SIAM J. NUMER. ANAL.
Vol. 52, No. 2, pp. 993–1016

DISCONTINUOUS GALERKIN FINITE ELEMENT APPROXIMATION OF HAMILTON–JACOBI–BELLMAN EQUATIONS WITH CORDES COEFFICIENTS*

IAIN SMEARS† AND ENDRE SÜLI†

Abstract. We propose an *hp*-version discontinuous Galerkin finite element method for fully nonlinear second-order elliptic Hamilton–Jacobi–Bellman equations with Cordes coefficients. The method is proved to be consistent and stable, with convergence rates that are optimal with respect to mesh size, and suboptimal in the polynomial degree by only half an order. Numerical experiments on problems with nonsmooth solutions and strongly anisotropic diffusion coefficients illustrate the accuracy and computational efficiency of the scheme. An existence and uniqueness result for strong solutions of the fully nonlinear problem and a semismoothness result for the nonlinear operator are also provided.

Key words. Hamilton–Jacobi–Bellman equations, *hp*-version discontinuous Galerkin finite element methods, Cordes condition, fully nonlinear equations, semismooth Newton methods

AMS subject classifications. 65N30, 65N12, 65N15, 35J15, 35J66, 35D35, 49M15, 47J25

DOI. 10.1137/130909536

1. Introduction.

We study the numerical analysis of fully nonlinear second-order elliptic Hamilton–Jacobi–Bellman (HJB) equations of the form

$$\sup_{\alpha\in\Lambda}\,[L^\alpha u - f^\alpha] = 0 \quad \text{in } \Omega, \tag{1.1}$$

where Ω is a convex domain in $\mathbb{R}^n$, $n \geq 2$, Λ is a compact metric space, and the L^α, $\alpha \in \Lambda$, are elliptic operators of the form

$$L^\alpha v = \sum_{i,j=1}^{n} a_{ij}^\alpha\, v_{x_i x_j} + \sum_{i=1}^{n} b_i^\alpha\, v_{x_i} - c^\alpha\, v. \tag{1.2}$$

HJB equations characterize the value functions of stochastic control problems, which arise from applications in engineering, physics, economics, and finance [11]. The solution of (1.1) leads to the best choices of controls from the set Λ for steering a stochastic process toward optimizing the expected value of a functional. We are interested in consistent, stable, convergent, and high-order methods for multidimensional uniformly elliptic HJB equations with anisotropic diffusions.

Discrete state Markov chain approximations to the underlying stochastic dynamics were among the earliest computational approaches to these problems [19]. Alongside the advent of the notion of a viscosity solution to a fully nonlinear second-order equation [6], it became apparent that these Markov chain approximations admit equivalent interpretations as *monotone* finite difference methods (FDM) [5, 11], i.e., that satisfy a discrete maximum principle. These methods feature a general convergence theory due to Barles and Souganidis [1], and are capable of approximating nonsmooth viscosity solutions of certain degenerate problems.

*Received by the editors February 14, 2013; accepted for publication (in revised form) February 21, 2014; published electronically April 24, 2014.
http://www.siam.org/journals/sinum/52-2/90953.html
†Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK (smears@maths.ox.ac.uk, suli@maths.ox.ac.uk).

Various authors have commented on the necessarily low-order convergence rates of monotone schemes [23] and on the restrictions that the choice of stencil imposes on the set of problems amenable to discretization by monotone FDM [7, 17]. For an analysis of convergence rates, see [9] and the references therein. Motzkin and Wasow [22] found that for any choice of stencil, there exists a uniformly elliptic operator with no consistent and monotone discretization; yet, for any set of nondegenerate diffusion coefficients with uniformly bounded ellipticity constants, there is a stencil providing a monotone and consistent discretization. Kocan studied the minimum size of such a stencil as a function of the ellipticity constant in [17]. Conversely, Bonnans and Zidani [5] examined the conditions that determine the set of problems that can be discretized with various stencils: they found that the number of conditions on the diffusion coefficient grows with both the stencil size and the problem dimension. An algorithm was developed in [4] to compute a monotone discretization of two-dimensional problems, with a consistency error depending on the stencil width.

While the above considerations concern the notion of consistency of FDM, convergent monotone methods for fully nonlinear problems can also employ notions of consistency from finite element methods (FEM); see [16] for the first convergent monotone FEM for viscosity solutions of parabolic HJB equations. Böhmer proposed in [2] a nonmonotone H^2-conforming FEM for fully nonlinear PDE with linearizations in divergence form; yet linearizations of the HJB operator are usually in nondivergence form with discontinuous coefficients and cannot be recast into divergence form.

Discontinuous Galerkin finite element methods (DGFEM) allow the approximate solution to be discontinuous between elements of the mesh, with the continuity conditions being enforced only weakly through the discretized problem [8]. This facilitates *hp*-refinement, which varies both mesh size and polynomial degree, thereby allowing for exponential convergence rates, even for problems with nonsmooth solutions [30]. For problems in nondivergence form, a challenge in the design of DGFEM is to obtain stable interelement communication. Nevertheless, the authors found new techniques in [27] to obtain stable discretizations of certain linear nondivergence form equations with discontinuous coefficients, and these techniques are taken further in this work.

We consider here uniformly elliptic HJB equations that satisfy the Cordes condition: we provide a concise and accessible proof of existence and uniqueness of a strong solution of (1.1) associated to a homogeneous Dirichlet boundary condition. Then, we construct a stable, consistent, and convergent *hp*-version DGFEM, for which we prove convergence rates in a discrete H^2-type norm that are optimal with respect to mesh size and suboptimal in the polynomial degree by only half an order. As opposed to the monotone methods considered above, our method is consistent regardless of the choice of mesh, thereby permitting *hp*-refinement on very general shape-regular sequences of meshes. An assumption on the solution of broken H^s-regularity, $s > 5/2$, is used for the analysis, but numerical evidence shows that this is not a necessary condition for convergence of the scheme. Our experiments show the gains in computational efficiency, flexibility, and accuracy over existing monotone methods.

The Cordes condition, defined in section 2 below, encompasses a large range of applications. For example, in two spatial dimensions, the condition amounts to simply requiring uniform ellipticity of the diffusion coefficient and coercivity of the lower-order terms; see Examples 1 and 2 of section 2. Let us now recount how the motivation for the Cordes condition stems from genuine PDE-theoretic considerations. There is a famous solution algorithm for (1.1), due to Bellman and Howard [3, 24], that may be understood as follows. Given an approximate solution u^k, $k \in \mathbb{N}$, to (1.1), one finds for each $x \in \Omega$ an $\Lambda \ni \alpha_k(x) = \mathrm{argmax}_\alpha(L^\alpha u^k - f^\alpha)(x)$. A new approximation

u^{k+1} is sought as the solution of $L^{\alpha_k} u^{k+1} = f^{\alpha_k}$, where $f^{\alpha_k}\colon x \mapsto f^{\alpha_k(x)}(x)$, and where the coefficients of the linear operator L^{α_k} are similarly defined; formally, a solution of (1.1) is a fixed point of this iteration. It has long been known that this method is in fact a Newton method for a nondifferentiable operator [3, 24], and we contribute to its analysis by showing the semismoothness in function spaces [29] of the HJB operator. The question of the well-posedness of the linear PDE to be solved at each iteration is instructive: these are nondivergence form elliptic equations *with discontinuous coefficients*, and it is known that well-posedness in the strong sense is not guaranteed by uniform ellipticity alone [12, 20, 26], although it is recovered under the Cordes condition [20]. Importantly, we show here that well-posedness of strong solutions extends to HJB equations, under the same condition. Inspired by the analysis of the PDE, the stability of our method is obtained by relating the residual of the equation to terms measuring the lack of H^2-conformity of the numerical solution.

The structure of this article is as follows. After defining the problem in section 2, we prove its well-posedness in section 3. The *hp*-version DGFEM framework is prepared in section 4 and is followed by the definition and consistency analysis of the method in section 5. We establish the stability of the scheme in section 6 and we determine its convergence rates in section 7. Section 8 analyzes a superlinearly convergent semismooth Newton method used to solve the discrete problem, and section 9 presents the results of numerical experiments that demonstrate the high accuracy and computational efficiency of the method.

2. Statement of the problem. Let Ω be a bounded convex polytopal open set in $\mathbb{R}^n$, $n \geq 2$, and let Λ be a compact metric space. It will always be assumed that Ω and Λ are nonempty. Convexity of Ω implies that the boundary $\partial\Omega$ is Lipschitz; see [13]. Let the real-valued functions $a_{ij} = a_{ji}$, b_i, c, and f belong to $C(\overline{\Omega} \times \Lambda)$ for all $i, j = 1, \ldots, n$. For each $\alpha \in \Lambda$, we consider the function $a_{ij}^{\alpha}\colon x \mapsto a_{ij}(x, \alpha)$, $x \in \overline{\Omega}$. The functions b_i^{α}, c^{α}, and f^{α} are defined in a similar way. Define the matrix-valued functions $a^{\alpha} := (a_{ij}^{\alpha})$ and define the vector-valued functions $b^{\alpha} := (b_1^{\alpha}, \ldots, b_n^{\alpha})$, where $\alpha \in \Lambda$. The bounded linear operators $L^{\alpha}\colon H^2(\Omega) \to L^2(\Omega)$ are defined by

$$L^{\alpha} v := \sum_{i,j=1}^{n} a_{ij}^{\alpha}\, v_{x_i x_j} + \sum_{i=1}^{n} b_i^{\alpha}\, v_{x_i} - c^{\alpha}\, v, \qquad v \in H^2(\Omega),\ \alpha \in \Lambda. \tag{2.1}$$

Compactness of Λ and continuity of the coefficients a, b, c, and f imply that the fully nonlinear operator F, defined by

$$F\colon v \mapsto F[v] := \sup_{\alpha \in \Lambda} \left[L^{\alpha} v - f^{\alpha} \right], \tag{2.2}$$

is well-defined as a mapping from $H^2(\Omega)$ to $L^2(\Omega)$. The problem considered is to find $u \in H^2(\Omega) \cap H_0^1(\Omega)$ that is a strong solution of the HJB equation subject to a homogeneous Dirichlet boundary condition

$$\begin{aligned} F[u] &= 0 \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2.3}$$

Well-posedness of problem (2.3) is established in section 3 under the following hypotheses. It is assumed that there exist positive constants $\nu \leq \bar{\nu}$ such that

$$\nu |\xi|^2 \leq \sum_{i,j=1}^{n} a_{ij}^{\alpha}(x)\, \xi_i\, \xi_j \leq \bar{\nu} |\xi|^2 \quad \forall\, \xi \in \mathbb{R}^n,\ \forall\, x \in \Omega,\ \forall\, \alpha \in \Lambda. \tag{2.4}$$

The function c^α is supposed to be nonnegative on $\overline{\Omega}$ for each $\alpha \in \Lambda$. We assume *the Cordes condition*: there exist $\lambda > 0$ and $\varepsilon \in (0,1)$ such that for each $\alpha \in \Lambda$,

$$\frac{|a^\alpha|^2 + |b^\alpha|^2/2\lambda + (c^\alpha/\lambda)^2}{(\operatorname{Tr} a^\alpha + c^\alpha/\lambda)^2} \leq \frac{1}{n+\varepsilon} \quad \text{in } \overline{\Omega}, \tag{2.5}$$

where $|\cdot|$ represents the Euclidian norm for vectors and the Frobenius norm for matrices. In the special case $b^\alpha \equiv 0$ and $c^\alpha \equiv 0$ for each $\alpha \in \Lambda$, we set $\lambda = 0$ and the Cordes condition (2.5) is replaced by the following: there exists $\varepsilon \in (0,1)$ such that for each $\alpha \in \Lambda$,

$$\frac{|a^\alpha|^2}{(\operatorname{Tr} a^\alpha)^2} \leq \frac{1}{n-1+\varepsilon} \quad \text{in } \overline{\Omega}. \tag{2.6}$$

Conditions (2.5) and (2.6) are related through the observation that the term c^α/λ may be viewed as the $(n+1,n+1)$ entry of an $(n+1)\times(n+1)$ matrix with principal $n\times n$ submatrix a^α, which explains the difference in the right-hand sides of the inequalities in (2.5) and (2.6). The parameter λ serves to make the Cordes condition invariant under rescaling the coordinates. It will be seen below that it is often easy to choose an appropriate value for λ.

Example 1. We show how the Cordes condition (2.5) arises in practice in an example from stochastic control problems [11]. We consider a problem where the controls permit the choice of orientation and angle between two Wiener diffusions. Let Ω be a domain in $\mathbb{R}^2$ and let $\Lambda = [0,\pi/3]\times \mathrm{SO}(2)$, where $\mathrm{SO}(2)$ is the set of 2×2 rotation matrices. The diffusions act along the directions σ_1^α and σ_2^α, where

$$\sigma^\alpha := (\sigma_1^\alpha\ \sigma_2^\alpha) := R^\top \begin{pmatrix} 1 & \sin\theta \\ 0 & \cos\theta \end{pmatrix}, \quad \alpha = (\theta, R) \in \Lambda. \tag{2.7}$$

In stochastic control problems, we have $a^\alpha := \sigma^\alpha(\sigma^\alpha)^\top/2$ and usually $c^\alpha \equiv c_0 > 0$ is a fixed constant [11]. Then, $\operatorname{Tr} a^\alpha = 1$ and $|a^\alpha|^2 = (1+\sin^2\theta)/2 \leq 7/8$; so condition (2.6) holds with $\varepsilon = 1/7$. Momentarily assuming that $b^\alpha \equiv 0$, by choosing the value $\lambda = \frac{8}{7}c_0$ that minimizes the left-hand side in (2.5), we find that condition (2.5) also holds with $\varepsilon = 1/7$. For nonzero b^α, the Cordes condition holds for $\varepsilon < 1/7$ whenever $|b^\alpha|^2/c_0$ is sufficiently small; this amounts to a standard coercivity assumption.

