SIAM J. NUMER. ANAL.
Vol. 51, No. 1, pp. 137–162

ON THE CONVERGENCE OF FINITE ELEMENT METHODS FOR HAMILTON–JACOBI–BELLMAN EQUATIONS*

MAX JENSEN† AND IAIN SMEARS‡

Abstract. We study the convergence of monotone $P1$ finite element methods on unstructured meshes for fully nonlinear Hamilton–Jacobi–Bellman equations arising from stochastic optimal control problems with possibly degenerate, isotropic diffusions. Using elliptic projection operators we treat discretizations which violate the consistency conditions of the framework by Barles and Souganidis. We obtain strong uniform convergence of the numerical solutions and, under nondegeneracy assumptions, strong L^2 convergence of the gradients.

Key words. finite element methods, partial differential equations, Hamilton–Jacobi–Bellman equations, viscosity solutions

AMS subject classifications. 65M60, 65M12, 35D40, 35K65, 35K55

DOI. 10.1137/110856198

1. Introduction.

Hamilton–Jacobi–Bellman (HJB) equations, which are of the form

$$-\partial_t v + \sup_{\alpha}(L^{\alpha} v - f^{\alpha}) = 0, \tag{1.1}$$

where the L^{α} are linear first- or second-order operators and $f^{\alpha} \in L^{\infty}$, characterize the value function of optimal control problems. Indeed, one possibility to introduce the notion of solution of (1.1) is via the underlying optimal control structure. An alternative approach is to use the monotonicity properties of the operator, which leads to the concept of viscosity solutions. While these perceptions are essentially equivalent [19, p. 72], both views have been instructive for the design and analysis of numerical methods.

The former approach, based on the discretization of the optimal control problem before employing the dynamic programming principle, has been proposed in the setting of finite elements in [33, 9, 10]; see also the review article [26] and the references therein. Regarding finite difference methods we refer to the book [27]. The latter approach, which is also adopted in this note, was firmly established with the contribution [3] by Barles and Souganidis in 1991, providing an abstract framework for the convergence to viscosity solutions. Starting with [24, 25], techniques were developed to quantify the rate of convergence; more recent works are [1, 16]. A third direction was opened by the method of vanishing moments which neither enforces discrete maximum principles nor makes use of the underlying optimal control structure but relies on a higher-order regularization [18]. For a more comprehensive review of the state-of-the-art in the numerical solution of fully nonlinear second-order equations we refer to [17]; see also recent results in [5, 6, 28, 31].

It is helpful to briefly recall the convergence argument in [3], formulated there in the setting of finite difference methods. Consider a sequence of abstract numerical

*Received by the editors November 22, 2011; accepted for publication (in revised form) November 7, 2012; published electronically January 17, 2013.
http://www.siam.org/journals/sinum/51-1/85619.html
†Department of Mathematical Sciences, University of Durham, Durham DH1 3LE, England (m.p.j.jensen@durham.ac.uk).
‡Mathematical Institute, University of Oxford, Oxford OX1 3LB, England (iain.smears@maths.ox.ac.uk).

schemes $F_i[v_i](s_i^k, y_i^\ell) = 0$ with numerical solutions v_i, where $\{y_i^\ell\}_\ell$ is the set of nodes or grid points and $\{s_i^k\}_k$ is the set of time levels of the ith refinement level. Under a stability condition, one can define the upper envelope v^* of the sequence v_i by

$$v^*(t,x) = \sup_{(s_i^k, y_i^\ell) \to (t,x)} \limsup_{i\to\infty} v_i(s_i^k, y_i^\ell).$$

One can also define the lower envelope v_* analogously to v^* by replacing the suprema with infima. Clearly, $v_* \le v^*$. It is shown in [3] that if $v^* - w$ has a strict local maximum for a smooth function w, then $v_i - \mathcal{I}_i w$ also has a strict local maximum at a nearby node (s_i^k, y_i^ℓ) for $i \in \mathbb{N}$ and with the nodal interpolation operator $\mathcal{I}_i$. A monotonicity assumption implies $0 = F_i[v_i](s_i^k, y_i^\ell) \ge F_i[\mathcal{I}_i w](s_i^k, y_i^\ell)$. Now the consistency condition at the point (t,x)

$$F_i[\mathcal{I}_i w](s_i^k, y_i^\ell) \to -w_t(t,x) + Hw(t,x) \tag{1.2}$$

implies that $-w_t(t,x) + Hw(t,x) \le 0$, where in our context the Hamiltonian H is defined pointwise by $Hw = \sup_\alpha(L^\alpha w - f^\alpha)$. Therefore, v^* is a subsolution. A similar argument shows that v_* is a supersolution. Finally, with a comparison principle, subsolutions are bounded from above by supersolutions; so $v^* \le v_*$, which gives convergence. To have a comparison principle, it is usual that the convergence properties on the parabolic boundary need to be studied.

Condition (1.2) raises the question of how to enforce consistency in a finite element setting. In the traditional finite element analysis, the multiplicative testing with hat functions is viewed as the discrete analogue of the multiplicative testing procedure to define weak solutions of the (variational) differential equation. While elements of this viewpoint are implicitly used in section 7 on gradient convergence, we would like to stress a second interpretation: multiplication with hat functions as regularization of the residual. Consider for a moment the linear problem $-a(x)\Delta u(x) = f(x)$ with smooth functions a and u as well as a hat function ϕ at the node y^ℓ. Let P be the orthogonal projection onto the approximation space with respect to the scalar product $\langle v, w\rangle = \int \nabla v \cdot \nabla w \,\mathrm{d}x$ (given suitable boundary conditions). If y is near y^ℓ, then on a fine mesh

$$\begin{aligned}-a(y)\Delta u(y) &= -\int a(y)\Delta u(y)\,\hat\phi(x)\,\mathrm{d}x \approx -a(y^\ell)\int \Delta u(x)\,\hat\phi(x)\,\mathrm{d}x\\ &= a(y^\ell)\int \nabla u(x)\cdot\nabla\hat\phi(x)\,\mathrm{d}x = a(y^\ell)\int \nabla P u(x)\cdot\nabla\hat\phi(x)\,\mathrm{d}x,\end{aligned}$$

since $\hat\phi := \phi/\|\phi\|_{L^1(\Omega)}$ approximates a Dirac delta as the element size is decreased. In contrast, on general meshes,

$$-a(y)\Delta u(y) \not\approx a(y^\ell)\int \nabla\mathcal{I}_i u(x)\cdot\nabla\hat\phi(x)\,\mathrm{d}x \qquad (\mathcal{I}_i \text{ nodal interpolant})$$

even in the limit as the mesh is refined (see Example 1 below). This indicates that the orthogonality properties of the projection of the exact solution into the approximation space play an important role for the understanding of the (pointwise) consistency of the finite element scheme. Furthermore, this interpretation may serve as a starting point in selecting a discretization of the HJB operator.

Viscosity solutions are a mathematical concept to select the value function v from the (possibly infinite) set of weak solutions of the HJB equation. Once the convergence to the viscosity solution is guaranteed, the attention turns to other convergence properties. For a fixed α and nonnegative v and f^α one uses (1.1) with

$$-\partial_t v + L^\alpha v - f^\alpha \leq 0 \implies \langle -\partial_t v, v\rangle + \langle L^\alpha v, v\rangle \leq \langle f^\alpha, v\rangle,$$

where $\langle \cdot, \cdot\rangle$ denotes an L^2 scalar product. Finite element methods whose test and trial spaces coincide lend themselves well to exploit this variational inequality together with coercivity properties of L^α to control gradient terms of v on unstructured meshes.

Our analysis combines the following features in a single finite element framework.

Treatment of nodally inconsistent discretizations and uniform convergence. The consistency condition (see [3, eq. (2.4)] or [19, p. 332]) of Barles and Souganidis is based on a limit involving pointwise values of smooth test functions. This condition is not satisfied by finite element methods, even for linear equations. Using an alternative consistency condition, we show the uniform convergence of finite element solutions to the viscosity solution.

Gradient convergence. For problems with coercive linear operators under the supremum, we demonstrate how the coercivity is recovered by the finite element method in order to control the gradient of the numerical solutions. In a uniformly parabolic setting, this leads to strong convergence in $L^2([0,T], H^1(\Omega))$.

Operators of nonnegative characteristic form. The analysis includes the treatment of HJB equations arising from partially and fully deterministic optimal control problems that correspond to degenerate elliptic operators under the supremum of the Hamiltonian.

Unstructured meshes. In the spirit of finite element methods the computational domain may be triangulated with an unstructured mesh, allowing the capture of complex domains more easily than in a finite difference setting. Typically, weaker conditions on the mesh than quasiuniformity can be made.

Regularization with second-order operators. We highlight that the regularization with second-order elliptic operators is sufficient to achieve convergence to the viscosity solution. Indeed, in the example of the method of artificial diffusion, we illustrate how the regularization in the second-order fully nonlinear case is of the same kind and order as for first-order linear operators.

Unconditional time step size. Our analysis permits explicit, semi-implicit and fully implicit discretizations in time. Fully implicit discretizations in time lead to unconditionally stable schemes.

The structure of the article is as follows: in section 2, we introduce a framework of monotone finite element methods for HJB equations. In section 3, we study the well-posedness of the discrete systems of equations and describe how these systems are solvable by a known globally convergent algorithm with local superlinear convergence. Section 4 establishes the consistency properties of the scheme with respect to elliptic projection operators. This enables us to demonstrate in section 5 that the upper and lower envelopes of the numerical solutions are sub- and supersolutions. Uniform convergence to the viscosity solution is derived in section 6 and is then built upon to analyze the convergence of the gradient in section 7. We provide a concrete specimen of a scheme belonging to our framework by describing the method of artificial diffusion in section 8. The scheme is put into practice in section 9, which presents the results of a numerical test of the convergence rates.

