Numerische Mathematik manuscript No.
(will be inserted by the editor)

Discontinuous Galerkin finite element methods for time-dependent Hamilton–Jacobi–Bellman equations with Cordes coefficients

Iain Smears · Endre Süli

June 9, 2015

Abstract We propose and analyse a fully discrete discontinuous Galerkin time-stepping method for parabolic Hamilton–Jacobi–Bellman equations with Cordes coefficients. The method is consistent and unconditionally stable on rather general unstructured meshes and time-partitions. Error bounds for both rough and regular solutions in terms of temporal regularity show that the method is arbitrarily high-order with optimal convergence rates with respect to the mesh size, time-interval length and temporal polynomial degree, and possibly suboptimal by an order and a half in the spatial polynomial degree. Numerical experiments on problems with strongly anisotropic diffusion coefficients and early-time singularities demonstrate the accuracy and computational efficiency of the method, with exponential convergence rates achieved under combined hp- and τq-refinement.

Keywords Fully nonlinear partial differential equations · Hamilton–Jacobi–Bellman equations · hp-version discontinuous Galerkin methods · Cordes condition

Mathematics Subject Classification (2000) 65N30 · 65N12 · 65N15 · 35K10 · 35K55 · 35D35

1 Introduction

We consider the numerical analysis of the Cauchy–Dirichlet problem for Hamilton–Jacobi–Bellman (HJB) equations of the form

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

where $\Omega \subset \mathbb{R}^d$ is a bounded convex domain, $I = (0,T)$, Λ is a compact metric space, and where the L^α are nondivergence form elliptic operators given by

$$L^\alpha v := a^\alpha : D^2 v + b^\alpha \cdot \nabla v - c^\alpha v, \qquad \alpha\in\Lambda. \tag{1.2}$$

HJB equations of the form (1.1) arise from problems of optimal control of stochastic processes over a finite-time horizon [12]. Note that the specific form of the HJB equation in (1.1) is obtained after reversing the time variable of the control problem, and thus it will be considered along with an initial-time Cauchy condition and a lateral Dirichlet boundary condition. The equation (1.1) is called *uniformly parabolic* if there exist positive constants $\nu \leq \bar{\nu}$ such that

$$\nu|\xi|^2 \leq \xi^\top a^\alpha(x,t)\,\xi \leq \bar{\nu}|\xi|^2 \quad \forall\,\xi\in\mathbb{R}^d,\ \forall\,(x,t)\in\Omega\times I,\ \forall\,\alpha\in\Lambda. \tag{1.3}$$

Mathematical Institute, University of Oxford, E-mail: smears@maths.ox.ac.uk, · E-mail: suli@maths.ox.ac.uk

In this case, the regularity theory founded on the celebrated Evans–Krylov Theorem establishes interior $C^{2,\beta}$-regularity of the viscosity solution of fully nonlinear uniformly elliptic and parabolic equations with convex nonlinearities [4,5,10,16,29]; this applies to HJB equations of the form (1.1). However, if the uniform parabolicity assumption is relaxed and the diffusion is allowed to become *degenerate*, then the viscosity solution is typically Lipschitz continuous. From a computational point of view, it is helpful to also consider the case where the diffusion coefficient a^{α} is *strongly anisotropic*, a typical example being when the diffusion is dominant in certain directions, and the eigenvectors of a^{α} are not well-aligned with the computational grid. Strongly anisotropic problems can occur in both the uniformly parabolic or degenerate cases.

Monotone schemes of finite difference (FD) type, which conserve the maximum principle in the discrete setting, represent a significant class of numerical methods for (1.1) since Barles and Souganidis [2] established a general convergence theory for these methods that is applicable to a broad class of possibly degenerate fully nonlinear equations. The history and early literature of these methods is discussed for example in [12,18], and there is a significant literature on the practical issues related to enforcing the monotonicity assumption required by the convergence theory [3,8,15,17,22]. The main upshot is that, for strongly anisotropic problems, monotonicity makes it necessary to use wide stencils, thus leading to significant consequences for the accuracy and computational complexity of these methods.

Among recent works on second order HJB and closely related Bellman–Isaacs equations, building on earlier work by Camilli and Falcone [6], Debrabant and Jakobsen developed in [9] a semi-Lagrangian framework, where the stencil width increases as the mesh is refined. Although [9] treats HJB and Bellman–Isaacs problems posed on the entire space $\mathbb{R}^d$, they point to some of the issues associated with these schemes on bounded domains in [9, Section 6.1], such as a possible loss of accuracy or monotonicity near the boundary. Uniform convergence to the viscosity solution of monotone finite element methods for isotropic but possibly degenerate HJB equations was shown by Jensen and the first author in [14] through an extension of the Barles–Souganidis framework, along with strong convergence results in $L^2(H^1)$ under nondegeneracy assumptions. The focus of this work is rather complementary to [14], as we concentrate here on anisotropic but uniformly parabolic problems.

Several authors have suggested different nonmonotone methods for various fully nonlinear second order PDE, most commonly for Monge–Ampère equations. Feng et al. summarise many of these works in the review paper [11], which points to current perspectives and challenges in the analysis of stability and convergence of these methods. Nevertheless, some methods have offered promising computational results in the absence of theoretical analysis; for instance, Lakkis and Pryer have successfully tested in [19,20] a FEM using Hessian reconstructions on a range of fully nonlinear elliptic equations including Monge–Ampère and Pucci equations.

One approach to developing a nonmonotone method that allows a full theoretical analysis in terms of consistency, stability, and error bounds, was proposed in [26,27]. This approach was first applied to linear nondivergence form elliptic equations [26] and then to elliptic HJB equations on convex domains [27]. The approach is founded on the Cordes condition, an algebraic assumption on the coefficients of the operators L^{α}, which comes from the study of nondivergence form elliptic and parabolic equations with discontinuous coefficients [7,21]. The diffusion coefficient can be strongly anisotropic, since, in the case of problems without lower order terms in two dimensions, the Cordes condition is equivalent to uniform ellipticity [27, Example 2].

As first shown in [27], the Cordes condition permits a straightforward proof of existence and uniqueness in H^2 of the solution of a fully nonlinear elliptic HJB equation on a convex domain. In the parabolic setting, the solution of (1.1) belongs to $L^2(H^2) \cap H^1(L^2)$, as shown in section 2. The continuous analysis in [26,27] highlighted the key ingredients for developing a consistent, stable and high-order hp-version discontinuous Galerkin finite element method (DGFEM) for uniformly elliptic but possibly strongly anisotropic problems. Indeed, the discrete form that defines the numerical scheme is constructed to reproduce the structural properties of its continuous counterpart; this is achieved by weakly enforcing an important integration by parts identity connected to the Miranda–Talenti Inequality. The accuracy and efficiency of the method was demonstrated through numerical experiments for a range of challenging problems, including boundary layers, corner singularities and strongly anisotropic diffusion coefficients.

This work extends our previous results to parabolic HJB equations by combining the spatial discretisation of [27] with a discontinuous Galerkin (DG) time-stepping scheme [28]. The resulting method is consistent, unconditionally stable and arbitrarily high-order, whilst permitting rather general unstructured meshes and time partitions. Moreover, the results of this work are applicable to other forms of HJB equations, such as the case where the supremum is replaced by an infimum in (1.1), and also to Bellman–Isaacs equations from stochastic differential games.

To simplify the presentation of the numerical method, we first introduce the essential ideas of the time-stepping scheme in a semidiscrete context in section 3. After defining the relevant finite element spaces in section 4, we present the fully discrete scheme in section 5 and we show its consistency. A key contribution of this work relates to the construction of the time-stepping scheme. Indeed, in order to treat the nonlinearity of the HJB operator, the scheme proposed here differs from standard discontinuous Galerkin time-stepping methods through testing the equation with time and spatial partial derivatives of the test function. We show in section 6 that this choice leads to stability in a discrete $H^1(L^2) \cap L^2(H^2)$-type norm. We emphasise however that this construction does not lead to a least squares method, as our method does not seek to minimise the norm of the residual.

Then, in section 7, we show that the consistency and good stability properties of method lead to optimal convergence rates in terms of the mesh size h, time-interval length τ, and temporal polynomial degrees q. The rates in the spatial polynomial degrees p are possibly suboptimal by an order and a half, as is common for DGFEM that are stable in discrete H^2-norms [23]. For example, in the case of sufficiently smooth solutions and quasi-uniform meshes and time-partitions with uniform polynomial degrees, we show an error bound of the form

$$\begin{aligned}\|u - u_h\|_h \lesssim \frac{h^{\min(s,\,p+1)-2}}{p^{s-7/2}} \|u\|_{L^2(H^s)} + \frac{h^{\min(\bar{s},\,p+1)}}{p^{\bar{s}}} \|u\|_{H^1(H^{\bar{s}})} \\ + \frac{h^{\min(\tilde{s},\,p+1)-1}}{p^{\tilde{s}-3/2}} \|u(0)\|_{H^{\tilde{s}}} + p^{3/2} \sum_{\ell \in \{0,2\}} \frac{\tau^{\min(\sigma_\ell,\,q+1)-1+\ell/2}}{q^{\sigma_\ell - 1 + \ell/2}} \|u\|_{H^{\sigma_\ell}(H^\ell)},\end{aligned} \tag{1.4}$$

where $\|\cdot\|_h$ is a discrete $H^1(L^2) \cap L^2(H^2)$-type norm defined in (7.1), and we assume that $s > 5/2$, $\bar{s} > 0$, $\tilde{s} > 3/2$, and $\sigma_\ell \geq 1$ for $\ell \in \{0, 2\}$. As mentioned above, the error bound (1.4) is thus of optimal orders in τ, h and q, and suboptimal in p.

We note that the techniques of error analysis in the literature on discontinuous Galerkin time discretisations of parabolic equations often require sufficient smoothness [1,25], which, in the present setting, would correspond to assuming $H^1(H^2)$-regularity over each time interval. In view of parabolic regularity theory, this appears as rather more restrictive than the spatial regularity assumptions. Therefore, we use Clément-type projection operators to obtain bounds under very weak temporal regularity assumptions of the form $H^\sigma(H^2)$ for any $\sigma \geq 0$. To help distinguish the special treatment of this case, we shall refer to such solutions as "low regularity" solutions, as opposed to the case of "regular" solutions when $\sigma \geq 1$. In particular, the error bounds for low regularity solutions are applicable to problems with early-time singularities induced by the initial datum.

In section 8.1, we test the numerical scheme on a problem with a strongly anisotropic diffusion and we demonstrate the scheme's accuracy and efficiency. Furthermore, a key reason for choosing DG time stepping methods is that Schötzau and Schwab showed in [25] that these schemes have the potential for exponential convergence rates under hp- and τq-refinement, even for low regularity solutions. Therefore, we show that our method retains this quality in the numerical experiment of section 8.2, which involves a solution with an early-time singularity.

2 Analysis of the problem

Let Ω be a bounded convex polytopal open set in $\mathbb{R}^d$, $d \geq 2$, let Λ be a compact metric space, and let $I := (0,T)$, with $T > 0$. It is assumed that Ω and Λ are non-empty. Convexity of Ω implies that the boundary $\partial\Omega$ of Ω is Lipschitz [13]. Let the symmetric $\mathbb{R}^{d\times d}$-valued function a, the $\mathbb{R}^d$-valued function b, and scalar-valued functions c and f be continuous on $\overline{\Omega} \times \overline{I} \times \Lambda$. For each $\alpha \in \Lambda$, define the functions $a^\alpha \colon (x,t) \mapsto a(x,t,\alpha)$, where $(x,t) \in \overline{\Omega} \times \overline{I}$; the functions b^α, c^α and f^α are similarly defined.

The operators $L^\alpha \colon L^2(I;H^2(\Omega)) \to L^2(I;L^2(\Omega))$ are given by

$$L^\alpha v := a^\alpha : D^2 v + b^\alpha \cdot \nabla v - c^\alpha v, \quad v \in L^2(I;H^2(\Omega)),\ \alpha \in \Lambda, \tag{2.1}$$

where $D^2 v$ denotes the Hessian matrix of v. Compactness of Λ and continuity of the functions a, b, c and f imply that the fully nonlinear operator F, given by

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

is well-defined as a mapping from $H(I;\Omega) := L^2(I;H^2(\Omega)\cap H^1_0(\Omega)) \cap H^1(I;L^2(\Omega))$ into $L^2(I;L^2(\Omega))$. The problem considered is to find a function $u \in H(I;\Omega)$ that is a strong solution of the parabolic HJB equation subject to Cauchy–Dirichlet boundary conditions:

$$\begin{aligned} F[u] &= 0 && \text{in } \Omega \times I, \\ u &= 0 && \text{on } \partial\Omega \times I, \\ u &= u_0 && \text{on } \Omega \times \{0\}, \end{aligned} \tag{2.3}$$

where $u_0 \in H^1_0(\Omega)$. Note that the lateral condition $u = 0$ on $\partial\Omega \times I$ is incorporated in the function space $H(I;\Omega)$. Well-posedness of (2.3) is established in section 2.1 under the following hypotheses.

We assume uniform parabolicity (1.3), nonnegativity of c, and the Cordes condition [26,27]: there exist $\varepsilon \in (0,1]$, $\lambda > 0$ and $\omega > 0$ such that

$$\frac{|a^\alpha|^2 + 1/\lambda^2 + 1/\omega^2}{(\operatorname{Tr} a^\alpha + 1/\lambda + 1/\omega)^2} \leq \frac{1}{d+1+\varepsilon} \quad \text{in } \overline{\Omega}\times\overline{I},\ \forall\, \alpha \in \Lambda, \tag{2.4}$$

where $|a^\alpha|$ denotes the Frobenius norm of the matrix a^α. In the special case where $b \equiv 0$ and $c \equiv 0$, we set $\lambda = 0$ and assume that there exist $\varepsilon \in (0,1]$ and $\omega > 0$ such that

$$\frac{|a^\alpha|^2 + 1/\omega^2}{(\operatorname{Tr} a^\alpha + 1/\omega)^2} \leq \frac{1}{d+\varepsilon} \quad \text{in } \overline{\Omega}\times\overline{I},\ \forall\, \alpha \in \Lambda. \tag{2.5}$$

As explained in [27],the parameters λ and ω serve to make the Cordes condition invariant under rescaling of the spatial and temporal domains. In the case of elliptic equations in two dimensions without lower order terms, the Cordes condition is equivalent to uniform ellipticity [27].

Given (2.4), by considering transformations of the unknown of the type $u = \mathrm{e}^{\mu t}\tilde{u}$, we can assume without loss of generality that

$$\frac{|a^\alpha|^2 + |b^\alpha|^2/2\lambda + (c^\alpha/\lambda)^2 + 1/\omega^2}{(\operatorname{Tr} a^\alpha + c^\alpha/\lambda + 1/\omega)^2} \leq \frac{1}{d+1+\varepsilon} \quad \text{in } \overline{\Omega}\times\overline{I},\ \forall\, \alpha \in \Lambda. \tag{2.6}$$

The relevance of (2.4) is to show that the Cordes condition is essentially independent of the lower order terms b^α and c^α, although it will be simpler to work with (2.6). Define the strictly positive function $\gamma \colon \Omega \times I \times \Lambda \to \mathbb{R}_{>0}$ by

$$\gamma(x,t,\alpha) := \frac{\operatorname{Tr} a^\alpha(x,t) + c^\alpha/\lambda + 1/\omega}{|a^\alpha(x,t)|^2 + |b^\alpha|^2/2\lambda + (c^\alpha/\lambda)^2 + 1/\omega^2}. \tag{2.7}$$

In the case of $b \equiv 0$ and $c \equiv 0$, the function γ is defined by

$$\gamma(x,t,\alpha) := \frac{\operatorname{Tr} a^\alpha(x,t) + 1/\omega}{|a^\alpha(x,t)|^2 + 1/\omega^2}. \tag{2.8}$$

Continuity of the data implies that $\gamma \in C(\overline{\Omega} \times \overline{I} \times \Lambda)$, and it follows from (1.3) that there exists a positive constant $\gamma_0 > 0$ such that $\gamma \geq \gamma_0$ on $\overline{\Omega} \times \overline{I} \times \Lambda$. For each $\alpha \in \Lambda$, define $\gamma^\alpha \colon (x,t) \mapsto \gamma(x,t,\alpha)$, and define the operator $F_\gamma \colon H(I;\Omega) \to L^2(I;L^2(\Omega))$ by

$$F_\gamma[v] := \inf_{\alpha\in\Lambda} \left[\gamma^\alpha \left(\partial_t v - L^\alpha v + f^\alpha\right)\right]. \tag{2.9}$$

For ω and λ as in (2.6), we introduce the operators L_λ and L_ω defined by

$$L_\lambda v := \Delta v - \lambda v \qquad\qquad L_\omega v := \omega\, \partial_t v - L_\lambda v. \tag{2.10}$$

The following result is similar to [27, Lemma 1], so the proof is omitted here.