Example 1 is considered further in the numerical experiments of section 9.1. Observe that for any choice of Cartesian coordinates on $\mathbb{R}^2$, for $\theta = \pi/3$ there is an $R \in \mathrm{SO}(2)$ such that a^α is not diagonally dominant. Therefore, the classical monotone Kushner–Dupuis FDM is not applicable here [5].

Example 2. For problems in two dimensions, i.e., $n = 2$, the uniform ellipticity condition (2.4) is sufficient for the Cordes condition (2.6). Indeed, for each $\alpha \in \Lambda$, we have $\nu^2 \leq \det a^\alpha$, and $a_{11}^\alpha + a_{22}^\alpha \leq 2\bar{\nu}$, so, for $\varepsilon = \nu^2/(2\bar{\nu}^2 - \nu^2)$, we get

$$\frac{(a_{11}^\alpha)^2 + 2(a_{12}^\alpha)^2 + (a_{22}^\alpha)^2}{(a_{11}^\alpha + a_{22}^\alpha)^2} \leq 1 - \frac{2\nu^2}{(a_{11}^\alpha + a_{22}^\alpha)^2} \leq 1 - \frac{\nu^2}{2\bar{\nu}^2} = \frac{1}{1+\varepsilon}. \tag{2.8}$$

The above examples demonstrate that the results of this paper are relevant to a very broad class of problems, including some that require large stencils for monotone FDM; significant further evidence for this observation is found in section 9. Define the strictly positive function $\gamma\colon \overline{\Omega}\times\Lambda \to \mathbb{R}_{>0}$ by

$$\gamma(x,\alpha) := \frac{\operatorname{Tr} a^\alpha(x) + c^\alpha(x)/\lambda}{|a^\alpha(x)|^2 + |b^\alpha(x)|^2/2\lambda + (c^\alpha(x)/\lambda)^2}. \tag{2.9}$$

In the special case $b^\alpha \equiv 0$ and $c^\alpha \equiv 0$ for all $\alpha \in \Lambda$, we take $\lambda = 0$ and define

$$\gamma(x,\alpha) \coloneqq \frac{\operatorname{Tr} a^\alpha(x)}{|a^\alpha(x)|^2}. \tag{2.10}$$

As above, for each $\alpha \in \Lambda$, we define $\gamma^\alpha \colon x \mapsto \gamma(x,\alpha)$, $x \in \overline{\Omega}$. It follows from the continuity assumptions on the coefficients and from the uniform ellipticity condition (2.4) that $\gamma \in C(\overline{\Omega}\times\Lambda)$. Furthermore, nonnegativity of c, continuity of the coefficients and (2.4) imply that there is a positive constant $\gamma_0 > 0$ such that $\gamma \geq \gamma_0$ on $\overline{\Omega} \times \Lambda$. Define the operator $F_\gamma \colon H^2(\Omega) \to L^2(\Omega)$ by

$$F_\gamma[v] \coloneqq \sup_{\alpha\in\Lambda} \left[\gamma^\alpha \ (L^\alpha v - f^\alpha)\right]. \tag{2.11}$$

It will be seen below that the HJB equation (2.3) is in fact equivalent to the problem $F_\gamma[u] = 0$ in Ω, $u = 0$ on $\partial\Omega$. For λ as in (2.5), let the operator L_λ be defined by

$$L_\lambda v \coloneqq \Delta v - \lambda v, \quad v \in H^2(\Omega). \tag{2.12}$$

The following inequality generalizes results in [20, 27] that were used to analyze linear PDE satisfying the Cordes condition. It is key to our analysis of HJB equations.

LEMMA 1. *Let Ω be a bounded open subset of $\mathbb{R}^n$ and suppose that* (2.4) *holds, and suppose that either* (2.5) *holds with $\lambda > 0$ or that* (2.6) *holds with $b^\alpha \equiv 0$, $c^\alpha \equiv 0$ for all α, and $\lambda = 0$. Then, for any open set $U \subset \Omega$ and $u, v \in H^2(U)$, $w \coloneqq u - v$, the following inequality holds a.e. in U:*

$$|F_\gamma[u] - F_\gamma[v] - L_\lambda(u-v)| \leq \sqrt{1-\varepsilon}\,\sqrt{|D^2 w|^2 + 2\lambda|\nabla w|^2 + \lambda^2|w|^2}. \tag{2.13}$$

Proof. It will be clear how to adapt the following arguments to treat the simpler situation where $b^\alpha \equiv 0$, $c^\alpha \equiv 0$, and $\lambda = 0$. So, we consider the case where (2.5) holds with $\lambda > 0$. First, set $w \coloneqq u - v$. Note that we have the identity $F_\gamma[u] - L_\lambda u = \sup_{\alpha\in\Lambda}\left[\gamma^\alpha L^\alpha u - L_\lambda u - \gamma^\alpha f^\alpha\right]$. Also, for bounded sets of real numbers, $\{x^\alpha\}_\alpha$ and $\{y^\alpha\}_\alpha$, we have $|\sup_\alpha x^\alpha - \sup_\alpha y^\alpha| \leq \sup_\alpha |x^\alpha - y^\alpha|$. Therefore,

$$\begin{aligned} |F_\gamma[u] - F_\gamma[v] - L_\lambda w| &\leq \sup_{\alpha\in\Lambda} |\gamma^\alpha L^\alpha w - L_\lambda w| \\ &\leq \sup_{\alpha\in\Lambda} |\gamma^\alpha a^\alpha - \mathrm{I}_n||D^2 w| + |\gamma^\alpha||b^\alpha||\nabla w| + |\lambda - c^\alpha\gamma^\alpha||w|, \end{aligned}$$

where I_n is the $n \times n$ identity matrix. The Cauchy–Schwarz inequality with a parameter gives

$$|F_\gamma[u] - F_\gamma[v] - L_\lambda w| \leq \left(\sup_{\alpha\in\Lambda}\sqrt{C^\alpha}\right)\sqrt{|D^2 w|^2 + 2\lambda|\nabla w|^2 + \lambda^2|w|^2},$$

where, for each $\alpha \in \Lambda$,

$$C^\alpha \coloneqq |\gamma^\alpha a^\alpha - \mathrm{I}_n|^2 + |\gamma^\alpha|^2\frac{|b^\alpha|^2}{2\lambda} + \frac{|\lambda - c^\alpha\gamma^\alpha|^2}{\lambda^2}. \tag{2.14}$$

Expanding the square terms in (2.14) gives

$$C^\alpha = n + 1 - 2\gamma^\alpha\left(\operatorname{Tr} a^\alpha + \frac{c^\alpha}{\lambda}\right) + |\gamma^\alpha|^2\left(|a^\alpha|^2 + \frac{|b^\alpha|^2}{2\lambda} + \frac{|c^\alpha|^2}{\lambda^2}\right).$$

The definition of γ in (2.9) and the Cordes condition (2.5) imply that $C^\alpha \leq 1 - \varepsilon$ on U for every $\alpha \in \Lambda$, thus completing the proof of (2.13). ∎

In the following analysis, we shall write $a \lesssim b$ for $a, b \in \mathbb{R}$ to signify that there exists a constant C such that $a \leq C\,b$, where C is independent of the mesh size and polynomial degrees used to define the finite element spaces below, but otherwise possibly dependent on other fixed quantities, such as the constants in (2.4) and (2.5) or the shape-regularity parameters of the mesh, for example.

3. Analysis of the PDE. For $\lambda \geq 0$ as above, define the seminorm $|\cdot|_{H^2(\Omega),\lambda}$ on $H^2(\Omega)$ by

$$|u|^2_{H^2(\Omega),\lambda} := |u|^2_{H^2(\Omega)} + 2\lambda|u|^2_{H^1(\Omega)} + \lambda^2\|u\|^2_{L^2(\Omega)}. \tag{3.1}$$

If $\lambda > 0$, then this defines a norm on $H^2(\Omega)$. The following result follows from the Miranda–Talenti estimate; see [13, 20, 27]. Recall that $L_\lambda u = \Delta u - \lambda u$.

THEOREM 2. *Let Ω be a bounded convex open subset of $\mathbb{R}^n$. Then, for any $\lambda \geq 0$ and any $u \in H^2(\Omega) \cap H^1_0(\Omega)$, the following inequalities hold:*

$$|u|_{H^2(\Omega),\lambda} \leq \|L_\lambda u\|_{L^2(\Omega)}, \tag{3.2a}$$

$$\|u\|_{H^2(\Omega)} \leq C\|L_\lambda u\|_{L^2(\Omega)}, \tag{3.2b}$$

where C is a positive constant depending only on n and $\operatorname{diam}\Omega$.

Proof. In [27, Theorem 2], it is shown that on bounded convex domains, we have the Miranda–Talenti estimate $|u|_{H^2(\Omega)} \leq \|\Delta u\|_{L^2(\Omega)}$ for any $u \in H^2(\Omega) \cap H^1_0(\Omega)$. The identity $\int_\Omega u\Delta u\,\mathrm{d}x = -\int_\Omega |\nabla u|^2 \mathrm{d}x$, based on integration by parts, gives

$$\|L_\lambda u\|^2_{L^2(\Omega)} = \int_\Omega (\Delta u - \lambda u)^2\,\mathrm{d}x = \|\Delta u\|^2_{L^2(\Omega)} + 2\lambda|u|^2_{H^1(\Omega)} + \lambda^2\|u\|^2_{L^2(\Omega)}. \tag{3.3}$$

The Miranda–Talenti estimate and (3.3) give (3.2a). The bound (3.2b) follows from (3.3) and the estimate $\|u\|_{H^2(\Omega)} \leq C(n, \operatorname{diam}\Omega)\|\Delta u\|_{L^2(\Omega)}$ shown in [27, Theorem 2]. ∎

THEOREM 3. *Let Ω be a bounded convex open subset of $\mathbb{R}^n$, and let Λ be a compact metric space. Let the data a, b, c, f be continuous on $\overline{\Omega} \times \Lambda$ and satisfy* (2.4) *and either* (2.5) *with $\lambda > 0$ or* (2.6) *with $c \equiv 0$, $b \equiv 0$, and $\lambda = 0$. Then, there exists a unique strong solution $u \in H^2(\Omega) \cap H^1_0(\Omega)$ of the HJB equation* (2.3). *Moreover, u is also the unique solution of $F_\gamma[u] = 0$ in Ω, $u = 0$ on $\partial\Omega$.*

Proof. First, set $H := H^2(\Omega) \cap H^1_0(\Omega)$; then H is a separable Hilbert space. The proof consists of showing solvability of the equation $F_\gamma[u] = 0$ in H by the method of Browder and Minty and establishing its equivalence with the HJB equation (2.3). Let the operator $\mathcal{A}\colon H \to H^*$ be defined by

$$\langle \mathcal{A}(u), v\rangle := \int_\Omega F_\gamma[u]\, L_\lambda v\,\mathrm{d}x, \quad u,\, v \in H. \tag{3.4}$$

We claim that $\mathcal{A}$ is Lipschitz continuous and strongly monotone. Indeed, let u, $v \in H$ and set $w := u - v$. Then, by adding and subtracting $L_\lambda w$, we get

$$\langle \mathcal{A}(u) - \mathcal{A}(v), u - v\rangle = \|L_\lambda w\|^2_{L^2(\Omega)} + \int_\Omega (F_\gamma[u] - F_\gamma[v] - L_\lambda w)\, L_\lambda w\,\mathrm{d}x. \tag{3.5}$$

Lemma 1 and the Cauchy–Schwarz inequality show that

$$\langle \mathcal{A}(u) - \mathcal{A}(v), u - v\rangle \geq \|L_\lambda w\|^2_{L^2(\Omega)} - \sqrt{1-\varepsilon}\,|w|_{H^2(\Omega),\lambda}\|L_\lambda w\|_{L^2(\Omega)}. \tag{3.6}$$

We then use (3.2a) to obtain $\langle \mathcal{A}(u) - \mathcal{A}(v), u - v\rangle \geq \left(1 - \sqrt{1-\varepsilon}\right) \|L_\lambda w\|^2_{L^2(\Omega)}$, so $\|u - v\|^2_{H^2(\Omega)} \lesssim \langle \mathcal{A}(u) - \mathcal{A}(v), u - v\rangle$ as a result of (3.2b), thus showing that $\mathcal{A}$ is strongly monotone. Compactness of Λ and continuity of the data imply that $\mathcal{A}$ is Lipschitz continuous: to see this, let u, v, $z \in H$. Then, we find that

$$|\langle \mathcal{A}(u) - \mathcal{A}(v), z\rangle| \leq \|F_\gamma[u] - F_\gamma[v]\|_{L^2(\Omega)} \|L_\lambda z\|_{L^2(\Omega)} \leq C\|u - v\|_{H^2(\Omega)} \|z\|_{H^2(\Omega)},$$

where the constant C depends only on λ and on the supremum norms of a_{ij}, b_i, c, and γ over $\overline{\Omega} \times \Lambda$ for i, $j = 1, \ldots, n$. Lipschitz continuity and strong monotonicity imply that $\mathcal{A}$ is bounded, continuous, coercive, and strongly monotone, so the Browder–Minty theorem [25] shows that there exists a unique $u \in H$ such that $\mathcal{A}(u) = 0$.

