SIAM J. NUMER. ANAL.
Vol. 57, No. 2, pp. 592–614

MIXED FINITE ELEMENT APPROXIMATION OF THE HAMILTON–JACOBI–BELLMAN EQUATION WITH CORDES COEFFICIENTS*

DIETMAR GALLISTL† AND ENDRE SÜLI‡

Abstract. A mixed finite element approximation of H^2 solutions to the fully nonlinear Hamilton–Jacobi–Bellman equation, with coefficients that satisfy the Cordes condition, is proposed and analyzed. A priori and a posteriori bounds on the approximation error are proved. The contributions from the a posteriori error estimator can be used as refinement indicators in an adaptive mesh-refinement algorithm. The convergence of this procedure is proved and empirically studied in numerical experiments.

Key words. mixed finite element methods, Cordes condition, Hamilton–Jacobi–Bellman equation, nondivergence form PDE, fully nonlinear PDE, a posteriori error analysis, adaptive algorithm

AMS subject classifications. 65N12, 65N15, 65N30

DOI. 10.1137/18M1192299

1. Introduction.

This work presents a mixed finite element approximation of H^2 solutions to the Hamilton–Jacobi–Bellman (HJB) equation

$$\sup_{\alpha\in\Lambda}(a_\alpha : D^2u + b_\alpha\cdot\nabla u - c_\alpha u - f_\alpha) = 0 \ \text{ in } \Omega, \qquad u = 0 \ \text{ on } \partial\Omega, \tag{1.1}$$

with uniformly continuous coefficients satisfying a Cordes condition, which are parametrized over a compact metric space Λ. Here D^2u and ∇u denote, respectively, the Hessian and the gradient of the real-valued function u. The domain $\Omega\subseteq\mathbb{R}^d$ is a bounded, open, convex Lipschitz domain, which, for ease of discretization and the sake of simplicity of the exposition, we shall henceforth consider to be a bounded, open, convex polytope. The equation arises in the theory of optimal stochastic control for continuous time Markov processes [19]; it is fully nonlinear and involves a parametrized family of second-order linear elliptic partial differential operators in nondivergence form. While second-order elliptic equations in divergence form possess a weak formulation in first-order Sobolev spaces, which can be directly discretized with finite elements, the theory of nondivergence form and, more generally, fully nonlinear PDEs, relies on different solution concepts such as strong solutions, viscosity solutions, or measure-valued solutions. The construction of numerical methods for these problems, and finite element methods in particular, is much less straightforward, the reason being that the leading-order term does not stem from an energy minimization procedure and thus there is no natural variational formulation.

The recent papers [39], [40], and [41] have identified a class of domains and coefficients for which an existence and uniqueness theory of strong solutions to HJB equations (and linear nondivergence form problems, in particular,) in the Sobolev space $H^2(\Omega)$ is available. Therein, the main condition on the uniformly elliptic coefficients $a_\alpha\in L^\infty(\Omega;\mathbb{R}^{d\times d})$, $\alpha\in\Lambda$, is the so-called *Cordes condition*, which dates back

*Received by the editors June 5, 2018; accepted for publication (in revised form) January 29, 2019; published electronically March 19, 2019.
http://www.siam.org/journals/sinum/57-2/M119229.html
Funding: The first author was supported by the DFG through SFB 1173.
†Department of Applied Mathematics, University of Twente, Enschede, 7500 AE, The Netherlands (d.gallistl@utwente.nl).
‡Mathematical Institute, University of Oxford, Oxford, OX2 6GG (suli@maths.ox.ac.uk).

to [11]. In the absence of lower-order terms it basically requires that the Frobenius norm of the tensor a_α is properly dominated by its trace for each $\alpha \in \Lambda$. For the analysis of PDEs with discontinuous coefficients under the Cordes condition the reader is referred to the monograph [33]. In the presence of lower-order terms as in (1.1), the Cordes condition [40] requires the existence of $\lambda > 0$ and $\varepsilon \in (0, 1)$ such that, for each $\alpha \in \Lambda$,

$$\frac{|a_\alpha|^2 + |b_\alpha|^2/(2\lambda) + (c_\alpha/\lambda)^2}{(\operatorname{tr} a_\alpha + c_\alpha/\lambda)^2} \leq 1/(d+\varepsilon) \quad \text{a.e. in } \Omega. \tag{1.2}$$

The existence and uniqueness of strong solutions to elliptic and parabolic HJB equations on convex domains, under the Cordes condition, were studied in [40] and [41], and, based on these, discontinuous Galerkin finite element approximations of H^2 solutions were constructed and analyzed.

The scheme presented here is a mixed finite element method, based on a splitting technique for equations involving the Hessian into systems of divergence-form PDEs [20]. Its characteristic feature is that the gradient $w = \nabla u$ is discretized by an additional independent variable, which bypasses the need for H^2-conforming finite elements. In the context of polyharmonic equations [20], the relation $w = \nabla u$ was enforced through a saddlepoint formulation, the use of which is, however, merely optional in this work on HJB equations. Such mixed discretizations were applied to linear problems in nondivergence form in [21] and generalized to oblique derivative problems in [23]. One advantage of this approach is that a priori as well as a posteriori error bounds can be derived in a relatively direct way, and the convergence of adaptive mesh-refinement algorithms can be proved. Key to the analysis of [39, 40] is the so-called *Miranda–Talenti* estimate $|u|_{H^2(\Omega)} \leq \|\Delta u\|_{L^2(\Omega)}$ for all $u \in H^2(\Omega) \cap H^1_0(\Omega)$. In the context of the mixed formulation considered herein, the corresponding relevant inequality reads

$$\|Dw\|^2_{L^2(\Omega)} \leq \|\operatorname{rot} w\|^2_{L^2(\Omega)} + \|\operatorname{div} w\|^2_{L^2(\Omega)} \tag{1.3}$$

for any vector field w all of whose components belong to $H^1(\Omega)$ and whose tangential trace vanishes on $\partial\Omega$. Here Dw denotes the gradient of the d-component vector-function w. If w equals the gradient of a real-valued function from $H^2(\Omega) \cap H^1_0(\Omega)$, the original Miranda–Talenti estimate is recovered. In mixed finite element methods, irrotationality of the finite element approximation to w is usually not imposed in a pointwise fashion and thus quantities approximating gradients need not be true gradients of discrete functions. It turns out that this is basically the reason why mixed finite element approximations require stabilizing terms [21]. In this work, a novel generalization of (1.3) will be introduced, which plays a key role in the construction and the stability analysis of mixed approximations of (1.1).

The new finite element discretization proposed here greatly simplifies the numerical approximation of (1.1) compared to prior contributions. In particular, standard H^1-conforming finite elements can be used. The numerical analysis of the scheme requires careful investigation of Miranda–Talenti-type estimates in the spirit of (1.3), thereby generalizing existing results, which, in turn, lead to the design of stabilization terms, resulting in a well-posed numerical method for which a priori and a posteriori error bounds will be derived. The latter enables the use of an adaptive mesh-refinement algorithm, which can be proved to converge to the unique strong solution of the problem. These theoretical results are supplemented by numerical experiments, which suggest that the adaptive scheme can improve the accuracy of the method compared with uniform mesh refinement.

For the PDE analysis of elliptic equations in nondivergence form the reader is referred to the monograph [33]. Finite element methods for the numerical solution of linear problems in nondivergence form can be found in [32, 39, 37, 18, 15]. Concerning the discretization of fully nonlinear problems, there are several monotone finite difference schemes [1, 29, 12, 38]. Key to these methods is monotonicity, which means that discrete maximum principles are respected. This allows a quite general convergence theory [1] on the one hand, and on the other hand it is known [35] that simultaneous monotonicity and consistency necessarily come at the expense of a finite difference stencil which increases (relative to the mesh size) under mesh refinement. Finite elements appear as one possibility of using nonmonotone methods, such as the method by [4] (when the linearizations are in divergence form) or approaches based on adding small perturbations via the bi-Laplacian [17]. A general overview can be found in the survey articles [14, 36]. Regarding H^1-conforming finite element methods for HJB equations, we refer the reader to the works [7, 8, 27, 28] and the references therein.

Finite element methods for HJB equations with Cordes coefficients on convex domains were introduced in [40]. As was noted above, that work establishes the existence of a unique strong solution to the problem and proposes a discontinuous Galerkin finite element discretization, based on prior work by the same authors [39]; an extension to parabolic problems was presented in [41]. The HJB equation can furthermore be used to reformulate the Monge–Ampère equation without the need to separately enforce the convexity of the solution. This result [30] was recently generalized to the case of viscosity solutions by [16], who proposed a semi-Lagrangian method.

The study of self-adaptive mesh refinement algorithms for fully nonlinear problems is still in its infancy. While the convergence and optimality of adaptive finite element approximations of elliptic problems in divergence form have been reasonably well understood during the past decade [9, 42], the only contributions to a posteriori error estimation and the convergence analysis of adaptive schemes for linear problems in nondivergence form seem to be [21, 23].

The article is organized as follows. Section 2 introduces the mixed formulation of the problem and proves its well-posedness. The numerical method is described and analyzed in section 3. The error analysis is comprised of a priori and a posteriori error bounds as well as the convergence analysis of an adaptive algorithm. Section 4 presents numerical experiments.

Standard notation for function spaces is used throughout the article. Lebesgue and Sobolev functions with values in $\mathbb{R}^d$ are denoted by $L^2(\Omega;\mathbb{R}^d)$ with $L^2(\Omega) := L^2(\Omega;\mathbb{R})$, $H^1(\Omega;\mathbb{R}^d)$ with $H^1(\Omega) := H^1(\Omega;\mathbb{R})$, etc. The symbol $H^1_t(\Omega;\mathbb{R}^d)$ denotes the subspace of $H^1(\Omega;\mathbb{R}^d)$ consisting of vector fields with vanishing tangential trace on $\partial\Omega$. The $n\times n$ identity matrix is denoted by $I_{n\times n}$. The inner product of real-valued $n\times n$ matrices A, B is denoted by $A:B := \sum_{j,k=1}^n A_{jk}B_{jk}$. The Frobenius norm of an $n\times n$ matrix A is denoted by $|A| := \sqrt{A:A}$; the trace of an $n\times n$ matrix A is denoted by $\operatorname{tr} A$. For vectors, $|\cdot|$ refers to the Euclidean length. The notation $a \lesssim b$ denotes the inequality $a \le Cb$ up to a multiplicative constant C that does not depend on the mesh-size. The results of this paper apply in any space dimension $d \ge 2$, but for ease of readability the results are presented here for $d \in \{2,3\}$, where we define

$$\operatorname{rot} v = \partial_2 v_1 - \partial_1 v_2 \text{ for } d=2 \quad \text{or} \quad \operatorname{rot} v = \begin{pmatrix} \partial_2 v_3 - \partial_3 v_2 \\ \partial_3 v_1 - \partial_1 v_3 \\ \partial_1 v_2 - \partial_2 v_1 \end{pmatrix} \text{ for } d=3.$$

Setting rot as the exterior derivative operator, the proofs also extend to the case $d \ge 4$.

2. Mixed formulation. This section consists of the following parts. After a brief review of H^2 solutions to the HJB equation in section 2.1, the stabilized saddle-point formulation is introduced in section 2.2. The lemmas of section 2.3 provide the results necessary for the proof of well-posedness in section 2.4.

2.1. Review of strong solutions to the HJB equation. The existence of a unique H^2 solution to the HJB equation was established in [40]. Since that setting will be assumed throughout this work, it is briefly summarized here.