Lemma 1 *Let Ω be a bounded open subset of $\mathbb{R}^d$, let $I = (0,T)$, and suppose that* (2.6) *holds, or that* (2.5) *holds if $b \equiv 0$ and $c \equiv 0$. Let $U \subset \Omega$ be an open set, let $J \subset I$ be an open interval, and let the functions u, $v \in L^2(J;H^2(U)) \cap H^1(J;L^2(U))$, and set $w := u - v$. Then, the following inequality holds a.e. in U, for a.e. $t \in J$:*

$$|F_\gamma[u] - F_\gamma[v] - L_\omega w| \leq \sqrt{1-\varepsilon}\left(\omega^2|\partial_t w|^2 + |D^2 w|^2 + 2\lambda|\nabla w|^2 + \lambda^2|w|^2\right)^{1/2}, \tag{2.11}$$

with $\lambda = 0$ if $b \equiv 0$ and $c \equiv 0$.

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 discretisation parameters such as the element sizes of the meshes and the polynomial degrees of the finite element spaces used below, but otherwise possibly dependent on other fixed quantities, such as, for example, the constants in (1.3) and (2.4) or the shape-regularity parameters of the mesh.

2.1 Well-posedness

For a bounded convex domain $\Omega \subset \mathbb{R}^d$, the Miranda–Talenti Inequality [13,21] states that $|v|_{H^2(\Omega)} \leq \|\Delta v\|_{L^2(\Omega)}$ for all $v \in H^2(\Omega) \cap H^1_0(\Omega)$. Along with the Poincaré Inequality, it implies that $H := H^2(\Omega) \cap H^1_0(\Omega)$ is a Hilbert space when equipped with the inner-product $\langle u, v\rangle_\Delta := \langle L_\lambda u, L_\lambda v\rangle_{L^2(\Omega)}$, where L_λ is from (2.10) and $\lambda \geq 0$ is from (2.6). It is possible to identify H^*, the dual space of H, with $L^2(\Omega)$ through the duality pairing

$$\langle f, v\rangle_{L^2\times H} := \int_\Omega f(-L_\lambda v)\,\mathrm{d}x, \quad f \in L^2(\Omega),\ v \in H. \tag{2.12}$$

Indeed, we clearly have $L^2(\Omega) \hookrightarrow H^*$, and H^2-regularity of solutions of Poisson's equation in convex domains [13] shows that this embedding is an isometry: for any $f \in L^2(\Omega)$, we have $\|f\|_{L^2(\Omega)} = \|f\|_{H^*}$. If $\varphi \in H^*$, then the Riesz Representation Theorem implies that there is a unique $w \in H$ such that $\langle w, v\rangle_\Delta = \varphi(v)$ for all $v \in H$. Then $f = -L_\lambda w \in L^2(\Omega)$ satisfies $\langle f, v\rangle_{L^2\times H} = \varphi(v)$ for all $v \in H$.

The space $H^1_0(\Omega)$ may be equipped with the inner-product $\langle u, v\rangle_{H^1_0} := \int_\Omega \nabla u\cdot\nabla v + \lambda u v\,\mathrm{d}x$ with associated norm $\|\cdot\|_{H^1_0}$; we note that the Poincaré Inequality implies positive definiteness of $\langle\cdot,\cdot\rangle_{H^1_0}$ in the case of $\lambda = 0$.

The relevance of these choices of duality pairing and inner-products is that the spaces H, $H_0^1(\Omega)$ and $L^2(\Omega)$ form a Gelfand triple as a result of the following integration by parts identity: for any $w \in H_0^1(\Omega)$ and $v \in H$, we have

$$\langle w, v\rangle_{L^2\times H} = \int_\Omega w(-L_\lambda v)\,\mathrm{d}x = \int_\Omega \nabla w\cdot\nabla v + \lambda\, w\, v\,\mathrm{d}x = \langle w, v\rangle_{H_0^1}. \tag{2.13}$$

Recall that $H(I;\Omega) \coloneqq L^2(I;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(I;L^2(\Omega))$. The general theory of Bochner spaces, see for instance [30], yields the following result.

Lemma 2 *Let $\Omega \subset \mathbb{R}^d$ be a bounded convex domain and let $I = (0,T)$. Then,*

$$H = H^2(\Omega)\cap H_0^1(\Omega) \hookrightarrow H_0^1(\Omega) \hookrightarrow L^2(\Omega)$$

form a Gelfand triple [30] *under the inner product $\langle\cdot,\cdot\rangle_{H_0^1}$ and the duality pairing $\langle\cdot,\cdot\rangle_{L^2\times H}$. The space $H(I;\Omega)$ is continuously embedded in $C(\overline{I};H_0^1(\Omega))$, and for every u, $v \in H(I;\Omega)$ and any $t\in\overline{I}$, we have*

$$\langle u(t), v(t)\rangle_{H_0^1} = \langle u(0), v(0)\rangle_{H_0^1} + \int_0^t \langle \partial_t u, v\rangle_{L^2\times H} + \langle \partial_t v, u\rangle_{L^2\times H}\,\mathrm{d}s. \tag{2.14}$$

Define the norms $\|\cdot\|_H$ on H and $\|\cdot\|_{H(I;\Omega)}$ on $H(I;\Omega)$ by

$$\|v\|_H^2 \coloneqq |v|_{H^2(\Omega)}^2 + 2\lambda|v|_{H^1(\Omega)}^2 + \lambda^2\|v\|_{L^2(\Omega)}^2, \qquad v\in H, \tag{2.15}$$

$$\|v\|_{H(I;\Omega)}^2 \coloneqq \int_0^T \omega^2\|\partial_t v\|_{L^2(\Omega)}^2 + \|v\|_H^2\,\mathrm{d}t, \qquad v \in H(I;\Omega). \tag{2.16}$$

We will make use of the following solvability result for the Cauchy–Dirichlet problem associated to the linear operator L_ω from (2.10).

Theorem 3 *Let $\Omega \subset \mathbb{R}^d$ be a bounded convex domain and let $I = (0,T)$. For each $g \in L^2(I;L^2(\Omega))$ and $v_0 \in H_0^1(\Omega)$, there exists a unique $v\in H(I;\Omega)$ such that*

$$\begin{aligned} L_\omega v &= g && \text{a.e. in } \Omega,\ \text{for a.e. } t\in I,\\ v(0) &= v_0 && \text{in } \Omega. \end{aligned} \tag{2.17}$$

Moreover, the function v satisfies

$$\|v\|_{H(I;\Omega)}^2 + \omega\|v(T)\|_{H_0^1}^2 \le \|g\|_{L^2(I;L^2(\Omega))}^2 + \omega\|v_0\|_{H_0^1}^2. \tag{2.18}$$

In Theorem 3, well-posedness of (2.17) is simply a special case of the general theory of Galerkin's method for parabolic equations, see [30]. The bound (2.18) is obtained by combining (2.14), integration by parts and the Miranda–Talenti Inequality.

Theorem 4 *Let $\Omega\subset\mathbb{R}^d$ be a bounded convex domain, let $I = (0,T)$, and let Λ be a compact metric space. Let the data a, b, c and f be continuous on $\overline{\Omega}\times\overline{I}\times\Lambda$ and satisfy* (1.3) *and* (2.6)*, or alternatively* (2.5) *in the case where $b\equiv 0$ and $c\equiv 0$. Then, there exists a unique strong solution $u\in H(I;\Omega)$ of the HJB equation* (2.3)*. Moreover, u is also the unique solution of $F_\gamma[u] = 0$ in $\Omega\times I$, $u = 0$ on $\partial\Omega\times I$ and $u = u_0$ on $\Omega\times\{0\}$.*

Proof The proof consists of establishing the equivalence of (2.3) with the problem of solving the equation $F_\gamma[u] = 0$ and $u(0) = u_0$, which can be analysed with the Browder–Minty Theorem. Let the operator $\mathcal{A}\colon H(I;\Omega) \to H(I;\Omega)^*$ be defined by

$$\langle \mathcal{A}(u), v\rangle := \int_I \int_\Omega F_\gamma[u]\, L_\omega v \,\mathrm{d}x\,\mathrm{d}t + \omega\langle u(0) - u_0, v(0)\rangle_{H_0^1}. \tag{2.19}$$

Compactness of Λ and continuity of the data imply that $\mathcal{A}$ is Lipschitz continuous. Indeed, letting u, v and $z \in H(I;\Omega)$, we find that

$$\begin{aligned} |\langle \mathcal{A}(u) - \mathcal{A}(v), z\rangle| &\leq \|F_\gamma[u] - F_\gamma[v]\|_{L^2(I;L^2(\Omega))}\|L_\omega z\|_{L^2(I;L^2(\Omega))} \\ &\quad + \omega\|u(0) - v(0)\|_{H_0^1}\|z(0)\|_{H_0^1} \leq C\|u - v\|_{H(I;\Omega)}\|z\|_{H(I;\Omega)}, \end{aligned} \tag{2.20}$$

where the constant C depends only on the dimension d, ω, T, and on the supremum norms of a, b, c and f and γ over $\overline{\Omega} \times \overline{I} \times \Lambda$. We also claim that $\mathcal{A}$ is strongly monotone. Define $w := u - v$. Addition and subtraction of $\int_{I_n} \langle L_\omega w, L_\omega w\rangle_{L^2}\,\mathrm{d}t$ shows that

$$\begin{aligned} \langle \mathcal{A}(u) - \mathcal{A}(v), w\rangle = \|L_\omega w\|^2_{L^2(I;L^2(\Omega))} + \omega\|w(0)\|^2_{H_0^1} \\ + \int_I \int_\Omega \left(F_\gamma[u] - F_\gamma[v] - L_\omega w\right) L_\omega w \,\mathrm{d}x\,\mathrm{d}t. \end{aligned}$$

Lemma 1, the bound (2.18) and the Cauchy–Schwarz Inequality show that

$$\begin{aligned} \langle \mathcal{A}(u) - \mathcal{A}(v), w\rangle &\geq \frac{1}{2}\|L_\omega w\|^2_{L^2(I;L^2(\Omega))} + \omega\|w(0)\|^2_{H_0^1} - \frac{1-\varepsilon}{2}\|w\|^2_{H(I;\Omega)} \\ &\geq \frac{\varepsilon}{2}\|w\|^2_{H(I;\Omega)} + \frac{\omega}{2}\|w(T)\|^2_{H_0^1} + \frac{\omega}{2}\|w(0)\|^2_{H_0^1}. \end{aligned} \tag{2.21}$$

The inequalities (2.20) and (2.21) imply that $\mathcal{A}$ is a bounded, continuous, coercive and strongly monotone operator, so the Browder–Minty Theorem [24] shows that there exists a unique $u \in H(I;\Omega)$ such that $\mathcal{A}(u) = 0$.

Theorem 3 shows that for each $g \in L^2(I;L^2(\Omega))$, there exists a $v \in H(I;\Omega)$ such that $L_\omega v = g$ and $v(0) = 0$. So, $\mathcal{A}(u) = 0$ implies that $\int_I \int_\Omega F_\gamma[u]\, g \,\mathrm{d}x\,\mathrm{d}t = 0$ for all $g \in L^2(I;L^2(\Omega))$, and since $F_\gamma[u] \in L^2(I;L^2(\Omega))$, we obtain $F_\gamma[u] = 0$. Theorem 3 also shows that $\langle u(0), v\rangle_{H_0^1} = \langle u_0, v\rangle_{H_0^1}$ for all $v \in H_0^1(\Omega)$, hence $u(0) = u_0$.

We claim that $u \in H(I;\Omega)$ solves $F_\gamma[u] = 0$ with $u(0) = u_0$ if and only if u solves (2.3). Since γ^α is positive, $\gamma^\alpha(\partial_t u - L^\alpha u + f^\alpha) \geq 0$ for all $\alpha \in \Lambda$ is equivalent to $\partial_t u - L^\alpha u + f^\alpha \geq 0$ for all $\alpha \in \Lambda$, so $F_\gamma[u] \geq 0$ is equivalent to $F[u] \geq 0$. Compactness of Λ and continuity of the data imply that for a.e. $t \in I$, for a.e. point of Ω, the extrema in the definitions of $F_\gamma[u]$ and $F[u]$ are attained by some elements of Λ, thereby giving $F_\gamma[u] \leq 0$ if and only if $F[u] \leq 0$. Therefore, existence and uniqueness in $H(I;\Omega)$ of a solution of $F_\gamma[u] = 0$ is equivalent to existence and uniqueness of a solution of (2.3). □

3 Temporal semi-discretisation

In this section, we explore some of the general principles underlying the numerical scheme for the parabolic problem (2.3). Before presenting the fully discrete scheme in section 5, we briefly consider in this section the temporal semi-discretisation of parabolic HJB equations, so as to highlight some key ideas in the derivation and analysis of a stable method. The fully discrete scheme will then combine these ideas with the methods from [27] used to discretise space.

The proof of Theorem 4 indicates that we should discretise the operator appearing in (2.19), and find stability in a norm that is analogous to $\|\cdot\|_{H(I;\Omega)}$ from (2.16). Although (2.19) expresses the global space-time problem, we will employ a temporal discontinuous Galerkin method, thus leading to a time-stepping scheme.

Let $\{\mathcal{J}_\tau\}_\tau$ be a sequence of partitions of $(0,T)$ into half-intervals $I_n := (t_{n-1}, t_n] \in \mathcal{J}_\tau$, with $1 \leq n \leq N = N(\tau)$. We say that $\mathcal{J}_\tau$ is regular provided that

$$[0,T] = \bigcup_{I_n \in \mathcal{J}_\tau} \overline{I_n}, \quad 0 = t_0 \leq t_{n-1} < t_n \leq t_N = T, \quad \forall\, n \leq N,\ \forall\, \tau. \tag{3.1}$$

For each interval $I_n \in \mathcal{J}_\tau$, let $\tau_n := |t_n - t_{n-1}|$. It is assumed that $\tau = \max_{1 \leq n \leq N} \tau_n$. For each τ, let $\mathbf{q} = (q_1, \dots, q_N)$ be a vector of positive integers, so $q_n \geq 1$ for all $I_n \in \mathcal{J}_\tau$. For a vector space V and $I_n \in \mathcal{J}_\tau$, let $\mathcal{Q}_{q_n}(V)$ denote the space of V-valued univariate polynomials of degree at most q_n. Recalling that $H := H^2(\Omega) \cap H^1_0(\Omega)$, we define the semi-discrete DG finite element space $V^{\tau,\mathbf{q}}$ by

$$V^{\tau,\mathbf{q}} := \left\{ v \in L^2(I;H),\ v|_{I_n} \in \mathcal{Q}_{q_n}(H)\ \ \forall\, I_n \in \mathcal{J}_\tau \right\}. \tag{3.2}$$

Functions from $V^{\tau,\mathbf{q}}$ are taken to be left-continuous, but are generally discontinuous at the partition points $\{t_n\}_{n=1}^{N-1}$. We denote the right-limit of $v \in V^{\tau,\mathbf{q}}$ at t_n by $v(t_n^+)$, where $0 \leq n < N$. The jump operators $(\!|\cdot|\!)_n$ and average operators $\langle\cdot\rangle_n$, $0 \leq n \leq N$, are defined by

$$\begin{aligned}
&(\!|v|\!)_n := -v(0^+), && \langle v\rangle_n := v(0^+), && \text{if } n = 0,\\
&(\!|v|\!)_n := v(t_n) - v(t_n^+), && \langle v\rangle_n := \tfrac{1}{2}v(t_n) + \tfrac{1}{2}v(t_n^+), && \text{if } 1 \leq n < N,\\
&(\!|v|\!)_n := v(T), && \langle v\rangle_n := v(T), && \text{if } n = N.
\end{aligned} \tag{3.3}$$

Define the nonlinear form $A_\tau\colon V^{\tau,\mathbf{q}} \times V^{\tau,\mathbf{q}} \to \mathbb{R}$ by

$$\begin{aligned}
A_\tau(u_\tau; v_\tau) := \sum_{n=1}^{N} \int_{I_n} \langle F_\gamma[u_\tau], L_\omega v_\tau\rangle_{L^2(\Omega)}\,\mathrm{d}t \\
- \omega \sum_{n=0}^{N-1} \langle (\!|u_\tau|\!)_n, \langle v_\tau\rangle_n\rangle_{H^1_0} + \frac{\omega}{2}\sum_{n=1}^{N-1} \langle (\!|u_\tau|\!)_n, (\!|v_\tau|\!)_n\rangle_{H^1_0}.
\end{aligned} \tag{3.4}$$

We note that $\frac{1}{2}(\!|v|\!)_n - \langle v\rangle_n = v(t_n^+)$ for $1 \leq n < N$. The semi-discrete scheme consists of finding a $u_\tau \in V^{\tau,\mathbf{q}}$ such that

$$A_\tau(u_\tau; v_\tau) = \omega\langle u_0, v_\tau(0^+)\rangle_{H^1_0} \qquad \forall\, v_\tau \in V^{\tau,\mathbf{q}}. \tag{3.5}$$

Since the solution $u \in H(I;\Omega)$ of (2.3) belongs to $C(\overline{I}; H^1_0(\Omega))$, it is clear that $A_\tau(u; v_\tau) = \omega\langle u_0, v_\tau(0^+)\rangle_{H^1_0}$ for all $v_\tau \in V^{\tau,\mathbf{q}}$, so the scheme is consistent. By considering test functions v_τ that have support on successive intervals $\overline{I_n} \in \mathcal{J}_\tau$, it is easily seen that $u_\tau|_{I_n}$ is determined only by the data and by $u(t_{n-1})$, thus (3.5) is a time-stepping scheme. The main ingredients required to show that the above scheme is stable are as follows. We introduce the bilinear form $C_\tau\colon V^{\tau,\mathbf{q}} \times V^{\tau,\mathbf{q}} \to \mathbb{R}$ defined by

$$\begin{aligned}
C_\tau(u_\tau, v_\tau) := \sum_{n=1}^{N} \int_{I_n} \langle L_\omega u_\tau, L_\omega v_\tau\rangle_{L^2(\Omega)}\,\mathrm{d}t \\
- \omega \sum_{n=0}^{N-1} \langle (\!|u_\tau|\!)_n, \langle v_\tau\rangle_n\rangle_{H^1_0} + \frac{\omega}{2}\sum_{n=1}^{N-1} \langle (\!|u_\tau|\!)_n, (\!|v_\tau|\!)_n\rangle_{H^1_0}.
\end{aligned} \tag{3.6}$$