For every $g \in L^2(\Omega)$, there is a $v \in H$ such that $L_\lambda v = g$. Therefore $\mathcal{A}(u) = 0$ implies $\int_\Omega F_\gamma[u]\, g \,\mathrm{d}x = 0$ for all $g \in L^2(\Omega)$, thus showing that $F_\gamma[u] = 0$ a.e. in Ω. We claim that $F_\gamma[u] = 0$ if and only if u solves (2.3). Since γ^α is positive, $\gamma^\alpha(L^\alpha u - f^\alpha) \leq 0$ for all $\alpha \in \Lambda$ is equivalent to $L^\alpha u - f^\alpha \leq 0$ for all $\alpha \in \Lambda$; i.e., $F[u] \leq 0$ if and only if $F_\gamma[u] \leq 0$. Compactness of Λ and continuity of a, b, c, f, and γ imply that at a.e. point of Ω, the suprema in the definitions of $F[u]$ and $F_\gamma[u]$ are attained by an element of Λ, thereby giving $F[u] \geq 0$ if and only if $F_\gamma[u] \geq 0$. Therefore, existence and uniqueness of the solution u of $F_\gamma[u] = 0$ in Ω is equivalent to existence and uniqueness of a solution of (2.3). ∎

4. Finite element spaces. Let $\{\mathcal{T}_h\}_h$ be a sequence of shape-regular meshes on Ω, consisting of simplices or parallelepipeds. For each element $K \in \mathcal{T}_h$, let $h_K \coloneqq \operatorname{diam} K$. It is assumed that $h = \max_{K \in \mathcal{T}_h} h_K$ for each mesh $\mathcal{T}_h$. Let $\mathcal{F}_h^i$ denote the set of interior faces of the mesh $\mathcal{T}_h$, and let $\mathcal{F}_h^{i,b}$ denote the set of boundary faces. The set of all faces is $\mathcal{F}_h^{i,b} \coloneqq \mathcal{F}_h^i \cup \mathcal{F}_h^b$. Since each element has piecewise flat boundary, the faces may also be chosen to be flat.

Mesh conditions. We shall make the following assumptions on the meshes. The meshes are allowed to be irregular, i.e., there may be hanging nodes. We assume that there is a uniform upper bound on the number of faces composing the boundary of any given element; in other words, there is a $c_{\mathcal{F}} > 0$, independent of h, such that

$$\max_{K \in \mathcal{T}_h} \operatorname{card}\left\{F \in \mathcal{F}_h^{i,b} \colon F \subset \partial K\right\} \leq c_{\mathcal{F}} \qquad \forall K \in \mathcal{T}_h,\ \forall h > 0. \tag{4.1}$$

It is also assumed that any two elements sharing a face have commensurate diameters, i.e., there is a $c_{\mathcal{T}} \geq 1$, independent of h, such that

$$\max(h_K, h_{K'}) \leq c_{\mathcal{T}} \min(h_K, h_{K'}), \tag{4.2}$$

for any K and K' in $\mathcal{T}_h$ that share a face. For each h, let $\mathbf{p} = (p_K \colon K \in \mathcal{T}_h)$ be a vector of positive integers. In order to let p_K appear in the denominator of various expressions, we shall assume that $p_K \geq 1$ for all $K \in \mathcal{T}_h$. We make the assumption that $\mathbf{p}$ has *local bounded variation* [15]: there is a $c_{\mathcal{P}} \geq 1$, independent of h, such that

$$\max(p_K, p_{K'}) \leq c_{\mathcal{P}} \min(p_K, p_{K'}), \tag{4.3}$$

for any K and K' in $\mathcal{T}_h$ that share a face.

Function spaces. For each $K \in \mathcal{T}_h$, let $\mathcal{P}_{p_K}(K)$ be the space of all polynomials with either total or partial degree less than or equal to p_K. The discontinuous Galerkin finite element space $V_{h,\mathbf{p}}$ is defined by

$$V_{h,\mathbf{p}} \coloneqq \left\{v \in L^2(\Omega),\ v|_K \in \mathcal{P}_{p_K}(K)\ \ \forall K \in \mathcal{T}_h\right\}. \tag{4.4}$$

Let $\mathbf{s} = (s_K : K \in \mathcal{T}_h)$ denote a vector of nonnegative real numbers, and let $r \in [1, \infty]$. The broken Sobolev space $W^{\mathbf{s},r}(\Omega; \mathcal{T}_h)$ is defined by

$$W^{\mathbf{s},r}(\Omega; \mathcal{T}_h) := \{v \in L^r(\Omega),\ v|_K \in W^{s_K,r}(K)\ \ \forall K \in \mathcal{T}_h\}. \tag{4.5}$$

For shorthand, define $H^{\mathbf{s}}(\Omega; \mathcal{T}_h) := W^{\mathbf{s},2}(\Omega; \mathcal{T}_h)$, and set $W^{s,r}(\Omega; \mathcal{T}_h) := W^{\mathbf{s},r}(\Omega; \mathcal{T}_h)$, where $s_K = s$, $s \geq 0$, for all $K \in \mathcal{T}_h$. For $v \in W^{1,r}(\Omega; \mathcal{T}_h)$, let $\nabla_h v \in L^r(\Omega; \mathbb{R}^n)$ denote the broken gradient of v, i.e., $(\nabla_h v)|_K = \nabla(v|_K)$ for all $K \in \mathcal{T}_h$. Higher broken derivatives are defined in a similar way. Define a norm on $W^{s,r}(\Omega; \mathcal{T}_h)$ by

$$\|v\|^r_{W^{s,r}(\Omega;\mathcal{T}_h)} := \sum_{K \in \mathcal{T}_h} \|v\|^r_{W^{s,r}(K)} \tag{4.6}$$

with the usual modification when $r = \infty$.

Jump, average, and tangential operators. For each face F, let $n_F \in \mathbb{R}^n$ denote a *fixed* choice of a unit normal vector to F. Since each face F is flat, the normal n_F is constant. For an element $K \in \mathcal{T}_h$ and a face $F \subset \partial K$, let $\tau_F : H^s(K) \to H^{s-1/2}(F)$, $s > 1/2$, denote the trace operator from K to F. The trace operator τ_F is extended componentwise to vector-valued functions. Define the jump operator $[\![\cdot]\!]$ and the average operator $\{\cdot\}$ by

$$\begin{aligned} &[\![\phi]\!] := \tau_F\left(\phi|_{K_{\text{ext}}}\right) - \tau_F\left(\phi|_{K_{\text{int}}}\right), && \{\phi\} := \tfrac{1}{2}\tau_F\left(\phi|_{K_{\text{ext}}}\right) + \tfrac{1}{2}\tau_F\left(\phi|_{K_{\text{int}}}\right) && \text{if } F \in \mathcal{F}_h^i, \\ &[\![\phi]\!] := \tau_F\left(\phi|_{K_{\text{ext}}}\right), && \{\phi\} := \tau_F\left(\phi|_{K_{\text{ext}}}\right) && \text{if } F \in \mathcal{F}_h^b, \end{aligned}$$

where ϕ is a sufficiently regular scalar or vector-valued function, and K_{ext} and K_{int} are the elements of which F is a face, i.e., $F = \partial K_{\text{ext}} \cap \partial K_{\text{int}}$. Here, the labeling is chosen so that n_F is outward pointing for K_{ext} and inward pointing for K_{int}. Using this notation, the jump and average of scalar-valued functions, respectively, vector-valued, are scalar-valued, respectively, vector-valued. For two matrices $A, B \in \mathbb{R}^{n\times n}$, we set $A : B = \sum_{i,j=1}^n A_{ij}B_{ij}$. For an element K, we define the inner product $\langle\cdot,\cdot\rangle_K$ by

$$\langle u, v\rangle_K := \begin{cases} \int_K u\, v \,\mathrm{d}x & \text{if } u, v \in L^2(K), \\ \int_K u \cdot v \,\mathrm{d}x & \text{if } u, v \in L^2(K; \mathbb{R}^n), \\ \int_K u : v \,\mathrm{d}x & \text{if } u, v \in L^2(K; \mathbb{R}^{n\times n}). \end{cases} \tag{4.7}$$

The abuse of notation will be resolved by the arguments of the inner product. The inner products $\langle\cdot,\cdot\rangle_{\partial K}$ and $\langle\cdot,\cdot\rangle_F$, $F \in \mathcal{F}_h^{i,b}$, are defined in a similar way.

For $F \in \mathcal{F}_h^{i,b}$, denote the space of H^s-regular tangential vector fields on F by $H^s_{\mathrm{T}}(F) := \{v \in H^s(F)^n : v \cdot n_F = 0 \text{ on } F\}$. We define below the tangential gradient $\nabla_{\mathrm{T}} : H^s(F) \to H^{s-1}_{\mathrm{T}}(F)$ and the tangential divergence $\mathrm{div}_{\mathrm{T}} : H^s_{\mathrm{T}}(F) \to H^{s-1}(F)$, where $s \geq 1$, following [13]. Let $\{t_i\}_{i=1}^{n-1} \subset \mathbb{R}^n$ be an orthonormal coordinate system on F. Then, for $u \in H^s(F)$ and $v = \sum_{i=1}^{n-1} v_i\, t_i$, with $v_i \in H^s(F)$ for $i = 1, \ldots, n-1$, we define

$$\nabla_{\mathrm{T}}\, u := \sum_{i=1}^{n-1} t_i \frac{\partial u}{\partial t_i}, \qquad \mathrm{div}_{\mathrm{T}}\, v := \sum_{i=1}^{n-1} \frac{\partial v_i}{\partial t_i}. \tag{4.8}$$

5. Numerical scheme. The definition of the numerical scheme requires the following bilinear and nonlinear forms. First, for $\lambda \geq 0$ as above, the symmetric bilinear form $B_{h,*} : V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ is defined by

$$
\begin{aligned}
B_{h,*}(u_h, v_h) \coloneqq & \sum_{K\in\mathcal{T}_h} \left[\langle D^2 u_h, D^2 v_h\rangle_K + 2\lambda\langle \nabla u_h, \nabla v_h\rangle_K + \lambda^2\langle u_h, v_h\rangle_K\right] \\
& + \sum_{F\in\mathcal{F}_h^i} \left[\langle \operatorname{div}_{\mathrm{T}} \nabla_{\mathrm{T}} \{u_h\}, [\![\nabla v_h \cdot n_F]\!]\rangle_F + \langle \operatorname{div}_{\mathrm{T}} \nabla_{\mathrm{T}} \{v_h\}, [\![\nabla u_h \cdot n_F]\!]\rangle_F\right] \\
& - \sum_{F\in\mathcal{F}_h^{i,b}} \left[\langle \nabla_{\mathrm{T}} \{\nabla u_h \cdot n_F\}, [\![\nabla_{\mathrm{T}} v_h]\!]\rangle_F + \langle \nabla_{\mathrm{T}} \{\nabla v_h \cdot n_F\}, [\![\nabla_{\mathrm{T}} u_h]\!]\rangle_F\right] \\
& - \lambda \sum_{F\in\mathcal{F}_h^{i,b}} \left[\langle \{\nabla u_h \cdot n_F\}, [\![v_h]\!]\rangle_F + \langle \{\nabla v_h \cdot n_F\}, [\![u_h]\!]\rangle_F\right] \\
& - \lambda \sum_{F\in\mathcal{F}_h^{i}} \left[\langle \{u_h\}, [\![\nabla v_h \cdot n_F]\!]\rangle_F + \langle \{v_h\}, [\![\nabla u_h \cdot n_F]\!]\rangle_F\right],
\end{aligned}
$$

where u_h and v_h will denote functions in $V_{h,\mathbf{p}}$ throughout this work. Then, for positive face-dependent quantities μ_F and η_F to be specified later, the jump stabilization bilinear form $J_h\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ is defined by

$$
\begin{aligned}
J_h(u_h, v_h) \coloneqq & \sum_{F\in\mathcal{F}_h^{i,b}} \left[\mu_F \langle [\![\nabla_{\mathrm{T}} u_h]\!], [\![\nabla_{\mathrm{T}} v_h]\!]\rangle_F + \eta_F \langle [\![u_h]\!], [\![v_h]\!]\rangle_F\right] \\
& + \sum_{F\in\mathcal{F}_h^{i}} \mu_F \langle [\![\nabla u_h \cdot n_F]\!], [\![\nabla v_h \cdot n_F]\!]\rangle_F.
\end{aligned} \tag{5.1}
$$

For each $\theta \in [0,1]$, define the bilinear form $B_{h,\theta}\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ by

$$
B_{h,\theta}(u_h, v_h) \coloneqq \theta B_{h,*}(u_h, v_h) + (1-\theta) \sum_{K\in\mathcal{T}_h} \langle L_\lambda u_h, L_\lambda v_h\rangle_K + J_h(u_h, v_h). \tag{5.2}
$$

The nonlinear form $A_h\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ is defined by

$$
A_h(u_h\,; v_h) \coloneqq \sum_{K\in\mathcal{T}_h} \langle F_\gamma[u_h], L_\lambda v_h\rangle_K + B_{h,1/2}(u_h, v_h) - \sum_{K\in\mathcal{T}_h} \langle L_\lambda u_h, L_\lambda v_h\rangle_K. \tag{5.3}
$$

The form A_h is linear in its second argument but nonlinear in its first argument. The scheme for approximating the solution of (2.3) is to find $u_h \in V_{h,\mathbf{p}}$ such that

$$
A_h(u_h\,; v_h) = 0 \qquad \forall\, v_h \in V_{h,\mathbf{p}}. \tag{5.4}
$$

The choice of nonlinear form in (5.3) is made to mirror the addition-substraction step of (3.5) in the proof of Theorem 3. It will be seen below that the last two terms of (5.3) cancel when the first argument of the form is smooth, and it is in this sense that this method relates the residual of the numerical solution to its lack of smoothness.

5.1. Consistency. The next result shows that the bilinear form $B_{h,\theta}$ is obtained from a discrete analogue of the identities that underpin Theorem 2.