Let $\Omega \subseteq \mathbb{R}^d$ be a bounded, open, convex polytope and let Λ be a compact metric space. Assume that we are given the functions

$$a : \Omega \times \Lambda \to \mathbb{R}^{d\times d}, \quad b : \Omega \times \Lambda \to \mathbb{R}^d, \quad c : \Omega \times \Lambda \to \mathbb{R}, \quad f : \Omega \times \Lambda \to \mathbb{R},$$

which are uniformly continuous, i.e., their components belong to $\mathcal{C}(\overline{\Omega} \times \Lambda)$. The set Λ serves as a parameter set. Thus, the following notation will be used throughout:

$$a_\alpha := a(\cdot, \alpha), \quad b_\alpha := b(\cdot, \alpha), \quad c_\alpha := c(\cdot, \alpha), \quad f_\alpha := f(\cdot, \alpha) \qquad \text{for any } \alpha \in \Lambda.$$

Define, for any $\alpha \in \Lambda$, the linear operator $\tilde{L}_\alpha : H^2(\Omega) \to L^2(\Omega)$ by

$$\tilde{L}_\alpha(v) := a_\alpha : D^2 v + b_\alpha \cdot \nabla v - c_\alpha v \quad \text{for any } v \in H^2(\Omega).$$

It is assumed throughout that c_α is nonnegative for all $\alpha \in \Lambda$. Assume that a_α is uniformly elliptic in the sense that there exist constants $0 < \zeta_1 \le \zeta_2 < \infty$ such that, for any $\alpha \in \Lambda$,

$$\zeta_1 \le \inf_{\substack{\xi\in\mathbb{R}^d \\ |\xi|=1}} \xi^{\mathrm{T}} a_\alpha \xi \le \sup_{\substack{\xi\in\mathbb{R}^d \\ |\xi|=1}} \xi^{\mathrm{T}} a_\alpha \xi \le \zeta_2 \quad \text{a.e. in } \Omega. \tag{2.1}$$

Assume furthermore that the Cordes condition (1.2) holds. It can be shown [40] that a relaxed Cordes condition can be assumed when b_α and c_α vanish for all $\alpha \in \Lambda$. For ease of reading, this possibility is disregarded here since the adaptation of the arguments to this case is straightforward. With this notation, the HJB equation (1.1) can be rewritten as follows: find a function $u \in H^2(\Omega) \cap H_0^1(\Omega)$ such that

$$\sup_{\alpha\in\Lambda}(\tilde{L}_\alpha u - f_\alpha) = 0 \quad \text{a.e. in } \Omega. \tag{2.2}$$

It is shown in [40, Thm. 3] that under the aforementioned assumptions there exists a unique solution $u \in H^2(\Omega) \cap H_0^1(\Omega)$ to (2.2).

2.2. The mixed formulation. In order to state the mixed formulation in the fashion of [20, 21, 23, 22], let $W := H_t^1(\Omega; \mathbb{R}^d)$ denote the linear space of all H^1 vector fields defined on Ω with vanishing tangential boundary trace on $\partial\Omega$, and let $U := H_0^1(\Omega)$. Let $X := W \times U$ and define on X the family of differential operators $(L_\alpha)_{\alpha\in\Lambda}$, for any $\alpha \in \Lambda$ and $(w, u) \in X$, by

$$L_\alpha(w, u) := a_\alpha : Dw + b_\alpha \cdot \nabla u - c_\alpha u. \tag{2.3}$$

This is a generalization of the operator $\tilde{L}_\alpha$ from section 2.1, which satisfies $\tilde{L}_\alpha(v) = L_\alpha(\nabla v, v)$ for any $v \in H^2(\Omega) \cap H_0^1(\Omega)$. The mixed formulation of the HJB equation will rely on a reformulation of (2.2) as

$$w = \nabla u \quad \text{and} \quad \sup_{\alpha\in\Lambda}(L_\alpha(w, u) - f_\alpha) = 0 \quad \text{in } \Omega.$$

As in [40] we define the function

$$\gamma_\alpha := \gamma(\cdot,\alpha) := \frac{\operatorname{tr} a_\alpha + c_\alpha/\lambda}{|a_\alpha|^2 + |b_\alpha|^2/(2\lambda) + (c_\alpha/\lambda)^2} \quad \text{for any } \alpha \in \Lambda. \tag{2.4}$$

The functions $(\gamma_\alpha)_{\alpha\in\Lambda}$ play the role of a scaling factor. Indeed, γ is strictly positive and uniformly continuous, $\gamma \in \mathcal{C}(\overline{\Omega} \times \Lambda)$, thanks to the assumed uniform continuity of the coefficients, the uniform ellipticity (2.1) of $(a_\alpha)_{\alpha\in\Lambda}$, and the Cordes condition (1.2). With γ thus defined, we consider the operator $F_\gamma : X \to L^2(\Omega)$, defined by

$$F_\gamma(w,u) := \sup_{\alpha\in\Lambda}[\gamma_\alpha(L_\alpha(w,u) - f_\alpha)]. \tag{2.5}$$

The uniform continuity of the coefficients guarantees that $F_\gamma(w,u) \in L^2(\Omega)$ for any pair $(w,u) \in X$. We further define the operator $\tau_\lambda : X \to L^2(\Omega)$, for any $(w,u) \in X$, by

$$\tau_\lambda(w,u) := \operatorname{div} w - \lambda u.$$

The map τ_λ will play the role of a surjective test-function operator. It is inspired by the map $u \mapsto \Delta u - \lambda u$ proposed in [40] for this purpose in the H^2 setting. Next, we define the operator $\mathcal{A} : X \to X^\star$ by

$$\langle \mathcal{A}[(w,u)], (w',u')\rangle := (F_\gamma[(w,u)], \tau_\lambda(w',u'))_{L^2(\Omega)} \tag{2.6}$$

for any $(w,u),(w',u') \in X$. The semilinear form $a : X \times X \to \mathbb{R}$ is defined by

$$\begin{aligned} a((w,u),(w',u')) := {} & \langle \mathcal{A}[(w,u)], (w',u')\rangle \\ & + \sigma_1(\operatorname{rot} w, \operatorname{rot} w')_{L^2(\Omega)} + \sigma_2(w - \nabla u, w' - \nabla u')_{L^2(\Omega)} \end{aligned} \tag{2.7}$$

for any $(w,u),(w',u') \in X$. Here, the positive real parameters σ_1, σ_2 are defined as follows:

$$\sigma_1 := 1 - \frac{1}{2}\sqrt{1-\varepsilon} \quad \text{and} \quad \sigma_2 := \frac{\lambda(1-\sqrt{1-\varepsilon})}{2} + \frac{\lambda}{4(1-\sqrt{1-\varepsilon})} \tag{2.8}$$

with ε and λ as in (1.2). These numbers σ_1, σ_2 will play the role of stabilization parameters in the mixed method. If we restrict our attention to pairs (w,u) and (w',u') with the property that $w = \nabla u$ and $w' = \nabla u'$, then the stabilization terms in (2.7) vanish. This, however, will not be true in the mixed discretization considered below, where the finite element approximation w_h to the vector field w will only be the gradient of u_h in a *discrete weak sense*. It is the stabilization that will guarantee well-posedness of the discrete problem. There are of course different ways of enforcing $w = \nabla u$ at the discrete level. It will turn out that the use of the aforementioned stabilization terms will suffice for formulating a stable and convergent finite element scheme. The reason is that the PDE is satisfied pointwise almost everywhere. Introducing a saddlepoint formulation may nevertheless be useful as it may help to reduce the number of (quasi-)Newton steps in the course of solving the systems of nonlinear algebraic equations resulting from the discretization. This is a purely empirical observation (see section 4), which is currently not explained by theoretical considerations. In order to state the mixed formulation, we let $M \subseteq H^1_0(\Omega)$ denote any closed linear subspace (in particular $M = \{0\}$ or $M = H^1_0(\Omega)$ are possible) and define the bilinear form $b : M \times X \to \mathbb{R}$ by

$$b(\mu,(w,u)) := (\nabla\mu, \nabla u - w)_{L^2(\Omega)} \quad \text{for any } \mu \in M \text{ and } (w,u) \in X. \tag{2.9}$$

Note that the freedom in the choice of $M \subseteq H_0^1(\Omega)$ will give rise to different numerical schemes, which will be compared in section 4.4. The mixed formulation of the HJB equation is then defined as the following nonlinear saddlepoint problem: seek $(w,u) \in X$ and $\mu \in M$ such that

$$a((w,u),(w',u')) + b(\mu,(w',u')) = 0 \qquad \text{for all } (w',u') \in X, \tag{2.10a}$$

$$b((w,u),\mu') = 0 \qquad \text{for all } \mu' \in M. \tag{2.10b}$$

For any open subset $\omega \subseteq \Omega$, we define the following seminorm on X:

$$\begin{aligned} &|\!|\!|(w,u)|\!|\!|_{\lambda,\omega} \\ &\quad := \Big(\|Dw\|_{L^2(\omega)}^2 + 2\lambda \|\nabla u\|_{L^2(\omega)}^2 + \lambda^2 \|u\|_{L^2(\omega)}^2 \Big)^{1/2} \quad \text{for any } (w,u) \in X \end{aligned} \tag{2.11}$$

and abbreviate the norm $|\!|\!|\cdot|\!|\!|_\lambda := |\!|\!|\cdot|\!|\!|_{\lambda,\Omega}$. A simple argument shows that the global version, $|\!|\!|\cdot|\!|\!|_\lambda$, does indeed define a norm: for any $w \in W$, $Dw = 0$ a.e. on Ω implies that the d-component vector function w is a constant vector on Ω; the vanishing tangential boundary trace of w on $\partial\Omega$ then implies that $w = 0$ a.e. on Ω, upon noting that the set of all vectors in $\mathbb{R}^d$ that are tangential to $\partial\Omega$ span the whole of $\mathbb{R}^d$. By a standard compactness argument [6], the L^2 norm of Dw is equivalent to the H^1 norm of w for any $w \in W$. The norm $|\!|\!|\cdot|\!|\!|_\lambda$ is obviously induced by the following scalar product on X:

$$((w,u),(w',u'))_\lambda := (Dw,Dw')_{L^2(\Omega)} + 2\lambda(\nabla u, \nabla u')_{L^2(\Omega)} + \lambda^2 (u,u')_{L^2(\Omega)} \tag{2.12}$$

for any $(w,u),(w',u') \in X$.

It is readily shown that any $(w,u) \in X$ satisfies

$$|\tau_\lambda(w,u)| \le \sqrt{2d}\,\big(|Dw|^2 + 2\lambda|\nabla u|^2 + \lambda^2|u|^2\big)^{1/2} \quad \text{almost everywhere on } \Omega.$$

Therefore, on every open subset $\omega \subseteq \Omega$, one has

$$\|\tau_\lambda(w,u)\|_{L^2(\omega)} \le \sqrt{2d}\,|\!|\!|(w,u)|\!|\!|_{\lambda,\omega} \quad \text{for any } (w,u) \in X. \tag{2.13}$$

Furthermore, on polytopes, the relation $\|\operatorname{div} w\|_{L^2(\Omega)} \le \|Dw\|_{L^2(\Omega)}$ for any $w \in W$ implies the sharper global estimate

$$\|\tau_\lambda(w,u)\|_{L^2(\Omega)} \le \sqrt{2}\,|\!|\!|(w,u)|\!|\!|_\lambda \quad \text{for any } (w,u) \in X. \tag{2.14}$$

2.3. Preparatory results. Since $\Omega \subseteq \mathbb{R}^d$ is a convex polytope, the Miranda–Talenti estimate (1.3) holds for any $w \in W$. The next lemma may be viewed as a further new generalization of this estimate.