Integration by parts shows that for any $u_\tau,\, v_\tau \in V^{\tau,\mathbf{q}}$, we have

$$C_\tau(u_\tau, v_\tau) = \sum_{n=1}^{N} \int_{I_n} \omega^2 \langle \partial_t u_\tau, \partial_t v_\tau \rangle_{L^2(\Omega)} + \langle L_\lambda u_\tau, L_\lambda v_\tau \rangle_{L^2(\Omega)} \,\mathrm{d}t \\ + \omega \sum_{n=1}^{N} \langle \langle u_\tau \rangle_n \,, \llparenthesis v_\tau \rrparenthesis_n \rangle_{H_0^1} + \frac{\omega}{2} \sum_{n=1}^{N-1} \langle \llparenthesis u_\tau \rrparenthesis_n, \llparenthesis v_\tau \rrparenthesis_n \rangle_{H_0^1}. \quad (3.7)$$

Combining (3.6) and (3.7) reveals the stability properties of C_τ when re-written as

$$C_\tau(u_\tau, v_\tau) = \frac{1}{2} \sum_{n=1}^{N} \int_{I_n} \omega^2 \langle \partial_t u_\tau, \partial_t v_\tau \rangle_{L^2} + \langle L_\lambda u_\tau, L_\lambda v_\tau \rangle_{L^2} + \langle L_\omega u_\tau, L_\omega v_\tau \rangle_{L^2} \,\mathrm{d}t \\ + \frac{\omega}{2} \sum_{n=1}^{N} \langle \langle u_\tau \rangle_n \,, \llparenthesis v_\tau \rrparenthesis_n \rangle_{H_0^1} - \frac{\omega}{2} \sum_{n=0}^{N-1} \langle \llparenthesis u_\tau \rrparenthesis_n, \langle v_\tau \rangle_n \rangle_{H_0^1} + \frac{\omega}{2} \sum_{n=1}^{N-1} \langle \llparenthesis u_\tau \rrparenthesis_n, \llparenthesis v_\tau \rrparenthesis_n \rangle_{H_0^1}. \quad (3.8)$$

Indeed, it follows from (3.8) and the Miranda–Talenti Inequality that, for any $u_\tau \in V^{\tau,\mathbf{q}}$,

$$C_\tau(u_\tau, u_\tau) \geq \frac{1}{2} \sum_{n=1}^{N} \int_{I_n} \omega^2 \|\partial_t u\|_{L^2(\Omega)}^2 + \|u_\tau\|_H^2 + \|L_\omega u_\tau\|_{L^2(\Omega)}^2 \mathrm{d}t \\ + \frac{\omega}{2} \|u_\tau(T)\|_{H_0^1}^2 + \frac{\omega}{2} \|u_\tau(0^+)\|_{H_0^1}^2 + \frac{\omega}{2} \sum_{n=1}^{N-1} \|\llparenthesis u_\tau \rrparenthesis_n\|_{H_0^1}^2. \quad (3.9)$$

The key observation here is that the antisymmetric terms in (3.8) cancel in $C_\tau(u_\tau, u_\tau)$, and this technique will be used again in section 6 for the analysis of stability of the fully discrete scheme. The above considerations imply stability of the scheme as follows: (3.6) implies that

$$A_\tau(u_\tau; v_\tau) = \sum_{n=1}^{N} \int_{I_n} \langle F_\gamma[u_\tau] - L_\omega u_\tau, L_\omega v_\tau \rangle_{L^2(\Omega)} \,\mathrm{d}t + C_\tau(u_\tau, v_\tau) \quad \forall\, u_\tau,\, v_\tau \in V^{\tau,\mathbf{q}};$$

which mirrors the addition-subtraction step of the proof of Theorem 4. Then, we use (3.9) to show that A_τ is strongly monotone: for any $u_\tau,\, v_\tau \in V^{\tau,\mathbf{q}}$, $w_\tau := u_\tau - v_\tau$, we have

$$A_\tau(u_\tau; w_\tau) - A_\tau(v_\tau; w_\tau) \geq \frac{\varepsilon}{2} \sum_{n=1}^{N} \int_{I_n} \omega^2 \|\partial_t w_\tau\|_{L^2(\Omega)}^2 + \|w_\tau\|_H^2 \mathrm{d}t + \frac{\omega}{2} \sum_{n=0}^{N} \|\llparenthesis w_\tau \rrparenthesis_n\|_{H_0^1}^2.$$

Therefore, the well-posedness of the semi-discrete scheme can be shown by an induction argument, based on the Browder–Minty Theorem, that is similar to the one given in the proof of Theorem 10 below, concerning the well-posedness of the fully discrete scheme. Instead of pursuing the analysis of the semi-discrete scheme further, we now turn towards the fully discrete method.

4 Finite element spaces

Let $\{\mathcal{T}_h\}_h$ be a sequence of shape-regular meshes on Ω, such that each element $K \in \mathcal{T}_h$ is a simplex or a parallelepiped. Let $h_K := \operatorname{diam} K$ for each $K \in \mathcal{T}_h$. 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 $\mathcal{F}_h^b$ denote the set of boundary faces. The set of all faces of $\mathcal{T}_h$ is denoted by $\mathcal{F}_h^{i,b} := \mathcal{F}_h^i \cup \mathcal{F}_h^b$. Since each element has piecewise flat boundary, the faces may be chosen to be flat.

Mesh conditions 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

$$\max_{K\in\mathcal{T}_h} \operatorname{card}\{F \in \mathcal{F}_h^{i,b} \colon F \subset \partial K\} \leq c_{\mathcal{F}}. \tag{4.1}$$

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, for any K, K' that share a face,

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

For each h, let $\mathbf{p} = (p_K;\ K \in \mathcal{T}_h)$ be a vector of positive integers, such that there is a $c_{\mathcal{P}} \geq 1$, independent of h, such that, for any K, K' that share a face,

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

Function spaces For each $K \in \mathcal{T}_h$, let $\mathcal{P}_{p_K}$ be the space of all real-valued polynomials in $\mathbb{R}^d$ with either total or partial degree at most p_K. In particular, we allow the combination of spaces of polynomials of fixed total degree on some parts of the mesh with spaces of polynomials of fixed partial degree on the remainder. We also allow the use of the space of polynomials of total degree at most p_K even when K is a parallelepiped. The spatial discontinuous Galerkin finite element space $V_{h,\mathbf{p}}$ is defined by

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

For $\mathcal{J}_\tau$ a regular partition of I, the space-time discontinuous Galerkin finite element space $V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ is defined by

$$V_{h,\mathbf{p}}^{\tau,\mathbf{q}} := \left\{ v \in L^2\left(I; V_{h,\mathbf{p}}\right),\ v|_{I_n} \in \mathcal{Q}_{q_n}(V_{h,\mathbf{p}}) \ \ \forall I_n \in \mathcal{J}_\tau \right\}. \tag{4.5}$$

As in section 3, we take functions from $V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ to be left-continuous. The support of a function $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$, denoted by $\operatorname{supp} v_h$, is a subset of $\overline{I}$, and is understood to be the support of $v_h \colon I \to V_{h,\mathbf{p}}$, i.e. when viewing v_h as a mapping from I into $V_{h,\mathbf{p}}$.

For $\mathbf{s} := (s_K \colon K \in \mathcal{T}_h)$ a vector of nonnegative real numbers, and $r \in [1,\infty]$, define the broken Sobolev space $W_r^{\mathbf{s}}(\Omega;\mathcal{T}_h) := \{v \in L^r(\Omega),\ v|_K \in W_r^{s_K}(K)\ \forall K \in \mathcal{T}_h\}$. For shorthand, define $H^{\mathbf{s}}(\Omega;\mathcal{T}_h) := W_2^{\mathbf{s}}(\Omega;\mathcal{T}_h)$, and, for $s \geq 0$, set $W_r^s(\Omega;\mathcal{T}_h) := W_r^{\mathbf{s}}(\Omega;\mathcal{T}_h)$, where $s_K = s$ for all $K \in \mathcal{T}_h$. Define the norm $\|\cdot\|_{W_r^{\mathbf{s}}(\Omega;\mathcal{T}_h)}$ on $W_r^{\mathbf{s}}(\Omega;\mathcal{T}_h)$ by $\|v\|_{W_r^{\mathbf{s}}(\Omega;\mathcal{T}_h)}^r := \sum_{K\in\mathcal{T}_h} \|v\|_{W_r^{s_K}(K)}^r$, with the usual modification when $r = \infty$.

Spatial jump, average, and tangential operators For each face F, let $n_F \in \mathbb{R}^d$ 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 $\tau_F \colon H^s(K) \to H^{s-1/2}(F)$, $s > 1/2$, 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

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

where ϕ is a sufficiently regular scalar or vector-valued function, and K_{ext} and K_{int} are the elements to which F is a face, i.e. $F = \partial K_{\mathrm{ext}} \cap \partial K_{\mathrm{int}}$. Here, the labelling 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, resp. vector-valued, are scalar-valued, resp. vector-valued. For a face F, let ∇_{T} and $\operatorname{div}_{\mathrm{T}}$ denote respectively the tangential gradient and tangential divergence operators on F; see [13,26] for further details.

5 Numerical Scheme

The definition of the numerical scheme requires the following bilinear forms, which were first introduced in the analysis of elliptic HJB equations in [27]. First, for $\lambda \geq 0$ as in section 2, the symmetric bilinear form $B_{h,*}\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ is defined by

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

Then, for face-dependent quantities $\mu_F > 0$ and $\eta_F > 0$, to be specified later, let the jump stabilisation term $J_h\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ be defined by

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

Recalling that $L_\lambda v \coloneqq \Delta v - \lambda v$, we introduce the one-parameter family of bilinear forms $B_{h,\theta}\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$, where $\theta \in [0,1]$, defined by

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

Define the bilinear form $a_h\colon V_{h,\mathbf{p}} \times V_{h,\mathbf{p}} \to \mathbb{R}$ by

$$\begin{aligned} a_h(u_h, v_h) \coloneqq {} & \sum_{K\in\mathcal{T}_h} \langle \nabla u_h, \nabla v_h\rangle_K + \lambda\langle u_h, v_h\rangle_K - \sum_{F\in\mathcal{F}_h^{i,b}} \langle \{\nabla u_h\cdot n_F\}, [\![v_h]\!]\rangle_F \\ & - \sum_{F\in\mathcal{F}_h^{i,b}} \langle \{\nabla v_h\cdot n_F\}, [\![u_h]\!]\rangle_F + \sum_{F\in\mathcal{F}_h^{i,b}} \mu_F\langle [\![u_h]\!], [\![v_h]\!]\rangle_F. \end{aligned} \tag{5.3}$$

Observe that the bilinear form a_h corresponds precisely to the standard symmetric interior penalty discretisation of the operator $-L_\lambda$, and its symmetry plays an imporant role in the subsequent analysis.

Define the bilinear forms $C_h^{\mathcal{F}}$ and $C_h\colon V_{h,\mathbf{p}}^{\tau,\mathbf{q}} \times V_{h,\mathbf{p}}^{\tau,\mathbf{q}} \to \mathbb{R}$ by

$$\begin{aligned} C_h^{\mathcal{F}}(u_h,v_h) &:= \omega \sum_{n=1}^{N} \int_{I_n} \sum_{F\in\mathcal{F}_h^i} \langle [\![\nabla u_h\cdot n_F]\!], \{\partial_t v_h\}\rangle_F \,\mathrm{d}t \\ &\quad + \omega \sum_{n=1}^{N} \int_{I_n} \sum_{F\in\mathcal{F}_h^{i,b}} \left[\mu_F \langle [\![u_h]\!], [\![\partial_t v_h]\!]\rangle_F - \langle [\![u_h]\!], \{\nabla \partial_t v_h\cdot n_F\}\rangle_F\right] \mathrm{d}t, \end{aligned} \tag{5.4}$$

$$\begin{aligned} C_h(u_h,v_h) &:= \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \langle L_\omega u_h, L_\omega v_h\rangle_K \,\mathrm{d}t + C_h^{\mathcal{F}}(u_h,v_h) \\ &\quad + \sum_{n=1}^{N} \int_{I_n} B_{h,1/2}(u_h,v_h) - \sum_{K\in\mathcal{T}_h} \langle L_\lambda u_h, L_\lambda v_h\rangle_K \,\mathrm{d}t \\ &\quad - \omega \sum_{n=0}^{N-1} a_h((\!|u_h|\!)_n, \langle v_h\rangle_n) + \frac{\omega}{2} \sum_{n=1}^{N-1} a_h((\!|u_h|\!)_n, (\!|v_h|\!)_n). \end{aligned} \tag{5.5}$$

Define the nonlinear form $A_h\colon V_{h,\mathbf{p}}^{\tau,\mathbf{q}} \times V_{h,\mathbf{p}}^{\tau,\mathbf{q}} \to \mathbb{R}$ by

$$A_h(u_h;v_h) := \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \left[\langle F_\gamma[u_h], L_\omega v_h\rangle_K - \langle L_\omega u_h, L_\omega v_h\rangle_K\right] \mathrm{d}t + C_h(u_h,v_h). \tag{5.6}$$

The form A_h is linear in its second argument, but it is nonlinear in its first argument. Supposing that u_0 is sufficiently regular, such as $u_0 \in H^s(\Omega;\mathcal{T}_h)$, with $s > 3/2$, the numerical scheme is to find $u_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ such that

$$A_h(u_h;v_h) = \omega\, a_h(u_0, v_h(0^+)) \qquad \forall\, v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}. \tag{5.7}$$

If u_0 fails to be sufficiently regular, we can replace u_0 in the right-hand side of (5.7) with a suitable projection into $V_{h,\mathbf{p}}$, at the expense of introducing a consistency error that vanishes in the limit. By testing with functions $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ that are supported on $\overline{I_n}$, it is found that (5.7) is equivalent to finding $u_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ such that

$$\begin{aligned} &\int_{I_n} \sum_{K\in\mathcal{T}_h} \langle F_\gamma[u_h], L_\omega v_h\rangle_K + B_{h,1/2}(u_h,v_h) - \sum_{K\in\mathcal{T}_h} \langle L_\lambda u_h, L_\lambda v_h\rangle_K \,\mathrm{d}t \\ &\qquad + \omega \int_{I_n} \sum_{F\in\mathcal{F}_h^i} \langle [\![\nabla u_h\cdot n_F]\!], \{\partial_t v_h\}\rangle_F + \sum_{F\in\mathcal{F}_h^{i,b}} \mu_F \langle [\![u_h]\!], [\![\partial_t v_h]\!]\rangle_F \,\mathrm{d}t \\ &\qquad - \omega \int_{I_n} \sum_{F\in\mathcal{F}_h^{i,b}} \langle [\![u_h]\!], \{\nabla\partial_t v_h\cdot n_F\}\rangle_F \,\mathrm{d}t + \omega\, a_h(u_h(t_{n-1}^+), v_h(t_{n-1}^+)) \\ &\qquad\qquad = \omega\, a_h(u_h(t_{n-1}), v_h(t_{n-1}^+)), \end{aligned} \tag{5.8}$$

for all $v_h \in \mathcal{Q}_{q_n}(V_{h,\mathbf{p}})$, with the convention $u_h(t_0) := u_0$. Therefore, (5.7) defines a time-stepping scheme, and in practice it is (5.8) that is used for computations.

Consistency The following result is shown in [26,27].