LEMMA 4. *Let Ω be a bounded Lipschitz polytopal domain and let $\mathcal{T}_h$ be a simplicial or parallelepipedal mesh on Ω. Let $w \in H^s(\Omega;\mathcal{T}_h) \cap H^2(\Omega) \cap H_0^1(\Omega)$, $s > 5/2$. Then, for every $v_h \in V_{h,\mathbf{p}}$, we have the identities*

$$
B_{h,*}(w, v_h) = \sum_{K\in\mathcal{T}_h} \langle L_\lambda w, L_\lambda v_h\rangle_K \quad \textit{and} \quad J_h(w, v_h) = 0. \tag{5.5}
$$

Proof. The second part of (5.5) is obvious. We also note that all terms in $B_{h,*}(w, v_h)$ that involve jumps of w or of its first derivates vanish. For the case $\lambda = 0$, the stated result reduces to [27, Lemma 5], which treats the consistency of the second-order terms, namely, $B_{h,*}(w, v_h) = \sum_K \langle \Delta w, \Delta v_h \rangle_K$ for all $v_h \in V_{h,\mathbf{p}}$. So, for $\lambda > 0$, the identities of (5.5) are deduced from the previous result and from the identities

$$-\lambda \sum_{K\in\mathcal{T}_h} \langle \Delta w, v_h\rangle_K = \lambda \sum_{K\in\mathcal{T}_h} \langle \nabla w, \nabla v_h\rangle_K - \lambda \sum_{F\in\mathcal{F}_h^{i,b}} \langle \{\nabla w \cdot n_F\}, [\![v_h]\!]\rangle_F, \tag{5.6}$$

$$-\lambda \sum_{K\in\mathcal{T}_h} \langle w, \Delta v_h\rangle_K = \lambda \sum_{K\in\mathcal{T}_h} \langle \nabla w, \nabla v_h\rangle_K - \lambda \sum_{F\in\mathcal{F}_h^{i}} \langle \{w\}, [\![\nabla v_h \cdot n_F]\!]\rangle_F \tag{5.7}$$

for all $v_h \in V_{h,\mathbf{p}}$, where we use the fact that $w|_F = 0$ for all $F \in \mathcal{F}_h^b$ in (5.7). ∎

If the function w satisfies the hypotheses of Lemma 4, then (5.5) implies that

$$B_{h,\theta}(w, v_h) = \sum_{K\in\mathcal{T}_h} \langle L_\lambda w, L_\lambda v_h\rangle_K \qquad \forall\, v_h \in V_{h,\mathbf{p}},\ \forall\, \theta \in [0,1]. \tag{5.8}$$

The following consistency result for the scheme (5.4) follows immediately from Theorem 3, (5.8), and from the definition of A_h in (5.3).

COROLLARY 5. *Let Ω be a bounded convex polytopal domain, let $\mathcal{T}_h$ be a simplicial or parallelepipedal mesh, and let $u \in H^2(\Omega) \cap H_0^1(\Omega)$ be the unique solution of* (2.3). *If $u \in H^s(\Omega; \mathcal{T}_h)$, $s > 5/2$, then u satisfies $A_h(u\,; v_h) = 0$ for every $v_h \in V_{h,\mathbf{p}}$.*

The above consistency result involves a regularity assumption on the solution. This assumption is also used in the error analysis of section 7. However, we refer the reader to the numerical experiment of section 9.3 for an example of convergence of the scheme when this assumption is relaxed.

6. Stability. For $\lambda \geq 0$ as above, define the seminorms $|\cdot|_{H^2(K),\lambda}$, $K \in \mathcal{T}_h$, and $|\cdot|_{H^2(\Omega;\mathcal{T}_h),\lambda}$ on $H^2(\Omega; \mathcal{T}_h)$ by

$$|v|^2_{H^2(K),\lambda} \coloneqq \|D^2 v\|^2_{L^2(K)} + 2\lambda \|\nabla v\|^2_{L^2(K)} + \lambda^2 \|v\|^2_{L^2(K)}, \tag{6.1}$$

$$|v|^2_{H^2(\Omega;\mathcal{T}_h),\lambda} \coloneqq \sum_{K\in\mathcal{T}_h} |v|^2_{H^2(K),\lambda}. \tag{6.2}$$

For each $\theta \in [0,1]$, define the functional $\|\cdot\|_{h,\theta} \colon V_{h,\mathbf{p}} \to \mathbb{R}_{\geq 0}$ by

$$\|v_h\|^2_{h,\theta} \coloneqq \sum_{K\in\mathcal{T}_h} \left[\theta |v_h|^2_{H^2(K),\lambda} + (1-\theta)\|L_\lambda v_h\|^2_{L^2(K)}\right] + J_h(v_h, v_h). \tag{6.3}$$

For each $\theta \in [0,1]$, $\|\cdot\|_{h,\theta}$ is a norm on $V_{h,\mathbf{p}}$. Indeed, homogeneity and the triangle inequality are clear. If $\|v_h\|_{h,\theta} = 0$, then $v_h \in H^2(\Omega) \cap H_0^1(\Omega)$ since $[\![\nabla v_h]\!] = 0$ for all $F \in \mathcal{F}_h^i$ and $[\![v_h]\!] = 0$ for all $F \in \mathcal{F}_h^{i,b}$. Moreover, $L_\lambda v_h \equiv 0$ (if $\theta = 1$, use $|v_h|_{H^2(K),\lambda} = 0$ for all K), so $v_h \equiv 0$ as a result of (3.2b).

For each face $F \in \mathcal{F}_h^{i,b}$, define

$$\tilde{h}_F \coloneqq \begin{cases} \min(h_K, h_{K'}) & \text{if } F \in \mathcal{F}_h^i, \\ h_K & \text{if } F \in \mathcal{F}_h^b, \end{cases} \qquad \tilde{p}_F \coloneqq \begin{cases} \max(p_K, p_{K'}) & \text{if } F \in \mathcal{F}_h^i, \\ p_K & \text{if } F \in \mathcal{F}_h^b, \end{cases} \tag{6.4}$$

where K and K' are such that $F = \partial K \cap \partial K'$ if $F \in \mathcal{F}_h^i$ or $F \subset \partial K \cap \partial\Omega$ if $F \in \mathcal{F}_h^b$. The assumptions on the mesh and the polynomial degrees, in particular (4.2) and (4.3), show that if F is a face of K, then

$$h_K \le c_{\mathcal{T}}\, \tilde{h}_F \quad \text{and} \quad \tilde{p}_F \le c_{\mathcal{P}}\, p_K. \tag{6.5}$$

LEMMA 6. *Let Ω be a bounded convex polytopal domain and let $\{\mathcal{T}_h\}_h$ be a shape-regular sequence of simplicial or parallelepipedal meshes satisfying* (4.1). *Then, for each constant $\kappa > 1$, there exists a positive constant c_{stab}, independent of h, $\mathbf{p}$, and θ, such that for any $v_h \in V_{h,\mathbf{p}}$ and any $\theta \in [0,1]$, we have*

$$B_{h,\theta}(v_h, v_h) \ge \frac{\theta}{\kappa} |v_h|^2_{H^2(\Omega;\mathcal{T}_h),\lambda} + (1-\theta) \sum_{K\in\mathcal{T}_h} \|L_\lambda v_h\|^2_{L^2(K)} + \tfrac{1}{2} J_h(v_h, v_h) \tag{6.6}$$

whenever, for any fixed constant $\sigma \ge 1$,

$$\mu_F = \sigma c_{\mathrm{stab}} \frac{\tilde{p}_F^2}{\tilde{h}_F} \quad \textit{and} \quad \eta_F > \sigma\lambda\, c_{\mathrm{stab}} \frac{\tilde{p}_F^2}{\tilde{h}_F}. \tag{6.7}$$

The strict inequality in the second part of (6.7) serves to cover the case $\lambda = 0$.

Proof. For $v_h \in V_{h,\mathbf{p}}$, we have

$$B_{h,\theta}(v_h, v_h) = \theta |v_h|^2_{H^2(\Omega;\mathcal{T}_h),\lambda} + (1-\theta) \sum_{K\in\mathcal{T}_h} \|L_\lambda v_h\|^2_{L^2(K)} + J_h(v_h, v_h) + \theta \sum_{i=1}^{4} I_i,$$

where

$$I_1 \coloneqq 2 \sum_{F\in\mathcal{F}_h^i} \langle \mathrm{div}_{\mathrm{T}}\, \nabla_{\mathrm{T}} \{v_h\}, [\![\nabla v_h \cdot n_F]\!] \rangle_F, \qquad I_3 \coloneqq -2\lambda \sum_{F\in\mathcal{F}_h^i} \langle \{v_h\}, [\![\nabla v_h \cdot n_F]\!] \rangle_F,$$

$$I_2 \coloneqq -2 \sum_{F\in\mathcal{F}_h^{i,b}} \langle \nabla_{\mathrm{T}} \{\nabla v_h \cdot n_F\}, [\![\nabla_{\mathrm{T}} v_h]\!] \rangle_F, \quad I_4 \coloneqq -2\lambda \sum_{F\in\mathcal{F}_h^{i,b}} \langle \{\nabla v_h \cdot n_F\}, [\![v_h]\!] \rangle_F.$$

In [27, Lemma 7], it is shown that there is a constant $C(n)$ depending only on n, such that for any $\delta > 0$,

$$|I_1| \le \delta\, C(n)\, C_{\mathrm{Tr}}\, c_{\mathcal{F}} \sum_{K\in\mathcal{T}_h} \|D^2 v_h\|^2_{L^2(K)} + \sum_{F\in\mathcal{F}_h^i} \frac{\tilde{p}_F^2}{\delta \tilde{h}_F} \|[\![\nabla v_h \cdot n_F]\!]\|^2_{L^2(F)}, \tag{6.8}$$

$$|I_2| \le \delta\, C(n)\, C_{\mathrm{Tr}}\, c_{\mathcal{F}} \sum_{K\in\mathcal{T}_h} \|D^2 v_h\|^2_{L^2(K)} + \sum_{F\in\mathcal{F}_h^{i,b}} \frac{\tilde{p}_F^2}{\delta \tilde{h}_F} \|[\![\nabla_{\mathrm{T}} v_h]\!]\|^2_{L^2(F)}, \tag{6.9}$$

where C_{Tr} is the combined constant of the trace and inverse inequalities, and $c_{\mathcal{F}}$ is given by (4.1). The inverse and trace inequalities also show that

$$\begin{aligned} |I_3| &\le 2\lambda \sqrt{\sum_{F\in\mathcal{F}_h^i} \frac{\delta \tilde{h}_F}{\tilde{p}_F^2} \|\{v_h\}\|^2_{L^2(F)}} \sqrt{\sum_{F\in\mathcal{F}_h^i} \frac{\tilde{p}_F^2}{\delta \tilde{h}_F} \|[\![\nabla v_h \cdot n_F]\!]\|^2_{L^2(F)}} \\ &\le \delta\, C(n)\, C_{\mathrm{Tr}}\, c_{\mathcal{F}} \sum_{K\in\mathcal{T}_h} \lambda^2 \|v_h\|^2_{L^2(K)} + \sum_{F\in\mathcal{F}_h^i} \frac{\tilde{p}_F^2}{\delta \tilde{h}_F} \|[\![\nabla v_h \cdot n_F]\!]\|^2_{L^2(F)}. \end{aligned} \tag{6.10}$$

Similarly, it is found that

$$|I_4| \leq \delta\, C(n)\, C_{\mathrm{Tr}}\, c_{\mathcal{F}} \sum_{K\in\mathcal{T}_h} 2\lambda \|\nabla v_h\|^2_{L^2(K)} + \sum_{F\in\mathcal{F}_h^{i,b}} \frac{\lambda \tilde{p}_F^2}{2\delta \tilde{h}_F} \|[\![v_h]\!]\|^2_{L^2(F)}. \tag{6.11}$$

We may take $C(n)$ to be the same constant in each of the above estimates. So,

$$\begin{aligned} B_{h,\theta}(v_h, v_h) &\geq \theta(1-\delta C(n) C_{\mathrm{Tr}} c_{\mathcal{F}}) |v_h|^2_{H^2(\Omega;\mathcal{T}_h),\lambda} + (1-\theta) \sum_{K\in\mathcal{T}_h} \|L_\lambda v_h\|^2_{L^2(K)} \\ &+ \sum_{F\in\mathcal{F}_h^i} \left(\mu_F - \frac{2\theta\tilde{p}_F^2}{\delta \tilde{h}_F} \right) \|[\![\nabla v_h \cdot n_F]\!]\|^2_{L^2(F)} + \sum_{F\in\mathcal{F}_h^{i,b}} \left(\mu_F - \frac{\theta\tilde{p}_F^2}{\delta \tilde{h}_F} \right) \|[\![\nabla_{\mathrm{T}}\, v_h]\!]\|^2_{L^2(F)} \\ &+ \sum_{F\in\mathcal{F}_h^{i,b}} \left(\eta_F - \frac{\lambda\theta\tilde{p}_F^2}{2\delta \tilde{h}_F} \right) \|[\![v_h]\!]\|^2_{L^2(F)}. \end{aligned}$$

For any given $\kappa > 1$, there is a $\delta > 0$ such that $1 - \delta C(n) C_{\mathrm{Tr}} c_{\mathcal{F}} > \kappa^{-1}$. Set $c_{\mathrm{stab}} = 4/\delta$, so that (6.6) holds whenever μ_F and η_F satisfy (6.7). ∎

THEOREM 7. *Let Ω be a bounded convex polytopal domain and let $\{\mathcal{T}_h\}_h$ be a shape-regular sequence of simplicial or parallelepipedal meshes satisfying* (4.1). *Let Λ be a compact metric space and let the data satisfy* (2.4) *and either* (2.5) *or* (2.6) *with $b \equiv 0$, $c \equiv 0$, $\lambda = 0$. Let c_{stab}, η_F, and μ_F be chosen so that Lemma* 6 *holds with $\kappa < (1-\varepsilon)^{-1}$. Then, for every u_h, $v_h \in V_{h,\mathbf{p}}$, we have*

$$\|u_h - v_h\|^2_{h,1} \leq C\big(A_h(u_h\,; u_h - v_h) - A_h(v_h\,; u_h - v_h) \big), \tag{6.12}$$

where the constant $C \coloneqq 2\kappa/\left(1 - \kappa\,(1-\varepsilon)\right)$. Moreover, there exists a constant C, independent of h and $\mathbf{p}$, such that for any u_h, v_h, and z_h in $V_{h,\mathbf{p}}$,

$$|A_h(u_h\,; z_h) - A_h(v_h\,; z_h)| \leq C \|u_h - v_h\|_{h,1} \|z_h\|_{h,1}. \tag{6.13}$$