LEMMA 2.1 (Miranda–Talenti-type estimate). *Let $\Omega \subseteq \mathbb{R}^d$ be a convex polytope. Then, any $(w,u) \in X$ satisfies, for any $0 < \rho < 2$, the following inequality:*

$$|\!|\!|(w,u)|\!|\!|_\lambda^2 \le \frac{2}{2-\rho}\left(\|\tau_\lambda(w,u)\|_{L^2(\Omega)}^2 + \frac{\lambda}{\rho}\|\nabla u - w\|_{L^2(\Omega)}^2 + \|\operatorname{rot} w\|_{L^2(\Omega)}^2 \right).$$

Proof. The definition of the norm $|\!|\!|\cdot|\!|\!|_\lambda$ in (2.11), the classical estimate (1.3), together with the integration-by-parts formula $-(w,\nabla u)_{L^2(\Omega)} = (\operatorname{div} w, u)_{L^2(\Omega)}$ for $(w,u) \in X$, and elementary algebraic manipulations reveal that

$$\begin{aligned} |\!|\!|(w,u)|\!|\!|_\lambda^2 &\le \|\operatorname{div} w\|_{L^2(\Omega)}^2 + \|\operatorname{rot} w\|_{L^2(\Omega)}^2 + 2\lambda(\nabla u - w, \nabla u)_{L^2(\Omega)} \\ &\quad - 2\lambda(\operatorname{div} w, u)_{L^2(\Omega)} + \lambda^2\|u\|_{L^2(\Omega)}^2 \\ &= \|\tau_\lambda(w,u)\|_{L^2(\Omega)}^2 + \|\operatorname{rot} w\|_{L^2(\Omega)}^2 + 2\lambda(\nabla u - w, \nabla u)_{L^2(\Omega)}. \end{aligned}$$

Young's inequality yields, for any $0 < \rho < 2$, that

$$2\lambda(\nabla u - w, \nabla u)_{L^2(\Omega)} \leq \frac{\lambda}{\rho}\|\nabla u - w\|^2_{L^2(\Omega)} + \rho\lambda\|\nabla u\|^2_{L^2(\Omega)}.$$

Thus,

$$|||(w,u)|||^2_\lambda \leq \|\tau_\lambda(w,u)\|^2_{L^2(\Omega)} + \|\operatorname{rot} w\|^2_{L^2(\Omega)} + \frac{\lambda}{\rho}\|\nabla u - w\|^2_{L^2(\Omega)} + \rho\lambda\|\nabla u\|^2_{L^2(\Omega)}.$$

After subtracting $\rho\lambda\|\nabla u\|^2_{L^2(\Omega)}$ from both sides, the left-hand side is still bounded from below by $(1-\rho/2)\,|||(w,u)|||^2_\lambda$, which proves the assertion. ∎

LEMMA 2.2. *Let $(w,u),(w',u') \in X$ and abbreviate $\delta_w := w - w'$, $\delta_u := u - u'$, and $\delta := (\delta_w, \delta_u)$. The following estimate holds almost everywhere in Ω:*

$$\begin{aligned} &|F_\gamma[(w,u)] - F_\gamma[(w',u')] - \tau_\lambda(\delta_w,\delta_u)| \\ &\qquad \leq \sqrt{1-\varepsilon}\,\left(|D\delta_w|^2 + 2\lambda|\nabla\delta_u|^2 + \lambda^2|\delta_u|^2\right)^{1/2}. \end{aligned} \tag{2.15}$$

Proof. The proof closely follows the lines of [40, Lem. 1]. The definition of F_γ, the fact that τ_λ is independent of α, and elementary properties of the supremum imply that the left-hand side of (2.15) is bounded from above by $\sup_{\alpha\in\Lambda}|\gamma_\alpha L_\alpha\delta - \tau_\lambda\delta|$. The Cauchy–Schwarz and triangle inequalities bound this term by

$$\sup_{\alpha\in\Lambda}\Big(|\gamma_\alpha a_\alpha - I_{d\times d}|\,|D(w-w')| + |\gamma_\alpha|\,|b_\alpha|\,|\nabla(u-u')| + |\lambda - c_\alpha\gamma_\alpha|\,|u-u'|\Big),$$

where we have used that $\operatorname{div}(w-w') = I_{d\times d} : D(w-w')$. The Cauchy–Schwarz inequality in $\mathbb{R}^3$ therefore eventually leads to

$$|F_\gamma[(w,u)] - F_\gamma[(w',u')] - \tau_\lambda\delta| \leq \sup_{\alpha\in\Lambda}\sqrt{C_\alpha}\,\left(|D\delta_w|^2 + 2\lambda|\nabla\delta_u|^2 + \lambda^2|\delta_u|^2\right)^{1/2},$$

where

$$C_\alpha := |\gamma_\alpha a_\alpha - I_{d\times d}|^2 + |\gamma_\alpha|^2|b_\alpha|^2/(2\lambda) + |\lambda - c_\alpha\gamma_\alpha|^2/\lambda^2.$$

An elementary calculation reveals that

$$C_\alpha = d + 1 - 2\gamma_\alpha(\operatorname{tr} a_\alpha + c_\alpha/\lambda) + |\gamma_\alpha|^2\big(|a_\alpha|^2 + |b_\alpha|^2/(2\lambda) + |c_\alpha|^2/\lambda^2\big).$$

The definition of γ_α and the Cordes condition (1.2) imply that $C_\alpha \leq 1-\varepsilon$. This completes the proof. ∎

The next lemma asserts monotonicity of a on the space X.

LEMMA 2.3 (lower bound). *For any $(w,u),(w',u') \in X$ with $\delta := (w-w', u-u')$ one has that*

$$c_{\mathrm{mon}}|||\delta|||^2_\lambda \leq a((w,u),\delta) - a((w',u'),\delta),$$

where $c_{\mathrm{mon}} := (1-\sqrt{1-\varepsilon})/4$.

Proof. We define $\delta_w := w-w'$ and $\delta_u := u-u'$ so that $\delta = (\delta_w,\delta_u)$. The definition of $\mathcal{A}$ from (2.6) and elementary algebraic manipulations lead to

$$\langle\mathcal{A}[(w,u)] - \mathcal{A}[(w',u')],\delta\rangle = \|\tau_\lambda\delta\|^2_{L^2(\Omega)} + (F_\gamma[(w,u)] - F_\gamma[(w',u')] - \tau_\lambda\delta, \tau_\lambda\delta)_{L^2(\Omega)}.$$

Lemma 2.2 and Young's inequality bound the right-hand side from below by

$$\|\tau_\lambda\delta\|^2_{L^2(\Omega)} - \sqrt{1-\varepsilon}|||\delta|||_\lambda\|\tau_\lambda\delta\|_{L^2(\Omega)} \geq \left(1 - \frac{\sqrt{1-\varepsilon}}{2}\right)\|\tau_\lambda\delta\|^2_{L^2(\Omega)} - \frac{\sqrt{1-\varepsilon}}{2}|||\delta|||^2_\lambda.$$

Lemma 2.1 with the choice $\rho := 2 - 2\sqrt{1-\varepsilon}$ yields

$$\frac{\sqrt{1-\varepsilon}}{2}|||\delta|||^2_\lambda \leq \frac{1}{2}\|\tau_\lambda\delta\|^2_{L^2(\Omega)} + \frac{\lambda}{4(1-\sqrt{1-\varepsilon})}\|\nabla\delta_u - \delta_w\|^2_{L^2(\Omega)} + \frac{1}{2}\|\operatorname{rot}\delta_w\|^2_{L^2(\Omega)}.$$

Hence, the combination of the foregoing displayed inequalities results in

$$\begin{aligned}&\langle\mathcal{A}[(w,u)] - \mathcal{A}[(w',u')],\delta\rangle\\ &\quad\geq \frac{1-\sqrt{1-\varepsilon}}{2}\|\tau_\lambda\delta\|^2_{L^2(\Omega)} - \frac{\lambda}{4(1-\sqrt{1-\varepsilon})}\|\nabla\delta_u - \delta_w\|^2_{L^2(\Omega)} - \frac{1}{2}\|\operatorname{rot}\delta_w\|^2_{L^2(\Omega)}.\end{aligned}$$

We add $\sigma_1\|\operatorname{rot}\delta_w\|^2_{L^2(\Omega)}$ and $\sigma_2\|\nabla\delta_u - \delta_w\|^2_{L^2(\Omega)}$ to both sides, where σ_1 and σ_2 are the parameters defined in (2.8). Then, with the definition of the form a from (2.7),

$$\begin{aligned}&a((w,u),\delta) - a((w',u'),\delta)\\ &\quad\geq \frac{1-\sqrt{1-\varepsilon}}{2}\big(\|\tau_\lambda\delta\|^2_{L^2(\Omega)} + \lambda\|\nabla\delta_u - \delta_w\|^2_{L^2(\Omega)} + \|\operatorname{rot}\delta_w\|^2_{L^2(\Omega)}\big).\end{aligned}$$

The application of Lemma 2.1 with $\rho := 1$ implies that the right-hand side of this is bounded from below by $(1-\sqrt{1-\varepsilon})|||\delta|||^2_\lambda/4$. That concludes the proof. ∎

The next lemma guarantees the Lipschitz continuity of a. The local version of the result on open subdomains $\omega \subseteq \Omega$ will rely on a modified norm, whereas the global version is stated with respect to the norm $|||\cdot|||_\lambda$. We define, for any open subdomain $\omega \subseteq \Omega$ and any $(w,u),(w',u') \in X$, the localized version of a by

$$\begin{aligned}a_\omega((w,u),(w',u')) := (F_\gamma[(w,u)],\tau_\lambda(w',u'))_{L^2(\omega)}&.\\ &+ \sigma_1(\operatorname{rot} w,\operatorname{rot} w')_{L^2(\omega)} + \sigma_2(w - \nabla u, w' - \nabla u')_{L^2(\omega)}.\end{aligned}$$

LEMMA 2.4 (Lipschitz continuity). *There exist positive constants C_{Lip} and $C^{\mathrm{loc}}_{\mathrm{Lip}}$ such that any $(w,u),(w',u'),(z,v) \in X$, with $\delta_w := w - w'$ and $\delta_u := u - u'$, satisfy*

$$a((w,u),(z,v)) - a((w',u'),(z,v)) \leq C_{\mathrm{Lip}}|||(\delta_w,\delta_u)|||_\lambda|||(z,v)|||_\lambda,$$

as well as

$$\begin{aligned}&a_\omega((w,u),(z,v)) - a_\omega((w',u'),(z,v))\\ &\quad\leq C^{\mathrm{loc}}_{\mathrm{Lip}}\Big(\big(|||(\delta_w,\delta_u)|||_{\lambda,\omega} + \|\delta_w\|_{L^2(\omega)}\big)\big(|||(z,v)|||_{\lambda,\omega} + \|z\|_{L^2(\omega)}\big)\Big)\end{aligned}$$

for any open subset $\omega \subseteq \Omega$. The constants C_{Lip} and $C^{\mathrm{loc}}_{\mathrm{Lip}}$ may depend on λ and ε from the Cordes condition (1.2).

Proof. Let $\omega \subseteq \Omega$ be an open subset. Consider the decomposition

$$a_\omega((w,u),(z,v)) - a_\omega((w',u'),(z,v)) = R_1 + R_2 + R_3,$$

where

$$R_1 := (F_\gamma[(w,u)],\tau_\lambda(z,v))_{L^2(\omega)} - (F_\gamma[(w',u')],\tau_\lambda(z,v))_{L^2(\omega)},$$
$$R_2 := \sigma_1(\operatorname{rot}\delta_w,\operatorname{rot} z)_{L^2(\omega)},$$
$$R_3 := \sigma_2(\delta_w - \nabla\delta_u, z-\nabla v)_{L^2(\omega)},$$

and abbreviate $\delta := (\delta_w,\delta_u)$. With the help of Lemma 2.2 and estimate (2.13), the term R_1 is bounded as follows:

$$\begin{aligned} R_1 &= (F_\gamma(w,u) - F_\gamma(w',u') - \tau_\lambda\delta,\tau_\lambda(z,v))_{L^2(\omega)} + (\tau_\lambda\delta,\tau_\lambda(z,v))_{L^2(\omega)} \\ &\le (2d+\sqrt{2d}\sqrt{1-\varepsilon})|\!|\!|\delta|\!|\!|_{\lambda,\omega}|\!|\!|(z,v)|\!|\!|_{\lambda,\omega}. \end{aligned}$$

The term R_2 can be directly bounded as follows:

$$R_2 \le \sigma_1|\!|\!|\delta|\!|\!|_{\lambda,\omega}|\!|\!|(z,v)|\!|\!|_{\lambda,\omega}.$$

The term R_3 can be bounded using the triangle inequality as follows:

$$\begin{aligned} R_3 &\le \sigma_2(\|\delta_w\|_{L^2(\omega)} + \|\nabla\delta_u\|_{L^2(\omega)})(\|z\|_{L^2(\omega)} + \|\nabla v\|_{L^2(\omega)}) \\ &\le \sigma_2 \max\{1,1/\sqrt{2\lambda}\}\big(|\!|\!|(\delta_w,\delta_u)|\!|\!|_{\lambda,\omega} + \|\delta_w\|_{L^2(\omega)}\big)\big(|\!|\!|(z,v)|\!|\!|_{\lambda,\omega} + \|z\|_{L^2(\omega)}\big). \end{aligned}$$

By combining the bounds on R_1, R_2, and R_3 we deduce the localized version of the asserted Lipschitz continuity. The global version for $\omega = \Omega$ follows from the Poincaré-type inequality $\|w\|_{L^2(\Omega)} \lesssim \|Dw\|_{L^2(\Omega)}$ for any $w \in W$. ∎

2.4. Well-posedness of the mixed problem. The following result asserts well-posedness of the mixed formulation and its equivalence to the original boundary-value problem for the HJB equation.

