 10^{-6}$	−1.574	$8.142 \cdot 10^{-9}$	−2.048	$8.959 \cdot 10^{-11}$	−2.541	$1.597 \cdot 10^{-6}$	−1.564
16,641	$1.673 \cdot 10^{-7}$	−1.507	$5.777 \cdot 10^{-10}$	−1.930	$5.611 \cdot 10^{-11}$	−0.341	$2.056 \cdot 10^{-7}$	−1.495

Table 6
$p = 4$, $a = 0.5$, adaptive refinement.

ndofs	$\|e_h\|_h$	EOC	$\|e_h\|_{H^1(\Omega)}$	EOC	$\|e_h\|_{L^2(\Omega)}$	EOC	η_h	EOC
81	$8.886 \cdot 10^{-4}$	0.000	$3.136 \cdot 10^{-5}$	0.000	$2.558 \cdot 10^{-6}$	0.000	$9.902 \cdot 10^{-4}$	0.000
139	$6.3 \cdot 10^{-4}$	−0.637	$2.18 \cdot 10^{-5}$	−0.673	$1.734 \cdot 10^{-6}$	−0.720	$7.346 \cdot 10^{-4}$	−0.553
275	$2.816 \cdot 10^{-4}$	−1.180	$7.504 \cdot 10^{-6}$	−1.563	$2.955 \cdot 10^{-7}$	−2.594	$2.856 \cdot 10^{-4}$	−1.385
495	$9.876 \cdot 10^{-5}$	−1.782	$1.534 \cdot 10^{-6}$	−2.701	$5.567 \cdot 10^{-8}$	−2.840	$1.028 \cdot 10^{-4}$	−1.739
1,107	$3.64 \cdot 10^{-5}$	−1.240	$4.15 \cdot 10^{-7}$	−1.625	$1.303 \cdot 10^{-8}$	−1.805	$3.794 \cdot 10^{-5}$	−1.238
2,531	$1.107 \cdot 10^{-5}$	−1.439	$1.022 \cdot 10^{-7}$	−1.695	$2.092 \cdot 10^{-9}$	−2.212	$1.242 \cdot 10^{-5}$	−1.351
4,915	$4.528 \cdot 10^{-6}$	−1.348	$3.541 \cdot 10^{-8}$	−1.596	$6.216 \cdot 10^{-10}$	−1.829	$4.419 \cdot 10^{-6}$	−1.557
9,285	$1.479 \cdot 10^{-6}$	−1.759	$7.005 \cdot 10^{-9}$	−2.547	$7.541 \cdot 10^{-11}$	−3.316	$1.598 \cdot 10^{-6}$	−1.600

Table 7
$p = 4$, uniform refinement.

ndofs	$\|\theta_h\|_h$	EOC	$\|\theta_h\|_{H^1(\Omega)}$	EOC	$\|\theta_h\|_{L^2(\Omega)}$	EOC	η_h	EOC
4,225	1.044	0.000	$1.214 \cdot 10^{-2}$	0.000	$4.9 \cdot 10^{-4}$	0.000	0.451	0.000
16,641	1.090	$3.174 \cdot 10^{-2}$	$6.871 \cdot 10^{-3}$	−0.415	$1.237 \cdot 10^{-4}$	−1.004	0.436	$-2.442 \cdot 10^{-2}$
66,049	1.035	$-3.77 \cdot 10^{-2}$	$3.303 \cdot 10^{-3}$	−0.531	$2.497 \cdot 10^{-5}$	−1.161	0.424	$-2.061 \cdot 10^{-2}$

$u_a \in W^{2,\infty}(\Omega)$, and we set the polynomial degree $p = 4$. Note that in this case we apply the adaptive finite element method given by Algorithm 2, in conjunction with the semismooth Newton's method given by Algorithm 1.

We also compare the adaptive scheme with that of uniform refinement. The exact results are provided in Tables 3–6. We observe that when $a = 0.4$, the adaptive scheme out performs uniform refinement, whereas when $a = 0.5$ the two approaches are comparable (we surmise this is due to the piecewise smoothness property of $u_{0.5}$).