Therefore, there exists a unique solution $u_h \in V_{h,\mathbf{p}}$ to the numerical scheme (5.4). *We have the bound*

$$\|u_h\|_{h,1} \leq \frac{2\kappa\,\sqrt{n+1}\,\|\gamma\|_{C(\overline{\Omega}\times\Lambda)}}{1 - \kappa\,(1-\varepsilon)} \,\|\sup_{\alpha\in\Lambda} |f^\alpha|\|_{L^2(\Omega)}. \tag{6.14}$$

Proof. First, note that since $\varepsilon \in (0,1)$, it is possible to choose the constants c_{stab}, μ_F, and η_F such that $\kappa < (1-\varepsilon)^{-1}$. Let u_h and v_h belong to $V_{h,\mathbf{p}}$ and set $w_h \coloneqq u_h - v_h$. Then, we have

$$A_h(u_h\,; w_h) - A_h(v_h\,; w_h) = B_{h,1/2}(w_h, w_h) + \sum_{K\in\mathcal{T}_h} \langle F_\gamma[u_h] - F_\gamma[v_h] - L_\lambda w_h, L_\lambda w_h \rangle_K.$$

Note that Lemma 1 gives

$$\begin{aligned} \sum_{K\in\mathcal{T}_h} |\langle F_\gamma[u_h] - F_\gamma[v_h] - L_\lambda w_h, L_\lambda w_h\rangle_K| &\leq \sqrt{1-\varepsilon} \sum_{K\in\mathcal{T}_h} |w_h|_{H^2(K),\lambda} \|L_\lambda w_h\|_{L^2(K)} \\ &\leq \frac{1-\varepsilon}{2} |w_h|^2_{H^2(\Omega;\mathcal{T}_h),\lambda} + \frac{1}{2} \sum_{K\in\mathcal{T}_h} \|L_\lambda w_h\|^2_{L^2(K)}. \end{aligned}$$

This estimate and Lemma 6 show that

$$A_h(u_h\,;w_h)-A_h(v_h\,;w_h)\geq\frac{1-\kappa\,(1-\varepsilon)}{2\kappa}|w_h|^2_{H^2(\Omega;\mathcal{T}_h),\lambda}+\tfrac{1}{2}J_h(w_h,w_h)\geq C^{-1}\|w_h\|^2_{h,1},$$

where $C\coloneqq 2\kappa/\,(1-\kappa\,(1-\varepsilon))$. Since $\kappa\,(1-\varepsilon)<1$, we obtain (6.12). Now, let $z_h\in V_{h,\mathbf{p}}$. Then, using linearity of $B_{h,\theta}$ and inverse inequalities, we find that there exists a constant C depending on the constants appearing in the proof of Lemma 6, but not on h or $\mathbf{p}$, such that $|B_{h,1/2}(u_h-v_h,z_h)|\leq C\|u_h-v_h\|_{h,1}\|z_h\|_{h,1}$. Using Lemma 1 and the above estimates, we deduce that there is a constant C depending only on n and ε such that

$$\sum_{K\in\mathcal{T}_h}|\langle F_\gamma[u_h]-F_\gamma[v_h]-L_\lambda(u_h-v_h),L_\lambda z_h\rangle_K|\leq C\|u_h-v_h\|_{h,1}\|z_h\|_{h,1}.$$

It then follows that A_h is Lipschitz continuous, as stated in (6.13). The Browder–Minty theorem [25] with (6.12) and (6.13) imply that there exists a unique $u_h\in V_{h,\mathbf{p}}$ such that $A_h(u_h\,;v_h)=0$ for all $v_h\in V_{h,\mathbf{p}}$. By taking $v_h=0$ in (6.12), we find that

$$\begin{aligned}\|u_h\|^2_{h,1}&\leq C|A_h(0\,;u_h)|\leq C\sum_{K\in\mathcal{T}_h}|\langle\sup_{\alpha\in\Lambda}\left[-\gamma^\alpha f^\alpha\right],L_\lambda u_h\rangle_K|\\&\leq C\|\gamma\|_{C(\overline{\Omega}\times\Lambda)}\|\sup_{\alpha\in\Lambda}|f^\alpha|\|_{L^2(\Omega)}\sqrt{n+1}\|u_h\|_{h,1},\end{aligned}$$

where $C=2\kappa/\,(1-\kappa\,(1-\varepsilon))$, thus showing the bound (6.14). ∎

In the above stability result, it was required that c_{stab} be chosen so that Lemma 6 holds for some $\kappa<(1-\varepsilon)^{-1}$. It can be seen from the proof of Lemma 6 that there exists a constant C, independent of the discretization parameters, such that this holds whenever $c_{\mathrm{stab}}\geq C/\varepsilon$. Then, c_{stab} and κ can be chosen so that the constant in (6.12) is of order $1/\varepsilon$ when ε is small.

7. Error analysis. The above stability results make use of the lower bound (6.7) on the jump penalty terms η_F, $F\in\mathcal{F}_h^{i,b}$. In the following, we require that

$$\eta_F\leq C\,\frac{\tilde{p}_F^4}{\tilde{h}_F^3}\qquad\forall\,F\in\mathcal{F}_h^{i,b},\tag{7.1}$$

where C is a fixed constant that is chosen sufficiently large to allow both (6.7) and (7.1). The good stability properties of the proposed method make it possible to obtain the following a priori error bound.

THEOREM 8. *Let Ω be a bounded convex polytopal domain, and let the shape-regular sequence of simplicial or parallelepipedal meshes $\{\mathcal{T}_h\}_h$ satisfy* (4.1) *and* (4.2)*, with $\mathbf{p}$ satisfying* (4.3) *for each h. Let Λ be a compact metric space, let the data satisfy* (2.4) *and either* (2.5) *or* (2.6) *when $b\equiv 0$, $c\equiv 0$, and $\lambda=0$, and let $u\in H^2(\Omega)\cap H^1_0(\Omega)$ be the unique solution of* (2.3)*. Assume that $u\in H^{\mathbf{s}}(\Omega;\mathcal{T}_h)$ with $s_K>5/2$ for each $K\in\mathcal{T}_h$. Let c_{stab}, μ_F, and η_F be chosen as in Theorem 7 for all $F\in\mathcal{F}_h^{i,b}$, and let η_F also satisfy* (7.1) *for each $F\in\mathcal{F}_h^{i,b}$. Then, there exists a positive constant C, independent of h, $\mathbf{p}$, and u, but depending on $\max_K s_K$, such that*

$$\|u-u_h\|^2_{h,1}\leq C\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K-4}}{p_K^{2s_K-5}}\|u\|^2_{H^{s_K}(K)},\tag{7.2}$$

where $t_K=\min(p_K+1,s_K)$ for each $K\in\mathcal{T}_h$.

Note that for the special case of quasi-uniform meshes and uniform polynomial degrees, if $u \in H^s(\Omega)$ with $s > 5/2$, the a priori estimate (7.2) simplifies to

$$\|u - u_h\|_{h,1} \leq C \frac{h^{\min(p+1,s)-2}}{p^{s-5/2}} \|u\|_{H^s(\Omega)}.$$

Therefore, the convergence rates are optimal with respect to the mesh size and suboptimal in the polynomial degree only by half an order.

Proof. Since the sequence of meshes is shape-regular, there is a $z_h \in V_{h,\mathbf{p}}$ and a constant C, independent of u, h_K, and p_K, but dependent on $\max_K s_K$, such that for each $K \in \mathcal{T}_h$, each nonnegative integer $q \leq s_K$, and each multi-index β with $|\beta| < s_K - 1/2$, we have

$$\|u - z_h\|_{H^q(K)} \leq C \frac{h_K^{t_K - q}}{p_K^{s_K - q}} \|u\|_{H^{s_K}(K)}, \tag{7.3}$$

$$\|D^\beta (u - z_h)\|_{L^2(\partial K)} \leq C \frac{h_K^{t_K - |\beta| - 1/2}}{p_K^{s_K - |\beta| - 1/2}} \|u\|_{H^{s_K}(K)}. \tag{7.4}$$

Set $\psi_h \coloneqq u_h - z_h$ and $\xi_h \coloneqq u - z_h$. By Corollary 5, we have $A_h(u\,; v_h) = 0$ for all $v_h \in V_{h,\mathbf{p}}$. Strong monotonicity of A_h on $V_{h,\mathbf{p}}$, as shown in Theorem 7, yields

$$\|\psi_h\|_{h,1}^2 \lesssim A_h(u_h\,; \psi_h) - A_h(z_h\,; \psi_h) = A_h(u\,; \psi_h) - A_h(z_h\,; \psi_h). \tag{7.5}$$

By applying the Cauchy–Schwarz inequality to the terms appearing on the right-hand side of (7.5) and applying inverse inequalities to $\psi_h \in V_{h,\mathbf{p}}$, we eventually obtain

$$A_h(u\,; \psi_h) - A_h(z_h\,; \psi_h) \leq \sqrt{\textstyle\sum_{i=1}^{10} E_i}\, \|\psi_h\|_{h,1}, \tag{7.6}$$

where the quantities E_i are defined by

$$E_1 \coloneqq \sum_{K \in \mathcal{T}_h} |\xi_h|_{H^2(K),\lambda}^2, \qquad E_2 \coloneqq \sum_{K \in \mathcal{T}_h} \|L_\lambda \xi_h\|_{L^2(K)}^2,$$

$$E_3 \coloneqq \sum_{K \in \mathcal{T}_h} \|F_\gamma[u] - F_\gamma[z_h]\|_{L^2(K)}^2, \qquad E_4 \coloneqq \sum_{F \in \mathcal{F}_h^i} \mu_F^{-1} \|\mathrm{div}_{\mathrm{T}}\, \nabla_{\mathrm{T}} \{\xi_h\}\|_{L^2(F)}^2,$$

$$E_5 \coloneqq \sum_{F \in \mathcal{F}_h^i} \mu_F \|[\![\nabla \xi_h \cdot n_F]\!]\|_{L^2(F)}^2, \qquad E_6 \coloneqq \sum_{F \in \mathcal{F}_h^{i,b}} \mu_F^{-1} \|\nabla_{\mathrm{T}} \{\nabla \xi_h \cdot n_F\}\|_{L^2(F)}^2,$$

$$E_7 \coloneqq \sum_{F \in \mathcal{F}_h^{i,b}} \mu_F \|[\![\nabla_{\mathrm{T}}\, \xi_h]\!]\|_{L^2(F)}^2, \qquad E_8 \coloneqq \sum_{F \in \mathcal{F}_h^{i,b}} \lambda^2 \eta_F^{-1} \|\{\nabla \xi_h \cdot n_F\}\|_{L^2(F)}^2,$$

$$E_9 \coloneqq \sum_{F \in \mathcal{F}_h^{i,b}} (\lambda \mu_F + \eta_F) \|[\![\xi_h]\!]\|_{L^2(F)}^2, \quad E_{10} \coloneqq \sum_{F \in \mathcal{F}_h^i} \lambda^2 \mu_F^{-1} \|\{\xi_h\}\|_{L^2(F)}^2.$$

The estimate (7.3) shows that

$$E_1 + E_2 \lesssim \sum_{K \in \mathcal{T}_h} \frac{h_K^{2t_K - 4}}{p_K^{2s_K - 4}} \|u\|_{H^{s_K}(K)}^2. \tag{7.7}$$

By compactness of Λ, continuity of the data and (2.4), F_γ is Lipschitz continuous, so

$$E_3 \lesssim \sum_{K \in \mathcal{T}_h} \|\xi_h\|_{H^2(K)}^2 \lesssim \sum_{K \in \mathcal{T}_h} \frac{h_K^{2t_K - 4}}{p_K^{2s_K - 4}} \|u\|_{H^{s_K}(K)}^2. \tag{7.8}$$

We use (4.1), (4.2), (4.3), (6.7), and (7.4) to obtain

$$E_4 + E_6 \lesssim \sum_{K\in\mathcal{T}_h} \frac{h_K}{p_K^2}\frac{h_K^{2t_K-5}}{p_K^{2s_K-5}}\|u\|^2_{H^{s_K}(K)} = \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-3}}\|u\|^2_{H^{s_K}(K)}, \tag{7.9}$$

$$E_5 + E_7 \lesssim \sum_{K\in\mathcal{T}_h} \frac{p_K^2}{h_K}\frac{h_K^{2t_K-3}}{p_K^{2s_K-3}}\|u\|^2_{H^{s_K}(K)} = \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-5}}\|u\|^2_{H^{s_K}(K)}. \tag{7.10}$$

Similarly, we use (6.7) to get

$$E_8 \lesssim \sum_{K\in\mathcal{T}_h} \frac{h_K}{p_K^2}\frac{h_K^{2t_K-3}}{p_K^{2s_K-3}}\|u\|^2_{H^{s_K}(K)} = \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K-2}}{p_K^{2s_K-1}}\|u\|^2_{H^{s_K}(K)}. \tag{7.11}$$

By hypothesis, $\eta_F \lesssim \tilde{p}_F^4/\tilde{h}_F^3$ by (7.1), so (4.2) and (4.3) imply that

$$E_9 \lesssim \sum_{K\in\mathcal{T}_h} \frac{p_K^4}{h_K^3}\frac{h_K^{2t_K-1}}{p_K^{2s_K-1}}\|u\|^2_{H^{s_K}(K)} = \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-5}}\|u\|^2_{H^{s_K}(K)}. \tag{7.12}$$

Finally, (7.4) yields

$$E_{10} \lesssim \sum_{K\in\mathcal{T}_h} \frac{h_K}{p_K^2}\frac{h_K^{2t_K-1}}{p_K^{2s_K-1}}\|u\|^2_{H^{s_K}(K)} = \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K}}{p_K^{2s_K+1}}\|u\|^2_{H^{s_K}(K)}. \tag{7.13}$$

The a priori bound (7.2) is obtained from $\|u-u_h\|_{h,1} \le \|\psi_h\|_{h,1} + \|\xi_h\|_{h,1}$ and the above estimates. ∎

8. Semismooth Newton method. We turn to the analysis of an algorithm for solving the discrete problem (5.4), which can be interpreted as a Newton method for nonsmooth operator equations [24]. After showing that the algorithm is well-posed, we obtain and then use a semismoothness result for HJB operators in function spaces to establish its superlinear convergence. The semismoothness of finite-dimensional HJB operators in a different form was studied in [3].