2. Problem statement and definition of the numerical scheme. Let Ω be a bounded Lipschitz domain in $\mathbb{R}^d$, $d \geq 2$. Let A be a compact metric space and let

$$A \to C(\overline{\Omega}) \times C(\overline{\Omega}, \mathbb{R}^d) \times C(\overline{\Omega}) \times C(\overline{\Omega}),\ \alpha \mapsto (a^\alpha, b^\alpha, c^\alpha, f^\alpha)$$

be continuous, such that the families of functions $\{a^\alpha\}_{\alpha\in A}$, $\{b^\alpha\}_{\alpha\in A}$, $\{c^\alpha\}_{\alpha\in A}$, and $\{f^\alpha\}_{\alpha\in A}$ are equicontinuous. Consider the bounded linear operators

$$L^\alpha :\ H^2(\Omega) \to L^2(\Omega),\ w \mapsto -a^\alpha\,\Delta w + b^\alpha \cdot \nabla w + c^\alpha\, w, \qquad \alpha \in A.$$

We assume that $a^\alpha \geq 0$, i.e., that all L^α are of nonnegative characteristic form [32]. Furthermore, suppose that pointwise $f^\alpha \geq 0$. Then

$$\sup_{\alpha\in A} \| \,(a^\alpha, b^\alpha, c^\alpha, f^\alpha)\, \|_{C(\overline{\Omega})\times C(\overline{\Omega},\mathbb{R}^d)\times C(\overline{\Omega})\times C(\overline{\Omega})} < \infty, \tag{2.1}$$

and also $\sup_{\alpha\in A}\|L^\alpha\|_{H^2(\Omega)\to L^2(\Omega)} < \infty$. Let the final-time data $v_T \in C(\overline{\Omega})$ be nonnegative, $v_T \geq 0$ on $\overline{\Omega}$, and let v_T satisfy homogeneous boundary conditions on $\partial\Omega$. For smooth w, let

$$Hw := \sup_\alpha (L^\alpha w - f^\alpha),$$

where the supremum is applied pointwise. The HJB equation considered is

$$-\partial_t v + Hv = 0 \qquad \text{in } (0,T)\times\Omega, \tag{2.2a}$$

$$v = 0 \qquad \text{on } (0,T)\times\partial\Omega, \tag{2.2b}$$

$$v = v_T \qquad \text{on } \{T\}\times\overline{\Omega}. \tag{2.2c}$$

Definition 2.1 (see [2, 19]). *An upper semicontinuous (lower semicontinuous) function $v : [0,T]\times\overline{\Omega}\to\mathbb{R}$ is a viscosity subsolution (supersolution) of*

$$-\partial_t v + Hv = 0 \quad \textit{on } (0,T)\times\Omega \tag{2.3}$$

if for any $w \in C^\infty(\mathbb{R}\times\mathbb{R}^d)$ such that $v - w$ has a strict local maximum (minimum) at $(t,x)\in(0,T)\times\Omega$ with $v(t,x) = w(t,x)$, it gives $-\partial_t w(t,x) + Hw(t,x) \leq 0$ (greater than or equal to 0). If $v \in C([0,T]\times\overline{\Omega})$ is both a viscosity subsolution and a supersolution of (2.3)*, then v is called a viscosity solution.*

The viscosity solution of (2.2) is understood to be a viscosity solution of the PDE (2.2a), in the sense of Definition 2.1, that satisfies pointwise the boundary conditions (2.2b) and (2.2c); see also Assumption 6.1 below.

2.1. The numerical scheme. We now specify a class of discretizations of the HJB equation that permit explicit and implicit schemes as well as regularization and approximation of the data. The conditions required for the analysis of the scheme are stated in Assumptions 2.1 and 2.2 below. Section 8 provides an example of a concrete method for putting this framework into practice.

Let V_i, $i\in\mathbb{N}$, be a sequence of piecewise linear shape-regular finite element spaces with nodes y_i^ℓ. Here ℓ is the index ranging over the nodes of the finite element mesh. Let $V_i^0 \subset V_i$ be the subspace of functions which satisfy homogeneous Dirichlet conditions on $\partial\Omega$. It is convenient to assume that $y_i^\ell \in \Omega$ for $\ell \leq N_i := \dim V_i^0$, i.e.,the index ℓ first ranges over internal nodes and then over boundary nodes. The associated hat functions are denoted ϕ_i^ℓ, that is, $\phi_i^\ell \in V_i$ and $\phi_i^\ell(y_i^l) = 1$ if $l = \ell$, otherwise $\phi_i^\ell(y_i^l) = 0$. Set $\hat{\phi}_i^\ell := \phi_i^\ell/\|\phi_i^\ell\|_{L^1(\Omega)}$. Thus, the ϕ_i^ℓ are normalized in the L^∞ norm while the $\hat{\phi}_i^\ell$ are normalized in the L^1 norm. The mesh size, i.e., the largest diameter of an element, is denoted Δx_i. It is assumed that $\Delta x_i \to 0$ as $i\to\infty$.

Let h_i be the (uniform) time step size used in conjunction with V_i with $T/h_i \in \mathbb{N}$, and let s_i^k be the kth time step at the refinement level i. It is assumed that $h_i \to 0$

as $i \to \infty$. The set of time steps is $S_i := \{s_i^k : k = 0, \dots, T/h_i\}$. Let the ℓth entry of $d_i w(s_i^k, \cdot)$ be

$$(d_i w(s_i^k, \cdot))_\ell = \frac{w(s_i^{k+1}, y_i^\ell) - w(s_i^k, y_i^\ell)}{h_i}.$$

For each α and i, we introduce operators E_i^α and I_i^α to break L^α into an explicit and an implicit part:

$$\begin{aligned} E_i^\alpha &: \; H^2(\Omega) \to L^2(\Omega), \; w \mapsto -\bar{a}_i^\alpha \, \Delta w + \bar{b}_i^\alpha \cdot \nabla w + \bar{c}_i^\alpha \, w, \\ I_i^\alpha &: \; H^2(\Omega) \to L^2(\Omega), \; w \mapsto -\bar{\bar{a}}_i^\alpha \, \Delta w + \bar{\bar{b}}_i^\alpha \cdot \nabla w + \bar{\bar{c}}_i^\alpha \, w \end{aligned}$$

with continuous

$$\begin{aligned} A &\to C(\overline{\Omega}) \times C(\overline{\Omega}, \mathbb{R}^d) \times C(\overline{\Omega}), \quad \alpha \mapsto (\bar{a}_i^\alpha, \bar{b}_i^\alpha, \bar{c}_i^\alpha), \\ A &\to C(\overline{\Omega}) \times C(\overline{\Omega}, \mathbb{R}^d) \times C(\overline{\Omega}), \quad \alpha \mapsto (\bar{\bar{a}}_i^\alpha, \bar{\bar{b}}_i^\alpha, \bar{\bar{c}}_i^\alpha). \end{aligned} \tag{2.4}$$

It is required that $\bar{c}_i^\alpha$ and $\bar{\bar{c}}_i^\alpha$ are nonnegative and that there is $\gamma \in \mathbb{R}$ such that

$$\|\bar{c}_i^\alpha\|_{L^\infty} + \|\bar{\bar{c}}_i^\alpha\|_{L^\infty} \le \gamma \qquad \forall\, i \in \mathbb{N}, \; \forall\, \alpha \in A. \tag{2.5}$$

Also, find for each i a nonnegative f_i^α which approximates f^α: $f_i^\alpha \approx f^\alpha$. The conceptual statements $L^\alpha \approx E_i^\alpha + I_i^\alpha$ and $f^\alpha \approx f_i^\alpha$ are made precise as follows.

Assumption 2.1. For all sequences of nodes $(y_i^\ell)_{i \in \mathbb{N}}$, where in general $\ell = \ell(i)$ depends on i,

$$\begin{aligned} \lim_{i \to \infty} \sup_{\alpha \in A} \Big(\big\| a^\alpha - \big(\bar{a}_i^\alpha(y_i^\ell) + \bar{\bar{a}}_i^\alpha(y_i^\ell) \big) \big\|_{L^\infty(\operatorname{supp} \hat{\phi}_i^\ell)} + \big\| b^\alpha - \big(\bar{b}_i^\alpha + \bar{\bar{b}}_i^\alpha \big) \big\|_{L^\infty(\Omega, \mathbb{R}^d)} \\ + \big\| c^\alpha - \big(\bar{c}_i^\alpha + \bar{\bar{c}}_i^\alpha \big) \big\|_{L^\infty(\Omega)} + \big\| f^\alpha - f_i^\alpha \big\|_{L^\infty(\Omega)} \Big) = 0. \end{aligned}$$

Let $\langle \cdot, \cdot \rangle$ denote the standard inner product for both of the spaces $L^2(\Omega)$ and $L^2(\Omega, \mathbb{R}^d)$, the two cases being distinguished by the arguments of the inner product. The operators E_i^α and I_i^α are in nondivergence form with the highest-order term having the form $-a(x)\Delta w$ with a continuous function a. We obtain a discretization that is consistent in the sense needed for the analysis by approximating

$$-a(x)\Delta w(x) \approx -a(y_i^\ell)\langle \Delta w, \hat{\phi}_i^\ell \rangle = a(y_i^\ell)\langle \nabla w, \nabla \hat{\phi}_i^\ell \rangle$$

for w sufficiently smooth and y_i^ℓ close to x—this corresponds to "freezing" the coefficient before integrating by parts. This approach leads to the following discretization of E_i^α and I_i^α by operators E_i^α and I_i^α that map $H^1(\Omega)$ to $\mathbb{R}^{N_i}$: for $w \in H^1(\Omega)$, $\ell \in \{1, \dots, N_i = \dim V_i^0\}$,

$$(\mathsf{E}_i^\alpha w)_\ell := \bar{a}_i^\alpha(y_i^\ell)\langle \nabla w, \nabla \hat{\phi}_i^\ell \rangle + \langle \bar{b}_i^\alpha \cdot \nabla w + \bar{c}_i^\alpha \, w, \hat{\phi}_i^\ell \rangle, \tag{2.6a}$$

$$(\mathsf{I}_i^\alpha w)_\ell := \bar{\bar{a}}_i^\alpha(y_i^\ell)\langle \nabla w, \nabla \hat{\phi}_i^\ell \rangle + \langle \bar{\bar{b}}_i^\alpha \cdot \nabla w + \bar{\bar{c}}_i^\alpha w, \hat{\phi}_i^\ell \rangle, \tag{2.6b}$$

$$(\mathsf{F}_i^\alpha)_\ell := \langle f_i^\alpha, \hat{\phi}_i^\ell \rangle. \tag{2.6c}$$

Throughout this work, we identify E_i^α and I_i^α, when restricted to V_i, with their matrix representations with respect to the nodal basis $\{\phi_i^\ell\}_\ell$. Under this basis, the nodal evaluation operator $w \mapsto w(y_i^\ell)$ corresponds to the identity matrix Id.

We will make use of the partial ordering of $\mathbb{R}^n$: for x, $y \in \mathbb{R}^n$, we write $x \geq y$ if and only if $x_\ell \geq y_\ell$ for all $\ell \in \{1, \ldots, n\}$. For a collection $\{x^\alpha\}_\alpha \subset \mathbb{R}^n$, we define the operator $\sup_\alpha$ componentwise: $(\sup_\alpha x^\alpha)_\ell = \sup_\alpha x^\alpha_\ell$.