Case 2: Here we take $f_a \equiv 1$, $g_a \equiv 0$. In this case the true solution is unknown, and so we rely on the error estimator, as well as the incremental solutions in order to indicate the performance of the numerical method (as in Experiment 6.2). We take $p = 4$ and compare the adaptive scheme with uniform refinement. We display the exact convergence results in Tables 7–8, and observe that the adaptive scheme outperforms the uniform scheme in all (semi)norms.

Table 8
$p = 4$, adaptive refinement.

ndofs	$\|\theta_h\|_h$	EOC	$\|\theta_h\|_{H^1(\Omega)}$	EOC	$\|\theta_h\|_{L^2(\Omega)}$	EOC	η_h	EOC
1,985	1.042	0.000	$1.193 \cdot 10^{-2}$	0.000	$4.373 \cdot 10^{-4}$	0.000	0.452	0.000
3,751	1.091	$7.198 \cdot 10^{-2}$	$6.915 \cdot 10^{-3}$	−0.857	$1.353 \cdot 10^{-4}$	−1.843	0.437	$-5.367 \cdot 10^{-2}$
7,063	1.038	$-7.881 \cdot 10^{-2}$	$3.415 \cdot 10^{-3}$	−1.115	$4.694 \cdot 10^{-5}$	−1.673	0.424	$-4.598 \cdot 10^{-2}$
14,775	1.003	$-4.612 \cdot 10^{-2}$	$1.808 \cdot 10^{-3}$	−0.862	$2.397 \cdot 10^{-5}$	−0.911	0.413	$-3.546 \cdot 10^{-2}$
29,009	0.975	$-4.211 \cdot 10^{-2}$	$8.518 \cdot 10^{-4}$	−1.116	$5.544 \cdot 10^{-6}$	−2.170	0.403	$-3.543 \cdot 10^{-2}$
59,189	0.953	$-3.144 \cdot 10^{-2}$	$4.245 \cdot 10^{-4}$	−0.977	$3.558 \cdot 10^{-6}$	−0.622	0.394	$-3.16 \cdot 10^{-2}$

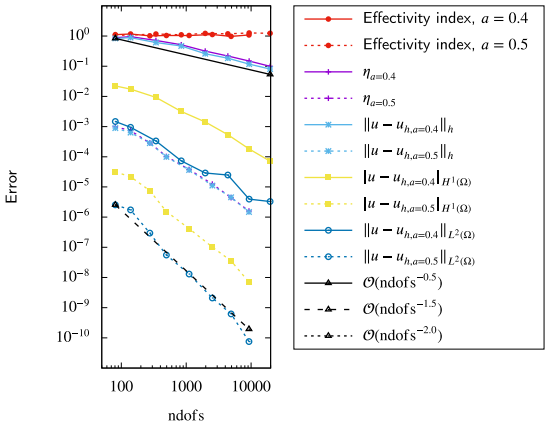

Fig. 4. Convergence rates for Experiment 6.3, we observe faster convergence rates for $u_{0.5}$ than for $u_{0.4}$.

7. Concluding remarks

In this paper, we were successful in proposing and analysing a C^0-interior penalty method for the approximation of the fully nonlinear Hamilton–Jacobi–Bellman equation with inhomogeneous Dirichlet boundary data. The analysis consisted of three parts: proving a stability estimate, a quasi-optimal a priori error estimate, and an a posteriori error estimate in a H^2-style norm. All of the aforementioned analysis was undertaken, assuming regularity no higher than $H^2(\Omega)$, as implied by the corresponding PDE theory. All of the theoretical results were confirmed in the experiments section, which included the implementation of an adaptive method, based upon the proven a posteriori error estimate. Furthermore, we were able to apply the proposed method to the fully nonlinear Monge–Ampère equation, providing a uniquely solvable, optimally convergent, and adaptive finite element method.