For $1 \le r \le \infty$, a function $u \in W^{2,r}(\Omega;\mathcal{T}_h)$ defines a vector-valued function $\mathbf{u} \in L^r(\Omega;\mathbb{R}^m)$ through $\mathbf{u} = \big(u, \nabla_h u, D_h^2 u\big)$, where $\nabla_h u$ and $D_h^2 u$ denote the broken gradient and broken Hessian of u; see section 4. For a vector $\mathbf{u} = (z,p,M) \in \mathbb{R}^m$, define the function $F_\gamma\colon \Omega\times\mathbb{R}^m \to \mathbb{R}$ by

$$F_\gamma(x,\mathbf{u}) \coloneqq \sup_{\alpha\in\Lambda} \left[\gamma^\alpha \left(a^\alpha\colon M + b^\alpha\cdot p - c^\alpha z - f^\alpha\right)\big|_x\right]. \tag{8.1}$$

For each $(x,\mathbf{u}) \in \Omega\times\mathbb{R}^m$, we define $\Lambda(x,\mathbf{u})$ as the set of all $\alpha\in\Lambda$ such that the supremum in (8.1) is attained. This defines a set-valued map $(x,\mathbf{u}) \mapsto \Lambda(x,\mathbf{u})$.

LEMMA 9. *Let Ω be a bounded open subset of $\mathbb{R}^n$, let Λ be a compact metric space, let the data a, b, c, and f be continuous on $\overline{\Omega}\times\Lambda$, and suppose that (2.4) holds. Then, for each $(x,\mathbf{u}) \in \Omega\times\mathbb{R}^m$, $\Lambda(x,\mathbf{u})$ is a nonempty closed subset of Λ. The set-valued map $(x,\mathbf{u}) \mapsto \Lambda(x,\mathbf{u})$ is upper semicontinuous; that is, for every $(x,\mathbf{u}) \in \Omega\times\mathbb{R}^m$, and any open neighborhood U of $\Lambda(x,\mathbf{u})$, there exists an open neighborhood V of $(x,\mathbf{u})$ such that $\Lambda(y,\mathbf{v}) \subset U$ for every $(y,\mathbf{v}) \in V$.*

We remark that the uniform ellipticity condition (2.4) is only used in Lemma 9 to guarantee that $\gamma \in C(\overline{\Omega}\times\Lambda)$.

Proof. For every $(x,\mathbf{u}) \in \Omega \times \mathbb{R}^m$, where $\mathbf{u} = (z,p,M)$, compactness of Λ and continuity of a, b, c, f, and γ imply the existence of a maximizer in (2.2); so $\Lambda(x,\mathbf{u})$ is nonempty. The set $\Lambda(x,\mathbf{u})$ is closed: if α is in the closure of $\Lambda(x,\mathbf{u})$, say, $\alpha_j \to \alpha$, with $\alpha_j \in \Lambda(x,\mathbf{u})$ for each $j \in \mathbb{N}$, then continuity of the data implies that

$$\gamma^{\alpha}\left(a^{\alpha}\colon M + b^{\alpha}\cdot p - c^{\alpha} z - f^{\alpha}\right)\big|_x = \lim_{j\to\infty} \gamma^{\alpha_j}\left(a^{\alpha_j}\colon M + b^{\alpha_j}\cdot p - c^{\alpha_j} z - f^{\alpha_j}\right)\big|_x . \tag{8.2}$$

Since $\alpha_j \in \Lambda(x,\mathbf{u})$ for each $j \in \mathbb{N}$, the right-hand side of (8.2) equals $F(x,\mathbf{u})$, thus giving $\alpha \in \Lambda(x,\mathbf{u})$ and showing that $\Lambda(x,\mathbf{u})$ is closed.

We prove upper semicontinuity of $(x,\mathbf{u}) \mapsto \Lambda(x,\mathbf{u})$ by contradiction. Suppose that there exists an $(x,\mathbf{u}) \in \Omega \times \mathbb{R}^m$, a neighborhood U of $\Lambda(x,\mathbf{u})$, and a sequence $\{(x_j,\mathbf{u}_j)\}_{j=1}^{\infty}$, $\mathbf{u}_j = (z_j,p_j,M_j)$, converging to $(x,\mathbf{u})$, together with $\alpha_j \in \Lambda(x_j,\mathbf{u}_j)\setminus U$ for all $j \in \mathbb{N}$. Because Λ is compact and $\Lambda \setminus U$ is closed, there exists a subsequence, to which we pass without change of notation, such that $\alpha_j \to \alpha \in \Lambda \setminus U$. On the one hand, $\Lambda(x,\mathbf{u})$ is nonempty so there is $\beta \in \Lambda(x,\mathbf{u})$. Then, by definition of F,

$$\gamma^{\alpha}\left(a^{\alpha}\colon M + b^{\alpha}\cdot p - c^{\alpha} z - f^{\alpha}\right)\big|_x \le F(x,\mathbf{u}). \tag{8.3}$$

On the other hand, $\alpha_j \in \Lambda(x_j,\mathbf{u}_j)$ implies that we have, for each $j \in \mathbb{N}$,

$$\gamma^{\alpha_j}\left(a^{\alpha_j}\colon M_j + b^{\alpha_j}\cdot p_j - c^{\alpha_j} z_j - f^{\alpha_j}\right)\big|_{x_j} \ge \gamma^{\beta}\left(a^{\beta}\colon M_j + b^{\beta}\cdot p_j - c^{\beta} z_j - f^{\beta}\right)\big|_{x_j} .$$

Taking the limit $j \to \infty$ in the above inequality shows that equality holds in (8.3) because $\beta \in \Lambda(x,\mathbf{u})$. Hence, $\alpha \in \Lambda(x,\mathbf{u})$; however, U is an open neighborhood of $\Lambda(x,\mathbf{u})$ and $\alpha \in \Lambda \setminus U$, so we have a contradiction. ∎

The following selection theorem, due to Kuratowski and Ryll-Nardzewski [18], is required for the analysis of the algorithm for solving (5.4).

THEOREM 10. *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set, let Λ be a compact metric space, and let $(x,\mathbf{u}) \mapsto \Lambda(x,\mathbf{u})$ be an upper semicontinuous set-valued function from $\Omega \times \mathbb{R}^m$ to the subsets of Λ, such that $\Lambda(x,\mathbf{u})$ is nonempty and closed for every $(x,\mathbf{u}) \in \Omega \times \mathbb{R}^m$. Then, for any Lebesgue measurable function $\mathbf{u}\colon \Omega \to \mathbb{R}^m$, there exists a Lebesgue measurable selection $\alpha\colon \Omega \to \Lambda$ such that $\alpha(x) \in \Lambda\big(x,\mathbf{u}(x)\big)$ for a.e. $x \in \Omega$.*

For $u \in W^{2,r}(\Omega;\mathcal{T}_h)$, let $\Lambda[u]$ be the set of all Lebesgue measurable functions $\alpha\colon \Omega \to \Lambda$ such that $\alpha(x) \in \Lambda(x,\mathbf{u}(x))$ for a.e. $x \in \Omega$, where $\mathbf{u} = \big(u, \nabla_h u, D_h^2 u\big)$. Lemma 9 and Theorem 10 show that $\Lambda[u]$ is nonempty for each $u \in W^{2,r}(\Omega;\mathcal{T}_h)$. For measurable $\alpha\colon \Omega \to \Lambda$, we define $\gamma^{\alpha}\colon \Omega \to \mathbb{R}_{>0}$ through $\gamma^{\alpha}(x) = \gamma(x,\alpha(x))$, where $\gamma\colon \Omega \times \Lambda \to \mathbb{R}_{>0}$ was defined by (2.9) or (2.10). It follows from uniform continuity of γ over $\Omega \times \Lambda$ that $\gamma^{\alpha} \in L^{\infty}(\Omega)$ with $\|\gamma^{\alpha}\|_{L^{\infty}(\Omega)} \le \|\gamma\|_{C(\overline{\Omega}\times\Lambda)}$. The functions a^{α}, b^{α}, c^{α}, and f^{α} and the operator L^{α} are defined in a similar way and are likewise bounded. It is clear that if $\alpha \in \Lambda[u]$, then $F_{\gamma}[u] = \gamma^{\alpha}(L^{\alpha}u - f^{\alpha})$ a.e. in Ω.

8.1. Algorithm. We now present the definition of the semismooth Newton method for solving (5.4) and state the main result concerning its convergence rate. Choose $u_h^0 \in V_{h,\mathbf{p}}$. Given $u_h^k \in V_{h,\mathbf{p}}$, $k \in \mathbb{N}$, choose $\alpha_k \in \Lambda[u_h^k]$. Then, obtain $u_h^{k+1} \in V_{h,\mathbf{p}}$ satisfying

$$A_h^k(u_h^{k+1}, v_h) = \sum_{K\in\mathcal{T}_h} \langle \gamma^{\alpha_k} f^{\alpha_k}, L_{\lambda} v_h\rangle_K \quad \forall\, v_h \in V_{h,\mathbf{p}}, \tag{8.4}$$

where the bilinear form $A_h^k \colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ is defined by

$$A_h^k(w_h, v_h) \coloneqq \sum_{K \in \mathcal{T}_h} \langle (\gamma^{\alpha_k} L^{\alpha_k} w_h, L_\lambda v_h \rangle_K + B_{h,1/2}(w_h, v_h) - \sum_{K \in \mathcal{T}_h} \langle L_\lambda w_h, L_\lambda v_h \rangle_K.$$

The fact that $\alpha_k \colon \Omega \to \Lambda$ is measurable ensures that A_h^k is well-defined. As in the proof of Theorem 7, it is found that the bilinear forms A_h^k, $k \in \mathbb{N}$, are coercive on $V_{h,\mathbf{p}}$. In fact, for each $k \in \mathbb{N}$, we have

$$\|v_h\|_{h,1}^2 \leq \frac{2\kappa}{1 - \kappa\,(1-\varepsilon)} A_h^k(v_h, v_h) \quad \forall\, v_h \in V_{h,\mathbf{p}}. \tag{8.5}$$

Therefore, the sequence of iterates $\{u_h^k\}_{k=1}^\infty$ is well-defined by (8.4) and remains bounded in $V_{h,\mathbf{p}}$. The main result of this section is the following.

THEOREM 11. *Under the hypotheses of Theorem* 7*, there exists a constant $R > 0$, possibly depending on h and $\mathbf{p}$, such that if $\|u_h - u_h^0\|_{h,1} < R$, where u_h solves* (5.4)*, then the sequence $\{u_h^k\}_{k=1}^\infty$ converges to u_h with a superlinear convergence rate.*

The proof of this theorem will be given in the next section. Despite the possible dependence of R on h and p in the above theorem, it is seen from the numerical experiments in section 9, in particular in Figure 2 later, that in practice, the convergence rates of the algorithm depend only weakly on the discretization parameters.

Since the bilinear forms A_h^k are stable in an H^2-type norm, the condition number of the resulting linear system is generally of order $\max_{K \in \mathcal{T}_h} p_K^8 / h_K^4$ for common choices of basis functions for $V_{h,\mathbf{p}}$. Therefore, there can be significant benefits in using preconditioners when solving the linear systems with certain iterative methods: see [10, 28] for the analysis and numerical study of nonoverlapping domain decomposition preconditioners for DGFEM that are stable in H^2-type norms.

8.2. Semismoothness of HJB operators. The proof of Theorem 11 rests upon the notion of semismoothness, as defined in [29]. We recall the definition below. For sets X and Y, we write $G \colon X \rightrightarrows Y$ if G is a set-valued map that maps X into the subsets of Y.

DEFINITION 12. *Let X and Y be Banach spaces, and let $F \colon U \subset X \to Y$ be a map defined on a nonempty open set U of X. Let $DF \colon U \rightrightarrows \mathcal{L}(X, Y)$ be a set-valued map with nonempty images. For $x \in U$, the map F is called* DF-semismooth at x *if*

$$\lim_{\|e\|_X \to 0} \frac{1}{\|e\|_X} \sup_{D \in DF[x+e]} \|F[x+e] - F[x] - De\|_Y = 0. \tag{8.6}$$

The map F is called DF-semismooth on U *if F is DF-semismooth at x for every $x \in U$. The set-valued map DF is then called a* generalized differential of F on U.