We now define the numerical scheme for (2.2). Define the numerical solution $v_i(T, \cdot) \in V_i^0$ by nodal interpolation of v_T. Then, for each $k \in \{0, \ldots, T/h_i - 1\}$, the numerical solution $v_i(s_i^k, \cdot) \in V_i^0$ is defined inductively by

$$-d_i v_i(s_i^k, \cdot) + \sup_{\alpha \in A} \big(\mathsf{E}_i^\alpha v_i(s_i^{k+1}, \cdot) + \mathsf{I}_i^\alpha v_i(s_i^k, \cdot) - \mathsf{F}_i^\alpha \big) = 0. \tag{2.7}$$

If all I_i^α vanish, then (2.7) is an explicit scheme; otherwise it is implicit.

2.2. Monotonicity. Monotonicity of the numerical scheme is important for the proof of convergence to the viscosity solution.

DEFINITION 2.2. *An operator $F : V_i \to \mathbb{R}^{N_i}$ is said to satisfy the local monotonicity property (LMP) if for all $v \in V_i$ such that v has a nonpositive local minimum at the internal node y_i^ℓ, we have $(Fv)_\ell \leq 0$. The operator F satisfies the weak discrete maximum principle (wDMP) provided that for any $v \in V_i$,*

$$\textit{if } \big(Fv\big)_\ell \geq 0 \;\forall\, \ell \in \{1, \ldots, N_i\}, \qquad \textit{then } \min_\Omega v \geq \min\{\min_{\partial\Omega} v, 0\}. \tag{2.8}$$

More explicit alternative formulations of the wDMP are discussed, for example, in [7] and [8]. Note that the identity Id and the null operator $\mathsf{0}$ satisfy the LMP. It is clear that if F satisfies the LMP and $v \in V_i$ has a *negative* local minimum at the internal node y_i^ℓ, then $((F + \varepsilon\,\mathsf{Id})v)_\ell < 0$ for all $\varepsilon > 0$. This implies for all $\varepsilon > 0$ that $F + \varepsilon\,\mathsf{Id}$ satisfies the wDMP.

Assumption 2.2. For each $\alpha \in A$, assume that E_i^α, restricted to V_i, has nonpositive off-diagonal entries. Let h_i be small enough so that $h_i \mathsf{E}_i^\alpha - \mathsf{Id}$ is monotone for every α, i.e., so that all entries of all $h_i \mathsf{E}_i^\alpha - \mathsf{Id}$ are nonpositive. For each α, suppose that I_i^α satisfies the LMP.

Notice that the monotonicity assumption on $h_i \mathsf{E}_i^\alpha - \mathsf{Id}$ is a time step restriction if E_i^α has positive diagonal entries. If the scheme is fully implicit, i.e., all E_i^α vanish, then there is no time step restriction.

2.3. An alternative formulation of the numerical method. To study the well-posedness of the numerical scheme, it is useful to reformulate it first. For a function $w : S_i \times \Omega \to \mathbb{R}$ that satisfies $w(s_i^k, \cdot) \in H^1(\Omega)$ for all $s_i^k \in S_i$, let $\alpha_i^{\ell,k}(w)$ be a control $\alpha \in A$ which maximizes

$$\sup_\alpha \big(\mathsf{E}_i^\alpha w(s_i^{k+1}, \cdot) + \mathsf{I}_i^\alpha w(s_i^k, \cdot) - \mathsf{F}_i^\alpha \big)_\ell . \tag{2.9}$$

The cost and complexity of the local maximization process in (2.9) depends strongly on the application at hand. Fortunately, as pointed out by Fleming and Soner in [19, p. 331], many applications give rise to explicit formulas that greatly simplify this task.

Let $\mathsf{I}_i^{k,w}$ and $\mathsf{E}_i^{k,w}$ be the matrices whose ℓth row is equal to that of

$$\mathsf{I}_i^{\alpha_i^{\ell,k}(w)} \quad \text{and} \quad \mathsf{E}_i^{\alpha_i^{\ell,k}(w)},$$

respectively. Also let the ℓth entry of $\mathsf{F}_i^{k,w}$ be

$$\left(\mathsf{F}_i^{\alpha_i^{\ell,k}(w)} \right)_\ell .$$

Thus, informally speaking, the $\mathsf{E}_i^{k,w}$, $\mathsf{I}_i^{k,w}$, and $\mathsf{F}_i^{k,w}$ are gained by "reshuffling" the rows of the E_i^{α}, I_i^{α}, and F_i^{α}, respectively. Notice that the maximizing control in (2.9) may be nonunique. Where no ambiguity can arise, we simply write I_i^{w}, E_i^{w}, and F_i^{w} without explicitly referring to k.

These definitions lead to an equivalent formulation of the numerical scheme (2.7): for each $k \in \{0, 1, \ldots, T/h_i - 1\}$, v_i solves

$$(h_i \mathsf{I}_i^{k,v_i} + \mathsf{Id})\, v_i(s_i^k, \cdot) + (h_i \mathsf{E}_i^{k,v_i} - \mathsf{Id})\, v_i(s_i^{k+1}, \cdot) - h_i \mathsf{F}_i^{k,v_i} = 0. \tag{2.10}$$

We will now prove several properties of $h_i \mathsf{I}_i^{k,w} + \mathsf{Id}$ and $h_i \mathsf{E}_i^{k,w} - \mathsf{Id}$ that lead to the well-posedness of (2.10) and, equivalently, of the scheme (2.7). The following lemma shows that for linear operators on V_i^0, the wDMP turns into an M-matrix property.

LEMMA 2.3. *Consider a $w : S_i \times \Omega \to \mathbb{R}$ so that $w(s_i^k, \cdot) \in H^1(\Omega)$ for all $s_i^k \in S_i$. Then, the matrices $h_i \mathsf{E}_i^{k,w} - \mathsf{Id}$ are monotone and the matrices of $h_i \mathsf{I}_i^{k,w} + \mathsf{Id}$ restricted to V_i^0 are diagonally dominant M-matrices. For fixed w, the operators $v \mapsto \mathsf{I}_i^{k,w} v$ and $v \mapsto (h_i \mathsf{I}_i^{k,w} + \mathsf{Id})\, v$ satisfy, respectively, the LMP and wDMP.*

Proof. Monotonicity of $h_i \mathsf{E}_i^{k,w} - \mathsf{Id}$ is a straightforward consequence of the nonpositivity of the entries of $h_i \mathsf{E}_i^{\alpha} - \mathsf{Id}$ for all $\alpha \in A$. The LMP of I_i^{α} for the node y_i^{ℓ} only imposes a condition on the ℓth row of the matrix of I_i^{α}. Hence it is easily checked that the $\mathsf{I}_i^{k,w}$ and the $h_i \mathsf{I}_i^{k,w} + \mathsf{Id}$, which are composed row-wise from the I_i^{α} and $h_i \mathsf{I}_i^{\alpha} + \mathsf{Id}$, satisfy the LMP and wDMP, respectively, when all I_i^{α} satisfy the LMP.

The LMP also implies that the matrix representations of the I_i^{α} restricted to V_i^0 are weakly diagonally dominant for all $\alpha \in A$. This is because taking $v = -\sum_{\ell=1}^{N_i} \phi_i^{\ell}$ yields

$$0 \geq (\mathsf{I}_i^{\alpha} v)_{\ell} = -(\mathsf{I}_i^{\alpha})_{\ell\ell} - \sum_{j \neq \ell}^{N_i} (\mathsf{I}_i^{\alpha})_{\ell j},$$

using the fact that v attains a nonpositive minimum at each internal node. For $j \neq \ell$ the hat function ϕ_i^j attains a nonpositive minimum at y_i^{ℓ}, giving $(\mathsf{I}_i^{\alpha})_{\ell j} \leq 0$. This shows that

$$(\mathsf{I}_i^{\alpha})_{\ell\ell} - \sum_{j \neq \ell}^{N_i} \left| (\mathsf{I}_i^{\alpha})_{\ell j} \right| \geq 0.$$

Because $\mathsf{I}_i^{k,w}$ is composed of the rows of various I_i^{α}, it follows that $h_i \mathsf{I}_i^{k,w} + \mathsf{Id}$ restricted to V_i^0 is strictly diagonally dominant, and is thus invertible, and additionally satisfies the wDMP. Furthermore, since $(h_i \mathsf{I}_i^{k,w} + \mathsf{Id}) + \varepsilon\, \mathsf{Id}$ is similarly invertible for all $\varepsilon \geq 0$ and all off-diagonal entries are nonpositive, [20, p. 114] shows that $h_i \mathsf{I}_i^{k,w} + \mathsf{Id}$, restricted to V_i^0, is represented by an invertible M-matrix. ∎

COROLLARY 2.4. *The nonlinear operators $w \mapsto \mathsf{I}_i^{k,w} w$ and $w \mapsto (h_i \mathsf{I}_i^{k,w} + \mathsf{Id})\, w$ satisfy the LMP and wDMP, respectively. Moreover, $w \mapsto -(h_i \mathsf{E}_i^{k,w} - \mathsf{Id})\, w$ is positive: if $w \geq 0$, then $-(h_i \mathsf{E}_i^{k,w} - \mathsf{Id})\, w \geq 0$.*

3. Well-posedness of the numerical method and a solution algorithm. We record a constructive proof of existence of a solution $v_i \colon S_i \to V_i^0$ to (2.7) for all $k \in \{0, \ldots, T/h_i - 1\}$ which uses Algorithm 1, described below. This algorithm, which can be traced back to [21], is found in the continuous setting in [29] which provides the proof of convergence and existence of solutions. In [4] it is shown that in the discrete setting it is a semismooth Newton method that converges superlinearly. We also refer to [28] for recent results on Newton methods for fully nonlinear equations.

The algorithm to solve the nonlinear problem (2.7) at a given time level is the following.