References

[1] H.O. Cordes, Über die erste Randwertaufgabe bei quasilinearen Differentialgleichungen zweiter Ordnung in mehr als zwei Variablen, Math. Ann. 131 (1956) 278–312, http://dx.doi.org/10.1007/BF01342965.
[2] A. Maugeri, D.K. Palagachev, L.G. Softova, Elliptic and Parabolic Equations with Discontinuous Coefficients, in: Mathematical Research, vol. 109, Wiley-VCH Verlag Berlin GmbH, Berlin, 2000, p. 256, http://dx.doi.org/10.1002/3527600868.
[3] D. Gilbarg, N.S. Trudinger, Elliptic partial differential equations of second order, in: Classics in Mathematics, Springer-Verlag, Berlin, 2001, p. xiv+517, Reprint of the 1998 edition.
[4] L.C. Evans, Classical solutions of the Hamilton–Jacobi–Bellman equation for uniformly elliptic operators, Trans. Amer. Math. Soc. 275 (1) (1983) 245–255, http://dx.doi.org/10.2307/1999016.
[5] W.H. Fleming, H.M. Soner, Controlled Markov processes and viscosity solutions, second ed., in: Stochastic Modelling and Applied Probability, vol. 25, Springer, New York, 2006, p. xviii+429.
[6] D. Gallistl, Numerical approximation of planar oblique derivative problems in nondivergence form, Math. Comp. 88 (317) (2019) 1091–1119, http://dx.doi.org/10.1090/mcom/3371.
[7] E.L. Kawecki, A DGFEM for nondivergence form elliptic equations with Cordes coefficients on curved domains, Numer. Methods Partial Differential Equations 35 (5) (2019) 1717–1744, http://dx.doi.org/10.1002/num.22372, 3985933.
[8] E.L. Kawecki, Finite element theory on curved domains with applications to discontinuous Galerkin finite element methods, Numer. Methods Partial Differential Equations 36 (6) (2020) 1492–1536, http://dx.doi.org/10.1002/num.22489.
[9] E.L. Kawecki, A discontinuous Galerkin finite element method for uniformly elliptic two dimensional oblique boundary-value problems, SIAM J. Numer. Anal. 57 (2) (2019) 751–778, http://dx.doi.org/10.1137/17M1155946.
[10] D. Gallistl, Variational formulation and numerical analysis of linear elliptic equations in nondivergence form with Cordes coefficients, SIAM J. Numer. Anal. 55 (2) (2017) 737–757.
[11] C. Wang, J. Wang, A primal-dual weak Galerkin finite element method for second order elliptic equations in non-divergence form, Math. Comp. 87 (310) (2018) 515–545.