For $1 \leq q < r \leq \infty$, the map $DF_\gamma \colon W^{2,r}(\Omega; \mathcal{T}_h) \rightrightarrows \mathcal{L}\left(W^{2,r}(\Omega; \mathcal{T}_h), L^q(\Omega)\right)$ is defined by

$$DF_\gamma[u] \coloneqq \left\{ \gamma^\alpha L^\alpha \coloneqq \gamma^\alpha \left(a^\alpha \colon D_h^2 + b^\alpha \cdot \nabla_h - c^\alpha \right) : \alpha \in \Lambda[u] \right\}. \tag{8.7}$$

THEOREM 13. *Let Ω be a bounded open subset of $\mathbb{R}^n$, let Λ be a compact metric space, let the data a, b, c, and f be continuous on $\overline{\Omega} \times \Lambda$ and suppose that* (2.4) *holds. Let $\mathcal{T}_h$ be a mesh on Ω. Then, for any $1 \leq q < r \leq \infty$, the operator $F_\gamma \colon W^{2,r}(\Omega; \mathcal{T}_h) \to L^q(\Omega)$ defined by $F_\gamma[u] = F_\gamma(\cdot, u, \nabla_h u, D_h^2 u)$ is DF_γ-semismooth on $W^{2,r}(\Omega; \mathcal{T}_h)$.*

Proof. Supposing the claim to be false, there exist a function $u \in W^{2,r}(\Omega;\mathcal{T}_h)$, a constant $\rho > 0$, and a sequence $\{e_j\}_{j=0}^{\infty} \subset W^{2,r}(\Omega;\mathcal{T}_h)$, with $\|e_j\|_{W^{2,r}(\Omega;\mathcal{T}_h)} \to 0$, and $\alpha_j \in \Lambda[u+e_j]$ such that, for each $j \in \mathbb{N}$,

$$\frac{1}{\|e_j\|_{W^{2,r}(\Omega;\mathcal{T}_h)}} \|F_\gamma[u+e_j] - F_\gamma[u] - \gamma^{\alpha_j} L^{\alpha_j} e_j\|_{L^q(\Omega)} > \rho. \tag{8.8}$$

We will show that there is a subsequence for which (8.8) is violated and thus obtain a contradiction. Since $\|e_j\|_{W^{2,r}(\Omega;\mathcal{T}_h)} \to 0$, by passing to a subsequence without change of notation, we may assume that e_j and its first and second broken derivatives tend to 0 pointwise a.e. in Ω. The following inequality will help to simplify the argument:

$$|F_\gamma[u+e_j] - F_\gamma[u] - \gamma^{\alpha_j} L^{\alpha_j} e_j| \lesssim G_j \left(|e_j| + |\nabla_h e_j| + |D_h^2 e_j| \right), \tag{8.9}$$

where $G_j : \Omega \to \mathbb{R}_{\geq 0}$ is defined by

$$G_j := \inf_{\alpha \in \Lambda(\cdot,\mathbf{u}(\cdot))} |\gamma^\alpha a^\alpha - \gamma^{\alpha_j} a^{\alpha_j}| + |\gamma^\alpha b^\alpha - \gamma^{\alpha_j} b^{\alpha_j}| + |\gamma^\alpha c^\alpha - \gamma^{\alpha_j} c^{\alpha_j}|. \tag{8.10}$$

It can be deduced from Lemma 9 that G_j is measurable, since it is the composition of a lower semicontinuous function with a measurable function; compactness of Λ and continuity of the data imply that $\|G_j\|_{L^\infty(\Omega)}$ is uniformly bounded for all $j \in \mathbb{N}$.

We prove (8.9): since $\alpha_j \in \Lambda[u+e_j]$, we have a.e. in Ω

$$F_\gamma[u+e_j] - F_\gamma[u] - \gamma^{\alpha_j} L^{\alpha_j} e_j = \gamma^{\alpha_j} \left(L^{\alpha_j} u - f^{\alpha_j} \right) - F_\gamma[u] \leq 0. \tag{8.11}$$

Now, for a.e. $x \in \Omega$, and arbitrary $\alpha \in \Lambda(x,\mathbf{u}(x))$, we have

$$\begin{aligned} 0 &\leq F_\gamma[u+e_j] - \gamma^\alpha \left(L^\alpha (u+e_j) - f^\alpha \right) \\ &= \gamma^{\alpha_j} \left(L^{\alpha_j} u - f^{\alpha_j} \right) - F_\gamma[u] + \left(\gamma^{\alpha_j} L^{\alpha_j} - \gamma^\alpha L^\alpha \right) e_j \\ &= F_\gamma[u+e_j] - F_\gamma[u] - \gamma^{\alpha_j} L^{\alpha_j} e_j + \left(\gamma^{\alpha_j} L^{\alpha_j} - \gamma^\alpha L^\alpha \right) e_j, \end{aligned} \tag{8.12}$$

where it is understood that the above expressions are evaluated at x. Rearranging (8.11) and (8.12) gives $\left(\gamma^\alpha L^\alpha - \gamma^{\alpha_j} L^{\alpha_j} \right) e_j \leq F_\gamma[u+e_j] - F_\gamma[u] - \gamma^{\alpha_j} L^{\alpha_j} e_j \leq 0$, so

$$|F_\gamma[u+e_j] - F_\gamma[u] - \gamma^{\alpha_j} L^{\alpha_j} e_j| \leq |\left(\gamma^\alpha L^\alpha - \gamma^{\alpha_j} L^{\alpha_j} \right) e_j|. \tag{8.13}$$

Since (8.13) holds for arbitrary $\alpha \in \Lambda(x,\mathbf{u}(x))$, we readily obtain (8.9).

We claim that $G_j \to 0$ pointwise a.e. in Ω. Recall that $\mathbf{e}_j := (e_j, \nabla_h e_j, D_h^2 e_j)$ tends to zero pointwise a.e. in Ω. Let $\varrho > 0$ and $x \in \Omega$ be such that $\mathbf{e}_j(x) \to 0$. Then, by continuity of the data on the compact metric space $\overline{\Omega} \times \Lambda$, there is a $\delta > 0$ such that for any α, $\beta \in \Lambda$ with $\operatorname{dist}(\alpha,\beta) < \delta$,

$$|\gamma^\alpha a^\alpha - \gamma^\beta a^\beta| + |\gamma^\alpha b^\alpha - \gamma^\beta b^\beta| + |\gamma^\alpha c^\alpha - \gamma^\beta c^\beta| < \varrho \quad \text{at } x \in \Omega.$$

Since $(x,\mathbf{u}) \mapsto \Lambda(x,\mathbf{u})$ is upper semicontinuous by Lemma 9, there is an $N \in \mathbb{N}$ such that for each $j \geq N$, there is an $\alpha \in \Lambda(x,\mathbf{u}(x))$ with $\operatorname{dist}(\alpha,\alpha_j(x)) < \delta$. Therefore $0 \leq G_j(x) < \varrho$ for all $j \geq N$, and hence $G_j \to 0$ pointwise a.e. in Ω.

Because $1 \leq q < r \leq \infty$, setting $s = r/q > 1$ and s' such that $1/s + 1/s' = 1$, we have $1 \leq s' < \infty$. Inequality (8.9) followed by an application of Hölder's inequality shows that

$$\frac{1}{\|e_j\|_{W^{2,r}(\Omega;\mathcal{T}_h)}} \|F_\gamma[u+e_j] - F_\gamma[u] - \gamma^{\alpha_j} L^{\alpha_j} e_j\|_{L^q(\Omega)} \lesssim \|G_j\|_{L^{qs'}(\Omega)}, \tag{8.14}$$

Since $G_j \to 0$ pointwise a.e. and $\{G_j\}_{j=0}^{\infty}$ is uniformly bounded in $L^\infty(\Omega)$, the dominated convergence theorem implies that $\|G_j\|_{L^{qs'}(\Omega)} \to 0$. Therefore, (8.14) contradicts (8.8), and F_γ is DF_γ-semismooth at u, thus completing the proof. ∎

Remark 1. The restriction $q < r$ in Theorem 13 cannot be relaxed in general, as evidenced by the counterexample in [14] involving a special case of the class of operators considered here.

Proof of Theorem 11. Since $\alpha_k \in \Lambda[u_h^k]$ for each k, we have $F_\gamma[u_h^k] = \gamma^{\alpha_k} L^{\alpha_k} u_h^k - \gamma^{\alpha_k} f^{\alpha_k}$. Therefore, (8.4) is equivalent to

$$A_h^k(u_h^{k+1}, v_h) = \sum_{K \in \mathcal{T}_h} \langle \gamma^{\alpha_k} L^{\alpha_k} u_h^k - F_\gamma[u_h^k], L_\lambda v_h \rangle_K \quad \forall\, v_h \in V_{h,\mathbf{p}}. \tag{8.15}$$

The definition of the numerical scheme (5.4) implies that u_h satisfies

$$A_h^k(u_h, v_h) = \sum_{K \in \mathcal{T}_h} \langle \gamma^{\alpha_k} L^{\alpha_k} u_h - F_\gamma[u_h], L_\lambda v_h \rangle_K \quad \forall\, v_h \in V_{h,\mathbf{p}}. \tag{8.16}$$

After subtracting (8.16) from (8.15), the bound (8.5) then shows that

$$\|u_h^{k+1} - u_h\|_{h,1} \le C_1 \|F_\gamma[u_h^k] - F_\gamma[u_h] - \gamma_k^\alpha L_k^\alpha (u_h^k - u_h)\|_{L^2(\Omega)}, \tag{8.17}$$

where the constant C_1 depends only on κ, ε, γ, and n as in (6.14), but not on k. Fix $r > 2$; since $V_{h,\mathbf{p}}$ is finite-dimensional, there is a constant C_2 depending on h and $\mathbf{p}$ such that $\|v_h\|_{W^{2,r}(\Omega;\mathcal{T}_h)} \le C_2 \|v_h\|_{h,1}$ for all $v_h \in V_{h,\mathbf{p}}$. Theorem 13 shows that for each $\rho \in (0,1)$, there is a $R_\rho > 0$ such that if $\|w_h - u_h\|_{h,1} < R_\rho$, then, for any $\alpha \in \Lambda[w_h]$,

$$\|F_\gamma[w_h] - F_\gamma[u_h] - \gamma^\alpha L^\alpha (w_h - u_h)\|_{L^2(\Omega)} \le \frac{\rho}{C_1 C_2} \|w_h - u_h\|_{W^{2,r}(\Omega;\mathcal{T}_h)}. \tag{8.18}$$

If $\|u_h^0 - u_h\|_{h,1} < R_\rho$ for some $\rho < 1$, then we use (8.17) and (8.18) to obtain

$$\|u_h^{k+1} - u_h\|_{h,1} \le \rho \|u_h^k - u_h\|_{h,1} \quad \forall\, k \ge 0,$$

which yields convergence of u_h^k to u_h. For each $\rho < 1$, $\|u_h^k - u_h\|_{h,1} < R_\rho$ is then eventually satisfied, thus implying a superlinear convergence rate. ∎

9. Numerical experiments. We provide the results of two tests of the scheme on problems with strongly anisotropic diffusion coefficients and an experiment for a solution that does not meet the regularity assumption of the analysis.

9.1. First experiment. We consider once again Example 1 for testing the accuracy of the scheme and the performance of the semismooth Newton method. Recalling that $\Lambda = [0, \pi/3] \times \mathrm{SO}(2)$ and $a^\alpha = \sigma^\alpha (\sigma^\alpha)^\top / 2$, with σ^α given by (2.7), let $\Omega = (0,1)^2$, let $b^\alpha \equiv 0$, $c^\alpha \equiv \pi^2$, and choose $f^\alpha \equiv \sqrt{3} \sin^2\theta/\pi^2 + g$, g independent of α, so that the exact solution of the HJB equation (2.3) is $u(x,y) = \exp(xy)\sin(\pi x)\sin(\pi y)$. These choices are made so that the optimal controls vary significantly throughout the domain and to ensure that the corresponding diffusion coefficient is not diagonally dominant in parts of Ω.

The numerical scheme (5.4) is applied with meshes obtained by regular subdivision of Ω into uniform quadrilateral elements of size $h = 2^{-k}$, $1 \le k \le 6$. The finite element spaces $V_{h,\mathbf{p}}$ are defined by employing the space of polynomials of fixed total degree p on each element with $2 \le p \le 5$. The penalty parameters are set to $c_{\mathrm{stab}} = 10$ and

$\eta_F = c_{\text{stab}}\,\tilde{p}_F^4/\tilde{h}_F^3$. Figure 1 confirms the optimal convergence rates with respect to mesh refinement that are predicted by Theorem 8.

The numerical solutions were obtained by the semismooth Newton method of section 8, for which we use a strict convergence criterion by requiring a relative residual below 5×10^{-12} and a step-increment L^2-norm below 1×10^{-11}. The initial guess used for each computation was $u_h^0 \equiv 0$. The convergence histories shown in Figure 2 demonstrate the fast convergence of the algorithm.

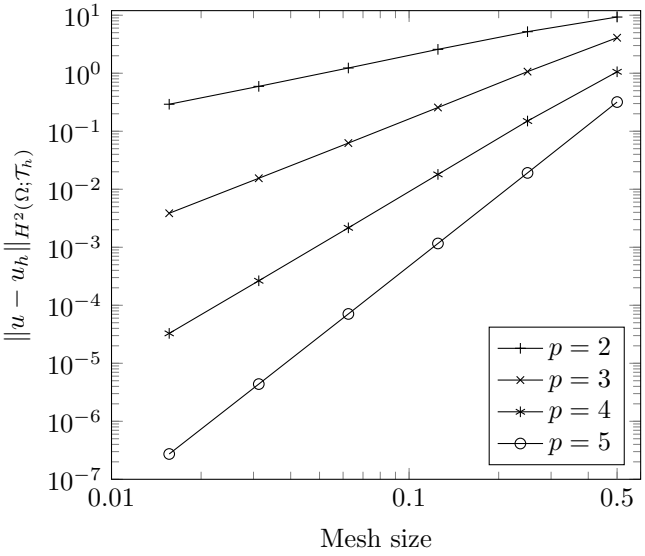

FIG. 1. *The errors in approximating the solution of the problem of section* 9.1 *for various mesh sizes and polynomial degrees. The optimal convergence rates* $\|u - u_h\|_{H^2(\Omega;\mathcal{T}_h)} = \mathsf{O}\left(h^{p-1}\right)$ *are observed.*

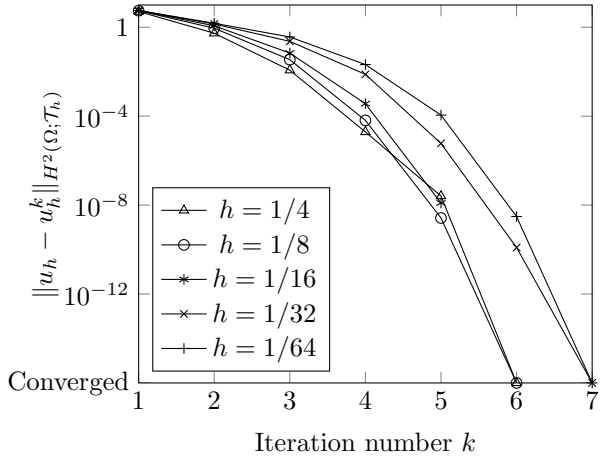

FIG. 2. *Convergence histories of the semismooth Newton method applied to the problem of section* 9.1 *on successively refined meshes, with* $p = 4$. *The predicted superlinear convergence rate is observed, and the number of iterations required for convergence shows little variation under refinement.*

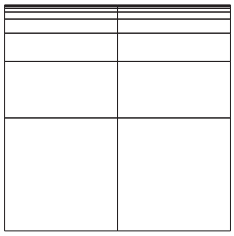

FIG. 3. *Mesh on* Ω *used for the approximation of* (9.2). *The origin is at the bottom left corner. The mesh has eight geometrically refined layers with grading factor* $1/2$.