Lemma 5 *Let Ω be a bounded Lipschitz polytopal domain and let $\mathcal{T}_h$ be a simplicial or parallelepipedal mesh on Ω. Let $w \in H^s(\Omega;\mathcal{T}_h) \cap H^2(\Omega) \cap H^1_0(\Omega)$, with $s > 5/2$. Then, for every $v_h \in V_{h,\mathbf{p}}$, we have the identities*

$$B_{h,*}(w, v_h) = \sum_{K\in\mathcal{T}_h} \langle L_\lambda w, L_\lambda v_h\rangle_K \quad \text{and} \quad J_h(w, v_h) = 0. \tag{5.9}$$

Lemma 6 *Let Ω be a bounded Lipschitz polytopal domain, let $\mathcal{T}_h$ be a simplicial or parallelepipedal mesh on Ω. Let $I = (0,T)$ and let $\mathcal{J}_\tau = \{I_n\}_{n=1}^N$ be a regular partition of I. Suppose that $u_0 \in H^1_0(\Omega) \cap H^r(\Omega;\mathcal{T}_h)$ with $r > 3/2$. Then, for any $w \in H(I;\Omega) \cap L^2(I;H^s(\Omega;\mathcal{T}_h))$, with $s > 5/2$, such that $w(0) = u_0$, we have*

$$C_h(w, v_h) = \sum_{n=1}^N \int_{I_n} \sum_{K\in\mathcal{T}_h} \langle L_\omega w, L_\omega v_h\rangle_K \,\mathrm{d}t + \omega\, a_h(u_0, v_h(0^+)) \qquad \forall\, v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}. \tag{5.10}$$

Proof Let the function w be as above, so that $w(t) \in H^2(\Omega)\cap H^1_0(\Omega)\cap H^s(\Omega;\mathcal{T}_h)$ for a.e. $t \in I$. Lemma 5 shows that $\int_{I_n} B_{h,1/2}(w, v_h)\,\mathrm{d}t = \int_{I_n} \sum_{K\in\mathcal{T}_h} \langle L_\lambda w, L_\lambda v_h\rangle_K \,\mathrm{d}t$ for all $I_n \in \mathcal{J}_\tau$ and all $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$. The spatial regularity of w also implies that $[\![\nabla w(t)\cdot n_F]\!]$ vanishes for all $F \in \mathcal{F}_h^i$ and a.e. $t \in I$, whilst $[\![w(t)]\!]$ and $[\![\nabla_{\mathrm{T}}\, w(t)]\!]$ vanish for all $F \in \mathcal{F}_h^{i,b}$ and a.e. $t \in I$. Therefore we have $C_h^{\mathcal{F}}(w, v) = 0$ for all $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$. Finally, since $H(I;\Omega) \hookrightarrow C(\overline{I};H^1_0(\Omega))$ by Lemma 2, the jump $(\!|w|\!)_n = 0$ for each $0 < n < N$, and thus $a_h((\!|w|\!)_n, v_h) = 0$ for all $v_h \in V_{h,\mathbf{p}}$, $0 < n < N$. The above identities and the definition of C_h in (5.5) imply (5.10). □

Lemma 6 and the definition of the nonlinear form A_h in (5.6) immediately imply the following consistency result for the numerical scheme.

Corollary 7 *Under the hypotheses of Lemma 6, suppose that the solution $u \in H(I;\Omega)$ of* (2.3) *belongs to $L^2(I;H^s(\Omega;\mathcal{T}_h))$, with $s > 5/2$. Then, u satisfies*

$$A_h(u; v_h) = \omega\, a_h(u_0, v_h(0^+)) \qquad \forall\, v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}. \tag{5.11}$$

6 Stability

It will be seen below that, for μ_F appropriately chosen, the symmetric bilinear form a_h is coercive on $V_{h,\mathbf{p}}$, and thus defines an inner-product on $V_{h,\mathbf{p}}$, with associated norm $\|v_h\|^2_{a_h} := a_h(v_h, v_h)$ for $v_h \in V_{h,\mathbf{p}}$. Define the functionals

$$|v_h|^2_{H^2(K),\lambda} := |v_h|^2_{H^2(K)} + 2\lambda |v_h|^2_{H^1(K)} + \lambda^2 \|v_h\|^2_{L^2(K)}, \quad v_h \in V_{h,\mathbf{p}},\ K \in \mathcal{T}_h, \tag{6.1}$$

$$|v_h|^2_{\mathrm{J}} := J_h(v_h, v_h), \qquad v_h \in V_{h,\mathbf{p}}. \tag{6.2}$$

For each $\theta \in [0,1]$, we introduce the functional $\|\cdot\|_{h,\theta} \colon V_{h,\mathbf{p}}^{\tau,\mathbf{q}} \to \mathbb{R}$ defined by

$$\begin{aligned}\|v_h\|^2_{h,\theta} := \sum_{n=1}^N \int_{I_n} \sum_{K\in\mathcal{T}_h} \theta \left[\omega^2 \|\partial_t v_h\|^2_{L^2(K)} + |v_h|^2_{H^2(K),\lambda}\right] + |v_h|^2_{\mathrm{J}}\,\mathrm{d}t \\ + \sum_{n=1}^N \int_{I_n} \sum_{K\in\mathcal{T}_h} (1-\theta)\, \|L_\omega v_h\|^2_{L^2(K)} \mathrm{d}t + \omega \sum_{n=0}^N \|(\!|v_h|\!)_n\|^2_{a_h}.\end{aligned} \tag{6.3}$$

It is shown below that, for an appropriate choice of μ_F, $\|\cdot\|_{h,\theta}$ defines a norm on $V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ for each $\theta \in [0,1]$. For each face $F \in \mathcal{F}_h^{i,b}$, define

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

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 following result is from [27, Lemma 6].

Lemma 8 *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_{s}, 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*

$$B_{h,\theta}(v_h, v_h) \geq \sum_{K\in\mathcal{T}_h} \left[\frac{\theta}{\kappa} |v_h|^2_{H^2(K),\lambda} + (1-\theta) \|L_\lambda v_h\|^2_{L^2(K)} \right] + \frac{1}{2} |v_h|^2_{\mathrm{J}}, \tag{6.5}$$

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

$$\mu_F = \sigma c_{\mathrm{s}} \frac{\tilde{p}_F^2}{\tilde{h}_F} \quad \textit{and} \quad \eta_F > \sigma\lambda\, c_{\mathrm{s}} \frac{\tilde{p}_F^2}{\tilde{h}_F}. \tag{6.6}$$

We note that μ_F may be chosen as in Lemma 8 whilst also guaranteeing the standard discrete Poincaré Inequality:

$$\sum_{K\in\mathcal{T}_h} \|v_h\|^2_{H^1(K)} + \sum_{F\in\mathcal{F}_h^{i,b}} \mu_F \|[\![v_h]\!]\|^2_{L^2(F)} \lesssim a_h(v_h, v_h) = \|v_h\|^2_{a_h} \qquad \forall\, v_h \in V_{h,\mathbf{p}}. \tag{6.7}$$

In the subsequent analysis, we shall choose μ_F and η_F to be given by

$$\mu_F := \sigma\, c_{\mathrm{s}} \frac{\tilde{p}_F^2}{\tilde{h}_F}, \qquad \eta_F := \sigma \max(1,\lambda)\, c_{\mathrm{s}} \frac{\tilde{p}_F^6}{\tilde{h}_F^3}, \tag{6.8}$$

where c_{s} is chosen so that Lemma 8 holds for $\kappa < (1-\varepsilon)^{-1}$, and where $\sigma \geq 1$ is a fixed constant chosen such that (6.7) also holds. Note that these orders of penalisation are the strongest that remain consistent with the discrete H^2-type norm appearing in the analysis of this work; see [23] for an example of a scheme for the biharmonic equation using the same penalisation orders.

To verify that the functional $\|\cdot\|_{h,\theta}$ defines a norm on $V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$, suppose that $\|v_h\|_{h,\theta} = 0$ for some $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$. Then, the jumps of v_h vanish across the mesh faces and across time intervals and, therefore, $v_h \in H(I;\Omega)$ with $v_h(0) = 0$. The fact that the volume terms in $\|v_h\|_{h,\theta}$ also vanish shows that $L_\omega v_h = 0$, so it follows from (2.18) that $v_h \equiv 0$. Hence, the functional $\|\cdot\|_{h,\theta}$ defines a norm on $V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$.

Lemma 9 *Under the hypotheses of Lemma 8, let $I = (0,T)$ and $\{\mathcal{J}_\tau\}_\tau$ be a sequence of regular partitions of I. Let μ_F and η_F satisfy* (6.8) *for each face F, so that Lemma 8 holds for a given $\kappa > 1$. Then, for every $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$, we have*

$$\begin{aligned} C_h(v_h, v_h) \geq \frac{1}{2} \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \omega^2 \|\partial_t v_h\|^2_{L^2(K)} + \frac{1}{\kappa} |v_h|^2_{H^2(K),\lambda} + |v_h|^2_{\mathrm{J}}\, \mathrm{d}t \\ + \frac{1}{2} \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \|L_\omega v_h\|^2_{L^2(K)} \mathrm{d}t + \frac{\omega}{2} \sum_{n=0}^{N} \|(\!|v_h|\!)_n\|^2_{a_h}. \end{aligned} \tag{6.9}$$

Proof We begin by showing that, for any $u_h,\, v_h \in V^{\tau,\mathbf{q}}_{h,\mathbf{p}}$, the bilinear form C_h satisfies the following identity:

$$
\begin{aligned}
C_h(u_h,v_h) = \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\omega^2\langle\partial_t u_h,\partial_t v_h\rangle_K + B_{h,1/2}(u_h,v_h)\,\mathrm{d}t - C_h^{\mathcal{F}}(v_h,u_h)\\
+\,\omega\sum_{n=1}^{N}a_h(\langle u_h\rangle_n,\llparenthesis v_h\rrparenthesis_n) + \frac{\omega}{2}\sum_{n=1}^{N-1}a_h(\llparenthesis u_h\rrparenthesis_n,\llparenthesis v_h\rrparenthesis_n). \qquad (6.10)
\end{aligned}
$$

The first step in deriving (6.10) is to show that for any $u_h,\, v_h \in V^{\tau,\mathbf{q}}_{h,\mathbf{p}}$, we have

$$
\begin{aligned}
&\sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\langle\omega\,\partial_t u_h,-L_\lambda v_h\rangle_K + \langle\omega\,\partial_t v_h,-L_\lambda u_h\rangle_K\,\mathrm{d}t\\
&= \omega\sum_{n=1}^{N}a_h(\langle u_h\rangle_n,\llparenthesis v_h\rrparenthesis_n) + \omega\sum_{n=0}^{N-1}a_h(\llparenthesis u_h\rrparenthesis_n,\langle v_h\rangle_n) - C_h^{\mathcal{F}}(u_h,v_h) - C_h^{\mathcal{F}}(v_h,u_h).
\end{aligned}
\qquad (6.11)
$$

Indeed, integration by parts over $\mathcal{T}_h$ shows that, for any $I_n\in\mathcal{J}_\tau$ and a.e. $t\in I_n$,

$$
\begin{aligned}
\sum_{K\in\mathcal{T}_h}\langle\omega\,\partial_t u_h,-L_\lambda v_h\rangle_K = \omega\sum_{K\in\mathcal{T}_h}\langle\nabla\partial_t u_h,\nabla v_h\rangle_K + \lambda\langle\partial_t u_h,v_h\rangle_K\\
-\,\omega\sum_{F\in\mathcal{F}_h^i}\langle\{\partial_t u_h\},[\![\nabla v_h\cdot n_F]\!]\rangle_F - \omega\sum_{F\in\mathcal{F}_h^{i,b}}\langle[\![\partial_t u_h]\!],\{\nabla v_h\cdot n_F\}\rangle_F. \qquad (6.12)
\end{aligned}
$$

Therefore, it is found that, for any $I_n\in\mathcal{J}_\tau$ and a.e. $t\in I_n$,

$$
\begin{aligned}
&\sum_{K\in\mathcal{T}_h}\langle\omega\,\partial_t u_h,-L_\lambda v_h\rangle_K + \langle\omega\,\partial_t v_h,-L_\lambda u_h\rangle_K\\
&\quad= \omega\frac{\mathrm{d}}{\mathrm{d}t}a_h(u_h,v_h) - \omega\sum_{F\in\mathcal{F}_h^{i,b}}\mu_F\left[\langle[\![\partial_t u_h]\!],[\![v_h]\!]\rangle_F + \langle[\![u_h]\!],[\![\partial_t v_h]\!]\rangle_F\right]\\
&\qquad-\omega\sum_{F\in\mathcal{F}_h^i}\left[\langle\{\partial_t u_h\},[\![\nabla v_h\cdot n_F]\!]\rangle_F + \langle\{\partial_t v_h\},[\![\nabla u_h\cdot n_F]\!]\rangle_F\right]\\
&\qquad\quad+\omega\sum_{F\in\mathcal{F}_h^{i,b}}\left[\langle[\![v_h]\!],\{\nabla\partial_t u_h\cdot n_F\}\rangle_F + \langle[\![u_h]\!],\{\nabla\partial_t v_h\cdot n_F\}\rangle_F\right]. \qquad (6.13)
\end{aligned}
$$

We obtain (6.11) upon integration and summation of (6.13) over all time intervals. So, we have

$$
\begin{aligned}
\sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\langle L_\omega u_h,L_\omega v_h\rangle_K\mathrm{d}t = \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\omega^2\langle\partial_t u_h,\partial_t v_h\rangle_K + \langle L_\lambda u_h,L_\lambda v_h\rangle_K\mathrm{d}t\\
+\,\omega\sum_{n=1}^{N}a_h(\langle u_h\rangle_n,\llparenthesis v_h\rrparenthesis_n) + \omega\sum_{n=0}^{N-1}a_h(\llparenthesis u_h\rrparenthesis_n,\langle v_h\rangle_n) - C_h^{\mathcal{F}}(u_h,v_h) - C_h^{\mathcal{F}}(v_h,u_h).
\end{aligned}
$$

The proof of (6.10) is then completed by substituting the above identity in the definition of C_h from (5.5). Expanding C_h with both (5.5) and (6.10) shows that

$$\begin{aligned} C_h(u_h,v_h) &= \frac{1}{2}\sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\omega^2\langle\partial_t u_h,\partial_t v_h\rangle_K + B_{h,1}(u_h,v_h) + J_h(u_h,v_h)\,\mathrm{d}t \\ &\quad+\frac{1}{2}\sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\langle L_\omega u_h, L_\omega v_h\rangle_K\,\mathrm{d}t + \frac{1}{2}C_h^{\mathcal{F}}(u_h,v_h) - \frac{1}{2}C_h^{\mathcal{F}}(v_h,u_h) \\ &\quad+\frac{\omega}{2}\sum_{n=1}^{N}a_h(\langle u_h\rangle_n, \llparenthesis v_h\rrparenthesis_n) - \frac{\omega}{2}\sum_{n=0}^{N-1}a_h(\llparenthesis u_h\rrparenthesis_n,\langle v_h\rangle_n) + \frac{\omega}{2}\sum_{n=1}^{N-1}a_h(\llparenthesis u_h\rrparenthesis_n,\llparenthesis v_h\rrparenthesis_n). \end{aligned} \tag{6.14}$$

Note that to get (6.14), we have used the identity

$$B_{h,1/2}(u_h,v_h) - \frac{1}{2}\sum_{K\in\mathcal{T}_h}\langle L_\lambda u_h, L_\lambda v_h\rangle_K = \frac{1}{2}B_{h,1}(u_h,v_h) + \frac{1}{2}J_h(u_h,v_h).$$

To show (6.9), we substitute $u_h = v_h$ in (6.14) and first observe that the flux terms involving $C_h^{\mathcal{F}}$ cancel. Furthermore, the symmetry of the bilinear form a_h implies that

$$\begin{aligned} \sum_{n=1}^{N}a_h(\langle v_h\rangle_n, \llparenthesis v_h\rrparenthesis_n) &- \sum_{n=0}^{N-1}a_h(\llparenthesis v_h\rrparenthesis_n,\langle v_h\rangle_n) + \sum_{n=1}^{N-1}\|\llparenthesis v_h\rrparenthesis_n\|_{a_h}^2 \\ &= a_h(v_h(T),v_h(T)) + a_h(v_h(0^+),v_h(0^+)) + \sum_{n=1}^{N-1}\|\llparenthesis v_h\rrparenthesis_n\|_{a_h}^2 = \sum_{n=0}^{N}\|\llparenthesis v_h\rrparenthesis_n\|_{a_h}^2. \end{aligned}$$

Then, we apply Lemma 8 for $\theta = 1$ to get $B_{h,1}(v_h,v_h) \geq \kappa^{-1}\sum_{K\in\mathcal{T}_h}|v_h|_{H^2(K),\lambda}^2$, thereby yielding (6.9). □

Recall that for a function $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$, the support of v_h is a subset of $\overline{I}$, since v_h is viewed as a mapping from I into $V_{h,\mathbf{p}}$.