PROPOSITION 2.5. *Let the domain Ω, the parameter set Λ, and the data a, b, c, f satisfy the conditions from section* 2.1. *Then, the system* (2.10) *has a unique solution $(w,u) \in X$, $\mu \in M$. Moreover, $\mu = 0$, $u \in H^2(\Omega)\cap H_0^1(\Omega)$ with $w = \nabla u$, and the function u satisfies* (2.2).

Proof. The result basically follows from the classical Brezzi splitting. For linear problems, this result is standard and is covered, for example, in [5, 3]. The application to the nonlinear problem (2.10) under consideration here follows by a similar reasoning and is briefly summarized below. We begin by defining the kernel of b as

$$Z := \{(w,u) \in W \times U : b(\mu,(w,u)) = 0 \text{ for all } \mu \in M\}.$$

Restricting (2.10) to the kernel Z leads to the problem of finding $(w,u) \in Z$ such that

$$a((w,u),(w',u')) = 0 \quad \text{for all } (w',u') \in Z. \tag{2.16}$$

This problem admits a unique solution: indeed, Lemmas 2.3 and 2.4 ensure that the nonlinear operator

$$(v,\eta) \mapsto a((v,\eta),\bullet) \in X^\star$$

is strongly monotone and Lipschitz continuous on the whole space X. The Browder–Minty theorem [45, Thm. 25B] therefore implies the existence of a unique solution $(w,u) \in Z$ to (2.16). In the case $M = \{0\}$, the identity $Z = X$ holds, and the

existence of a unique solution in X directly follows. In the more general case when $M \subseteq H^1_0(\Omega)$, the form b satisfies the following inf-sup condition for some $\beta > 0$:

$$\beta \leq \inf_{\xi \in M\setminus\{0\}} \sup_{(v,\eta)\in X\setminus\{0\}} \frac{b(\xi,(v,\eta))}{\|\nabla\xi\|_{L^2(\Omega)}|\!|\!|(v,\eta)|\!|\!|_\lambda}. \tag{2.17}$$

This follows from the fact that, for any $\xi \in M$, the Poisson-type problem of finding $\eta \in M$ such that (recall the definition of b in (2.9))

$$b(\zeta,(0,\eta)) = (\nabla\xi,\nabla\zeta)_{L^2(\Omega)} \quad \text{for all } \zeta \in M$$

admits a unique solution $\eta \in M \subseteq U$. Since $M \subseteq U$ is a closed linear subspace, this may be viewed as a Galerkin approximation in U from M. The choice of η implies that

$$b(\xi,(0,\eta)) = \|\nabla\xi\|^2_{L^2(\Omega)}.$$

Clearly, $(0,\eta) \in X$ satisfies

$$|\!|\!|(0,\eta)|\!|\!|_\lambda = \sqrt{2\lambda\|\nabla\eta\|^2_{L^2(\Omega)} + \lambda^2\|\eta\|^2_{L^2(\Omega)}} \lesssim \|\nabla\xi\|_{L^2(\Omega)}.$$

This proves the inf-sup condition (2.17). By a standard application of the closed range theorem as in [5, Lem. 4.2], the inf-sup condition (2.17) implies that the map

$$B : M \to Z^0, \qquad \eta \mapsto B(\eta) := b(\eta,\cdot)$$

is an isomorphism from M to the polar set

$$Z^0 := \{F \in X^\star : F(z) = 0 \text{ for all } z \in Z\}.$$

Since (2.16) implies that $a((w,u),\cdot) \in Z^0$, there exists a unique $\mu \in M$ such that

$$a((w,u),(w',u')) + b(\mu,(w',u')) = 0 \quad \text{for all } (w',u') \in X.$$

Since $(w,u) \in Z$, (2.10b) is obviously satisfied. This establishes the existence of a unique solution $(w,u) \in X$ to (2.10). Since Ω is convex, the operator $(\Delta - \lambda) : H^2(\Omega)\cap H^1_0(\Omega) \to L^2(\Omega)$ is surjective (cf. [24, Thm. 3.2.1.2]). This implies that τ_λ is a surjective map from the subset

$$Y := \{(w',u') \in X : w = \nabla u\} \subseteq X$$

onto $L^2(\Omega)$. Testing (2.10a) with pairs $(w',u') \in Y$ shows that

$$(F_\gamma[(w,u)],\tau_\lambda(w',u'))_{L^2(\Omega)} = 0 \quad \text{for all } (w',u') \in Y.$$

Thus, $F_\gamma[(w,u)] = 0$ as an equality in $L^2(\Omega)$. Testing (2.10a) with (w,u) therefore results in

$$\sigma_1\|\operatorname{rot} w\|^2_{L^2(\Omega)} + \sigma_2\|w - \nabla u\|^2_{L^2(\Omega)} = 0.$$

This implies that $w = \nabla u$ and therefore in particular $u \in H^2(\Omega)\cap H^1_0(\Omega)$. Hence, the identity $F_\gamma[(w,u)] = 0$ can be written as

$$\sup_{\alpha\in\Lambda}[\gamma_\alpha(a_\alpha : D^2u + b_\alpha\cdot\nabla u - c_\alpha u - f_\alpha)] = 0 \quad \text{a.e. in } \Omega.$$

Consequently, u solves the HJB equation. The strong solution property also shows that $\mu = 0$ because (2.16) holds for all $(w',u') \in X$. ∎

Remark 2.6. While the reasoning for the saddlepoint problem in the proof of Proposition 2.5 is somewhat artificial ($\mu = 0$ is known), the same arguments will be required in the discussion of the discrete problem; see Proposition 3.1 below.

3. Discretization. This section is devoted to the description and analysis of the numerical scheme.

3.1. Numerical scheme. Suppose that $W_h \subseteq W$, $U_h \subseteq U$, and $M_h \subseteq U_h \subseteq M$ are closed linear subspaces and define the space $X_h \subseteq X$ by $X_h := W_h \times U_h$. The discrete problem seeks $(w_h, u_h) \in X_h$ and $\mu_h \in M_h$ such that

$$a((w_h,u_h),(w_h',u_h')) + b(\mu_h,(w_h',u_h')) = 0 \qquad \text{for all } (w_h',u_h') \in X_h, \tag{3.1a}$$

$$b((w_h,u_h),\mu_h') = 0 \qquad \text{for all } \mu_h' \in M_h. \tag{3.1b}$$

PROPOSITION 3.1. *There exists a unique solution* $(w_h, u_h) \in X_h$, $\mu_h \in M_h$ *to* (3.1).

Proof. The arguments in the proof are analogous to those in Proposition 2.5. The existence of a unique solution to the problem, restricted to the discrete kernel

$$Z_h := \{(w_h,u_h) \in W_h \times U_h : b(\mu_h,(w_h,u_h)) = 0 \text{ for all } \mu \in M_h\},$$

follows from the strong monotonicity and Lipschitz continuity of a. The arguments from the proof of Proposition 2.5 show that the following discrete inf-sup condition is satisfied:

$$\beta \le \inf_{\xi_h \in M_h \setminus \{0\}} \sup_{(v_h,\eta_h) \in X_h \setminus \{0\}} \frac{b(\xi_h,(v_h,\eta_h))}{\|\nabla \xi_h\|_{L^2(\Omega)} |\!|\!| (v_h,\eta_h) |\!|\!|_\lambda}. \tag{3.2}$$

This and the arguments of Proposition 2.5 conclude the proof. ∎

3.2. Error analysis. The next result states an a priori error estimate.

THEOREM 3.2 (a priori error estimate). *Let* $(w,u) \in X$, $\mu \in M$ *solve* (2.10) *and* $(w_h,u_h) \in X_h$, $\mu_h \in M_h$ *solve* (3.1), *respectively, and define* $e := (w - w_h, u - u_h)$. *Then, the following a priori error bound holds:*

$$\begin{aligned} |\!|\!| e |\!|\!|_\lambda &\le (c_{\mathrm{mon}}^{-1} C_{\mathrm{Lip}}) \inf_{(v_h,\eta_h) \in Z_h} |\!|\!| (w,u) - (v_h,\eta_h) |\!|\!|_\lambda \\ &\lesssim \inf_{(v_h,\eta_h) \in X_h} |\!|\!| (w,u) - (v_h,\eta_h) |\!|\!|_\lambda. \end{aligned}$$

Proof. The monotonicity property from Lemma 2.3 yields that

$$c_{\mathrm{mon}} |\!|\!| e |\!|\!|_\lambda^2 \le a((w,u),e) - a((w_h,u_h),e).$$

The fact that $(w,u) \in X$ is a solution to (2.10) implies that the first term on the right-hand side is equal to zero, and so is the expression $a((w,u),(w,u) - (v_h,\eta_h))$ for any $(v_h,\eta_h) \in Z_h$. Note that, although $Z_h \not\subseteq Z$, the last assertion follows because (w,u) is a strong solution. As $(w_h,u_h) \in X_h$ is a solution to the discrete problem (3.1), the second term is equal to $a((w_h,u_h),(w,u) - (v_h,\eta_h))$ for arbitrary $(v_h,\eta_h) \in Z_h$. Altogether, we infer from the Lipschitz continuity stated in Lemma 2.4 that

$$\begin{aligned} c_{\mathrm{mon}} |\!|\!| e |\!|\!|_\lambda^2 &\le a((w,u),(w,u) - (v_h,\eta_h)) - a((w_h,u_h),(w,u) - (v_h,\eta_h)) \\ &\le C_{\mathrm{Lip}} |\!|\!| e |\!|\!|_\lambda |\!|\!| (w,u) - (v_h,\eta_h) |\!|\!|_\lambda. \end{aligned} \tag{3.3}$$

This implies the first inequality in the statement of the theorem.

The remaining part of the proof bounds the expression on the right-hand side of (3.3) by some constant times the best-approximation in X_h. Let $(w_\star,\eta_\star) \in X_h$

be the best-approximation to (w,u) from the closed linear subspace $Z_h \subseteq X$ with respect to the norm $|||\cdot|||_\lambda$. The discrete inf-sup condition (3.2) implies the existence of a multiplier $\tilde{\mu}_\star \in M_h$ such that the following linear variational problem is satisfied (recall the scalar product $(\cdot,\cdot)_\lambda$ from (2.12)):

$$\begin{aligned}((w_\star,u_\star),(v_h,\eta_h))_\lambda + (v_h - \nabla\eta_h, \nabla\tilde{\mu}_\star)_{L^2(\Omega)} &= ((w,u),(v_h,\eta_h))_\lambda,\\ (w_\star - \nabla u_\star, \nabla\xi_h)_{L^2(\Omega)} &= 0\end{aligned}$$

for all $(v_h,\eta_h) \in X_h$ and all $\xi_h \in M_h$. Obviously, (w,u) satisfies the same system with multiplier $\mu = 0$ for all test functions $(v,\eta) \in X$ and all $\xi \in M$ because $w = \nabla u$ holds as an equality in $L^2(\Omega)$. Moreover, the stability condition (3.2) and Brezzi's splitting theorem imply that the linear system satisfies a global inf-sup condition, and, thus, the theory of mixed finite elements [3, Thm. 5.2.2] shows that there exists a constant $C(\lambda)$ such that

$$\begin{aligned}&|||(w-w_\star,u-u_\star)|||_\lambda + \|\nabla(\mu-\mu_\star)\|_{L^2(\Omega)}\\ &\quad\le C(\lambda) \inf_{\substack{(v_h,\eta_h)\in X_h\\ \xi_h\in M_h}} \big(|||(w-v_h,u-\eta_h)|||_\lambda + \|\nabla(\mu-\xi_h)\|_{L^2(\Omega)}\big)\\ &\quad= C(\lambda) \inf_{(v_h,\eta_h)\in X_h} |||(w-v_h,u-\eta_h)|||_\lambda,\end{aligned}$$

where the last equality holds because $\mu = 0$. This concludes the proof. ∎

Let us denote by $S^k(\mathcal{T})$ the Lagrange finite element space of degree k over a shape-regular simplicial triangulation $\mathcal{T}$ of Ω and denote by $S^k(\mathcal{T};\mathbb{R}^d)$ the space of d-component vector fields whose components belong to $S^k(\mathcal{T})$.