ALGORITHM 1. *Given* $v_i(s_i^{k+1},\cdot) \in V_i^0$ *for* $k \in \{0,\dots,T/h_i-1\}$*, choose an arbitrary* $\alpha \in A$ *and find* $w_0 \in V_i^0$ *such that*

$$(h_i \mathsf{I}_i^\alpha + \mathsf{Id})\, w_0 = h_i \mathsf{F}_i^\alpha - (h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v_i(s_i^{k+1},\cdot).$$

For $m \in \{0,1,2,\dots\}$*, inductively find* $w_{m+1} \in V_i^0$ *such that*

$$(h_i \mathsf{I}_i^{w_m} + \mathsf{Id})\, w_{m+1} = h_i \mathsf{F}_i^{w_m} - (h_i \mathsf{E}_i^{w_m} - \mathsf{Id})\, v_i(s_i^{k+1},\cdot). \tag{3.1}$$

For each $\alpha \in A$, we define $v_i^\alpha \colon S_i \to V_i^0$ to be the numerical solution of the linear evolution problem associated to the control α with homogeneous Dirichlet conditions, that is, $v_i^\alpha(T,\cdot) = v_i(T,\cdot)$, the interpolant of v_T, and for each $k \in \{0,\dots,T/h_i-1\}$,

$$(h_i \mathsf{I}_i^\alpha + \mathsf{Id})\, v_i^\alpha(s_i^k,\cdot) = -(h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v_i^\alpha(s_i^{k+1},\cdot) + h_i \mathsf{F}_i^\alpha. \tag{3.2}$$

The wDMP for $h_i \mathsf{I}_i^\alpha + \mathsf{Id}$ implies that v_i^α is well-defined. The following result shows the well-posedness of the numerical scheme and relates v_i to v_i^α.

THEOREM 3.1. *There exists a unique numerical solution* $v_i \colon S_i \to V_i^0$ *that solves* (2.7) *and* (2.10)*. Moreover,* $0 \le v_i \le v_i^\alpha$ *for each* $\alpha \in A$*. Given* $v_i(s_i^{k+1},\cdot) \in V_i^0$ *for* $k \in \{0,\dots,T/h_i-1\}$*, the iterates of Algorithm* 1 *converge superlinearly to the unique solution* $v_i(s_i^k,\cdot)$ *of* (2.7)*, i.e.,* $w_m \to v_i(s_i^k,\cdot)$ *as* $m \to \infty$*.*

Proof. Bokanowski, Maroso, and Zidani [4, Thm. 2.1] show the existence and uniqueness of a solution $v_i(s_i^k,\cdot)$ given k and $v_i(s_i^{k+1},\cdot)$ and superlinear convergence of the algorithm: their Assumption (H1) is ensured by Lemma 2.3 and their Assumption (H2) is guaranteed by the continuity of the maps of (2.4) and the map $\alpha \mapsto f_i^\alpha$. The existence and uniqueness of a solution v_i is then obtained by induction over k.

We now show that $v_i \ge 0$ on $S_i \times \overline{\Omega}$ by induction over k. Recall that $v_T \ge 0$ on $\overline{\Omega}$, hence $v_i(T,\cdot) \ge 0$ since $v_i(T,\cdot)$ interpolates v_T. Now, suppose that $v_i(s_i^{k+1},\cdot) \ge 0$ on Ω for some $k \le T/h_i - 1$. Recall that all entries of $h_i \mathsf{E}_i^{v_i} - \mathsf{Id}$ are nonpositive and that all entries of $\mathsf{F}_i^{v_i}$ are nonnegative. Therefore, (2.10) shows that

$$(h_i \mathsf{I}_i^{v_i} + \mathsf{Id}) v_i(s_i^k,\cdot) = -(h_i \mathsf{E}_i^{v_i} - \mathsf{Id}) v_i(s_i^{k+1},\cdot) + h_i \mathsf{F}_i^{v_i} \ge 0.$$

We then deduce that $v_i(s_i^k,\cdot) \ge 0$ on $\overline{\Omega}$ by using inverse positivity of $h_i \mathsf{I}_i^{v_i} + \mathsf{Id}$, thus completing the inductive step.

Finally, we prove that $v_i \le v_i^\alpha$ for all $\alpha \in A$ by induction. Consider any $\alpha \in A$. First, $v_i(T,\cdot) = v_i^\alpha(T,\cdot)$ by definition of v_i and v_i^α. Now, for given k, assume that $v_i(s_i^{k+1},\cdot) \le v_i^\alpha(s_i^{k+1},\cdot)$. Then, the numerical scheme (2.7) implies that

$$(h_i \mathsf{I}_i^\alpha + \mathsf{Id})\, v_i(s_i^k,\cdot) \le h_i \mathsf{F}_i^\alpha - (h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v_i(s_i^{k+1},\cdot).$$

After subtracting (3.2) from the above inequality, we see that monotonicity of $h_i \mathsf{E}_i^\alpha - \mathsf{Id}$ gives

$$(h_i \mathsf{I}_i^\alpha + \mathsf{Id}) \left(v_i(s_i^k,\cdot) - v_i^\alpha(s_i^k,\cdot) \right) \le (h_i \mathsf{E}_i^\alpha - \mathsf{Id}) \left(v_i^\alpha(s_i^{k+1},\cdot) - v_i(s_i^{k+1},\cdot) \right) \le 0.$$

Thus, by inverse positivity of $h_i \mathsf{I}_i^\alpha + \mathsf{Id}$, we conclude that $v_i(s_i^k,\cdot) \le v_i^\alpha(s_i^k,\cdot)$ on $\overline{\Omega}$, which completes the induction. $\square$

Monotonicity properties and mass-lumping are suitable conditions to enforce the L^∞ bounds of parabolic Galerkin methods; see, for instance, [34, Chap. 15]. With the next lemma we assure ourselves that these bounds also hold in our setting.

LEMMA 3.2. *For all $i \in \mathbb{N}$ one has $\|(h_i \mathsf{I}_i^\alpha + \mathsf{Id})^{-1}\|_\infty \leq 1$ and $\|h_i \mathsf{E}_i^\alpha - \mathsf{Id}\|_\infty \leq 1$, where the norms are the matrix ∞-norms.*

Proof. Define $v = \sum_{\ell=1}^{\dim V_i} \phi_i^\ell \equiv 1$, and $v_0 = \sum_{\ell=1}^{N_i} \phi_i^\ell \in V_i^0$. By Lemma 2.3, $h_i \mathsf{I}_i^\alpha + \mathsf{Id}$ is an invertible M-matrix on V_i^0. Thus, $(h_i \mathsf{I}_i^\alpha + \mathsf{Id})^{-1} \geq 0$ entrywise, so

$$\|(h_i \mathsf{I}_i^\alpha + \mathsf{Id})^{-1}\|_\infty = \max_{1 \leq \ell \leq N_i} \sum_{j=1}^{N_i} (h_i \mathsf{I}_i^\alpha + \mathsf{Id})^{-1}_{\ell j} = \max_{1 \leq \ell \leq N_i} \left((h_i \mathsf{I}_i^\alpha + \mathsf{Id})^{-1} \mathbf{1} \right)_\ell, \tag{3.3}$$

where $\mathbf{1} \in \mathbb{R}^{N_i}$ is the vector with all entries equal to 1. Since $\nabla v \equiv 0$ (as $v \equiv 1$) we have for each $1 \leq \ell \leq N_i$ that $((h_i \mathsf{I}_i^\alpha + \mathsf{Id})v)_\ell = 1 + h_i \langle \bar{\bar{c}}_i^\alpha, \hat{\phi}_i^\ell \rangle \geq 1$, where we have used nonnegativity of $\bar{\bar{c}}_i^\alpha$. Moreover, since $1 \leq \ell \leq N_i$ and I_i^α satisfies the LMP,

$$\left((h_i \mathsf{I}_i^\alpha + \mathsf{Id})\, v \right)_\ell = \left((h_i \mathsf{I}_i^\alpha + \mathsf{Id})\, v_0 \right)_\ell + \sum_{j=N_i+1}^{\dim V_i} \underbrace{(h_i \mathsf{I}_i^\alpha)_{\ell j}}_{\leq 0} \leq \left((h_i \mathsf{I}_i^\alpha + \mathsf{Id})\, v_0 \right)_\ell.$$

Because $(h_i \mathsf{I}_i^\alpha + \mathsf{Id})v \geq \mathbf{1}$, we obtain $(h_i \mathsf{I}_i^\alpha + \mathsf{Id})v_0 \geq \mathbf{1}$. So, after applying $(h_i \mathsf{I}_i^\alpha + \mathsf{Id})^{-1}$ to both sides of this inequality, inverse positivity of $h_i \mathsf{I}_i^\alpha + \mathsf{Id}$ gives $1 \equiv v \geq v_0 \geq (h_i \mathsf{I}_i^\alpha + \mathsf{Id})^{-1}\mathbf{1}$ on $\overline{\Omega}$. This inequality and (3.3) imply $\|(h_i \mathsf{I}_i^\alpha + \mathsf{Id})^{-1}\|_\infty \leq 1$.

One has $\|h_i \mathsf{E}_i^\alpha - \mathsf{Id}\|_\infty = \max_{1 \leq \ell \leq N_i} (-(h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v_0)_\ell$ because all entries of $h_i \mathsf{E}_i^\alpha - \mathsf{Id}$ are nonpositive. For each $1 \leq \ell \leq N_i$,

$$\left((h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v \right)_\ell = \left((h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v_0 \right)_\ell + \sum_{j=N_i+1}^{\dim V_i} (h_i \mathsf{E}_i^\alpha)_{\ell j} \leq \left((h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v_0 \right)_\ell,$$

so $(-(h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v_0)_\ell \leq (-(h_i \mathsf{E}_i^\alpha - \mathsf{Id})\, v)_\ell = 1 - h_i \langle \bar{c}_i^\alpha, \hat{\phi}_i^\ell \rangle \leq 1$ because $\bar{c}_i^\alpha \geq 0$. Therefore, $-(h_i \mathsf{E}_i^\alpha - \mathsf{Id})v_0 \leq \mathbf{1}$. So $\|h_i \mathsf{E}_i^\alpha - \mathsf{Id}\|_\infty \leq 1$. ∎

COROLLARY 3.3. *The numerical solutions v_i are uniformly bounded in the L^∞ norm. In particular, there is a finite $C > 0$ such that for all $i \in \mathbb{N}$ and $\alpha \in A$,*

$$\|v_i\|_{L^\infty(S_i \times \Omega)} \leq \|v_i^\alpha\|_{L^\infty(S_i \times \Omega)} \leq \|v_T\|_{L^\infty(\Omega)} + T\, \|f_i^\alpha\|_{L^\infty(\Omega)} \leq C.$$

Proof. Applying Lemma 3.2 to (3.2) shows that for each $k \in \{0, \ldots, T/h_i - 1\}$,

$$\|v_i^\alpha(s_i^k, \cdot)\|_{L^\infty(\Omega)} \leq \|v_i^\alpha(s_i^{k+1}, \cdot)\|_{L^\infty(\Omega)} + h_i \|\mathsf{F}_i^\alpha\|_\infty. \tag{3.4}$$

The definition of F_i^α in (2.6) gives $\|\mathsf{F}_i^\alpha\|_\infty \leq \|f_i^\alpha\|_{L^\infty(\Omega)}$. Induction over k shows that $\|v_i^\alpha\|_{L^\infty(S_i \times \Omega)}$ is bounded by $\|v_i^\alpha(T, \cdot)\|_{L^\infty(\Omega)} + T \|f_i^\alpha\|_{L^\infty(\Omega)}$. Recall that $v_i^\alpha(T, \cdot)$ is the interpolant of $v_T \in C(\overline{\Omega})$, so $\|v_i^\alpha(T, \cdot)\|_{L^\infty(\Omega)} \leq \|v_T\|_{L^\infty(\Omega)}$. By Assumption 2.1, $f_i^\alpha \to f^\alpha$ in $L^\infty(\Omega)$ uniformly in α. Finally, $0 \leq v_i \leq v_i^\alpha$ on $S_i \times \Omega$ by Theorem 3.1, so $\|v_i\|_{L^\infty(S_i \times \Omega)} \leq \|v_i^\alpha\|_{L^\infty(S_i \times \Omega)}$. ∎

4. Consistency properties of elliptic projections. The argument by Barles and Souganidis [3] takes advantage of the fact that classical finite difference methods are pointwise consistent (1.2) when applied to nodal interpolants of smooth functions. However, in the case of FEM, the nodal interpolant may fail to satisfy this consistency condition, even for reasonable meshes. Example 1 below illustrates this fact.