Theorem 10 *Let Ω be a bounded convex polytopal domain and let $\{\mathcal{T}_h\}_h$ be a shape-regular sequence of meshes satisfying* (4.1)*. Let $I = (0,T)$ and let $\{\mathcal{J}_\tau\}_\tau$ be a sequence of regular partitions of I. Let Λ be a compact metric space and let the data a, b, c and f be continuous on $\overline{\Omega}\times\overline{I}\times\Lambda$ and satisfy* (1.3) *and* (2.6)*, or alternatively* (2.5) *in the case where $b\equiv 0$ and $c \equiv 0$. Assume that the initial data $u_0 \in H_0^1(\Omega)\cap H^s(\Omega;\mathcal{T}_h)$ with $s > 3/2$. Let μ_F and η_F satisfy* (6.8)*, with c_{s} chosen so that Lemmas* 8 *and* 9 *hold with $\kappa < (1-\varepsilon)^{-1}$. Then, for every z_h, $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$, we have*

$$\|z_h - v_h\|_{h,1}^2 \leq \frac{2\,\kappa}{1-\kappa\,(1-\varepsilon)}\left(A_h(z_h; z_h - v_h) - A_h(v_h; z_h - v_h)\right). \tag{6.15}$$

Moreover, A_h is interval-wise Lipschitz continuous, in the sense that, for any $I_n \in \mathcal{J}_\tau$ and any u_h, v_h and $z_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ with support contained in $\overline{I_n}$, we have

$$|A_h(u_h;z_h) - A_h(v_h;z_h)| \lesssim \|u_h - v_h\|_{h,1}\|z_h\|_{h,1}. \tag{6.16}$$

Therefore, there exists a unique solution $u_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ of the numerical scheme (5.7)*.*

Proof We begin by showing strong monotonicity of the nonlinear form A_h. Let $z_h, v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ and set $w_h := z_h - v_h$. Then, by (5.6) and Lemma 9, we have

$$A_h(z_h;w_h) - A_h(v_h;w_h) = C_h(w_h,w_h) \\ + \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\langle F_\gamma[z_h] - F_\gamma[v_h] - L_\omega w_h, L_\omega w_h\rangle_K\,\mathrm{d}t.$$

Lemma 1 and Young's Inequality show that

$$\sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}|\langle F_\gamma[z_h]-F_\gamma[v_h]-L_\omega w_h, L_\omega w_h\rangle_K|\mathrm{d}t \leq \frac{1}{2}\sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\|L_\omega w_h\|_{L^2(K)}^2\mathrm{d}t \\ + \frac{1-\varepsilon}{2}\sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\omega^2\|\partial_t w_h\|_{L^2(K)}^2 + |w_h|_{H^2(K),\lambda}^2\mathrm{d}t.$$

Since $1<\kappa<(1-\varepsilon)^{-1}$, Lemma 9 implies that

$$A_h(z_h;w_h) - A_h(v_h;w_h) \geq \frac{1}{C}\sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\omega^2\|\partial_t w_h\|_{L^2(K)}^2 + |w_h|_{H^2(K),\lambda}^2\mathrm{d}t \\ + \frac{1}{2}\sum_{n=1}^{N}\int_{I_n}|w_h|_{\mathrm{J}}^2\,\mathrm{d}t + \frac{\omega}{2}\sum_{n=0}^{N}\|(\!|w_h|\!)_n\|_{a_h}^2, \qquad (6.17)$$

where $C = 2\,\kappa/(1-\kappa\,(1-\varepsilon)) \geq 2$, thus showing (6.15).

To show (6.16), consider u_h, v_h and $z_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ that all have support in $\overline{I_n}$, and set $w_h := u_h - v_h$. It then follows from $\operatorname{supp} v_h \subset \overline{I_n}$ that

$$\|v_h\|_{h,1}^2 = \int_{I_n}\sum_{K\in\mathcal{T}_h}\left[\omega^2\|\partial_t v_h\|_{L^2(K)}^2 + |v_h|_{H^2(K),\lambda}^2\right] + |v_h|_{\mathrm{J}}^2\,\mathrm{d}t \\ + \omega\|v_h(t_n)\|_{a_h}^2 + \omega\|v_h(t_{n-1}^+)\|_{a_h}^2,$$

and similarly for u_h and z_h. We also have

$$A_h(u_h;z_h) - A_h(v_h;z_h) = \int_{I_n}\sum_{K\in\mathcal{T}_h}\langle F_\gamma[u_h] - F_\gamma[v_h], L_\omega z_h\rangle_K\,\mathrm{d}t + C_h^{\mathcal{F}}(w_h,z_h) \\ + \int_{I_n} B_{h,1/2}(w_h,z_h) - \sum_{K\in\mathcal{T}_h}\langle L_\lambda w_h, L_\lambda z_h\rangle_K\,\mathrm{d}t + \omega\,a_h(w_h(t_{n-1}^+), z_h(t_{n-1}^+)).$$

Lipschitz continuity of F_γ implies that

$$\int_{I_n}\sum_{K\in\mathcal{T}_h}|\langle F_\gamma[u_h] - F_\gamma[v_h], L_\omega z_h\rangle_K|\,\mathrm{d}t \lesssim \|w_h\|_{h,1}\|z_h\|_{h,1}.$$

Furthermore, we have $|C_h^{\mathcal{F}}(w_h,z_h)| \leq E_1 + E_2$, where

$$E_1 := \omega\int_{I_n}\sum_{F\in\mathcal{F}_h^i}|\langle[\![\nabla w_h\cdot n_F]\!], \{\partial_t z_h\}\rangle_F|\,\mathrm{d}t,$$

$$E_2 := \omega\int_{I_n}\sum_{F\in\mathcal{F}_h^{i,b}}\mu_F|\langle[\![w_h]\!], [\![\partial_t z_h]\!]\rangle_F| + |\langle[\![w_h]\!], \{\nabla\partial_t z_h\cdot n_F\}\rangle_F|\,\mathrm{d}t.$$

The shape-regularity of the meshes $\{\mathcal{T}\}_h$, the mesh assumption (4.1) and the trace and inverse inequalities show that

$$E_1 \lesssim \left(\int_{I_n} \sum_{K\in\mathcal{T}_h} \omega^2 \, \|\partial_t z_h\|_{L^2(K)}^2 \mathrm{d}t \right)^{1/2} \left(\int_{I_n} \sum_{F\in\mathcal{F}_h^i} \frac{\tilde{p}_F^2}{\tilde{h}_F} \|[\![\nabla w_h \cdot n_F]\!]\|_{L^2(F)}^2 \mathrm{d}t \right)^{1/2},$$

$$E_2 \lesssim \left(\int_{I_n} \sum_{K\in\mathcal{T}_h} \omega^2 \, \|\partial_t z_h\|_{L^2(K)}^2 \mathrm{d}t \right)^{1/2} \left(\int_{I_n} \sum_{F\in\mathcal{F}_h^{i,b}} \frac{\tilde{p}_F^6}{\tilde{h}_F^3} \|[\![w_h]\!]\|_{L^2(F)}^2 \mathrm{d}t \right)^{1/2}.$$

Since μ_F and η_F satisfy (6.8), we conclude that $|C_h^{\mathcal{F}}(w_h, z_h)| \lesssim \|w_h\|_{h,1} \, \|z_h\|_{h,1}$. By applying trace and inverse inequalities on the flux terms of the bilinear form $B_{h,*}$, it is found that

$$|B_{h,*}(w_h, z_h)| \lesssim \left(\sum_{K\in\mathcal{T}_h} |w_h|_{H^2(K),\lambda}^2 + |w_h|_{\mathrm{J}}^2 \right)^{1/2} \left(\sum_{K\in\mathcal{T}_h} |z_h|_{H^2(K),\lambda}^2 + |z_h|_{\mathrm{J}}^2 \right)^{1/2}.$$

Therefore, $\int_{I_n} |B_{h,1/2}(w_h, z_h)| + \sum_{K\in\mathcal{T}_h} |\langle L_\lambda w_h, L_\lambda z_h\rangle_K| \,\mathrm{d}t \lesssim \|u_h - v_h\|_{h,1} \, \|z_h\|_{h,1}$, thus completing the proof of (6.16).

Since the numerical scheme (5.7) is equivalent to solving (5.8) for each $I_n \in \mathcal{J}_\tau$, and since A_h is strongly monotone and Lipschitz continuous on the subspace of $V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ of functions with support in $\overline{I_n}$, for each $I_n \in \mathcal{J}_\tau$, repeated applications of the Browder–Minty Theorem show that there exists a unique $u_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ that solves (5.7). □

7 Error analysis

In the first part of this section, we present error bounds for regular solutions, i.e. when the solution is in $H^1(I_n; H)$ for each $I_n \in \mathcal{J}_\tau$. It is found that the method has convergence orders that are optimal with respect to h, τ and $\mathbf{q}$, and that are possibly suboptimal with respect to $\mathbf{p}$ by an order and a half. In a second part, we use Clément approximation operators in Bochner spaces to extend the analysis under weaker regularity assumptions and to cover the case where $u \notin H^1(I_n; H)$.

There are two reasons for presenting the error analysis in two parts. First, the error analysis for regular solutions is simpler and permits the use of known approximation theory from [25], whereas the case of low regularity solutions requires the additional construction of a Clément quasi-interpolation operator. Second, the Clément operator is generally suboptimal by one order in τ when applied to very regular solutions. Thus, the results given here for regular and low regularity solutions are complementary to each other.

We will present error bounds in the norm $\|\cdot\|_h$ defined by

$$\|v\|_h^2 := \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \left[\omega^2 \|\partial_t v\|_{L^2(K)}^2 + |v|_{H^2(K),\lambda}^2 \right] + |v|_{\mathrm{J}}^2 \,\mathrm{d}t + \omega \sum_{n=0}^{N-1} \|(\!|v|\!)_n\|_{a_h}^2. \quad (7.1)$$

We remark that for $v_h \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$, we have $\|v_h\|_{h,1}^2 = \|v_h\|_h^2 + \omega \|(\!|v_h|\!)_N\|_{a_h}^2$. Error bounds in the norm $\|\cdot\|_{h,1}$ can be shown under additional regularity assumptions for the solution at time T. To simplify the notation in this section, let

$$X_0 := L^2(\Omega), \quad X_1 := H_0^1(\Omega), \quad X_2 := H = H^2(\Omega) \cap H_0^1(\Omega). \quad (7.2)$$

Similarly to the definition of the broken Sobolev spaces $H^s(\Omega; \mathcal{T}_h)$, for a Hilbert space X, we define the broken Bochner space $H^\sigma(I; X; \mathcal{J}_\tau)$ to be the space of functions $u \in L^2(I; X)$ with restrictions $u|_{I_n} \in H^\sigma(I_n; X)$ for each $I_n \in \mathcal{J}_\tau$. We equip $H^\sigma(I; X; \mathcal{J}_\tau)$ with the obvious norm.

7.1 Regular solutions

If the solution u of (2.3) belongs to $H^1(I;H,\mathcal{J}_\tau)$, then the error analysis may be based on the following approximation result, found for instance in [25], albeit presented here in a form amenable to our purposes.

Theorem 11 *Let $\Omega\subset\mathbb{R}^d$ be a bounded convex domain, and let $\{\mathcal{J}_\tau\}_\tau$ be a sequence of regular partitions of $I=(0,T)$. For each τ, let $\mathbf{q}=(q_1,\dots,q_N)$ be a vector of positive integers. Then, for each τ, there exists a linear operator $\Pi_\tau^{\mathbf{q}}\colon H(I;\Omega)\cap H^1(I;H;\mathcal{J}_\tau)\to V^{\tau,\mathbf{q}}$ such that the following holds. The operator $\Pi_\tau^{\mathbf{q}}$ is an interpolant at the interval endpoints, i.e. for any $u\in H(I;\Omega)\cap H^1(I;H;\mathcal{J}_\tau)$, we have $\Pi_\tau^{\mathbf{q}}u(t_n)=\Pi_\tau^{\mathbf{q}}u(t_n^+)=u(t_n)$ for each $0\leq n\leq N$. For any $I_n\in\mathcal{J}_\tau$, any $\ell\in\{0,1,2\}$, any real number $\sigma_{n,\ell}\geq 1$ and any $j\in\{0,1\}$, we have*

$$\|u-\Pi_\tau^{\mathbf{q}}u\|_{H^j(I_n;X_\ell)}\lesssim\frac{\tau_n^{\varrho_{n,\ell}-j}}{q_n^{\sigma_{n,\ell}-j}}\|u\|_{H^{\sigma_{n,\ell}}(I_n;X_\ell)}\qquad\forall\,u\in H^{\sigma_{n,\ell}}(I_n;X_\ell),\tag{7.3}$$

where $\varrho_{n,\ell}\coloneqq\min(\sigma_{n,\ell},q_n+1)$, and where the constant depends only on $\sigma_{n,\ell}$ and $\max\tau$.

The construction of $\Pi_\tau^{\mathbf{q}}$ in the proof of Theorem 11 involves the truncated Legendre series of $\partial_t u$ and the values of u at the partition points. Therefore, the requirement of $H^1(I;H;\mathcal{J}_\tau)$ regularity is used to ensure that $\Pi_\tau^{\mathbf{q}}|_{I_n}$ maps into $\mathcal{Q}_{q_n}(H)$. A different approximation operator is used in section 7.2 to perform an analysis under weaker regularity assumptions.

Theorem 12 *Let $\Omega\subset\mathbb{R}^d$ 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), (4.2), *and let $\mathbf{p}=(p_K;\ K\in\mathcal{T}_h)$ be a vector of positive integers such that* (4.3) *holds for each h, and such that $p_K\geq 2$ for all $K\in\mathcal{T}_h$. Let $I=(0,T)$ and let $\{\mathcal{J}_\tau\}_\tau$ be a sequence of regular partitions of I, and, for each τ, let $\mathbf{q}=(q_1,\dots,q_N)$ be a vector of positive integers. Let Λ be a compact metric space and let the data a, b, c and f be continuous on $\overline{\Omega}\times\overline{I}\times\Lambda$ and satisfy* (1.3) *and* (2.6), *or alternatively* (2.5) *in the case where $b\equiv 0$ and $c\equiv 0$. Let μ_F and η_F satisfy* (6.8), *with c_{s} chosen so that Lemmas 8 and 9 hold with $\kappa<(1-\varepsilon)^{-1}$.*

Let $u\in H(I;\Omega)$ be the unique solution of the HJB equation (2.3), *and assume that $u\in L^2(I;H^{\mathbf{s}}(\Omega;\mathcal{T}_h))$ and $\partial_t u\in L^2(I;H^{\overline{\mathbf{s}}}(\Omega,\mathcal{T}_h))$ for each h, with $s_K>5/2$ and $\overline{s}_K>0$ for each $K\in\mathcal{T}_h$. Suppose also that, for each τ, each $\ell\in\{0,2\}$ and each $I_n\in\mathcal{J}_\tau$, the function $u|_{I_n}\in H^{\sigma_{n,\ell}}(I_n;X_\ell)$ for some $\sigma_{n,\ell}\geq 1$. Assume that $u_0\in H_0^1(\Omega)\cap H^{\tilde{\mathbf{s}}}(\Omega;\mathcal{T}_h)$ with $\tilde{s}_K>3/2$ for each $K\in\mathcal{T}_h$. Then, we have*

$$\begin{aligned}\|u-u_h\|_h^2&\lesssim\sum_{n=1}^N\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K-4}}{p_K^{2s_K-7}}\|u\|_{H^{s_K}(K)}^2+\frac{h_K^{2\bar{t}_K}}{p_K^{2\bar{s}_K}}\|\partial_t u\|_{H^{\bar{s}_K}(K)}^2\mathrm{d}t\\&+\max_{K\in\mathcal{T}_h}p_K^3\sum_{n=1}^N\sum_{\ell\in\{0,2\}}\frac{\tau_n^{2\varrho_{n,\ell}-2+\ell}}{q_n^{2\sigma_{n,\ell}-2+\ell}}\|u\|_{H^{\sigma_{n,\ell}}(I_n;X_\ell)}^2+\sum_{K\in\mathcal{T}_h}\frac{h_K^{2\tilde{t}_K-2}}{p_K^{2\tilde{s}_K-3}}\|u_0\|_{H^{\tilde{s}_K}(K)}^2,\end{aligned}\tag{7.4}$$

with a constant independent of u, h, $\mathbf{p}$, τ and $\mathbf{q}$, and where $t_K\coloneqq\min(s_K,p_K+1)$, $\bar{t}_K\coloneqq\min(\bar{s}_K,p_K+1)$ and $\tilde{t}_K\coloneqq\min(\tilde{s}_K,p_K+1)$ for each $K\in\mathcal{T}_h$, and where $\varrho_{n,\ell}\coloneqq\min(\sigma_{n,\ell},q_n+1)$ for each $1\leq n\leq N$ and each $\ell\in\{0,2\}$.

Since the norm $\|\cdot\|_h$ comprises the broken H^2-seminorm in space and a broken H^1-norm in time, it is seen that the error bound is optimal with respect to h, τ and $\mathbf{q}$, but is suboptimal with respect to $\mathbf{p}$ by an order and a half. We remark that since Theorem 12 assumes $u\in H^1(I_1;H)$, the initial data satisfies $u_0\in H$, so we may take $\tilde{s}_K\geq 2$ for each $K\in\mathcal{T}_h$.

Proof The approximation theory for hp-version discontinuous Galerkin finite element spaces shows that there exists a sequence of linear projection operators $\{\Pi_h^{\mathbf{p}}\}_h$, with $\Pi_h^{\mathbf{p}}\colon L^2(\Omega) \to V_{h,\mathbf{p}}$ and such that for each $K \in \mathcal{T}_h$, for each nonnegative real number $r_K \leq \max(s_K, \overline{s}_K, \tilde{s}_K)$ and for each nonnegative integer $j \leq r_K$, and if $r_K > 1/2$, for each multi-index β such that $|\beta| < r_K - 1/2$, we have

$$\|u - \Pi_h^{\mathbf{p}} u\|_{H^j(K)} \lesssim \frac{h_K^{\min(r_K,\, p_K+1)-j}}{(p_K+1)^{r_K-j}} \|u\|_{H^{r_K}(K)} \qquad \forall\, u \in H^{r_K}(K), \quad (7.5)$$

$$\|D^\beta(u - \Pi_h^{\mathbf{p}} u)\|_{L^2(\partial K)} \lesssim \frac{h_K^{\min(r_K,\, p_K+1)-|\beta|-1/2}}{(p_K+1)^{r_K-|\beta|-1/2}} \|u\|_{H^{r_K}(K)} \quad \forall\, u \in H^{r_K}(K), \quad (7.6)$$

where the constant is independent of r_K, h_K, p_K but possibly dependent on s_K, $\overline{s}_K$ and $\tilde{s}_K$. The technical form of this approximation result expresses the optimality and stability of $\Pi_h^{\mathbf{p}}$ for functions in $H^{r_K}(K)$, $0 \leq r_K \leq \max(s_K, \overline{s}_K, \tilde{s}_K)$. In particular, we will use the fact that $\Pi_h^{\mathbf{p}}$ is elementwise L^2-stable, H^1-stable and H^2-stable in the analysis below.