COROLLARY 3.3. *Let $\mathcal{T}$ be a simplicial triangulation of Ω, let $k,m \ge 1$ be integers, let $X_h := (S^k(\mathcal{T};\mathbb{R}^d) \cap H^1_t(\Omega;\mathbb{R}^d)) \times S^m_0(\mathcal{T})$, and let $M_h \subseteq S^m(\mathcal{T})$ be an arbitrary linear subspace. Assume that $u \in H^{2+s}(\Omega)$ for some $s \ge 0$. Then, with the notation of Theorem* 3.2,

$$|||e|||_\lambda \lesssim \|h\|^r_{L^\infty(\Omega)} \|u\|_{H^{2+s}(\Omega)},$$

where $r := \min\{k,m,s\}$.

Proof. This follows from standard interpolation error bounds [6, p. 123]. ∎

The strong monotonicity and the localized Lipschitz continuity properties of the form $a(\cdot,\cdot)$ result in the following a posteriori error bound.

THEOREM 3.4 (a posteriori error bound). *Let $(w,u) \in X$, $\mu \in M$ and $(w_h,u_h) \in X_h$, $\mu_h \in M_h$ solve* (2.10) *and* (3.1)*, respectively, and abbreviate $e := (w-w_h, u-u_h)$. Then, the following reliable a posteriori error bound holds:*

$$\begin{aligned}&|||e|||^2_\lambda\\ &\quad\le \frac{2}{c_{\mathrm{mon}}}\Big(\frac{1}{c_{\mathrm{mon}}}\|F_\gamma[(w_h,u_h)]\|^2_{L^2(\Omega)} + \sigma_1\|\mathrm{rot}\, w_h\|^2_{L^2(\Omega)} + \sigma_2\|w_h - \nabla u_h\|^2_{L^2(\Omega)}\Big).\end{aligned}$$

Furthermore, for any open subset $\omega \subseteq \Omega$, the following local efficiency estimate holds:

$$\begin{aligned}&\|F_\gamma[(w_h,u_h)]\|^2_{L^2(\omega)} + 2\sigma_1\|\mathrm{rot}\, w_h\|^2_{L^2(\omega)} + 2\sigma_2\|w_h - \nabla u_h\|^2_{L^2(\omega)}\\ &\quad\le (4C^{\mathrm{loc}}_{\mathrm{Lip}} + 1 - \varepsilon)(|||e|||^2_{\lambda,\omega} + \|w-w_h\|^2_{L^2(\omega)}).\end{aligned}$$

Proof. The monotonicity from Lemma 2.3, the fact that $(w,u) \in X$ is a strong solution to (2.10), and the Cauchy–Schwarz inequality imply that

$$\begin{aligned} c_{\mathrm{mon}} &|\!|\!| e |\!|\!|_\lambda^2 \\ &\le a((w,u),e) - a((w_h,u_h),e) = -a((w_h,u_h),e) \\ &\le \|F_\gamma[(w_h,u_h)]\|_{L^2(\Omega)} \|\tau_\lambda e\|_{L^2(\Omega)} + \sigma_1 \|\mathrm{rot}\, w_h\|_{L^2(\Omega)}^2 + \sigma_2 \|w_h - \nabla u_h\|_{L^2(\Omega)}^2, \end{aligned}$$

where we have used that $\mathrm{rot}\, w = 0$ and $w = \nabla u$. Since, by (2.14), $\|\tau_\lambda e\|_{L^2(\Omega)} \le \sqrt{2}\, |\!|\!| e |\!|\!|_\lambda$, Young's inequality yields

$$\begin{aligned} \|F_\gamma[(w_h,u_h)]\|_{L^2(\Omega)} \|\tau_\lambda e\|_{L^2(\Omega)} &\le \sqrt{2} \|F_\gamma[(w_h,u_h)]\|_{L^2(\Omega)} |\!|\!| e |\!|\!|_\lambda \\ &\le c_{\mathrm{mon}}^{-1} \|F_\gamma[(w_h,u_h)]\|_{L^2(\Omega)}^2 + 2^{-1} c_{\mathrm{mon}} |\!|\!| e |\!|\!|_\lambda^2. \end{aligned}$$

The combination of the foregoing estimates concludes the proof of reliability. The proof of efficiency follows from the Lipschitz continuity stated in Lemma 2.4. Indeed, since

$$\|F_\gamma[(w,u)]\|_{L^2(\omega)}^2 = \sigma_1 \|\mathrm{rot}\, w\|_{L^2(\omega)}^2 = \sigma_2 \|w - \nabla u\|_{L^2(\omega)}^2 = 0$$

and, in particular, the quantities under the norms are equal to zero almost everywhere, it follows from Lemmas 2.2 and 2.4 that

$$\begin{aligned} &\|F_\gamma[(w_h,u_h)]\|_{L^2(\omega)}^2 + \sigma_1 \|\mathrm{rot}\, w_h\|_{L^2(\omega)}^2 + \sigma_2 \|w_h - \nabla u_h\|_{L^2(\omega)}^2 \\ &\quad = a_\omega((w,u),e) - a_\omega((w_h,u_h),e) \\ &\qquad - (F_\gamma[(w_h,u_h)], F_\gamma[(w,u)] - F_\gamma[(w_h,u_h)] - \tau_\lambda e)_{L^2(\omega)} \\ &\quad \le 2C_{\mathrm{Lip}}^{\mathrm{loc}} (|\!|\!| e |\!|\!|_{\lambda,\omega}^2 + \|w - w_h\|_{L^2(\omega)}^2) + \|F_\gamma[(w_h,u_h)]\|_{L^2(\omega)} \sqrt{1-\varepsilon}\, |\!|\!| e |\!|\!|_{\lambda,\omega}. \end{aligned} \tag{3.4}$$

Thanks to Young's inequality we have that

$$\|F_\gamma[(w_h,u_h)]\|_{L^2(\omega)} \sqrt{1-\varepsilon}\, |\!|\!| e |\!|\!|_{\lambda,\omega} \le \frac{1}{2} \|F_\gamma[(w_h,u_h)]\|_{L^2(\omega)}^2 + \frac{1}{2}(1-\varepsilon) |\!|\!| e |\!|\!|_{\lambda,\omega}^2,$$

which then allows us to absorb the term $\|F_\gamma[(w_h,u_h)]\|_{L^2(\omega)}$ into the left-hand side of (3.4), so that

$$\begin{aligned} &\|F_\gamma[(w_h,u_h)]\|_{L^2(\omega)}^2 + 2\sigma_1 \|\mathrm{rot}\, w_h\|_{L^2(\omega)}^2 + 2\sigma_2 \|w_h - \nabla u_h\|_{L^2(\omega)}^2 \\ &\quad \le 4C_{\mathrm{Lip}}^{\mathrm{loc}} (|\!|\!| e |\!|\!|_{\lambda,\omega}^2 + \|w - w_h\|_{L^2(\omega)}^2) + (1-\varepsilon) |\!|\!| e |\!|\!|_{\lambda,\omega}^2. \end{aligned}$$

This concludes the proof. ∎

3.3. Convergence of an adaptive algorithm. The remainder of this section is devoted to the convergence analysis of an adaptive algorithm. The arguments are similar to those in [21] and are based on the framework of [34]. Since this appears to be the first convergence proof of an adaptive algorithm applied to a fully nonlinear problem of HJB-type, we have included the details of the argument. For ease of the exposition the choice $M = \{0\}$ has been made so that the problem is positive definite. For any triangulation $\mathcal{T}_\ell$ in the adaptively refined sequence, the space X_h is chosen as some fixed-order Lagrange finite element space as in Corollary 3.3. In the notation of this section, quantities related to the level $\ell \in \mathbb{N}_0$ and the triangulation $\mathcal{T}_\ell$ are labelled by the index ℓ (instead of h from earlier sections).

The algorithm is as follows. The input of the algorithm consists of an initial mesh $\mathcal{T}_0$ and a marking parameter $0 < \theta \le 1$. The algorithm runs the usual SOLVE → ESTIMATE → MARK → REFINE loop for $\ell = 0, 1, 2, \ldots$ as follows.

SOLVE. Solve the discrete problem (3.1) with respect to the mesh $\mathcal{T}_\ell$ and the space $X(\mathcal{T}_\ell)$ and denote the corresponding solution by (w_ℓ, u_ℓ).
ESTIMATE. Compute, for any $T \in \mathcal{T}_\ell$, the local error estimator contributions

$$\eta_\ell^2(T) = \|F_\gamma[(w_\ell, u_\ell)]\|_{L^2(T)}^2 + \sigma_1 \|\mathrm{rot}\, w_\ell\|_{L^2(T)}^2 + \sigma_2 \|w_\ell - \nabla u_\ell\|_{L^2(T)}^2$$

and set $\eta_\ell^2 := \sum_{T\in\mathcal{T}_\ell} \eta_\ell^2(T)$.
MARK. Mark a minimal subset $\mathcal{M} \subseteq \mathcal{T}_\ell$ such that $\theta\eta_\ell^2 \le \sum_{T\in\mathcal{M}} \eta_\ell^2(T)$.
REFINE. Use the refinement rules from [2, 43] to compute a refined admissible partition $\mathcal{T}_{\ell+1}$ of $\mathcal{T}_\ell$ such that at least all elements of $\mathcal{M}$ are refined.

In the marking step, the bulk selection rule from [13] is chosen, but some other strategies are possible as well.

THEOREM 3.5. *The sequence $(w_\ell, u_\ell) \in X(\mathcal{T}_\ell)$ produced by the adaptive algorithm converges to the exact solution $(w, u) \in X$, i.e., $|||(w, u) - (w_\ell, u_\ell)|||_\lambda \to 0$ as $\ell \to \infty$.*

Proof. The proof begins with the observation that the sequence of discrete solutions $x_\ell := (w_\ell, u_\ell)$ converges to some limit $x_\star = (w_\star, u_\star) \in X_\star$, which solves, for all $\ell \in \mathbb{N}_0$,

$$a(x_\star, x'_\ell) = 0 \quad \text{for all } x'_\ell \in X(\mathcal{T}_\ell). \tag{3.5}$$

For the proof of this claim it suffices to consider the closure $X_\star$ of $\cup_{\ell\in\mathbb{N}_0} X(\mathcal{T}_\ell)$ with respect to the norm $|||\cdot|||_\lambda$. Lemmas 2.3 and 2.4 and the density of $\cup_{\ell\in\mathbb{N}_0} X(\mathcal{T}_\ell)$ in $X_\star$ shows that (3.5) is uniquely solvable. The stated convergence follows from the monotonicity (Lemma 2.3)

$$|||x_\star - x_\ell|||_\lambda^2 \lesssim a(x_\star, x_\star - x_\ell) - a(x_\ell, x_\star - x_\ell),$$

the Galerkin property

$$a(x_\star, x_\star - x_\ell) - a(x_\ell, x_\star - x_\ell) = a(x_\star, x_\star - x'_\ell) - a(x_\ell, x_\star - x'_\ell) \quad \text{for any } x'_\ell \in X(\mathcal{T}_\ell),$$

and the Lipschitz continuity (Lemma 2.4)

$$a(x_\star, x_\star - x'_\ell) - a(x_\ell, x_\star - x'_\ell) \lesssim |||x_\star - x_\ell|||\,|||x_\star - x'_\ell||| \quad \text{for any } x'_\ell \in X(\mathcal{T}_\ell).$$

The density of $\cup_{\ell\in\mathbb{N}_0} X(\mathcal{T}_\ell)$ in $X_\star$ implies the convergence of x_ℓ to $x_\star$ as $\ell \to \infty$.