(a) Patch of consistent method (b) Patch of inconsistent method

FIG. 4.1. (a) *illustrates a mesh that leads to a FEM discretization of the Laplacian that is pointwise consistent with respect to the interpolant. This is no longer the case for the mesh depicted by* (b).

Therefore, in the context of finite element methods, the construction of an alternative to nodal interpolation becomes an essential step of the analysis.

Example 1. For a fixed point x in a domain, consider two sequences of meshes, such that the elements neighboring x are as depicted in Figure 4.1. Denote $\hat{\phi}_i$ and $\hat{\varphi}_i$ the L^1-normalized hat functions associated with the node x for the meshes depicted, respectively, by Figure 4.1(a) and Figure 4.1(b). Let w be a smooth function; let $\mathcal{I}_a w$ and $\mathcal{I}_b w$ be the nodal interpolants of w, respectively, on the two meshes. For the mesh of Figure 4.1(a), it is well known that the FEM discretization of the Laplacian coincides with a finite difference discretization and that

$$\langle \nabla \mathcal{I}_a w, \nabla \hat{\phi}_i \rangle = -\Delta w(x) + \mathsf{O}\big(\Delta x_i^2\big).$$

For the mesh of Figure 4.1(b), a simple calculation shows that

$$\langle \nabla \mathcal{I}_b w, \nabla \hat{\varphi}_i \rangle = -\frac{3}{2}\Delta w(x) + \mathsf{O}\big(\Delta x_i^2\big).$$

Therefore, the mesh type of 4.1(a) leads to a FEM discretization of the Laplacian that is strongly consistent with respect to interpolation, whereas the mesh type of 4.1(b) does not.

We overcome this difficulty by using a different projection operator in the Barles-Souganidis argument. Given $w \in C([0,T], H^1(\Omega))$, denote by $P_i w$ a linear mapping into $[0,T] \times V_i$ which satisfies for all $\hat{\phi}_i^\ell \in V_i^0$

$$\langle \nabla P_i w(t,\cdot), \nabla \hat{\phi}_i^\ell \rangle = \langle \nabla w(t,\cdot), \nabla \hat{\phi}_i^\ell \rangle \quad \forall t \in [0,T]. \tag{4.1}$$

Notice that P_i coincides with the classical elliptic projection of the Laplacian if $P_i w$ is chosen to interpolate w on the boundary.

Assumption 4.1. There are linear mappings P_i satisfying (4.1), and there is a constant $C \geq 0$ such that for every $w \in C^\infty(\mathbb{R}^d)$ and $i \in \mathbb{N}$,

$$\|P_i w\|_{W^{1,\infty}(\Omega)} \leq C \|w\|_{W^{1,\infty}(\Omega)} \quad \text{and} \quad \lim_{i\to\infty} \|P_i w - w\|_{W^{1,\infty}(\Omega)} = 0. \tag{4.2}$$

The settings under which the above assumption holds for the elliptic projection typically include a condition on the mesh grading and on the domain. In [15], it is shown that (4.1) holds when Ω is a bounded convex polyhedral domain in $\mathbb{R}^d$, $d \in \{2,3\}$, when the mesh satisfies a local quasi-uniformity condition and when the test functions vanish on the boundary. To apply the result for nonconvex domains Ω and general $w \in C^\infty(\mathbb{R} \times \mathbb{R}^d)$, consider for example a convex polyhedral domain B

containing Ω and assume there is a locally quasi-uniform mesh on B which coincides with the original mesh on Ω. Let η be a smooth cutoff function with compact support in B such that $\eta \equiv 1$ on Ω. Then the classical elliptic projection on B, acting on $\eta w : B \to \mathbb{R}$, has the required properties. Given this construction for P_i, it is natural to refer to it as an elliptic projection.

LEMMA 4.1. *Let $w \in C^\infty(\mathbb{R} \times \mathbb{R}^d)$ and let $s_i^{k(i)} \to t \in [0,T)$ as $i \to \infty$. Then*

$$\lim_{i\to\infty} d_i P_i w(s_i^{k(i)}, \cdot) = \partial_t w(t, \cdot) \text{ in } W^{1,\infty}(\Omega). \tag{4.3}$$

The proof is left to the reader; the main trick is to use the triangle inequality and the identity $d_i P_i w = P_i d_i w$, followed by stability and convergence of P_i.

LEMMA 4.2. *Let $w \in C^\infty(\mathbb{R} \times \mathbb{R}^d)$ and let $s_i^{k(i)} \to t \in [0,T]$, $y_i^{\ell(i)} \to x \in \Omega$ as $i \to \infty$. Then*

$$\lim_{i\to\infty} \Big(\mathsf{E}_i^\alpha P_i w(s_i^{k(i)+1}, \cdot) + \mathsf{I}_i^\alpha P_i w(s_i^{k(i)}, \cdot) - \mathsf{F}_i^\alpha \Big)_{\ell(i)} = L^\alpha w(t,x) - f^\alpha(x), \tag{4.4}$$

where convergence to the limit is uniform over all $\alpha \in A$.

Proof. For ease of notation, the dependence of k and ℓ on i is made implicit. First we show consistency in the second-order terms; see (4.8) below. From the definition of P_i and integration by parts,

$$\begin{aligned}
&\Big| \bar{\bar{a}}_i^\alpha(y_i^\ell) \langle \nabla P_i w(s_i^k, \cdot), \nabla \hat{\phi}_i^\ell \rangle + \bar{a}_i^\alpha(y_i^\ell) \langle \nabla P_i w(s_i^{k+1}, \cdot), \nabla \hat{\phi}_i^\ell \rangle - a^\alpha(y_i^\ell) \langle \nabla w(t, \cdot), \nabla \hat{\phi}_i^\ell \rangle \Big| \\
&\quad = \Big| \bar{\bar{a}}_i^\alpha(y_i^\ell) \langle \nabla w(s_i^k, \cdot), \nabla \hat{\phi}_i^\ell \rangle + \bar{a}_i^\alpha(y_i^\ell) \langle \nabla w(s_i^{k+1}, \cdot), \nabla \hat{\phi}_i^\ell \rangle - a^\alpha(y_i^\ell) \langle \nabla w(t, \cdot), \nabla \hat{\phi}_i^\ell \rangle \Big| \\
&\quad \le \Big| \big(a^\alpha(y_i^\ell) - \bar{\bar{a}}_i^\alpha(y_i^\ell) - \bar{a}_i^\alpha(y_i^\ell) \big) \langle -\Delta w(t, \cdot), \hat{\phi}_i^\ell \rangle \Big| + \Big| \bar{\bar{a}}_i^\alpha(y_i^\ell) \langle \Delta w(t, \cdot) - \Delta w(s_i^k, \cdot), \hat{\phi}_i^\ell \rangle \Big| \\
&\qquad + \Big| \bar{a}_i^\alpha(y_i^\ell) \langle \Delta w(t, \cdot) - \Delta w(s_i^{k+1}, \cdot), \hat{\phi}_i^\ell \rangle \Big| .
\end{aligned} \tag{4.5}$$

Using Assumption 2.1 and smoothness of w together with uniform boundedness of the $|\bar{\bar{a}}_i^\alpha(y_i^\ell)|$ and $|\bar{a}_i^\alpha(y_i^\ell)|$ over $\alpha \in A$, we conclude from the above inequality that

$$\begin{aligned}
\lim_{i\to\infty} \sup_{\alpha\in A} \Big| &\bar{\bar{a}}_i^\alpha(y_i^\ell) \langle \nabla P_i w(s_i^k, \cdot), \nabla \hat{\phi}_i^\ell \rangle + \bar{a}_i^\alpha(y_i^\ell) \langle \nabla P_i w(s_i^{k+1}, \cdot), \nabla \hat{\phi}_i^\ell \rangle \\
&- a^\alpha(y_i^\ell) \langle \nabla w(t, \cdot), \nabla \hat{\phi}_i^\ell \rangle \Big| = 0.
\end{aligned} \tag{4.6}$$

Owing to the Heine–Cantor theorem, for all $\varepsilon > 0$, there is a $\delta > 0$ such that $|\Delta w(t,x) - \Delta w(t,y)| < \varepsilon$ if $|x - y| < \delta$. For i sufficiently large, the support of $\hat{\phi}_i^\ell$ is contained in the ball $B(x, \delta)$. Also, $\|\hat{\phi}_i^\ell\|_{L^1(\Omega)} = 1$ and $\hat{\phi}_i^\ell \ge 0$. Thus, $|\Delta w(t,x) - \langle \Delta w(t, \cdot), \hat{\phi}_i^\ell \rangle| < \varepsilon$. Recall that $\{a^\alpha\}_{\alpha \in A}$ is an equicontinuous family of functions, so integration by parts shows that

$$\lim_{i\to\infty} \sup_{\alpha\in A} \Big| a^\alpha(y_i^\ell) \langle \nabla w(t, \cdot), \nabla \hat{\phi}_i^\ell \rangle - \big(a^\alpha(x) \Delta w(t,x) \big) \Big| = 0. \tag{4.7}$$