[12] X. Feng, L. Hennings, M. Neilan, Finite element methods for second order linear elliptic partial differential equations in non-divergence form, Math. Comp. 86 (307) (2017) 2025–2051.
[13] X. Feng, M. Neilan, S. Schnake, Interior penalty discontinuous Galerkin methods for second order linear non-divergence form elliptic PDEs, J. Sci. Comput. 74 (3) (2018) 1651–1676.
[14] I. Smears, E. Süli, Discontinuous Galerkin finite element approximation of nondivergence form elliptic equations with Cordès coefficients, SIAM J. Numer. Anal. 51 (4) (2013) 2088–2106, http://dx.doi.org/10.1137/120899613.
[15] D. Gallistl, E. Süli, Mixed finite element approximation of the Hamilton–Jacobi–Bellman equation with cordes coefficients, SIAM J. Numer. Anal. 57 (2) (2019) 592–614, http://dx.doi.org/10.1137/18M1192299.
[16] X. Feng, M. Jensen, Convergent semi-Lagrangian methods for the Monge–Ampère equation on unstructured grids, SIAM J. Numer. Anal. 55 (2) (2017) 691–712.
[17] M. Jensen, I. Smears, Finite element methods with artificial diffusion for Hamilton–Jacobi–Bellman equations, in: Numerical Mathematics and Advanced Applications 2011, Springer, Heidelberg, 2013, pp. 267–274.
[18] M. Jensen, I. Smears, On the convergence of finite element methods for Hamilton–Jacobi–Bellman equations, SIAM J. Numer. Anal. 51 (1) (2013) 137–162.
[19] I. Smears, E. Süli, Discontinuous Galerkin finite element approximation of Hamilton–Jacobi–Bellman equations with Cordès coefficients, SIAM J. Numer. Anal. 52 (2) (2014) 993–1016, http://dx.doi.org/10.1137/130909536.
[20] M. Neilan, M. Wu, Discrete Miranda–Talenti estimates and applications to linear and nonlinear PDEs, J. Comput. Appl. Math. 356 (2019) 358–376.
[21] I. Smears, E. Süli, Discontinuous Galerkin finite element methods for time-dependent Hamilton–Jacobi–Bellman equations with Cordes coefficients, Numer. Math. 133 (1) (2016) 141–176, http://dx.doi.org/10.1007/s00211-015-0741-6.
[22] S.C. Brenner, L.-Y. Sung, Virtual enriching operators, Calcolo 56 (4) (2019) http://dx.doi.org/10.1007/s10092-019-0338-z, Paper No. 44.
[23] X. Feng, R. Glowinski, M. Neilan, Recent developments in numerical methods for fully nonlinear second order partial differential equations, SIAM Rev. 55 (2) (2013) 205–267, http://dx.doi.org/10.1137/110825960.
[24] J.-D. Benamou, B.D. Froese, A.M. Oberman, Two numerical methods for the elliptic Monge–Ampère equation, M2AN Math. Model. Numer. Anal. 44 (4) (2010) 737–758, http://dx.doi.org/10.1051/m2an/2010017.
[25] N.V. Krylov, Nonlinear elliptic and parabolic equations of the second order, in: Mathematics and its Applications (Soviet Series), vol. 7, D. Reidel Publishing Co., Dordrecht, 1987, p. xiv+462, http://dx.doi.org/10.1007/978-94-010-9557-0, Translated from the Russian by P. L. Buzytsky [P. L. Buzytskiĭ].
[26] C. Budd, M. Cullen, E. Walsh, Monge–Ampére Based moving mesh methods for numerical weather prediction, with applications to the Eady problem, J. Comput. Phys. 236 (2013) 247–270, http://dx.doi.org/10.1016/j.jcp.2012.11.014.
[27] E.L. Kawecki, O. Lakkis, T. Pryer, A finite element method for the Monge–Ampère equation with transport boundary conditions, 2018, arXiv preprint arXiv:1807.03535.
[28] M. Neilan, Finite element methods for fully nonlinear second order PDEs based on a discrete Hessian with applications to the Monge-Ampère equation, J. Comput. Appl. Math. 263 (2014) 351–369, http://dx.doi.org/10.1016/j.cam.2013.12.027.
[29] M.S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, G.N. Wells, The fenics project version 1.5, Arch. Numer. Softw. 3 (100) (2015) http://dx.doi.org/10.11588/ans.2015.100.20553.
[30] S.C. Brenner, M. Neilan, A. Reiser, L.-Y. Sung, A C^0 interior penalty method for a von Kármán plate, Numer. Math. 135 (3) (2017) 803–832, http://dx.doi.org/10.1007/s00211-016-0817-y.
[31] S.C. Brenner, K. Wang, J. Zhao, Poincaré–Friedrichs Inequalities for piecewise H^2 functions, Numer. Funct. Anal. Optim. 25 (5–6) (2004) 463–478, http://dx.doi.org/10.1081/NFA-200042165.
[32] O. Bokanowski, S. Maroso, H. Zidani, Some convergence results for Howard's algorithm, SIAM J. Numer. Anal. 47 (4) (2009) 3001–3026.
[33] R.A. Howard, Dynamic Programming and Markov Processes, John Wiley, 1960.
[34] M. Neilan, A.J. Salgado, W. Zhang, Numerical analysis of strongly nonlinear PDEs, Acta Numer. 26 (2017) 137–303.
[35] W. Dörfler, A convergent adaptive algorithm for Poisson's equation, SIAM J. Numer. Anal. 33 (3) (1996) 1106–1124.
[36] E. Kawecki, Finite Element Methods for Monge–Ampère Type Equations (Thesis D.Phil), The University of Oxford, 2018.