As in [34], the convergence proof employs the subset $\mathcal{K} \subseteq \cup_{\ell\ge 0}\mathcal{T}_\ell$ of never refined elements, defined by

$$\mathcal{K} := \bigcup_{\ell\ge 0} \bigcap_{m\ge\ell} \mathcal{T}_m.$$

For any $\ell \ge 0$, the partition $\mathcal{T}_\ell$ can be written as the following disjoint union:

$$\mathcal{T}_\ell = \mathcal{K}_\ell \cup \mathcal{R}_\ell \qquad \text{for } \mathcal{K}_\ell := \mathcal{K} \cap \mathcal{T}_\ell \text{ and } \mathcal{R}_\ell := \mathcal{T}_\ell \setminus \mathcal{K}_\ell. \tag{3.6}$$

This means that each triangulation $\mathcal{T}_\ell$ is decomposed in a set $\mathcal{K}_\ell$ of never refined elements and a set $\mathcal{R}_\ell$ of elements that are eventually refined. For any $T \in \mathcal{T}_\ell$, the local efficiency from Theorem 3.4 and the triangle inequality yield that

$$\eta_\ell^2(T) \lesssim |||(w, u) - x_\star|||_{\lambda,T}^2 + \|w - w_\star\|_{L^2(T)}^2 + |||x_\star - x_\ell|||_{\lambda,T}^2 + \|w_\star - w_\ell\|_{L^2(T)}^2. \tag{3.7}$$

As, by definition, every element of $\mathcal{R}_\ell$ is eventually refined, it can be seen [34, 20] that for any $\rho > 0$ there exists an $\ell_0 \ge 0$ such that, for all $\ell \ge \ell_0$,

$$\max_{T\in\mathcal{R}_\ell} \mathrm{meas}(T) < \rho. \tag{3.8}$$

From the convergence of x_ℓ to $x_\star$ and the observation (3.8) that the elements in $\mathcal{R}_\ell$ become, uniformly, arbitrarily small for sufficiently large ℓ, one therefore deduces using (3.7) that

$$\max_{T\in\mathcal{R}_\ell}\eta_\ell^2(T)\to 0 \quad \text{as } \ell\to\infty.$$

The marking strategy ensures that

$$\max_{T\in\mathcal{K}_\ell}\eta_\ell^2(T)\le\max_{T\in\mathcal{R}_\ell}\eta_\ell^2(T).$$

Thus, for any $T\in\mathcal{K}:=\bigcup_{\ell\in\mathbb{N}_0}\mathcal{K}_\ell$,

$$\|F_\gamma[(w_\ell,u_\ell)]\|_{L^2(T)}^2+\sigma_1\|\mathrm{rot}\,w_\ell\|_{L^2(T)}^2+\sigma_2\|w_\ell-\nabla u_\ell\|_{L^2(T)}^2\to 0 \quad \text{as } \ell\to\infty.$$

The convergence of the sum of these contributions over all elements of $\mathcal{K}$ follows from the dominated convergence theorem: since $\mathrm{meas}(\cup\mathcal{K}\setminus\cup\mathcal{K}_m)\to 0$ as $m\to\infty$, the dominated convergence theorem implies that

$$\|F_\gamma[(w_\ell,u_\ell)]\|_{L^2(\cup\mathcal{K}\setminus\cup\mathcal{K}_m)}^2+\sigma_1\|\mathrm{rot}\,w_\ell\|_{L^2(\cup\mathcal{K}\setminus\cup\mathcal{K}_m)}^2+\sigma_2\|w_\ell-\nabla u_\ell\|_{L^2(\cup\mathcal{K}\setminus\cup\mathcal{K}_m)}^2\to 0$$

as $m\to\infty$. Thus, with the triangle inequality, estimate (2.13), and Lemma 2.2, it follows that

$$\begin{aligned}\|F_\gamma[x_\ell]\|_{L^2(\cup\mathcal{K})}&\lesssim\|F_\gamma[x_\ell]\|_{L^2(\cup\mathcal{K}_m)}+\|F_\gamma[x_\star]-F_\gamma[x_\ell]-\tau_\lambda(x_\star-x_\ell)\|_{L^2(\cup\mathcal{K}\setminus\cup\mathcal{K}_m)}\\&\quad+\|F_\gamma[x_\star]\|_{L^2(\cup\mathcal{K}\setminus\cup\mathcal{K}_m)}+|\!|\!|x_\star-x_\ell|\!|\!|_{\lambda,\cup\mathcal{K}\setminus\cup\mathcal{K}_m}\\&\lesssim\|F_\gamma[x_\ell]\|_{L^2(\cup\mathcal{K}_m)}+\|F_\gamma[x_\star]\|_{L^2(\cup\mathcal{K}\setminus\cup\mathcal{K}_m)}+|\!|\!|x_\star-x_\ell|\!|\!|_{\lambda,\cup\mathcal{K}\setminus\cup\mathcal{K}_m}\end{aligned}$$

for every $m\in\mathbb{N}_0$. Since $\|F_\gamma[x_\ell]\|_{L^2(\cup\mathcal{K}_m)}\to 0$ as $\ell\to\infty$ because this term is composed of error estimator contributions on a finite subset of $\mathcal{K}$, one has from the convergence of x_ℓ to $x_\star$ that

$$\limsup_{\ell\to\infty}\|F_\gamma[x_\ell]\|_{L^2(\cup\mathcal{K})}\le\|F_\gamma[x_\star]\|_{L^2(\cup\mathcal{K}\setminus\cup\mathcal{K}_m)}.$$

The right-hand side becomes arbitrarily small for large m, as can be seen using the dominated convergence theorem. Similar arguments for the remaining error estimator contributions yield that

$$\sum_{T\in\mathcal{K}}\Big(\|F_\gamma[(w_\ell,u_\ell)]\|_{L^2(T)}^2+\sigma_1\|\mathrm{rot}\,w_\ell\|_{L^2(T)}^2+\sigma_2\|w_\ell-\nabla u_\ell\|_{L^2(T)}^2\Big)\to 0$$

as $\ell\to\infty$.

The global efficiency of the error estimator (Theorem 3.4) and the monotonicity from Lemma 2.3 imply, with the Galerkin property, that

$$\begin{aligned}\sum_{T\in\mathcal{T}_\ell}\eta_\ell^2(T)&\lesssim a((w,u),(w,u)-(w_\ell,u_\ell))-a((w_\ell,u_\ell),(w,u)-(w_\ell,u_\ell))\\&=a((w,u),(w,u)-(I_\ell w,J_\ell u))-a((w_\ell,u_\ell),(w,u)-(I_\ell w,J_\ell u))\end{aligned}\tag{3.9}$$

for quasi-interpolation operators I_ℓ and J_ℓ [10] which yield quasi-local quasi-best approximations in W_ℓ and U_ℓ, respectively. The strong solution property of (w,u) shows that the term

$$a_\omega((w,u),(w,u)-(I_\ell w,J_\ell u))-a_\omega((w_\ell,u_\ell),(w,u)-(I_\ell w,J_\ell u))\tag{3.10}$$

for $\omega = \cup\mathcal{K}$ is controlled by error estimator contributions times quasi-interpolation errors. It converges to zero because the error estimator contributions are driven to zero in that region while the quasi-interpolation is stable. The term (3.10) converges to zero for the choice $\omega = \cup\mathcal{R}_\ell$, too. The reason is that (w, u) is approximated by its quasi-interpolant on elements whose diameter becomes arbitrarily small. In conclusion, the expression (3.10) converges to zero as $\ell \to \infty$ for the domain $\omega = \Omega$. The estimate (3.9) and the reliability of the error estimator therefore conclude the convergence proof. ∎

4. Numerical results. In this section we present numerical results in planar domains. The spaces W_h, U_h, and M_h consist of piecewise affine functions on a triangulation $\mathcal{T}$ of mesh size h of Ω. The coefficients in the partial differential equation are approximated with piecewise constant functions over the same triangulation in the sense that the pointwise-in-Ω supremization over Λ is replaced by an elementwise-in-$\mathcal{T}$ supremization where u_h is replaced by its integral mean over every element in $\mathcal{T}$. We shall compare uniform and adaptive mesh refinements (with $\theta = 0.3$). The discrete problems are solved using the semismooth Newton described in the next subsection.

4.1. Semismooth Newton method. The solution algorithm used to solve the nonlinear problems belongs to the class of semismooth Newton methods [44]. The use of such methods for problems of HJB-type dates back to the early reference [26]. The presentation most relevant to ours is [40], where, in particular, semismoothness of the HJB operator was shown and the proof of local (mesh-dependent) convergence was given. The arguments in [40] transfer to the present situation. Since adapting those proofs to our setting is quite straightforward, this section merely describes the algorithm and briefly highlights some of the steps in the convergence analysis.

In each iteration step of the semismooth Newton scheme, the parameter $\alpha \in \Lambda$ is supremized pointwise in Ω. It determines the PDE coefficients for the solution of a linear problem that defines the updated approximation to the PDE solution. Given any $(w, u) \in X$, the set of admissible maximizers is denoted by

$$\Lambda[(w,u)] := \left\{ \begin{array}{l|l} g : \Omega \to \Lambda & g(x) \in \arg\max\limits_{\alpha\in\Lambda}(L_\alpha(u,w)(x) - f_\alpha) \\ \text{measurable} & \text{for almost every } x \in \Omega \end{array} \right\}. \tag{4.1}$$

As discussed in [40], an application of a result from [31] shows that the sets $\Lambda[(w, u)]$ are indeed nonempty. The semismooth Newton algorithm is defined as follows:

Input: Initial guess $(w_h^0, u_h^0) \in X_h$ and a termination criterion.
for $k = 0, 1, 2, \ldots$ **until** termination **do**

Choose any $\alpha_k \in \Lambda[(w_h^k, u_h^k)]$ and compute $(w_h^{k+1}, u_h^{k+1}) \in X_h$ and $\mu_h^{k+1} \in M_h$ as the solution to the linear problem

$$\begin{aligned}(\gamma_{\alpha_k}(L_{\alpha_k}(w_h^{k+1}, u_h^{k+1}) - f_{\alpha_k}), \tau_\lambda(w_h', u_h'))_{L^2(\Omega)} + b(\mu_h^{k+1}, (w_h', u_h')) = 0,\\ b((w_h^{k+1}, u_h^{k+1}), \mu_h') = 0\end{aligned}$$

for all $(w_h', u_h') \in X_h$ and all $\mu_h' \in M_h$.
end do

Some comments are in order to explain why this procedure can be seen as a semismooth Newton iteration. In the notation of [40, Def. 12], the definition of semismoothness [44] is as follows.

DEFINITION 4.1. *Let* $\mathbf{X}$, $\mathbf{Y}$ *be Banach spaces, let* $\mathbf{U} \subseteq \mathbf{X}$ *be an open nonempty subset, and let* $F : \mathbf{X} \to \mathbf{Y}$. *Let* $DF : \mathbf{U} \to 2^{\mathcal{L}(\mathbf{X},\mathbf{Y})}$ *be a set-valued map from* $\mathbf{U}$ *to the space of bounded linear operators from* $\mathbf{X}$ *to* $\mathbf{Y}$. *Given* $\mathbf{x} \in \mathbf{U}$, *the map* F *is said to be* DF*-semismooth at* $\mathbf{x}$ *if*

$$\lim_{\|\mathbf{s}\|_{\mathbf{X}} \to 0} \|\mathbf{s}\|_{\mathbf{X}}^{-1} \sup_{B \in DF(\mathbf{x}+\mathbf{s})} \|F(\mathbf{x}+\mathbf{s}) - F(\mathbf{x}) - B\mathbf{s}\|_{\mathbf{Y}} = 0.$$

The map F *is called* DF*-semismooth on* $\mathbf{U}$ *if it is* DF*-semismooth at every* $\mathbf{x} \in \mathbf{U}$. *In this case,* DF *is called a generalized differential of* F *on* $\mathbf{U}$.