For each h and τ, let $z_\tau := \Pi_\tau^{\mathbf{q}} u \in V^{\tau,\mathbf{q}}$, and let $z_h := \Pi_h^{\mathbf{p}} z_\tau \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$. Continuity of z_τ implies continuity of z_h, so that $(\!|z_h|\!)_n = 0$ for each $1 \leq n < N$. Furthermore, we have $z_\tau(0^+) = u_0$, so $z_h(0^+) = \Pi_h^{\mathbf{p}} u_0$. Let $\xi_h := u - z_h$ and let $\psi_h := u_h - z_h$, so that $u - u_h = \xi_h - \psi_h$. Recall that $\|\psi_h\|_h \leq \|\psi_h\|_{h,1}$.

Theorem 10, the scheme (5.7) and Corollary 7 show that

$$\begin{aligned}
\|\psi_h\|_{h,1}^2 &\lesssim A_h(u_h;\psi_h) - A_h(z_h;\psi_h) = A_h(u;\psi_h) - A_h(z_h;\psi_h) \\
&= \sum_{n=1}^N \int_{I_n} \sum_{K\in\mathcal{T}_h} \langle F_\gamma[u] - F_\gamma[z_h], L_\omega \psi_h\rangle_K + B_{h,1/2}(\xi_h,\psi_h)\,\mathrm{d}t \\
&- \sum_{n=1}^N \int_{I_n} \sum_{K\in\mathcal{T}_h} \langle L_\lambda \xi_h, L_\lambda \psi_h\rangle_K \,\mathrm{d}t + C_h^{\mathcal{F}}(\xi_h,\psi_h) + \omega\, a_h(\xi_h(t_0^+), \psi_h(t_0^+)). \quad (7.7)
\end{aligned}$$

Therefore $\|\psi_h\|_h^2 \leq \|\psi_h\|_{h,1}^2 \leq \sum_{i=1}^4 D_i$, where the quantities D_i, $1 \leq i \leq 4$, are defined by

$$D_1 := \sum_{n=1}^N \int_{I_n} \sum_{K\in\mathcal{T}_h} |\langle F_\gamma[u] - F_\gamma[z_h], L_\omega\psi_h\rangle_K| + |\langle L_\lambda \xi_h, L_\lambda\psi_h\rangle_K|\mathrm{d}t,$$

$$D_2 := \sum_{n=1}^N \int_{I_n} |B_{h,1/2}(\xi_h,\psi_h)|\mathrm{d}t, \quad D_3 := |C_h^{\mathcal{F}}(\xi_h,\psi_h)|, \quad D_4 := \omega|a_h(\xi_h(0^+),\psi_h(0^+))|.$$

Lipschitz continuity of F_γ implies that $D_1 \lesssim \sqrt{E_1 + E_2}\ \|\psi_h\|_{h,1}$, where E_1 and E_2 are defined by

$$E_1 := \sum_{n=1}^N \int_{I_n} \sum_{K\in\mathcal{T}_h} \|\partial_t \xi_h\|_{L^2(K)}^2 \mathrm{d}t, \quad E_2 := \sum_{n=1}^N \int_{I_n} \sum_{K\in\mathcal{T}_h} \|\xi_h\|_{H^2(K)}^2 \mathrm{d}t.$$

Since the sequence of meshes $\{\mathcal{T}_h\}_h$ is shape-regular and since $\psi_h|_{I_n} \in \mathcal{Q}_{q_n}(V_{h,\mathbf{p}})$ for each $I_n \in \mathcal{J}_\tau$, the use of trace and inverse inequalities on the flux terms appearing in $B_{h,1/2}(\xi_h,\psi_h)$

yields $D_2 \lesssim \sqrt{\sum_{i=2}^{6} E_i}\ \|\psi_h\|_{h,1}$, where the quantities E_i, $3 \leq i \leq 5$, are defined by

$$E_3 := \sum_{n=1}^{N} \int_{I_n} \sum_{F \in \mathcal{F}_h^i} \mu_F^{-1} \|\mathrm{div}_{\mathrm{T}}\, \nabla_{\mathrm{T}} \{\xi_h\}\|_{L^2(F)}^2 + \sum_{F \in \mathcal{F}_h^{i,b}} \mu_F^{-1} \|\nabla_{\mathrm{T}} \{\nabla \xi_h \cdot n_F\}\|_{L^2(F)}^2 \mathrm{d}t,$$

$$E_4 := \sum_{n=1}^{N} \int_{I_n} \sum_{F \in \mathcal{F}_h^{i,b}} \eta_F^{-1} \|\{\nabla \xi_h \cdot n_F\}\|_{L^2(F)}^2 + \sum_{F \in \mathcal{F}_h^i} \mu_F^{-1} \|\{\xi_h\}\|_{L^2(F)}^2 \mathrm{d}t,$$

$$E_5 := \sum_{n=1}^{N} \int_{I_n} \sum_{F \in \mathcal{F}_h^i} \mu_F \|[\![\nabla \xi_h \cdot n_F]\!]\|_{L^2(F)}^2 + \sum_{F \in \mathcal{F}_h^{i,b}} \mu_F \|[\![\nabla_{\mathrm{T}}\, \xi_h]\!]\|_{L^2(F)}^2 \mathrm{d}t,$$

$$E_6 := \sum_{n=1}^{N} \int_{I_n} \sum_{F \in \mathcal{F}_h^{i,b}} \eta_F \|[\![\xi_h]\!]\|_{L^2(F)}^2 \mathrm{d}t.$$

Note that $\partial_t \psi_h|_{I_n} \in \mathcal{Q}_{q_n-1}(V_{h,\mathbf{p}})$ for each $I_n \in \mathcal{J}_\tau$. Thus, similarly to the proof of Theorem 10, the use of trace and inverse inequalities leads to $D_3 \lesssim \sqrt{E_4 + E_5}\ \|\psi_h\|_{h,1}$. It follows from (6.7) that we have $D_4 \lesssim \sqrt{E_6 + E_7 + E_8}\ \|\psi_h\|_{h,1}$, where the quantities E_i, $7 \leq i \leq 9$, are defined by

$$E_7 := \sum_{K \in \mathcal{T}_h} \|u_0 - \Pi_h^{\mathbf{p}} u_0\|_{H^1(K)}^2, \qquad E_8 := \sum_{F \in \mathcal{F}_h^{i,b}} \mu_F \|u_0 - \Pi_h^{\mathbf{p}} u_0\|_{L^2(F)}^2,$$

$$E_9 := \sum_{F \in \mathcal{F}_h^{i,b}} \mu_F^{-1} \|\{\nabla(u_0 - \Pi_h^{\mathbf{p}} u_0) \cdot n_F\}\|_{L^2(F)}^2.$$

Therefore, (7.7) implies that $\|\psi_h\|_h^2 \lesssim \sum_{i=1}^{9} E_i$. The properties of the operator $\Pi_h^{\mathbf{p}}$, namely its linearity, L^2-stability and approximation properties (7.5), together with (7.3), imply that

$$\begin{aligned} E_1 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K \in \mathcal{T}_h} \|\partial_t u - \Pi_h^{\mathbf{p}} \partial_t u\|_{L^2(K)}^2 + \|\Pi_h^{\mathbf{p}}(\partial_t u - \partial_t z_\tau)\|_{L^2(K)}^2 \mathrm{d}t \\ &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K \in \mathcal{T}_h} \|\partial_t u - \Pi_h^{\mathbf{p}} \partial_t u\|_{L^2(K)}^2 \mathrm{d}t + \sum_{n=1}^{N} \|u - z_\tau\|_{H^1(I_n;X_0)}^2 \\ &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K \in \mathcal{T}_h} \frac{h_K^{2\bar{t}_K}}{p_K^{2\bar{s}_K}} \|\partial_t u\|_{H^{\bar{s}_K}(K)}^2 \mathrm{d}t + \sum_{n=1}^{N} \frac{\tau_n^{2\varrho_{n,0}-2}}{q_n^{2\sigma_{n,0}-2}} \|u\|_{H^{\sigma_{n,0}}(I_n;X_0)}^2. \end{aligned} \tag{7.8}$$

Since the operator $\Pi_h^{\mathbf{p}}$ is elementwise H^2-stable, it is found that

$$\begin{aligned} E_2 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K \in \mathcal{T}_h} \|u - \Pi_h^{\mathbf{p}} u\|_{H^2(K)}^2 + \|\Pi_h^{\mathbf{p}}(u - z_\tau)\|_{H^2(K)}^2 \mathrm{d}t \\ &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K \in \mathcal{T}_h} \|u - \Pi_h^{\mathbf{p}} u\|_{H^2(K)}^2 \mathrm{d}t + \sum_{n=1}^{N} \|u - z_\tau\|_{L^2(I_n;X_2)}^2 \\ &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K \in \mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-4}} \|u\|_{H^{s_K}(K)}^2 \mathrm{d}t + \sum_{n=1}^{N} \frac{\tau_n^{2\varrho_{n,2}}}{q_n^{2\sigma_{n,2}}} \|u\|_{H^{\sigma_{n,2}}(I_n;X_2)}^2. \end{aligned} \tag{7.9}$$

The mesh assumptions (4.1), (4.2) and (4.3), the bound (7.6), and the application of trace and inverse inequalities on $\Pi_h^{\mathbf{p}}(u-z_\tau)|_{I_n} \in \mathcal{Q}_{q_n}(V_{h,\mathbf{p}})$, imply that

$$\begin{aligned} E_3 &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K}{p_K^2}\|D^2(u-\Pi_h^{\mathbf{p}}z_\tau)\|_{L^2(\partial K)}^2\,\mathrm{d}t \\ &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K}{p_K^2}\left[\|D^2(u-\Pi_h^{\mathbf{p}}u)+D^2\Pi_h^{\mathbf{p}}(u-z_\tau)\|_{L^2(\partial K)}^2\right]\mathrm{d}t \\ &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K-4}}{p_K^{2s_K-3}}\|u\|_{H^{s_K}(K)}^2+\sum_{K\in\mathcal{T}_h}\|u-z_\tau\|_{H^2(K)}^2\mathrm{d}t \\ &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K-4}}{p_K^{2s_K-3}}\|u\|_{H^{s_K}(K)}^2\mathrm{d}t+\sum_{n=1}^{N}\frac{\tau_n^{2\varrho_{n,2}}}{q_n^{2\sigma_{n,2}}}\|u\|_{H^{\sigma_{n,2}}(I_n;X_2)}^2. \end{aligned} \tag{7.10}$$

Similarly to E_3, we find that

$$E_4 \lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K}}{p_K^{2s_K+1}}\|u\|_{H^{s_K}(K)}^2\,\mathrm{d}t+\sum_{n=1}^{N}\frac{\tau_n^{2\varrho_{n,0}}}{q_n^{2\sigma_{n,0}}}\|u\|_{H^{\sigma_{n,0}}(I_n;X_0)}^2. \tag{7.11}$$

The spatial regularity of u and z_τ imply that

$$\begin{aligned} E_5 = &\sum_{n=1}^{N}\int_{I_n}\sum_{F\in\mathcal{F}_h^i}\mu_F\|[\![\nabla\left[u-\Pi_h^{\mathbf{p}}u+\Pi_h^{\mathbf{p}}(u-z_\tau)-(u-z_\tau)\right]\cdot n_F]\!]\|_{L^2(F)}^2\mathrm{d}t \\ &+\sum_{n=1}^{N}\int_{I_n}\sum_{F\in\mathcal{F}_h^{i,b}}\mu_F\|[\![\nabla_{\mathrm{T}}\left[u-\Pi_h^{\mathbf{p}}u+\Pi_h^{\mathbf{p}}(u-z_\tau)-(u-z_\tau)\right]]\!]\|_{L^2(F)}^2\mathrm{d}t. \end{aligned}$$

Therefore, the mesh assumptions (4.1), (4.2) and (4.3) and the approximation bound (7.6) yield

$$\begin{aligned} E_5 &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{p_K^2}{h_K}\|\nabla(u-\Pi_h^{\mathbf{p}}u)+\nabla\left[u-z_\tau-\Pi_h^{\mathbf{p}}(u-z_\tau)\right]\|_{L^2(\partial K)}^2\mathrm{d}t \\ &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K-4}}{p_K^{2s_K-5}}\|u\|_{H^{s_K}(K)}^2+\sum_{K\in\mathcal{T}_h}p_K\|u-z_\tau\|_{H^2(K)}^2\mathrm{d}t \\ &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K-4}}{p_K^{2s_K-5}}\|u\|_{H^{s_K}(K)}^2\mathrm{d}t+\max_{K\in\mathcal{T}_h}p_K\sum_{n=1}^{N}\frac{\tau_n^{2\varrho_{n,2}}}{q_n^{2\sigma_{n,2}}}\|u\|_{H^{\sigma_{n,2}}(I_n;X_2)}^2. \end{aligned} \tag{7.12}$$

Likewise, it follows from the spatial regularity of z_τ, the mesh assumptions, and the approximation bound (7.6) that

$$\begin{aligned} E_6 &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{p_K^6}{h_K^3}\|u-\Pi_h^{\mathbf{p}}u+\Pi_h^{\mathbf{p}}(u-z_\tau)-(u-z_\tau)\|_{L^2(\partial K)}^2\mathrm{d}t \\ &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K-4}}{p_K^{2s_K-7}}\|u\|_{H^{s_K}(K)}^2+\sum_{K\in\mathcal{T}_h}p_K^3\|u-z_\tau\|_{H^2(K)}^2\mathrm{d}t \\ &\lesssim \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}_h}\frac{h_K^{2t_K-4}}{p_K^{2s_K-7}}\|u\|_{H^{s_K}(K)}^2\mathrm{d}t+\max_{K\in\mathcal{T}_h}p_K^3\sum_{n=1}^{N}\frac{\tau_n^{2\varrho_{n,2}}}{q_n^{2\sigma_{n,2}}}\|u\|_{H^{\sigma_{n,2}}(I_n;X_2)}^2. \end{aligned} \tag{7.13}$$

Finally, it is readily shown that

$$\sum_{i=7}^{9} E_i \lesssim \sum_{K\in\mathcal{T}_h} \frac{h_K^{2\tilde{t}_K-2}}{p_K^{2\tilde{s}_K-3}} \|u_0\|_{H^{\tilde{s}_K}(K)}^2. \tag{7.14}$$

Since $\|\xi_h\|_h^2 \leq \sum_{i=1}^{9} E_i$, the above bounds and the triangle inequality $\|u-u_h\|_h \leq \|\xi_h\|_h + \|\psi_h\|_h$ complete the proof of (7.4). □

7.2 Low regularity solutions

The proof of Theorem 12 depends on the approximation result from Theorem 11, which requires that the solution u belongs to $H^1(I;H;\mathcal{J}_\tau)$. In this section, we relax this condition by using a Clément quasi-interpolation result instead of Theorem 11.

For $\mathcal{J}_\tau$ a regular partition of $(0,T)$, let $\{\phi_m\}_{m=0}^N$ denote the set of hat functions of $\mathcal{J}_\tau$, i.e. ϕ_m is the unique piecewise-affine function on $\mathcal{J}_\tau$ such that $\phi_m(t_n)=\delta_{nm}$ for $0\leq n,m\leq N$. For $0\leq m\leq N$, let $J_m := \operatorname{supp}\phi_m$, and note that $J_m = \overline{I_m}\cup\overline{I_{m+1}}$ for $1\leq m<N$, whilst $J_0=\overline{I_1}$ and $J_N=\overline{I_N}$.

Theorem 13 *Let $\Omega\subset\mathbb{R}^d$ be a bounded convex domain, and let $\{\mathcal{J}_\tau\}_\tau$ be a sequence of regular partitions of $I=(0,T)$. For each τ, let $\mathbf{q}=(q_1,\dots,q_N)$ be a vector of positive integers. Suppose that there exist positive constants $c_\mathcal{T}$ and c_q such that, for each τ, we have*

$$\frac{1}{c_\mathcal{T}} \leq \frac{\tau_{n-1}}{\tau_n} \leq c_\mathcal{T}, \quad \frac{1}{c_q} \leq \frac{q_{n-1}}{q_n} \leq c_q, \quad 2\leq n\leq N. \tag{7.15}$$

Let $u\in L^2(I;H)$ and suppose that $u|_{J_m}\in H^{\sigma_{m,\ell}}(J_m;X_\ell)$ for some $\sigma_{m,\ell}\in\mathbb{R}_{\geq 0}$ for each $\ell\in\{0,1,2\}$ and each $0\leq m\leq N$. Then, there exists a sequence of functions $\{z_\tau\}_\tau$, such that $z_\tau\in V^{\tau,\mathbf{q}}$ for each τ, and such that the following properties hold. The functions z_τ are continuous on I, i.e. $(\!|z_\tau|\!)_n=0$ for each $1\leq n<N$. For each $\ell\in\{0,1,2\}$ and each $I_n\in\mathcal{J}_\tau$, we have

$$\|z_\tau\|_{L^2(I_n;X_\ell)} \lesssim \sum_{J_m\supset I_n} \|u\|_{L^2(J_m;X_\ell)}, \tag{7.16}$$

where the constant is independent of all other quantities. For each $\ell\in\{0,1,2\}$, each $I_n\in\mathcal{J}_\tau$ and each nonnegative integer $j\leq\min_{J_m\supset I_n}\sigma_{m,\ell}$, we have

$$\|u-z_\tau\|_{H^j(I_n;X_\ell)} \lesssim \sum_{J_m\supset I_n} \frac{\tau_n^{\varrho_{m,\ell}-j}}{q_n^{\sigma_{m,\ell}-j}} \|u\|_{H^{\sigma_{m,\ell}}(J_m;X_\ell)}, \tag{7.17}$$

where $\varrho_{m,\ell}:=\min(\sigma_{m,\ell},\min_{I_n\subset J_m} q_n)$, and the constant depends only on $\max\sigma_{m,\ell}$, $\max\tau$, $c_\mathcal{T}$ and c_q.