Equations (4.6) and (4.7) imply consistency of the second-order terms:

$$\begin{aligned}
\lim_{i\to\infty} \sup_{\alpha\in A} \Big| &\bar{\bar{a}}_i^\alpha(y_i^\ell) \langle \nabla P_i w(s_i^k, \cdot), \nabla \hat{\phi}_i^\ell \rangle + \bar{a}_i^\alpha(y_i^\ell) \langle \nabla P_i w(s_i^{k+1}, \cdot), \nabla \hat{\phi}_i^\ell \rangle \\
&- \big(-a^\alpha(x) \Delta w(t,x) \big) \Big| = 0.
\end{aligned} \tag{4.8}$$

Using Assumption 4.1 and regularity of w, we see that $P_i w(s_i^k,\cdot)$ and $P_i w(s_i^{k+1},\cdot)$ converge to $w(t,\cdot)$ in $W^{1,\infty}(\Omega)$. Analogous estimates to the ones above and equicontinuity of $\{b^\alpha\}_{\alpha\in A}$, $\{c^\alpha\}_{\alpha\in A}$ and $\{f^\alpha\}_{\alpha\in A}$ imply that

$$\begin{aligned}
\lim_{i\to\infty}\sup_{\alpha\in A}\left|\langle\bar{\bar{b}}_i^\alpha\cdot\nabla P_i w(s_i^k,\cdot),\hat{\phi}_i^\ell\rangle+\langle\bar{b}_i^\alpha\cdot\nabla P_i w(s_i^{k+1},\cdot),\hat{\phi}_i^\ell\rangle-b^\alpha(x)\cdot\nabla w(t,x)\right|&=0,\\
\lim_{i\to\infty}\sup_{\alpha\in A}\left|\langle\bar{\bar{c}}_i^\alpha P_i w(s_i^k,\cdot),\hat{\phi}_i^\ell\rangle+\langle\bar{c}_i^\alpha P_i w(s_i^{k+1},\cdot),\hat{\phi}_i^\ell\rangle-c^\alpha(x)w(t,x)\right|&=0,\\
\lim_{i\to\infty}\sup_{\alpha\in A}\left|\langle f_i^\alpha,\hat{\phi}_i^\ell\rangle-f^\alpha(x)\right|&=0.
\end{aligned}\tag{4.9}$$

Combining equations (4.8) and (4.9) yields (4.4). ∎

The orthogonality condition (4.1), used in (4.5), ensures the consistency of the discretization for linear operators L^α with isotropic diffusion terms of the form $-a(x)\Delta w$ but apparently not for operators with anisotropic diffusion of the form $\mathcal{A}(x):D^2w$, $\mathcal{A}\in C(\Omega,\mathbb{R}^{d\times d})$. This restriction does not arise for finite difference methods because the discretization of second-order derivatives is consistent under nodal interpolation. We note that there are linear elliptic equations in nondivergence form for which it is not possible to construct monotone, pointwise consistent compact stencil schemes. For estimates on the stencil width see [23] and also [13]. This observation goes back to [30]; see also [31] for recent results. As pointed out in Example 2 below, in some cases our method coincides with a finite difference scheme, indicating that similar constraints are likely to arise in a finite element setting for anisotropic equations as well.

5. Sub- and supersolution.

Set

$$v^*(t,x)=\sup_{(s_i^k,y_i^\ell)\to(t,x)}\limsup_{i\to\infty}v_i(s_i^k,y_i^\ell),\qquad v_*(t,x)=\inf_{(s_i^k,y_i^\ell)\to(t,x)}\liminf_{i\to\infty}v_i(s_i^k,y_i^\ell),$$

where the limit superior and limit inferior are taken over all sequences of nodes in $[0,T]\times\overline{\Omega}$ which converge to $(t,x)\in[0,T]\times\overline{\Omega}$. Owing to Theorem 3.1 and Corollary 3.3, v^* and v_* attain nonnegative finite values. By construction, v^* is upper and v_* lower semicontinuous and $v_*\le v^*$. With the use of elliptic projection operators, key steps of the convergence proof in [3], which is stated there in a suitable form for finite difference methods, are transferred to finite element schemes, which do not satisfy the consistency condition in [3].

THEOREM 5.1. *The function v^* is a viscosity subsolution of* (2.3)*, and v_* is a viscosity supersolution of* (2.3).

Proof. Step 1 *(v^* is a subsolution).* To show that v^* is a viscosity subsolution, suppose that $w\in C^\infty(\mathbb{R}\times\mathbb{R}^d)$ is a test function such that v^*-w has a strict local maximum at $(s,y)\in(0,T)\times\Omega$ with $v^*(s,y)=w(s,y)$. Consider a closed neighborhood $B:=\{(t,x)\in(0,T)\times\Omega\,:\,|t-s|+|x-y|\le\delta\}$ with $\delta>0$ such that

$$v^*(s,y)-w(s,y)>v^*(t,x)-w(t,x)\quad\forall(t,x)\in B\setminus(s,y).$$

Choose i sufficiently large for B to contain nodes. Let (s_i^k,y_i^ℓ) denote the position where $v_i(s_i^\kappa,y_i^\lambda)-P_iw(s_i^\kappa,y_i^\lambda)$ attains a maximum among all nodes $(s_i^\kappa,y_i^\lambda)\in B$. Let us pass to a subsequence $\{(s_{i(j)}^k,y_{i(j)}^\ell)\}_j$ of $\{(s_i^k,y_i^\ell)\}_i$ for which $\{v_i(s_{i(j)}^k,y_{i(j)}^\ell)\}_j$ converges to the limit superior of $\{v_i(s_i^k,y_i^\ell)\}_i$. By compactness of B, there is a subsequence of $\{(s_{i(j)}^k,y_{i(j)}^\ell)\}_j$, to which we pass without change of notation, that converges to a point $(\tilde{s},\tilde{y}) \in B$. Then $P_i w(s^k_{i(j)}, y^\ell_{i(j)}) \to w(\tilde{s},\tilde{y})$ due to (4.2) and continuity of w. Since the (s_i^k, y_i^ℓ) are maximizers, one has

$$v^*(\tilde{s},\tilde{y}) - w(\tilde{s},\tilde{y}) = \limsup_{j\to\infty} v_i(s^k_{i(j)}, y^k_{i(j)}) - P_i w(s^k_{i(j)}, y^k_{i(j)}) = v^*(s,y) - w(s,y);$$

hence $(\tilde{s},\tilde{y}) = (s,y)$ since (s,y) is a strict maximizer of $v^* - w$ on B. Thus there is a subsequence of maximizing nodes converging to (s,y) to which we now pass without change of notation: $(s_i^k, y_i^\ell) \to (s,y)$. It follows that

$$v_i(s_i^k, y_i^\ell) - P_i w(s_i^k, y_i^\ell) \to v^*(s,y) - w(s,y) = 0. \tag{5.1}$$

Moreover, because of $(s_i^k, y_i^\ell) \to (s,y) \in \operatorname{int} B$, the neighbors of the (s_i^k, y_i^ℓ) eventually also belong to B: for i sufficiently large, we have $(s_i^\kappa, y_i^\lambda) \in B$ if $\kappa \in \{k, k+1\}$ and $y_i^\lambda \in \operatorname{supp} \hat{\phi}_i^\ell$, in which case

$$v_i(s_i^\kappa, y_i^\lambda) - P_i w(s_i^\kappa, y_i^\lambda) \le v_i(s_i^k, y_i^\ell) - P_i w(s_i^k, y_i^\ell) \;\Leftrightarrow\; P_i w(s_i^\kappa, y_i^\lambda) + \mu_i \ge v_i(s_i^\kappa, y_i^\lambda)$$

with $\mu_i = v_i(s_i^k, y_i^\ell) - P_i w(s_i^k, y_i^\ell)$. Notice that $\mu_i \to 0$ as $i \to \infty$ because of (5.1).

Recall that the matrices E_i^α have nonzero off diagonal entries $(\mathsf{E}_i^\alpha)_{\ell\lambda}$ only if $y_i^\lambda \in \operatorname{supp} \hat{\phi}_i^\ell$ and that $v_i(s_i^{k+1}, \cdot) \le P_i w(s_i^{k+1}, \cdot) + \mu_i$ on $\operatorname{supp} \hat{\phi}_i^\ell$. Therefore, monotonicity of $h_i \mathsf{E}_i^\alpha - \mathsf{Id}$ for all $\alpha \in A$ implies that

$$\big((h_i \mathsf{E}_i^\alpha - \mathsf{Id}) \left[P_i w(s_i^{k+1}, \cdot) + \mu_i\right]\big)_\ell \le \big((h_i \mathsf{E}_i^\alpha - \mathsf{Id}) v_i(s_i^{k+1}, \cdot)\big)_\ell .$$

Applying the LMP and linearity of I_i^α to $P_i w(s_i^k, \cdot) + \mu_i - v_i(s_i^k, \cdot)$, which has a non-positive local minimum at y_i^ℓ, yields

$$\big((h_i \mathsf{I}_i^\alpha + \mathsf{Id}) \left[P_i w(s_i^k, \cdot) + \mu_i\right]\big)_\ell \le \big((h_i \mathsf{I}_i^\alpha + \mathsf{Id}) v_i(s_i^k, \cdot)\big)_\ell .$$

From the definition of the scheme (2.7),

$$\begin{aligned}
0 &= - d_i v_i(s_i^k, y_i^\ell) + \sup_{\alpha\in A} \big(\mathsf{E}_i^\alpha v_i(s_i^{k+1}, \cdot) + \mathsf{I}_i^\alpha v_i(s_i^k, \cdot) - \mathsf{F}_i^\alpha\big)_\ell \\
&\ge - d_i \big(P_i w(s_i^k, y_i^\ell) + \mu_i\big) + \sup_{\alpha\in A} \big(\mathsf{E}_i^\alpha \big(P_i w(s_i^{k+1}, \cdot) + \mu_i\big) + \mathsf{I}_i^\alpha \big(P_i w(s_i^k, \cdot) + \mu_i\big) - \mathsf{F}_i^\alpha\big)_\ell \\
&= - d_i P_i w(s_i^k, y_i^\ell) + \sup_{\alpha\in A} \Big[\big(\mathsf{E}_i^\alpha P_i w(s_i^{k+1}, \cdot) + \mathsf{I}_i^\alpha P_i w(s_i^k, \cdot) - \mathsf{F}_i^\alpha\big)_\ell + \mu_i \langle \bar{c}_i^\alpha + \bar{\bar{c}}_i^\alpha, \hat{\phi}_i^\ell \rangle\Big] \\
&\ge - d_i P_i w(s_i^k, y_i^\ell) + \sup_{\alpha\in A} \big(\mathsf{E}_i^\alpha P_i w(s_i^{k+1}, \cdot) + \mathsf{I}_i^\alpha P_i w(s_i^k, \cdot) - \mathsf{F}_i^\alpha\big)_\ell - \gamma \left|\mu_i\right| .
\end{aligned}$$

(5.2)

Since

$$\begin{aligned}
&\left|\sup_{\alpha\in A} \big(\mathsf{E}_i^\alpha P_i w(s_i^{k+1}, \cdot) + \mathsf{I}_i^\alpha P_i w(s_i^k, \cdot) - \mathsf{F}_i^\alpha\big)_\ell - \sup_{\alpha\in A} \left(L^\alpha w(s,y) - f^\alpha(y)\right)\right| \\
&\quad \le \sup_{\alpha\in A} \left|\big(\mathsf{E}_i^\alpha P_i w(s_i^{k+1}, \cdot) + \mathsf{I}_i^\alpha P_i w(s_i^k, \cdot) - \mathsf{F}_i^\alpha\big)_\ell - \left(L^\alpha w(s,y) - f^\alpha(y)\right)\right| ,
\end{aligned}$$

Lemmas 4.1 and 4.2 show that we may take the limit $i \to \infty$ in inequality (5.2) and recall that $\mu_i \to 0$ to obtain

$$0 \ge -\partial_t w(s,y) + \sup_{\alpha\in A} \left(L^\alpha w(s,y) - f^\alpha(y)\right) . \tag{5.3}$$

Therefore v^* is a viscosity subsolution.