Let $1 \le q < r \le \infty$ be integrability indices, and consider the Banach spaces $\mathbf{X} := W^{1,r}(\Omega;\mathbb{R}^d) \times W^{1,r}(\Omega)$ and $\mathbf{Y} := L^q(\Omega))$ and the map

$$DF_\gamma : \mathbf{X} \to 2^{\mathcal{L}(\mathbf{X},\mathbf{Y})}$$

defined, for any $(w,u) \in \mathbf{X}$, by

$$DF_\gamma[(w,u)] := \{\gamma_\alpha L^\alpha : \alpha \in \Lambda[(w,u)]\},$$

where $\Lambda[(w,u)]$ is defined in (4.1). This property explains the structure of the linear problems in the above iterative loop. It can be shown with the arguments of [39, Thm. 13] that the operator F_γ from (2.5) is DF_γ semismooth as a map from $\mathbf{X}$ to $\mathbf{Y}$, but this property requires the stronger assumption $q < r$, which is generally not valid when DF_γ is viewed as a map from $X \supseteq \mathbf{X}$ to $L^2(\Omega) \subseteq \mathbf{Y}$ (see [25, 40]). Thus, the scheme cannot be directly analyzed on the infinite-dimensional level. On finite-dimensional subspaces of X, equivalence of norms can be employed so that the required mapping property is satisfied. However, the constants involved in the norm-equivalence will generally depend on the mesh-size. This is the reason why in the convergence analysis of the nonlinear solver, the closeness requirement for the choice of an initial guess is mesh-dependent. The result is as follows.

PROPOSITION 4.2. *Let* $\mathcal{T}$ *be a simplicial triangulation of* Ω *of mesh size* h, *and let* X_h *and* M_h *be finite-dimensional subspaces based on piecewise polynomials, as in Corollary* 3.3. *Then, there exists a constant* $R > 0$ *that may depend on the mesh-size of* $\mathcal{T}$ *as well as on the polynomial degree, such that for* $|||(w_h,u_h)-(w_{h,0},u_{h,0})|||_\lambda < R$, *the sequence* $((w_{h,k},u_{h,k}))_{k\in\mathbb{N}}$ *generated by the semismooth Newton algorithm converges, with a superlinear rate, to the unique solution* $(w_h,u_h) \in X_h$ *of the discrete problem* (3.1).

Proof. The proof is very similar to [40, Thm. 11] and it is therefore omitted. ∎

4.2. Experiment 1. The first example considers a test case from [40] with near-degenerate diffusion and a boundary layer in the solution. Let $\Omega := (0,1)^2$ be the unit square and let $\Lambda := \mathrm{SO}(2)$ be the special orthogonal group (describing rotations in the plane). The elementwise supremization problems are solved by sampling over a sufficiently fine subdivision of Λ. Let $b_\alpha = (0,1)$, $c_\alpha = 10$, and

$$a_\alpha := \alpha^{\mathrm{T}} \begin{pmatrix} 20 & 1 \\ 1 & 1/10 \end{pmatrix} \alpha \quad \text{for } \alpha \in \Lambda.$$

For this choice of parameters and $\lambda = 1/2$, the Cordes condition (1.2) is satisfied for $\varepsilon = 0.0024$ [40]. Let $\delta = 0.01$ and let $f_\alpha := \tilde{L}_\alpha(u)$ for the exact solution

$$u(x) = (2x_1 - 1)\Big(\exp(1 - |2x_1 - 1|) - 1\Big)\left(x_2 + \frac{1 - \exp(x_2/\delta)}{\exp(1/\delta) - 1}\right).$$

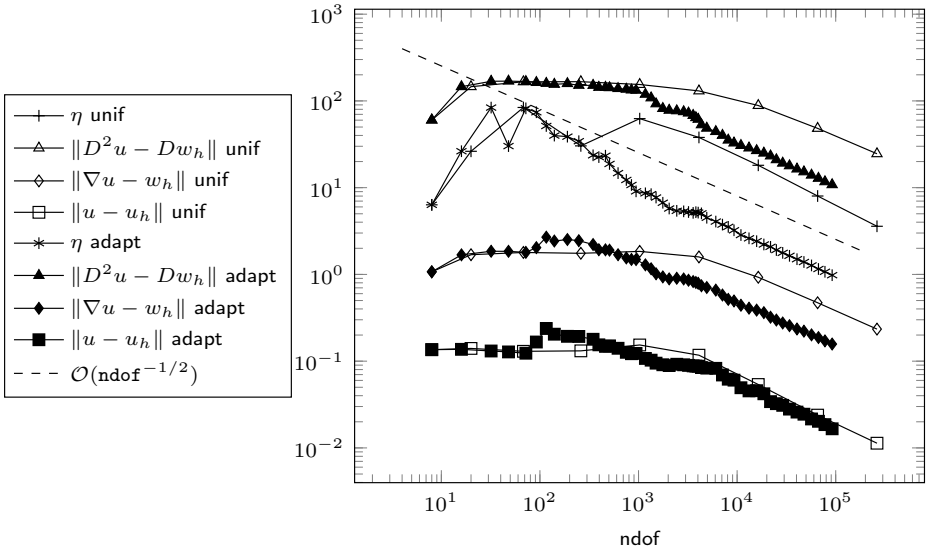

FIG. 1. *Convergence history for Experiment* 1.

The solution exhibits a sharp boundary layer near the line $\bar{\Omega} \cap \{x_2 = 1\}$. Ideally, such problems are approximated with anisotropic meshes, as was done in [40] (even with $\delta = 0.005$) by using an exponentially accurate *hp*-version discontinuous Galerkin scheme. Here, the focus is on *automatic* mesh refinement driven by the a posteriori error estimator from Theorem 3.4 with isotropic meshes. The solution u in this example belongs to $C_0^1(\overline{\Omega}) \cap H^2(\Omega)$ (but $u \notin H^3(\Omega)$), so uniform mesh refinement might not be expected to be optimal in terms of asymptotic rates in general. In addition, the preasymptotic range can be arbitrarily large and, indeed, the convergence rates for uniform meshes displayed in Figure 1 are only visible for very fine meshes. In particular, the error is observed to increase when the coarsest meshes are refined. The adaptive algorithm can improve the approximation in this example and leads to approximations that exhibit a convergence rate beginning from approximately 700 degrees of freedom. Accordingly, the adaptive mesh in Figure 2 shows a strong refinement toward the layer. While the observed convergence is justified by Theorem 3.5, these additional empirical findings indicate that adaptivity may significantly improve efficiency. The (square-root of the) constant in the reliable error estimate from Theorem 3.4 scales like c_{mon}^{-1}, and for this value of ε it is of the order of magnitude of 4×10^3, which means an overestimation of the actual error. From the successful mesh adaptation we can, however, infer that the error estimation still adequately indicates the error distribution over the domain.

In the convergence history reported in Figure 1, no higher-order convergence in weaker norms is observed. This could possibly be due to the fact that in the implementation the quadrature is chosen so that u_h is replaced by its piecewise integral mean over the triangulation in the supremization process over Λ. On the other hand, there is no proof of higher-order rates even in an idealized version of the algorithm with exact quadrature, and the absence of improved L^2 rates could also be caused by other features of the nonlinear problem. We also do not observe an improvement of the L^2 approximation through adaptivity in this example. This might be due to the fact that the solution is piecewise smooth and the mesh is aligned with the discontinuity

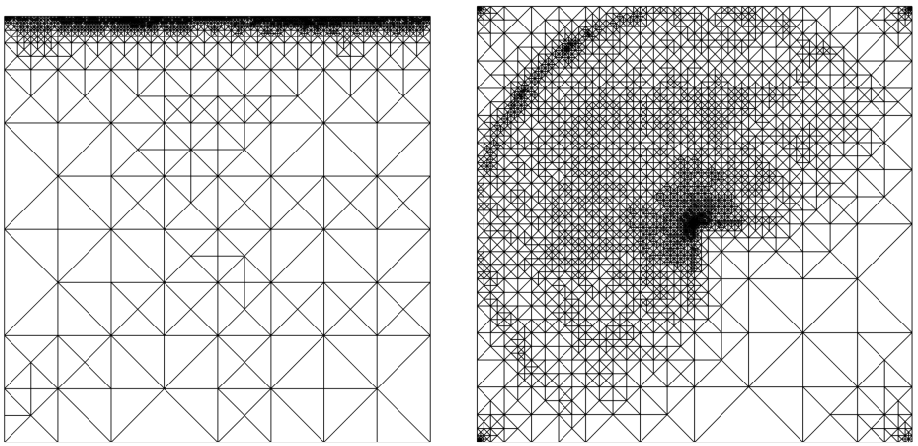

FIG. 2. *Adaptive meshes. Left: Experiment* 1, 2,905 *vertices,* 10,932 *degrees of freedom, level* 36. *Right: Experiment* 2; 3,431 *vertices,* 13,464 *degrees of freedom, level* 16.

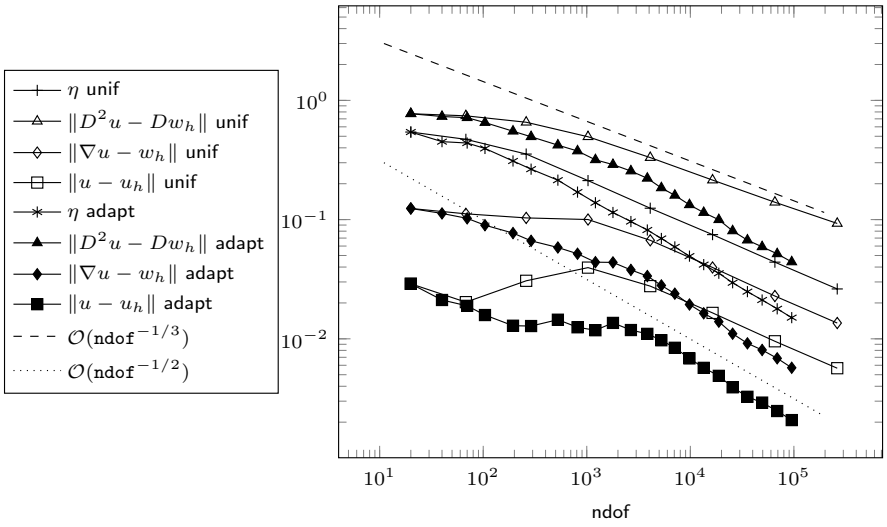

FIG. 3. *Convergence history for Experiment* 2.

of the second derivative, so that the adaptivity is primarily of importance in the preasymptotic regime.