Proof For $0\leq m\leq N$, define $\bar{q}_m:=\min_{I_n\subset J_m} q_n$, and note $\bar{q}_m\geq 1$ for all m since $q_n\geq 1$ for all n. Since $u\in L^2(J_m;X_2)$ for each m, standard approximation theory for Bochner spaces implies that there exist functions $v_m\in\mathcal{Q}_{\bar{q}_m-1}(H)$, $0\leq m\leq N$, with the following properties. For each $\ell\in\{0,1,2\}$, we have $\|v_m\|_{L^2(J_m;X_\ell)}\lesssim\|u\|_{L^2(J_m;X_\ell)}$, with a constant independent of all other quantities. For each $\ell\in\{0,1,2\}$ and each nonnegative integer $j\leq\sigma_{m,\ell}$, we have

$$\|u-v_m\|_{H^j(J_m;X_\ell)} \lesssim \frac{|J_m|^{\varrho_{m,\ell}-j}}{\bar{q}_m^{\sigma_{m,\ell}-j}} \|u\|_{H^{\sigma_{m,\ell}}(J_m;X_\ell)}, \tag{7.18}$$

where $\varrho_{m,\ell}:=\min(\sigma_{m,\ell},\bar{q}_m)$, where $|J_m|$ is the length of the interval J_m, and where the constant depends only on $\max\sigma_{m,\ell}$ and $\max\tau$.

The hypothesis (7.15) and the bound (7.18) imply that, for each $I_n \subset J_m$, each $\ell \in \{0,1,2\}$ and each nonnegative integer $j \le \sigma_{m,\ell}$,

$$\|u - v_m\|_{H^j(I_n;X_\ell)} \lesssim \frac{\tau_n^{\varrho_{m,\ell}-j}}{q_n^{\sigma_{m,\ell}-j}} \|u\|_{H^{\sigma_{m,\ell}}(J_m;X_\ell)}, \tag{7.19}$$

where the constant depends only on $\max \sigma_{m,\ell}$, $\max \tau$, $c_{\mathcal{T}}$ and c_q.

Define $z_\tau := \sum_{m=0}^{N} \phi_m v_m$, where ϕ_m is the hat function over the interval J_m. Note that we have $v_m|_{I_n} \in \mathcal{Q}_{q_n-1}(H)$ for each $I_n \in \mathcal{J}_\tau$ since $\bar{q}_m \le q_n$ for each $I_n \subset J_m$. Since ϕ_m is piecewise affine, it follows that $z_\tau|_{I_n} \in \mathcal{Q}_{q_n}(H)$ for each $I_n \in \mathcal{J}_\tau$, thereby showing that $z_\tau \in V^{\tau,\mathbf{q}}$. Furthermore, it is clear that z_τ is continuous on I, i.e. $(\!| z_\tau |\!)_n = 0$ for each $1 \le n \le N-1$. The bound (7.16) follows from $\|v_m\|_{L^2(J_m;X_\ell)} \lesssim \|u\|_{L^2(J_m;X_\ell)}$ and from the fact that $\|\phi_m\|_{L^\infty(I)} = 1$ for each $0 \le m \le N$. Since $\{\phi_m\}_{m=0}^N$ forms a partition of unity, the bound (7.19) implies that, for each $I_n \in \mathcal{J}_\tau$ and each $\ell \in \{0,1,2\}$,

$$\begin{aligned}\|u - z_\tau\|_{L^2(I_n;X_\ell)} &\le \sum_{J_m \supset I_n} \|\phi_m(u - v_m)\|_{L^2(I_n;X_\ell)} \\ &\lesssim \sum_{J_m \supset I_n} \|u - v_m\|_{L^2(I_n;X_\ell)} \lesssim \sum_{J_m \supset I_n} \frac{\tau_n^{\varrho_{m,\ell}}}{q_n^{\sigma_{m,\ell}}} \|u\|_{H^{\sigma_{m,\ell}}(J_m;X_\ell)},\end{aligned}$$

and, for each integer $1 \le j \le \min_{J_m \supset I_n} \sigma_{m,\ell}$,

$$\begin{aligned}|u - z_\tau|_{H^j(I_n;X_\ell)} &\le \sum_{J_m \supset I_n} |\phi_m(u - v_m)|_{H^j(I_n;X_\ell)} \\ \lesssim \sum_{J_m \supset I_n} |u - v_m|_{H^j(I_n;X_\ell)} &+ \frac{1}{\tau_n} |u - v_m|_{H^{j-1}(I_n;X_\ell)} \lesssim \sum_{J_m \supset I_n} \frac{\tau_n^{\varrho_{m,\ell}-j}}{q_n^{\sigma_{m,\ell}-j}} \|u\|_{H^{\sigma_{m,\ell}}(J_m;X_\ell)}.\end{aligned}$$

This completes the proof of (7.17). □

Theorem 14 *Let $\Omega \subset \mathbb{R}^d$ 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), (4.2), *and let $\mathbf{p} = (p_K;\ K \in \mathcal{T}_h)$ be a vector of positive integers satisfying* (4.3) *for each h and such that $p_K \ge 2$ for each $K \in \mathcal{T}_h$. Let $I = (0,T)$ and let $\{\mathcal{J}_\tau\}_\tau$ be a sequence of regular partitions of I, and, for each τ, let $\mathbf{q}$ be a vector of positive integers such that* (7.15) *holds. Let Λ be a compact metric space and let the data a, b, c and f be continuous on $\overline{\Omega} \times \overline{I} \times \Lambda$ and satisfy* (1.3) *and* (2.6), *or alternatively* (2.5) *in the case where $b \equiv 0$ and $c \equiv 0$. Let μ_F and η_F satisfy* (6.8), *with c_{s} chosen so that Lemmas 8 and 9 hold with $\kappa < (1-\varepsilon)^{-1}$.*

Let $u \in H(I;\Omega)$ be the unique solution of the HJB equation (2.3), *and assume that $u \in L^2(I;H^{\mathbf{s}}(\Omega;\mathcal{T}_h))$ and $\partial_t u \in L^2(I;H^{\bar{\mathbf{s}}}(\Omega,\mathcal{T}_h))$ for each h, with $s_K > 5/2$ and $\bar{s}_K > 0$ for each $K \in \mathcal{T}_h$. Suppose also that, for each τ, $\ell \in \{0,1,2\}$, and each $0 \le m \le N$, the function $u|_{J_m} \in H^{\sigma_{m,\ell}}(J_m;X_\ell)$ for some real $\sigma_{m,\ell} \ge 0$, with $\sigma_{m,0} \ge 1$ for all m. Assume that $u_0 \in H_0^1(\Omega) \cap H^{\tilde{\mathbf{s}}}(\Omega;\mathcal{T}_h)$ with $\tilde{s}_K > 3/2$ for each $K \in \mathcal{T}_h$. Then, we have*

$$\begin{aligned}\|u - u_h\|_h^2 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K \in \mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-7}} \|u\|_{H^{s_K}(K)}^2 + \frac{h_K^{2\bar{t}_K}}{p_K^{2\bar{s}_K}} \|\partial_t u\|_{H^{\bar{s}_K}(K)}^2 \mathrm{d}t \\ &+ \max_{K \in \mathcal{T}_h} p_K^3 \sum_{n=1}^{N} \sum_{\ell=0}^{2} \sum_{J_m \supset I_n} \frac{\tau_n^{2\varrho_{m,\ell}-2+\ell}}{q_n^{2\sigma_{m,\ell}-2+\ell}} \|u\|_{H^{\sigma_{m,\ell}}(J_m;X_\ell)}^2 + \sum_{K \in \mathcal{T}_h} \frac{h_K^{2\tilde{t}_K-2}}{p_K^{2\tilde{s}_K-3}} \|u_0\|_{H^{\tilde{s}_K}(K)}^2,\end{aligned} \tag{7.20}$$

with a constant independent of h, $\mathbf{p}$, τ, $\mathbf{q}$, and u, and where $t_K := \min(s_K, p_K+1)$, $\bar{t}_K := \min(\bar{s}_K, p_K+1)$, and $\tilde{t}_K := \min(\tilde{s}_K, p_K+1)$ for each $K \in \mathcal{T}_h$, and where $\varrho_{m,\ell} := \min(\sigma_{m,\ell}, \min_{I_n \subset J_m} q_n)$ for each $0 \le m \le N$ and each $\ell \in \{0,1,2\}$.

Proof For each h, let $\Pi_h^{\mathbf{p}} \colon L^2(\Omega) \to V_{h,\mathbf{p}}$ denote the approximation operator of the proof of Theorem 12; for each τ, let $z_\tau \in V^{\tau,\mathbf{q}}$ denote the approximation of u given by Theorem 13; then define $z_h := \Pi_h^{\mathbf{p}} z_\tau \in V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$. The fact that z_τ is continuous on $(0,T)$ implies that z_h is also continuous on $(0,T)$, so $(\!| z_h |\!)_n = 0$ for $1 \le n < N$. Let $\xi_h := u - z_h$ and $\psi_h := u_h - z_h$, so that $u - u_h = \xi_h - \psi_h$. As in the proof of Theorem 12, it is found that $\|\psi_h\|_h^2 \le \|\psi_h\|_{h,1}^2 \lesssim \sum_{i=1}^{9} E_i$, where the quantities E_i, $1 \le i \le 9$, are defined as before. Note that since $\sigma_{m,0} \ge 1$ for all m, the bound (7.17) is applicable for $j=1$ and $\ell = 0$. Therefore, the arguments from the proof of Theorem 12 and the approximation properties of z_τ from Theorem 13 imply that

$$
\begin{aligned}
E_1 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \frac{h_K^{2\bar{t}_K}}{p_K^{2\bar{s}_K}} \|\partial_t u\|_{H^{\bar{s}_K}(K)}^2 \mathrm{d}t + \sum_{n=1}^{N} \sum_{J_m \supset I_n} \frac{t_n^{2\varrho_{m,0}-2}}{q_n^{2\sigma_{m,0}-2}} \|u\|_{H^{\sigma_{m,0}}(J_m;X_0)}^2, \\
E_2 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-4}} \|u\|_{H^{s_K}(K)}^2 \mathrm{d}t + \sum_{n=1}^{N} \sum_{J_m \supset I_n} \frac{\tau_n^{2\varrho_{m,2}}}{q_n^{2\sigma_{m,2}}} \|u\|_{H^{\sigma_{m,2}}(J_m;X_2)}^2, \\
E_3 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-3}} \|u\|_{H^{s_K}(K)}^2 \mathrm{d}t + \sum_{n=1}^{N} \sum_{J_m \supset I_n} \frac{\tau_n^{2\varrho_{m,2}}}{q_n^{2\sigma_{m,2}}} \|u\|_{H^{\sigma_{m,2}}(J_m;X_2)}^2, \\
E_4 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K}}{p_K^{2s_K+1}} \|u\|_{H^{s_K}(K)}^2 \mathrm{d}t + \sum_{n=1}^{N} \sum_{J_m \supset I_n} \frac{\tau_n^{2\varrho_{m,0}}}{q_n^{2\sigma_{m,0}}} \|u\|_{H^{\sigma_{m,0}}(J_m;X_0)}^2, \\
E_5 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-5}} \|u\|_{H^{s_K}(K)}^2 \mathrm{d}t + \max_{K\in\mathcal{T}_h} p_K \sum_{n=1}^{N} \sum_{J_m \supset I_n} \frac{\tau_n^{2\varrho_{m,2}}}{q_n^{2\sigma_{m,2}}} \|u\|_{H^{\sigma_{m,2}}(J_m;X_2)}^2, \\
E_6 &\lesssim \sum_{n=1}^{N} \int_{I_n} \sum_{K\in\mathcal{T}_h} \frac{h_K^{2t_K-4}}{p_K^{2s_K-7}} \|u\|_{H^{s_K}(K)}^2 \mathrm{d}t + \max_{K\in\mathcal{T}_h} p_K^3 \sum_{n=1}^{N} \sum_{J_m \supset I_n} \frac{\tau_n^{2\varrho_{m,2}}}{q_n^{2\sigma_{m,2}}} \|u\|_{H^{\sigma_{m,2}}(J_m;X_2)}^2.
\end{aligned}
$$

Using inverse inequalities and H^1-stability of $\Pi_h^{\mathbf{p}}$, we find that

$$
\begin{aligned}
E_7 + E_8 &= \sum_{K\in\mathcal{T}_h} \|u_0 - \Pi_h^{\mathbf{p}} z_\tau(0^+)\|_{H^1(K)}^2 + \sum_{F\in\mathcal{F}_h^{i,b}} \mu_F^{-1} \|\{\nabla(u_0 - \Pi_h^{\mathbf{p}} z_\tau(0^+)) \cdot n_F\}\|_{L^2(F)}^2 \\
&\lesssim \sum_{K\in\mathcal{T}_h} \|u_0 - \Pi_h^{\mathbf{p}} u_0\|_{H^1(K)}^2 + \|u_0 - z_\tau(0^+)\|_{H^1(\Omega)}^2 \\
&\lesssim \sum_{K\in\mathcal{T}_h} \frac{h_K^{2\tilde{t}_K-2}}{p_K^{2\tilde{s}_K-2}} \|u_0\|_{H^{\tilde{s}_K}(K)}^2 + \|u_0 - z_\tau(0^+)\|_{H^1(\Omega)}^2.
\end{aligned}
$$

Since $z_\tau|_{I_1} \in \mathcal{Q}_{q_n}(H)$, we have $z_\tau(0^+) \in H_0^1(\Omega)$, so

$$
\begin{aligned}
E_9 &= \sum_{F\in\mathcal{F}_h^{i,b}} \mu_F \|[\![u_0 - \Pi_h^{\mathbf{p}} z_\tau(0^+)]\!]\|_{L^2(F)}^2 \\
&= \sum_{F\in\mathcal{F}_h^{i,b}} \mu_F \|[\![u_0 - \Pi_h^{\mathbf{p}} u_0 + \Pi_h^{\mathbf{p}}(u_0 - z_\tau(0^+)) - (u_0 - z_\tau(0^+))]\!]\|_{L^2(F)}^2 \\
&\lesssim \sum_{K\in\mathcal{T}_h} \frac{h_K^{2\tilde{t}_K-2}}{p_K^{2\tilde{s}_K-3}} \|u_0\|_{H^{\tilde{s}_K}(K)}^2 + \max_{K\in\mathcal{T}_h} p_K \|u_0 - z_\tau(0^+)\|_{H^1(\Omega)}^2.
\end{aligned}
\tag{7.21}
$$

Poincaré's Inequality and (7.17) then show that

$$
\begin{aligned}
\|u_0 - z_\tau(0^+)\|_{H^1(\Omega)}^2 &\lesssim \|u - z_\tau\|_{L^2(I_1;X_2)} \|u - z_\tau\|_{H^1(I_1;X_0)} + \frac{1}{\tau_1}\|u - z_\tau\|_{L^2(I_1;X_1)}^2 \\
&\lesssim \sum_{J_m \supset I_1} \frac{\tau_1^{2\varrho_{m,2}}}{q_1^{2\sigma_{m,2}}} \|u\|_{H^{\sigma_{m,2}}(J_m;X_2)}^2 + \frac{\tau_1^{2\varrho_{m,0}-2}}{q_1^{2\sigma_{m,0}-2}} \|u\|_{H^{\sigma_{m,0}}(J_m;X_0)}^2 \\
&\qquad + \sum_{J_m \supset I_1} \frac{\tau_1^{2\varrho_{m,1}-1}}{q_1^{2\sigma_{m,1}}} \|u\|_{H^{\sigma_{m,1}}(J_m;X_1)}^2 .
\end{aligned}
$$

Since $\|\xi_h\|_h^2 \lesssim \sum_{i=1}^9 E_i$, the combination of the above bounds with the triangle inequality $\|u - u_h\|_h \leq \|\xi_h\|_h + \|\psi_h\|_h$ completes the proof of (7.20). □

8 Numerical experiments

In the first experiment, we study the performance of the method on a fully nonlinear problem with strongly anisotropic diffusion coefficients, and observe optimal convergence rates for smooth solutions. In the second experiment, we obtain exponential convergence rates when combining hp-refinement and τq-refinement, even for problems with low regularity solutions.