9.2. Second experiment. We investigate the robustness of the scheme against a combination of near-degenerate diffusions, nonsmooth solutions, and boundary layers. This example is to our knowledge the first fully nonlinear second-order problem solved with an exponentially accurate scheme. Let $\Omega = (0,1)^2$, $b^\alpha \equiv (0,1)$, $c^\alpha \equiv 10$ and define

$$a^\alpha \coloneqq \alpha^\top \begin{pmatrix} 20 & 1 \\ 1 & 0.1 \end{pmatrix} \alpha, \quad \alpha \in \Lambda \coloneqq \mathrm{SO}(2). \tag{9.1}$$

For $\lambda = 1/2$, the Cordes condition (2.5) holds with $\varepsilon \approx 0.0024$. We choose f^α so that the solution of the corresponding HJB equation is

$$u(x,y) = (2x-1)\left(\mathrm{e}^{1-|2x-1|} - 1\right)\left(y + \frac{1-\mathrm{e}^{y/\delta}}{\mathrm{e}^{1/\delta}-1}\right), \quad \delta > 0. \tag{9.2}$$

We choose $\delta = 0.005$ to be of the same order as ε, thus leading to a sharp boundary layer in a neighborhood of $\{(x,y) \in \overline{\Omega} \colon y = 1\}$.

The results of [5] show that a very large stencil would be necessary to obtain a consistent monotone FD discretization of this problem. On uniform grids, these low-order methods would require a fine grid to resolve the boundary layer, while the use of locally refined grids is complicated by consistency and monotonicity requirements.

Our method features no such constraints, so we are free to take advantage of *hp*-refinement techniques that are capable of delivering highly accurate approximations for a smaller computational cost. Following a suggestion in [21], we perform a sequence of computations by increasing the uniform polynomial degrees p from 2 to 10 on a fixed mesh shown in Figure 3. The number of degrees of freedom ranges from 100 to 1320 and the following results were obtained with $c_{\mathrm{stab}} = 10$, as in section 9.1. Here, we use $\eta_F = \lambda\, c_{\mathrm{stab}}\, \tilde{p}_F^4 / \tilde{h}_F^3$. Figure 4 shows that the error converges with a rate of $\mathsf{O}\big(\exp(-c\sqrt[3]{\mathrm{DoF}})\big)$, as expected from the results in [30], which leads to high accuracy with few degrees of freedom.

9.3. Third experiment. We consider the convergence of the scheme when relaxing the assumption on the solution of broken H^s-regularity for some $s > 5/2$. We also treat a problem with an inhomogeneous Dirichlet boundary condition, which can be handled by a straightforward extension of the scheme; see [27] for further details.

Consider a hexagonal domain $\Omega \subset \mathbb{R}^2$ with unit face length, as shown in Figure 5. Laplace's equation, $\Delta u = 0$ in Ω, is a special case of the HJB equation at hand, and the boundary condition $u = g$ on $\partial\Omega$ is chosen such that $u = r^{3/2}\sin(\frac{3}{2}\theta)$, where r is the distance to the upper vertex of Ω, and θ is the counterclockwise angle from the upper left face of Ω. It follows that broken H^s-regularity of u fails for $s \geq 5/2$.

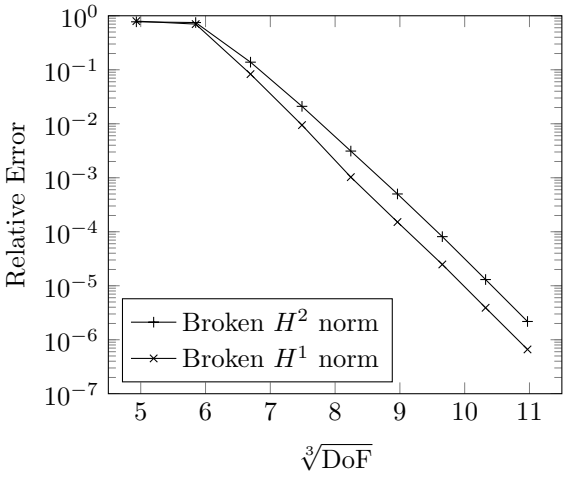

FIG. 4. *Exponential convergence in the broken H^1- and H^2-norms of the approximations to the solution defined by* (9.2). *The relative errors $\|u-u_h\|/\|u\|$ are plotted against the cube root of the number of degrees of freedom, with each data point corresponding to a computation using a total polynomial degree $p=2,\dots,10$.*

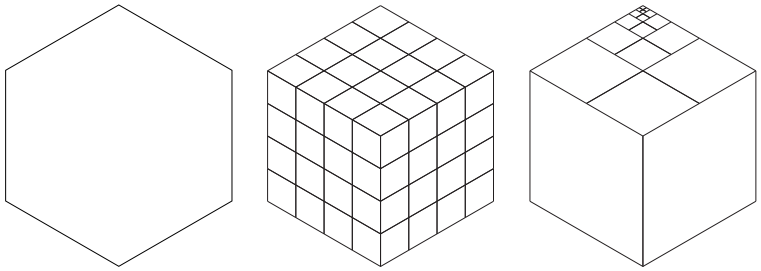

FIG. 5. *The planar hexagonal domain $\Omega\subset\mathbb{R}^2$ with uniformly refined and geometrically graded parallelepipedal meshes.*

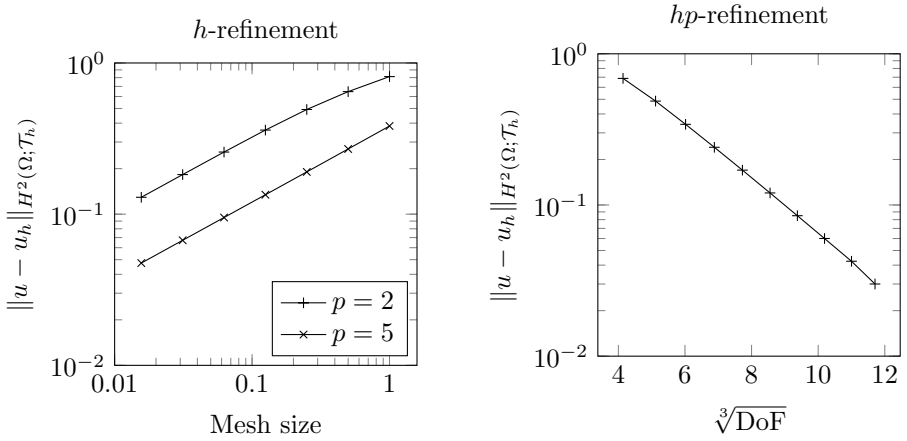

FIG. 6. *Convergence in the broken H^2-norm of the approximations to the singular solution on the hexagonal domain. We observe approximate convergence rates of $\mathsf{O}(h^{1/2})$ for uniform h-refinement, while rates of $\mathsf{O}(\exp(-c\sqrt[3]{\text{DoF}}))$ are obtained under hp-refinement. The number of degrees of freedom used for the largest computations were* 1604 *(hp-refinement),* 73,728 *(h-refinement, $p=2$), and* 258,048 *(h-refinement, $p=5$), thus showing the efficiency of hp-refinement.*

Two sets of computations were performed, the first using uniform h-refinement and the second using hp-refinement on geometrically graded meshes with linearly increasing polynomial degrees away from the singularity of the solution. Figure 5 illustrates the respective sequences of meshes. We use $c_{\text{stab}} = 10$ and $\eta_F = c_{\text{stab}}\,\tilde{p}_F^4/\tilde{h}_F^3$. Figure 6 shows that the method is convergent under both refinement strategies. In the case of h-refinement, the convergence rate is approximately of $\mathsf{O}(h^{1/2})$, irrespective of the polynomial degree, as may be expected given the limited regularity of the solution. In the case of hp-refinement, the rate is of $\mathsf{O}(\exp(-c\sqrt[3]{\text{DoF}}))$, where DoF is the number of degrees of freedom.

These results show that the regularity assumption on the solution used in the analysis of this work is not a necessary condition for the convergence of the numerical scheme. Furthermore, the ability of the method to handle hp-refinement is a significant advantage for computational efficiency.

10. Conclusion. We have considered the PDE analysis and numerical analysis of HJB equations that satisfy the Cordes condition. Our contributions include an existence and uniqueness result for strong solutions to the fully nonlinear problem, the construction of a consistent and stable hp-version DGFEM with proven convergence rates, and a study of the semismoothness of HJB operators. The numerical experiments demonstrated the high efficiency and accuracy of the scheme and the fast convergence of the semismooth Newton method, while highlighting the wide applicability of the results of this work.

REFERENCES

[1] G. Barles and P. Souganidis, *Convergence of approximation schemes for fully nonlinear second-order equations*, Asymptot. Anal., 4 (1991), pp. 271–283.

[2] K. Böhmer, *On finite element methods for fully nonlinear elliptic equations of second order*, SIAM J. Numer. Anal., 46 (2008), pp. 1212–1249.

[3] O. Bokanowski, S. Maroso, and H. Zidani, *Some convergence results for Howard's algorithm*, SIAM J. Numer. Anal., 47 (2009), pp. 3001–3026.

[4] J. F. Bonnans, É. Ottenwaelter, and H. Zidani, *A fast algorithm for the two dimensional HJB equation of stochastic control*, M2AN Math. Model. Numer. Anal., 38 (2004), pp. 723–735.

[5] J. F. Bonnans and H. Zidani, *Consistency of generalized finite difference schemes for the stochastic HJB equation*, SIAM J. Numer. Anal., 41 (2003), pp. 1008–1021.

[6] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second-order partial differential equations*, Bull. Amer. Math. Soc. (N.S.), 27 (1992), pp. 1–67.

[7] M. G. Crandall and P.-L. Lions, *Convergent difference schemes for nonlinear parabolic equations and mean curvature motion*, Numer. Math., 75 (1996), pp. 17–41.

[8] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Math. Appl. 69, Springer, Heidelberg, 2012.

[9] H. Dong and N. V. Krylov, *The rate of convergence of finite-difference approximations for parabolic Bellman equations with Lipschitz coefficients in cylindrical domains*, Appl. Math. Optim., 56 (2007), pp. 37–66.

[10] X. Feng and O. A. Karakashian, *Two-level non-overlapping Schwarz preconditioners for a discontinuous Galerkin approximation of the biharmonic equation*, J. Sci. Comput., 22/23 (2005), pp. 289–314.

[11] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Stoch. Model. Appl. Probab. 25, Springer, New York, 2006.

[12] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Classics in Math., Springer, Berlin, 2001.

[13] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Classics in Appl. Math. 69, SIAM, Philadelphia, 2011.

[14] M. Hintermuller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888.

[15] P. Houston, Ch. Schwab, and E. Süli, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 39 (2002), pp. 2133–2163.
[16] M. Jensen and I. Smears, *On the convergence of finite element methods for Hamilton–Jacobi–Bellman equations*, SIAM J. Numer. Anal., 51 (2013), pp. 137–162.
[17] M. Kocan, *Approximation of viscosity solutions of elliptic partial differential equations on minimal grids*, Numer. Math., 72 (1995), pp. 73–92.
[18] K. Kuratowski and C. Ryll-Nardzewski, *A general theorem on selectors*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys., 13 (1965), pp. 397–403.
[19] H. J. Kushner, *Numerical methods for stochastic control problems in continuous time*, SIAM J. Control Optim., 28 (1990), pp. 999–1048.
[20] A. Maugeri, D. K. Palagachev, and L. G. Softova, *Elliptic and Parabolic Equations with Discontinuous Coefficients*, Math. Res. 109, Wiley, Berlin, 2000.
[21] J. M. Melenk, *hp-Finite Element Methods for Singular Perturbations*, Lecture Notes in Math. 1796, Springer, Berlin, 2002.
[22] T. S. Motzkin and W. Wasow, *On the approximation of linear elliptic differential equations by difference equations with positive coefficients*, J. Math. Phys., 31 (1953), pp. 253–259.
[23] A. M. Oberman, *Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–Jacobi equations and free boundary problems*, SIAM J. Numer. Anal., 44 (2006), pp. 879–895.
[24] M. L. Puterman and S. L. Brumelle, *On the convergence of policy iteration in stationary dynamic programming*, Math. Oper. Res., 4 (1979), pp. 60–69.
[25] M. Renardy and R. C. Rogers, *An Introduction to Partial Differential Equations*, 2nd ed., Texts in Appl. Math. 13, Springer, New York, 2004.
[26] M. V. Safonov, *Nonuniqueness for second-order elliptic equations with measurable coefficients*, SIAM J. Math. Anal., 30 (1999), pp. 879–895.
[27] I. Smears and E. Süli, *Discontinuous Galerkin finite element approximation of nondivergence form elliptic equations with Cordès coefficients*, SIAM J. Numer. Anal., 51 (2013), pp. 2088–2106.
[28] I. Smears, *On Non-overlapping Domain Decomposition Preconditioners for Discontinuous Galerkin Finite Element Methods in H^2-Type Norms*, Tech. report 13/13, University of Oxford, 2013; available online from http://eprints.maths.ox.ac.uk/1711.
[29] M. Ulbrich, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2002), pp. 805–841.
[30] T. P. Wihler, P. Frauenfelder, and Ch. Schwab, *Exponential convergence of the hp-DGFEM for diffusion problems*, Comput. Math. Appl., 46 (2003), pp. 183–205.