Step 2 *(v_* is a supersolution).* Arguments similar to those above show that v_* is a viscosity supersolution, where the principal changes to the proof are that one considers $w \in C^\infty(\mathbb{R} \times \mathbb{R}^d)$ such that $v_* - w$ has a strict local minimum at some $(s, y) \in (0, T) \times \Omega$ with $v_*(s, y) = w(s, y)$. With analogous notation, the last line in (5.2) corresponds to

$$0 \leq -d_i P_i w(s_i^k, y_i^\ell) + \sup_{\alpha \in A} \left(\mathsf{E}_i^\alpha P_i w(s_i^{k+1}, \cdot) + \mathsf{I}_i^\alpha P_i w(s_i^k, \cdot) - \mathsf{F}_i^\alpha \right)_\ell + \gamma \left| \mu_i \right|,$$

i.e., there is a slight asymmetry in the argument due to the last sign in (5.2). Nevertheless, it is then deduced that

$$0 \leq -\partial_t w(s, y) + \sup_{\alpha \in A} \left(L^\alpha w(s, y) - f^\alpha(y) \right).$$

Thus v_* is a viscosity supersolution. ∎

6. Uniform convergence. We now turn to the initial and boundary conditions. Together with the sub- and supersolution property we appeal to a comparison principle to obtain uniform convergence of the numerical solutions.

For each $\alpha \in A$, define

$$v^{\alpha,*}(t, x) = \sup_{(s_i^k, y_i^\ell) \to (t,x)} \limsup_{i \to \infty} v_i^\alpha(s_i^k, y_i^\ell),$$

where the v_i^α are as in (3.2) and the limit superior is taken over all sequences of nodes which converge to $(t, x) \in [0, T] \times \overline{\Omega}$. Because of Corollary 3.3 it is clear that $v^{\alpha,*}$ attains finite values.

Assumption 6.1. *Suppose that for each $(t, x) \in [0, T] \times \partial\Omega$*

$$\inf_{\alpha \in A} v^{\alpha,*}(t, x) = 0. \tag{6.1}$$

Before further considerations, let us motivate Assumption 6.1 with a simple example. As a side remark, this example also illustrates how in some settings Kushner–Dupuis finite difference schemes, as described in [27, 19], may be interpreted as finite element methods in the framework of this paper.

Example 2. Consider the backward time-dependent equation in one spatial dimension

$$-v_t + |v_x| = 1 \quad \text{on } (0, 1) \times (-1, 1), \tag{6.2}$$

with boundary conditions $v = 0$ on $[0, 1] \times \{-1, 1\} \cup \{1\} \times [-1, 1]$. Equation (6.2) may be rewritten in HJB form as

$$-v_t + \sup_{\alpha \in \{-1,1\}} (\alpha v_x - 1) = 0.$$

The viscosity solution is $v = \min(1 - t, 1 - |x|)$. We choose a uniform mesh with element size Δx_i, and we use a fully explicit discretization, where monotonicity will be achieved by using the method of artificial diffusion, as described in [7]. Thus we have $(\mathsf{E}_i^\alpha w)_\ell = \varepsilon \langle \partial_x w, \partial_x \hat{\phi}_i^\ell \rangle + \alpha \langle \partial_x w, \hat{\phi}_i^\ell \rangle$, where ε is the artificial diffusion parameter to be chosen to obtain a monotone scheme. Calculating the entries shows that

$$(\mathsf{E}_i^\alpha)_{\ell j} = \begin{cases} -\alpha/2\Delta x_i - \varepsilon/\Delta x_i^2 & \text{if } j = \ell - 1, \\ 2\varepsilon/\Delta x_i^2 & \text{if } j = \ell, \\ \alpha/2\Delta x_i - \varepsilon/\Delta x_i^2 & \text{if } j = \ell + 1, \\ 0 & \text{otherwise.} \end{cases}$$

For monotonicity we require that all off-diagonal terms of the E_i^α be nonpositive, i.e., we require $\varepsilon \geq \Delta x_i/2$, because $|\alpha| \leq 1$. With the special choice $\varepsilon = \Delta x_i/2$ the matrices E_i^1 and E_i^{-1} are triangular. This is equivalent to discretizing the spatial part of $-v_t+v_x$ with backward finite differences and discretizing the spatial part of $-v_t-v_x$ with forward finite differences, as can be done in applying a Kushner–Dupuis scheme. If appropriate time steps sizes are used, then it can be deduced that v_i^1 approximates the solution of

$$-v_t + v_x = 1 \text{ on } (0,1)\times(-1,1), \qquad v = 0 \text{ on } (0,T)\times\{-1\}\cup\{1\}\times(-1,1),$$

while v_i^{-1} approximates the solution of

$$-v_t - v_x = 1 \text{ on } (0,1)\times(-1,1), \qquad v = 0 \text{ on } (0,T)\times\{1\}\cup\{1\}\times(-1,1).$$

Consequently, Assumption 6.1 is enforced by $v^{1,*}$ on $[0,1]\times\{-1\}$ and by $v^{-1,*}$ on $[0,1]\times\{1\}$.

Recall from Theorem 3.1 that

$$0 \leq v_i \leq v_i^\alpha \quad \forall \alpha \in A,$$

and note that by construction $0 \leq v_* \leq v^*$. Since $v^* \leq v^{\alpha,*}$ for all α, Assumption 6.1 implies $v_*|_{[0,T]\times\partial\Omega} = v^*|_{[0,T]\times\partial\Omega} = 0$. Observe that because (6.1) holds in particular for all $(t,x) \in \{T\}\times\partial\Omega$, Assumption 6.1 implicitly enforces that the initial condition v_T vanishes on $\partial\Omega$ as the v_i^α interpolate v_T at the final time.

LEMMA 6.1. *The sub- and supersolutions v^* and v_* satisfy*

$$v^*(T,\cdot) = v_*(T,\cdot) = v_T \quad \text{on } \overline{\Omega}. \tag{6.3}$$

Proof. Fix $\varepsilon > 0$ and choose a $v_T^\varepsilon \in C^\infty(\mathbb{R}^d)$ such that $v_T - 2\varepsilon \geq v_T^\varepsilon \geq v_T - 3\varepsilon$. Owing to Assumption 4.1 there is $n \in \mathbb{N}$ such that $\|P_i v_T^\varepsilon - v_T^\varepsilon\|_{L^\infty(\Omega)} \leq \varepsilon$ and $\|\mathcal{I}_i v_T - v_T\|_{L^\infty(\Omega)} \leq \varepsilon$ for all $i \geq n$. Hence, for $i \geq n$,

$$v_i(T,\cdot) = \mathcal{I}_i v_T \geq P_i v_T^\varepsilon \geq v_T - 4\varepsilon. \tag{6.4}$$

Recalling (4.1) and as $v_T^\varepsilon \in C^\infty(\mathbb{R}^d)$, it is clear that there exists $K = K(\varepsilon) \geq 0$ which bounds

$$\begin{aligned}
&\left|\left((\mathsf{E}_i^\alpha + \mathsf{I}_i^\alpha)P_i v_T^\varepsilon - \mathsf{F}_i^\alpha\right)_\ell\right| \\
&\quad \overset{(*)}{=} \left|-\left(\bar{a}_i^\alpha(y_i^\ell) + \bar{\bar{a}}_i^\alpha(y_i^\ell)\right)\langle \Delta v_T^\varepsilon, \hat{\phi}_i^\ell\rangle\right. \\
&\qquad \left. + \left\langle \left(\bar{b}_i^\alpha(y_i^\ell) + \bar{\bar{b}}_i^\alpha(y_i^\ell)\right)\cdot \nabla P_i v_T^\varepsilon + \left(\bar{c}_i^\alpha(y_i^\ell) + \bar{\bar{c}}_i^\alpha(y_i^\ell)\right)P_i v_T^\varepsilon, \hat{\phi}_i^\ell\right\rangle - \left(\mathsf{F}_i^\alpha\right)_\ell\right|
\end{aligned}$$

for all $i \in \mathbb{N}$, $\ell \in \{1,\ldots,N_i\}$ and $\alpha \in A$. Notice that $(*)$ uses (4.1). Define $w_i = P_i v_T^\varepsilon - K(T-t)$. To show inductively that $v_i(s_i^k,\cdot) \geq w_i(s_i^k,\cdot)$ assume $v_i(s_i^{k+1},\cdot) \geq w_i(s_i^{k+1},\cdot)$, noting (6.4) for $s_i^{k+1} = T$. Fix an i and ℓ and let $\alpha = \alpha_i^{k,\ell}(v_i)$ as for (2.9). From Lemma 2.3 and

$$\begin{aligned}
&-d_i w_i(s_i^k, y_i^\ell) + \left(\mathsf{E}_i^\alpha w_i(s_i^{k+1},\cdot) + \mathsf{I}_i^\alpha w_i(s_i^k,\cdot)\right)_\ell \\
&\quad = -K + \left((\mathsf{E}_i^\alpha + \mathsf{I}_i^\alpha)\, P_i v_T^\varepsilon\right)_\ell - K(T - s_i^{k+1})\langle \bar{c}_i^\alpha, \hat{\phi}_i^\ell\rangle - K(T - s_i^k)\langle \bar{\bar{c}}_i^\alpha, \hat{\phi}_i^\ell\rangle \\
&\quad \leq (\mathsf{F}_i^\alpha)_\ell \overset{(2.10)}{=} -d_i v_i(s_i^k, y_i^\ell) + \left(\mathsf{E}_i^{v_i} v_i(s_i^{k+1},\cdot) + \mathsf{I}_i^{v_i} v_i(s_i^k,\cdot)\right)_\ell
\end{aligned}$$

we may deduce that

$$\left((h_i \mathsf{I}_i^{v_i} + \mathsf{Id})\left[v_i(s_i^k,\cdot) - w_i(s_i^k,\cdot)\right]\right)_\ell \geq -\left((h_i \mathsf{E}_i^{v_i} - \mathsf{Id})\left[v_i(s_i^{k+1},\cdot) - w_i(s_i^{k+1},\cdot)\right]\right)_\ell \geq 0.$$