8.1 First experiment

We examine the orders of convergence of the method for a problem with strongly anisotropic diffusion coefficients and a smooth solution. Let $\Omega = (0,1)^2$, $I = (0,1)$, let $b^\alpha \equiv 0$, $c^\alpha \equiv 0$ and let the a^α be defined by

$$
a^\alpha := \alpha \begin{pmatrix} 1 & 1/40 \\ 1/40 & 1/800 \end{pmatrix} \alpha^\top, \quad \alpha \in \Lambda := \mathrm{SO}(2), \tag{8.1}
$$

where $\mathrm{SO}(2)$ is the special orthogonal group of 2×2 matrices. For $\omega = 1$, $\lambda = 0$, it is found that the Cordes condition (2.5) holds with $\varepsilon \approx 1.25 \times 10^{-3}$. We choose f^α so that the exact solution is $u = \left(1 - \mathrm{e}^{-t}\right) \mathrm{e}^{xy} \sin(\pi x) \sin(\pi y)$. The strong anisotropy of the diffusion coefficient in this problem implies that monotone finite difference discretisations would require very large stencils in order to achieve consistency [3].

The numerical scheme (5.7) is applied on a sequence of uniform meshes obtained by regular subdivision of Ω into quadrilateral elements of width $h = 2^{-k}$, $1 \leq k \leq 5$. The corresponding time partitions $\mathcal{J}_\tau$ are obtained by regular subdivision of the time interval $(0,1)$ into intervals of length $\tau = 2^{-k+1}$, $1 \leq k \leq 5$. The finite element spaces $V_{h,\mathbf{p}}^{\tau,\mathbf{q}}$ are defined using polynomials of total degree p in space and degree $q = p - 1$ in time, for $p \in \{2,3,4\}$. We set the penalty parameter $c_{\mathrm{s}} = 5/2$ and $\sigma = 1$ in (6.8). The semismooth Newton method analysed in [26] is used to compute the numerical solution at each timestep.

In order to study the accuracy of the method, we measure the global error in the norm $|||\cdot|||_h$ defined by

$$
|||v|||_h^2 := \sum_{n=1}^{N} \int_{I_n} \sum_{K \in \mathcal{T}_h} \left[\omega^2 \|\partial_t v\|_{L^2(K)}^2 + \|v\|_{H^2(K)}^2 \right] \mathrm{d}t. \tag{8.2}
$$

Figure 1 presents the global relative errors achieved by the method, where it is seen that the optimal orders of convergence $|||u - u_h|||_h \simeq h^{p-1} + \tau^q$ are achieved. The relative end-time errors, naturally measured in the broken H^1-norm, are also presented in Figure 1, which shows the optimal convergence rates $\|u(T) - u_h(T)\|_{H^1(\Omega;\mathcal{T}_h)} \simeq h^p$. These results show that the method can deliver high accuracy despite the strong anisotropy of the problem and the very small value of the constant ε appearing in the Cordes condition.

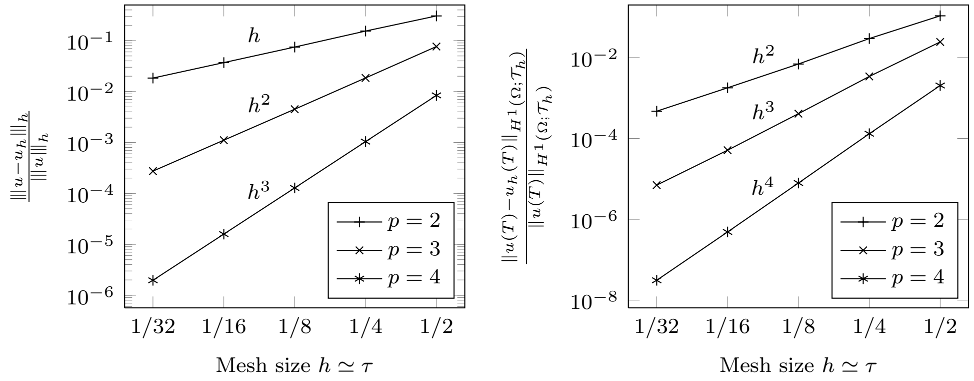

Fig. 1 Relative errors in approximating the solution of the problem of section 8.1 using uniform meshes and time partitions with $\tau \simeq h$ and $p = q + 1$. It is seen that the optimal convergence rates $|||u - u_h|||_h \simeq h^{p-1} + \tau^q$ are achieved. The final time error, as measured in the broken H^1-norm, also converges with the optimal rate $\|u(T) - u_h(T)\|_{H^1(\Omega;\mathcal{T}_h)} \simeq h^p$.

8.2 Second experiment

In section 7.2, we considered error bounds for solutions with limited regularity. The significance of these results stems from the fact that the solutions of many parabolic HJB equations possess limited regularity as a result of early-time singularities induced by the initial datum. This difficulty appears even in the simplest special case of the HJB equation (2.3), namely the heat equation: indeed, consider $\partial_t u = \Delta u$ in $\Omega \times (0,T)$, $\Omega = (0,1)^2$, with homogeneous lateral boundary condition $u = 0$ on $\partial\Omega \times (0,T)$ and initial datum $u_0(x,y) := x\,(1-x)\sin(\pi y)$. Then, the solution is

$$u(x,y,t) = \frac{4}{\pi^3}\sum_{k=1}^{\infty} \frac{1-(-1)^k}{k^3} \exp(-(k^2+1)\,\pi^2 t)\sin(k\,\pi x)\sin(\pi y). \tag{8.3}$$

It can be shown that for sufficiently small $t > 0$ and nonnegative integers σ and ℓ such that $2\sigma + \ell \geq 3$, we have $\|\partial_t^\sigma u\|_{X_\ell}^2 \simeq t^{-(2\sigma+\ell-5/2)}$, with the constants of these lower and upper bounds both depending on σ and ℓ, but not on t. Therefore, $u \notin H^1(I;H)$, rather $u \in H^{7/4-\delta}(I;L^2(\Omega)) \cap H^{5/4-\delta}(I;H_0^1(\Omega)) \cap H^{3/4-\delta}(I;H)$ for arbitrarily small $\delta > 0$. It is noted that a linear problem is chosen here so that the solution may be found explicity through (8.3). Nevertheless, this example exhibits many features that are typical of more general parabolic problems, so that the following results remain relevant to more general HJB equations.

Despite the limited regularity of the solution, accurate results can be obtained by using geometrically-graded time partitions with varying temporal polynomial degrees; see [25]. A combination of τq-refinement in time and hp-refinement in space can lead to a rate

$$|||u - u_h|||_h \lesssim \exp(-c_1\sqrt[3]{\mathrm{DoF}_x}) + \exp(-c_2\sqrt{\mathrm{DoF}_\tau}), \tag{8.4}$$

where $\mathrm{DoF}_x := \dim V_{h,\mathbf{p}}$, where $\mathrm{DoF}_\tau = \sum_{n=1}^{N}(q_n+1)$ is the number of degrees of freedom of the temporal finite element space, and where c_1 and c_2 are positive constants. We give here an experimental confirmation of these expectations.

The method is applied on a sequence of geometrically-graded partitions $\{\mathcal{J}_\tau\}_\tau$ constructed as follows. Let $T = 0.05$, and let $t_n = \sigma^{N-n}\,T$ for $n = 1,\dots,N$, for a chosen $\sigma \in (0,1)$, and $N = 2,\dots,6$. As suggested in [25], we choose $\sigma = 0.2$. The temporal polynomial degrees are linearly increasing with n, with $q_n := n+1$. We choose T to be small, because in practice it is natural to use τq-refinement on a small initial time segment, and then apply uniform

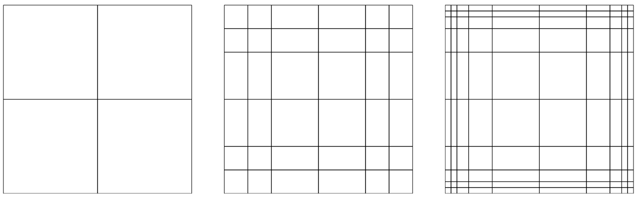

Fig. 2 Geometrically-graded spatial meshes used in conjunction with the geometrically-graded temporal meshes for the problem of section 8.2. From left to right, the meshes are those used for the first, third and fifth computations. The corresponding number of spatial degrees of freedom DoF_x are respectively 100, 1128, and 3980.

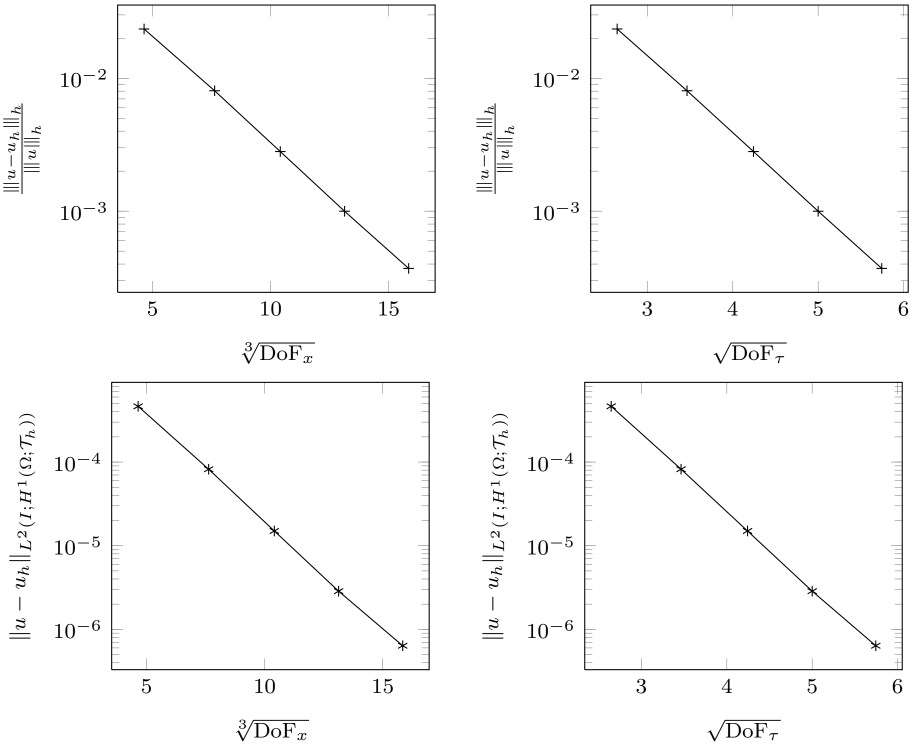

Fig. 3 Exponential convergence rates under hp-τq refinement for the problem of section 8.2. The errors in the norms $|||\cdot|||_h$ and $\|\cdot\|_{L^2(I;H^1(\Omega;\mathcal{T}_h))}$ are plotted against $\sqrt[3]{\mathrm{DoF}_x}$ and $\sqrt{\mathrm{DoF}_\tau}$, where DoF_x is the number of spatial degrees of freedom and DoF_τ is the number of temporal degrees of freedom. Exponential convergence rates of the form of (8.4) are confirmed.

or spectral refinement on the remaining time interval, see [25]. The spatial meshes are defined as follows: starting with a regular partition of Ω into four quadrilateral elements, for each successive computation, we refine the meshes geometrically towards the boundary, thereby yielding the meshes shown in Figure 2. The polynomial degrees $p_K \geq 3$ are chosen to be linearly increasing away from the boundary. Figure 3 presents the resulting errors in the norms $|||\cdot|||_h$ and $\|\cdot\|_{L^2(I;H^1(\Omega;\mathcal{T}_h))}$, plotted against $\sqrt[3]{\mathrm{DoF}_x}$ and $\sqrt{\mathrm{DoF}_\tau}$. It is found that the convergence rates of (8.4) are attained, with higher accuracies being achieved in lower order norms. These results show the computational efficiency of the method for problems with limited regularity.

9 Conclusion

We have introduced and analysed a fully discrete hp- and τq-version DGFEM for parabolic HJB equations with Cordes coefficients. The method is consistent and unconditionally stable, with proven convergence rates. The numerical experiments demonstrated the efficiency and accuracy of the method on problems with strongly anisotropic diffusion coefficients, and illustrated exponential convergence rates for solutions with limited regularity under hp- and τq-refinement.

References

1. Akrivis, G., Makridakis, C.: Galerkin time-stepping methods for nonlinear parabolic equations. M2AN Math. Model. Numer. Anal. **38**(2), 261–289 (2004).
2. Barles, G., Souganidis, P.: Convergence of approximation schemes for fully nonlinear second-order equations. Asymptotic Anal. **4**(3), 271–283 (1991).
3. Bonnans, J.F., Zidani, H.: Consistency of generalized finite difference schemes for the stochastic HJB equation. SIAM J. Numer. Anal. **41**(3), 1008–1021 (2003).
4. Caffarelli, L.A., Cabré, X.: Fully nonlinear elliptic equations, *American Mathematical Society Colloquium Publications*, vol. 43. American Mathematical Society, Providence, RI (1995).
5. Caffarelli, L.A., Silvestre, L.: On the Evans Krylov Theorem. Proceedings of the American Mathematical Society **138**(1), 263–265 (2009).
6. Camilli, F., Falcone, M.: An approximation scheme for the optimal control of diffusion processes. RAIRO Modél. Math. Anal. Numér. **29**, 97–122 (1995).
7. Cordes, H.O.: Über die erste Randwertaufgabe bei quasilinearen Differentialgleichungen zweiter Ordnung in mehr als zwei Variablen. Math. Ann. **131**, 278–312 (1956).
8. Crandall, M.G., Lions, P.L.: Convergent difference schemes for nonlinear parabolic equations and mean curvature motion. Numer. Math. **75**(1), 17–41 (1996).
9. Debrabant, K., Jakobsen, E.R.: Semi-Lagrangian schemes for linear and fully nonlinear diffusion equations. Math. Comp. **82**(283), 1433–1462 (2013).
10. Evans, L.C.: Classical solutions of the Hamilton–Jacobi–Bellman equation for uniformly elliptic operators. Transactions of the American Mathematical Society **275**(1), 245–255 (2008).
11. Feng, X., Glowinski, R., Neilan, M.: Recent developments in numerical methods for fully nonlinear second order partial differential equations. SIAM Rev. **55**(2), 205–267 (2013).
12. Fleming, W.H., Soner, H.M.: Controlled Markov processes and viscosity solutions, *Stochastic Modelling and Applied Probability*, vol. 25, second edn. Springer, New York (2006).
13. Grisvard, P.: Elliptic problems in nonsmooth domains, *Classics in Applied Mathematics*, vol. 69. SIAM, Philadelphia (2011).
14. Jensen, M., Smears, I.: On the convergence of finite element methods for Hamilton–Jacobi–Bellman equations. SIAM Journal on Numerical Analysis **51**(1), 137–162 (2013).
15. Kocan, M.: Approximation of viscosity solutions of elliptic partial differential equations on minimal grids. Numer. Math. **72**(1), 73–92 (1995).
16. Krylov, N.V.: Boundedly inhomogeneous elliptic and parabolic equations. Izv. Akad. Nauk SSSR Ser. Mat. **46**(3), 487–523, 670 (1982).
17. Kuo, H.J., Trudinger, N.S.: Discrete methods for fully nonlinear elliptic equations. SIAM J. Numer. Anal. **29**(1), 123–135 (1992).
18. Kushner, H.J.: Numerical methods for stochastic control problems in continuous time. SIAM J. Control Optim. **28**(5), 999–1048 (1990).
19. Lakkis, O., Pryer, T.: A finite element method for second order nonvariational elliptic problems. SIAM J. Sci. Comput. **33**(2), 786–801 (2011).
20. Lakkis, O., Pryer, T.: A finite element method for nonlinear elliptic problems. SIAM J. Sci. Comput. **35**(4), A2025–A2045 (2013).
21. Maugeri, A., Palagachev, D.K., Softova, L.G.: Elliptic and parabolic equations with discontinuous coefficients, *Mathematical Research*, vol. 109. Wiley-VCH Verlag Berlin GmbH, Berlin (2000).
22. Motzkin, T.S., Wasow, W.: On the approximation of linear elliptic differential equations by difference equations with positive coefficients. J. Math. Physics **31**, 253–259 (1953).
23. Mozolevski, I., Süli, E., Bösing, P.R.: hp-version a priori error analysis of interior penalty discontinuous Galerkin finite element approximations to the biharmonic equation. J. Sci. Comput. **30**(3), 465–491 (2007).
24. Renardy, M., Rogers, R.C.: An introduction to partial differential equations, *Texts in Applied Mathematics*, vol. 13, second edn. Springer-Verlag, New York (2004).
25. Schötzau, D., Schwab, C.: Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method. SIAM J. Numer. Anal. **38**(3), 837–875 (2000).

26. Smears, I., Süli, E.: Discontinuous Galerkin finite element approximation of nondivergence form elliptic equations with Cordes coefficients. SIAM J. Numer. Anal. **51**, 2088–2106 (2013).
27. Smears, I., Süli, E.: Discontinuous Galerkin finite element approximation of Hamilton–Jacobi–Bellman equations with Cordes coefficients. SIAM J. Numer. Anal. **52**(2), 993–1016 (2014).
28. Thomée, V.: Galerkin finite element methods for parabolic problems, *Springer Series in Computational Mathematics*, vol. 25, second edn. Springer-Verlag, Berlin (2006).
29. Wang, L.: On the regularity theory of fully nonlinear parabolic equations. I. Comm. Pure Appl. Math. **45**(1), 27–76 (1992).
30. Wloka, J.: Partial differential equations. Cambridge University Press, Cambridge (1987).