4.3. Experiment 2. Let Ω be the square $\Omega := (-1,1)^2$ and let again $\Lambda := \mathrm{SO}(2)$ and a_α, b_α, c_α be as in Experiment 1. Let $f_\alpha := \tilde{L}_\alpha(u)$ for the exact solution u given in polar coordinates as

$$u(r,\theta) = \begin{cases} r^{5/3}(1-r)^{5/2}\sin(2\theta/3)^{5/2} & \text{if } 0 < r \le 1 \text{ and } 0 < \theta < 3\pi/2, \\ 0 & \text{otherwise.} \end{cases}$$

In contrast with the first experiment, the asymptotic approximation rate on uniform meshes is suboptimal in this example because the solution has a point singularity at $(0,0)$. Figure 3 compares the convergence rates for uniform and adaptive mesh refine-

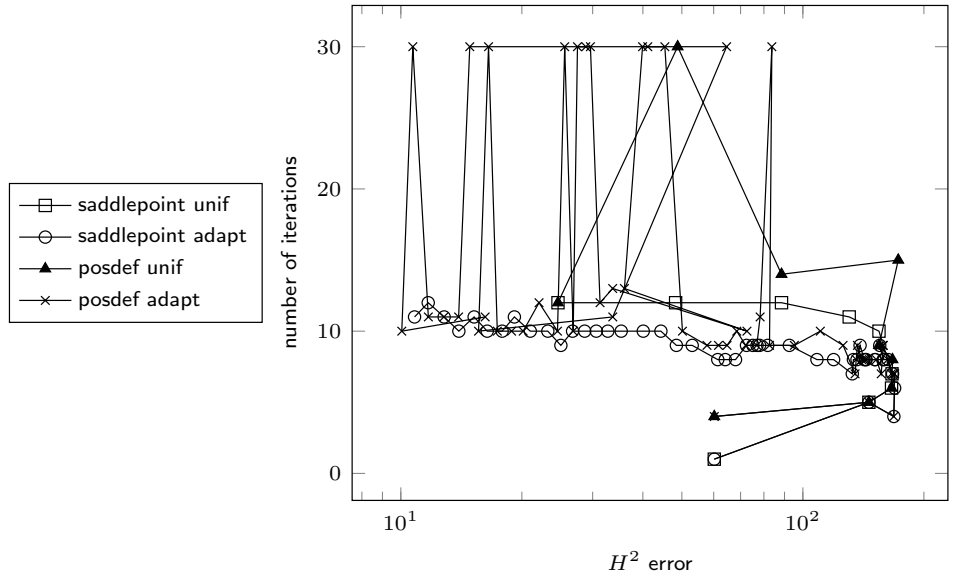

FIG. 4. *Comparison of semismooth Newton iterations in Experiment* 1 *dependent on the* H^2 *error: saddlepoint formulation (saddlepoint) and positive definite formulation (posdef).*

ments. As in the first example, a clear improvement by adaptivity can be observed. Uniform mesh-refinement leads to a convergence rate of $1/3$ while for adaptive mesh-refinement the optimal rate of $1/2$ is observed for the $H^2(\Omega)$ norm error. In contrast to the first experiment, in this example we see an improved L^2 approximation by adaptivity. The adaptive mesh displayed in Figure 2 shows strong refinement around the singularity.

4.4. Comparison of semismooth Newton iterations. As mentioned earlier, the discrete formulation does not require a nontrivial space M of Lagrange multipliers, and the choice $M = \{0\}$ leading to a positive definite problem is admissible. However, when comparing the number of semismooth Newton iterations required to reach the desired tolerance, the saddlepoint formulation corresponding to a nontrivial M appears to be more robust. In the following, the choices $M_h = \{0\}$ and $M_h = U_h$ are compared. The termination criterion in all examples was chosen as

$$\max\{\|\alpha_k - \alpha_{k-1}\|_{L^2(\Omega)}, \|Dw_k - Dw_{k-1}\|_{L^2(\Omega)}\} < 10^{-8}$$

with a maximum number of iterations $k_{\max} = 30$. The initial guess α_0 on the coarse mesh was chosen as $\alpha_0 = 0$. On finer meshes, the initial guess α_0 was taken as the solution on the previous mesh (nested iteration).

Figure 4 compares the number of iterations in the semismooth Newton method against the H^2 error for Experiment 1. It can be observed that in the saddlepoint formulation with $M_h = U_h$ these numbers robustly stay in a moderate range. In the positive definite formulation with $M_h = \{0\}$, especially in the adaptive method, the termination by 30 iterations is reached several times. A similar behavior is observable in Experiment 2 (see Figure 5). While on uniform meshes the iteration numbers are comparable, the semismooth Newton method for the positive definite formulation seems to be less robust when adaptive mesh-refinement is used. In conclusion, in this nonlinear problem the saddlepoint formulation may have advantages that are not

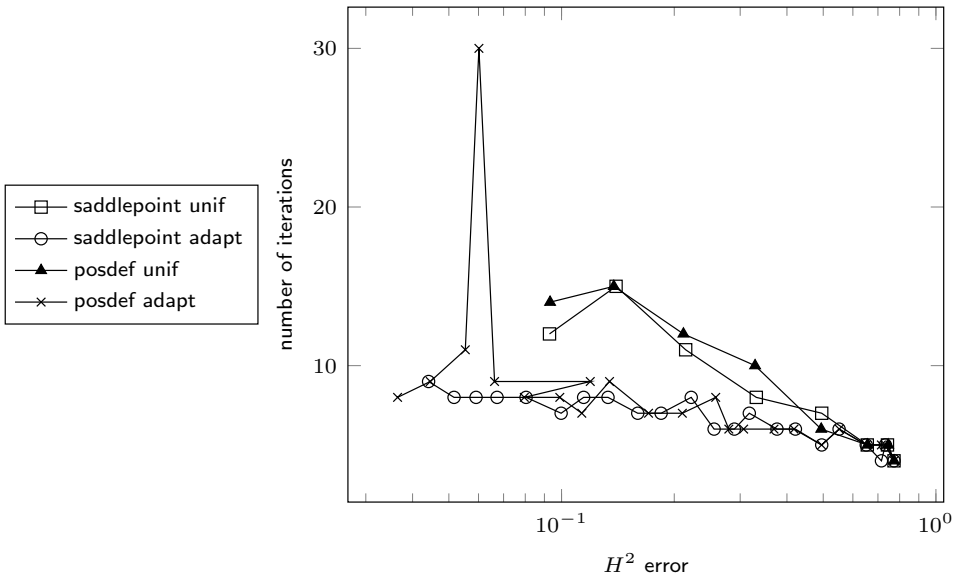

FIG. 5. *Comparison of semismooth Newton iterations in Experiment* 2 *dependent on the* H^2 *error; saddlepoint formulation (saddlepoint) and positive definite formulation (posdef).*

apparent from the numerical analysis of the scheme we have performed, but can be observed in practical computations employing an iterative scheme.

REFERENCES

[1] G. BARLES AND P. E. SOUGANIDIS, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptot. Anal., 4 (1991), pp. 271–283.
[2] P. BINEV, W. DAHMEN, AND R. DEVORE, *Adaptive finite element methods with convergence rates*, Numer. Math., 97 (2004), pp. 219–268.
[3] D. BOFFI, F. BREZZI, AND M. FORTIN, *Mixed Finite Element Methods and Applications*, Springer Ser. Comput. Math. 44, Springer, New York, 2013.
[4] K. BÖHMER, *Numerical Methods for Nonlinear Elliptic Differential Equations. A Synopsis*, Num. Math. Sci. Comput., Oxford University Press, Oxford, 2010.
[5] D. BRAESS, *Finite Elements. Theory, Fast Solvers, and Applications in Elasticity Theory*, 3rd ed., Cambridge University Press, Cambridge, 2007.
[6] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Texts Appl. Math. 44, Springer, New York, 2008.
[7] F. CAMILLI AND M. FALCONE, *An approximation scheme for the optimal control of diffusion processes*, RAIRO Modél. Math. Anal. Numér., 29 (1995), pp. 97–122.
[8] F. CAMILLI AND E. R. JAKOBSEN, *A finite element like scheme for integro-partial differential Hamilton-Jacobi-Bellman equations*, SIAM J. Numer. Anal., 47 (2009), pp. 2407–2431.
[9] J. CASCON, C. KREUZER, R. H. NOCHETTO, AND K. G. SIEBERT, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal., 46 (2008), pp. 2524–2550.
[10] P. CLÉMENT, *Approximation by finite element functions using local regularization*, Rev. Française Automat. Informat. Rech. Oper., 9 (1975), pp. 77–84.
[11] H. O. CORDES, *Über die erste Randwertaufgabe bei quasilinearen Differentialgleichungen zweiter Ordnung in mehr als zwei Variablen*, Math. Ann., 131 (1956), pp. 278–312.
[12] 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.
[13] W. DÖRFLER, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124.

[14] X. Feng, R. Glowinski, and M. Neilan, *Recent developments in numerical methods for fully nonlinear second order partial differential equations*, SIAM Rev., 55 (2013), pp. 205–267.
[15] X. Feng, L. Hennings, and M. Neilan, *Finite element methods for second order linear elliptic partial differential equations in non-divergence form*, Math. Comp., 86 (2017), pp. 2025–2051.
[16] X. Feng and M. Jensen, *Convergent semi-Lagrangian methods for the Monge-Ampère equation on unstructured grids*, SIAM J. Numer. Anal., 55 (2017), pp. 691–712.
[17] X. Feng and M. Neilan, *Analysis of Galerkin methods for the fully nonlinear Monge-Ampère equation*, J. Sci. Comput., 47 (2011), pp. 303–327.
[18] X. Feng, M. Neilan, and S. Schnake, *Interior Penalty Discontinuous Galerkin Methods for Second Order Linear Non-Divergence Form Elliptic PDEs*, arXiv:1605.04364, 2016.
[19] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Appl. Math. 25, Springer, New York, 1993.
[20] D. Gallistl, *Stable splitting of polyharmonic operators by generalized Stokes systems*, Math. Comp., 86 (2017), pp. 2555–2577.
[21] D. Gallistl, *Variational formulation and numerical analysis of linear elliptic equations in nondivergence form with Cordes coefficients*, SIAM J. Numer. Anal., 55 (2017), pp. 737–757.
[22] D. Gallistl, *Mixed Finite Element Approximation of Elliptic Equations Involving High-Order Derivatives*, Habilitation thesis, Karlsruher Institut für Technologie, Karlsruhe, 2018.
[23] D. Gallistl, *Numerical approximation of planar oblique derivative problems in nondivergence form*, Math. Comp., 88 (2019), pp. 1091–1119.
[24] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Classics Appl. Math. 69, SIAM, Philadelphia, 2011.
[25] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888.
[26] R. A. Howard, *Dynamic Programming and Markov Processes*, MIT Press, Cambridge, MA, 1960.
[27] M. Jensen, *$L^2(H^1_\gamma)$ finite element convergence for degenerate isotropic Hamilton-Jacobi-Bellman equations*, IMA J. Numer. Anal., 37 (2017), pp. 1300–1316.
[28] 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.
[29] M. Kocan, *Approximation of viscosity solutions of elliptic partial differential equations on minimal grids*, Numer. Math., 72 (1995), pp. 73–92.
[30] N. V. Krylov, *Nonlinear Elliptic and Parabolic Equations of the Second Order*, Math. Appl. 7, D. Reidel, Dordrecht, the Netherlands, 1987.
[31] 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.
[32] O. Lakkis and T. Pryer, *A finite element method for second order nonvariational elliptic problems*, SIAM J. Sci. Comput., 33 (2011), pp. 786–801.
[33] A. Maugeri, D. K. Palagachev, and L. G. Softova, *Elliptic and Parabolic Equations with Discontinuous Coefficients*, Wiley, Berlin, 2000.
[34] P. Morin, K. G. Siebert, and A. Veeser, *A basic convergence result for conforming adaptive finite elements*, Math. Models Methods Appl. Sci., 18 (2008), pp. 707–737.
[35] 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.
[36] M. Neilan, A. J. Salgado, and W. Zhang, *Numerical analysis of strongly nonlinear PDEs*, Acta Numer., 26 (2017), pp. 137–303.
[37] R. H. Nochetto and W. Zhang, *Discrete ABP Estimate and Convergence Rates for Linear Elliptic Equations in Non-Divergence Form*, preprint, arXiv:1411.6036, 2014.
[38] 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.
[39] 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.
[40] I. Smears and E. Süli, *Discontinuous Galerkin finite element approximation of Hamilton–Jacobi–Bellman equations with Cordes coefficients*, SIAM J. Numer. Anal., 52 (2014), pp. 993–1016.
[41] I. Smears and E. Süli, *Discontinuous Galerkin finite element methods for time-dependent Hamilton–Jacobi–Bellman equations with Cordes coefficients*, Numer. Math., 133 (2016), pp. 141–176.

[42] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math., 7 (2007), pp. 245–269.
[43] R. Stevenson, *The completion of locally refined simplicial partitions created by bisection*, Math. Comp., 77 (2008), pp. 227–241.
[44] M. Ulbrich, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2002), pp. 805–842.
[45] E. Zeidler, *Nonlinear Functional Analysis and its Applications.* II/B, Springer, New York, 1990.