Note that $v_i(s_i^k,\cdot) \in V_i^0$ vanishes on $\partial\Omega$ and $w_i(s_i^k,\cdot) \leq 0$ on $\partial\Omega$. Thus (2.8) and Lemma 2.3 imply $v_i(s_i^k,\cdot) \geq w_i(s_i^k,\cdot)$ on $\overline{\Omega}$. Because K is independent of i and $P_i v_T^\varepsilon \to v_T^\varepsilon$ as $i \to \infty$, we have for any sequence $(s_i^k, y_i^\ell) \to (T,x)$, $x \in \Omega$,

$$\liminf_{i\to\infty} v_i\big(s_i^k, y_i^\ell\big) \geq \liminf_{i\to\infty} w_i\big(s_i^k, y_i^\ell\big) \geq v_T(x) - 4\varepsilon.$$

So $v_*(T,\cdot) \geq v_T - 4\varepsilon$. Since ε was arbitrary, $v_*(T,\cdot) \geq v_T$. The argument for showing that $v^* \leq v_T$ is analogous with $w_i = P_i v_T^\varepsilon + K(T-t)$ and $v_T + 2\varepsilon \leq v_T^\varepsilon \leq v_T + 3\varepsilon$. To conclude, $v_T \leq v_*(T,\cdot) \leq v^*(T,\cdot) \leq v_T$, which proves (6.3). ∎

The proof of Lemma 6.1 is related to the arguments in [19, p. 335]. In the next assumption we draw upon one of the building blocks of the theory of viscosity solutions, namely, the extension of classical comparison principles to spaces of semicontinuous functions; cf. [12, sect. 5] and [19, p. 219].

Assumption 6.2. Let $\overline{v}$ be a lower semicontinuous supersolution with $\overline{v}(T,\cdot) = v_T$ and $\overline{v}|_{[0,T]\times\partial\Omega} = 0$. Similarly, let $\underline{v}$ be an upper semicontinuous subsolution with $\underline{v}|_{[0,T]\times\partial\Omega} = 0$ and $\underline{v}(T,\cdot) = v_T$. Then $\underline{v} \leq \overline{v}$.

Let $t = \vartheta s_i^k + (1-\vartheta)s_i^{k+1} \in [s_i^k, s_i^{k+1}]$ lie between two time steps, $\vartheta \in [0,1]$. Then we interpret $v_i(t,\cdot)$ as the linear interpolant between $v_i(s_i^k,\cdot)$ and $v_i(s_i^{k+1},\cdot)$:

$$v_i(t,\cdot) = \vartheta v_i(s_i^k,\cdot) + (1-\vartheta) v_i(s_i^{k+1},\cdot). \tag{6.5}$$

THEOREM 6.2. *One has $v_* = v^* = v$, where v is the unique viscosity solution of (2.3) with $v(T,\cdot) = v_T$ and $v|_{[0,T]\times\partial\Omega} = 0$. Furthermore*

$$\lim_{i\to\infty} \|v_i - v\|_{L^\infty((0,T)\times\Omega)} = 0. \tag{6.6}$$

Proof. The previous assumption implies that $v_* \geq v^*$ on $[0,T]\times\overline{\Omega}$ and thus $v^* = v_* = v$. It also follows that v is continuous and is the unique viscosity solution. Select for each $i \in \mathbb{N}$ a point $(t_i, x_i) \in [0,T]\times\overline{\Omega}$ such that

$$\|v_i - v\|_{L^\infty((0,T)\times\Omega)} = |v_i - v|(t_i, x_i).$$

Such (t_i,x_i) exist as $v_i - v$ is a continuous function on a compact domain. Let x_i belong to (the closure of) the element T of the finite element mesh and $t \in [s_i^\kappa, s_i^{\kappa+1}]$; then $v_i(t_i,x_i)$ is a weighted average of the values of v_i at the corners of the slab $[s_i^\kappa, s_i^{\kappa+1}] \times \overline{T}$. Thus there is a corner (s_i^k, y_i^ℓ) of the slab such that

$$\|v_i - v\|_{L^\infty((0,T)\times\Omega)} \leq |v_i(s_i^k, y_i^\ell) - v(t_i, x_i)|.$$

If (6.6) was wrong we could select a subsequence and an $\varepsilon > 0$ such that

$$\liminf_{j\to\infty} \left|v_{i(j)}(s_{i(j)}^k, y_{i(j)}^\ell) - v(t_{i(j)}, x_{i(j)})\right| \geq \varepsilon.$$

By possibly passing to a further subsequence we may assume that $\{(t_{i(j)}, x_{i(j)})\}_j$ converges to an $(t,x) \in [0,T]\times\overline{\Omega}$. However, this contradicts

$$\begin{aligned} v(t,x) = v_*(t,x) &\leq \liminf_{j\to\infty} v_{i(j)}(s_{i(j)}^k, y_{i(j)}^\ell) \\ &\leq \limsup_{j\to\infty} v_{i(j)}(s_{i(j)}^k, y_{i(j)}^\ell) \leq v^*(t,x) = v(t,x). \end{aligned}$$

Thus (6.6) holds. ∎

7. Gradient convergence. For the proof of convergence in $L^2([0,T],H^1(\Omega))$, we suppose condition (7.1) below, which points towards uniform ellipticity of at least one L^α. See also Example 3 at the end of this section.

For shorthand, let $W = W^{1,\infty}((0,T)\times\Omega)$. It is convenient to introduce the discrete spaces $W_i := \{v \in C([0,T],V_i^0) : v|_{[s_i^k,s_i^{k+1}]\times\Omega}$ is affine in time$\}$, which means that functions in W_i have between two time steps the form of (6.5). Observe that $W_i \subset W$ for all $i \in \mathbb{N}$.

Fix an arbitrary $\alpha \in A$. It is convenient to view E_i^α and I_i^α as bilinear forms on $H^1(\Omega)\times V_i$. Functions $u \in V_i$ have the nodal representation

$$u(y) = \sum_\ell u(y_i^\ell)\,\phi_i^\ell(y).$$

To test with functions other than $\hat\phi_i^\ell$ we introduce the following bilinear form as a partially discrete pivot: for $w \in H^1(\Omega)$ and $u \in V_i$

$$\langle\!\langle \mathsf{E}_i^\alpha w, u\rangle\!\rangle := \sum_\ell u(y_i^\ell)\big(\bar a_i^\alpha(y_i^\ell)\langle\nabla w,\nabla\phi_i^\ell\rangle + \langle \bar b_i^\alpha\cdot\nabla w + \bar c_i^\alpha\, w, \phi_i^\ell\rangle\big).$$

We use the corresponding interpretation for $\langle\!\langle \mathsf{I}_i^\alpha w, u\rangle\!\rangle$ and also

$$\langle\!\langle w, u\rangle\!\rangle = \langle\!\langle \mathsf{Id}\, w, u\rangle\!\rangle = \sum_\ell w(y_i^\ell)\,u(y_i^\ell)\|\phi_i^\ell\|_{L^1(\Omega)},$$
$$\langle\!\langle \mathsf{F}_i^\alpha, u\rangle\!\rangle = \sum_\ell u(y_i^\ell)\,\langle f_i^\alpha,\phi_i^\ell\rangle = \langle f_i^\alpha, u\rangle.$$

Assume that there is $\alpha \in A$ such that

$$\begin{aligned}
|w|^2_{L^2([0,T],H^1(\Omega))} &\lesssim \sum_{k=0}^{\frac{T}{h_i}-1}\Big(\big\langle\!\big\langle \big(h_i\mathsf{E}_i^\alpha - \mathsf{Id}\big)w(s_i^{k+1},\cdot) + \big(h_i\mathsf{I}_i^\alpha + \mathsf{Id}\big)w(s_i^k,\cdot), w(s_i^k,\cdot)\big\rangle\!\big\rangle\Big)\\
&\qquad + \tfrac12\langle\!\langle w(T,\cdot),w(T,\cdot)\rangle\!\rangle + h_i\|w(T,\cdot)\|^2_{H^1(\Omega)}\\
&\overset{(*)}{=} \sum_{k=0}^{\frac{T}{h_i}-1}\Big(h_i\big\langle\!\big\langle \mathsf{E}_i^\alpha w(s_i^{k+1},\cdot) + \mathsf{I}_i^\alpha w(s_i^k,\cdot), w(s_i^k,\cdot)\big\rangle\!\big\rangle\\
&\qquad\qquad + \tfrac12\langle\!\langle w(s_i^{k+1},\cdot)-w(s_i^k,\cdot), w(s_i^{k+1},\cdot)-w(s_i^k,\cdot)\rangle\!\rangle\Big)\\
&\qquad + \tfrac12\langle\!\langle w(0,\cdot),w(0,\cdot)\rangle\!\rangle + h_i\|w(T,\cdot)\|^2_{H^1(\Omega)}
\end{aligned}\tag{7.1}$$

for all $w \in W_i$ with $w \geq 0$ and $i \in \mathbb{N}$, where $(*)$ is a simple reformulation in terms of a telescope sum.

Due to the definition of the numerical method and the nonnegativity of the v_i,

$$\begin{aligned}
|v_i|^2_{L^2([0,T],H^1(\Omega))} &\lesssim \sum_{k=0}^{\frac{T}{h_i}-1}\Big(\big\langle\!\big\langle \big(h_i\mathsf{E}_i^\alpha - \mathsf{Id}\big)v_i(s_i^{k+1},\cdot) + \big(h_i\mathsf{I}_i^\alpha + \mathsf{Id}\big)v_i(s_i^k,\cdot), v_i(s_i^k,\cdot)\big\rangle\!\big\rangle\Big)\\
&\qquad + \tfrac12\langle\!\langle v_i(T,\cdot),v_i(T,\cdot)\rangle\!\rangle + h_i\|v_i(T,\cdot)\|^2_{H^1(\Omega)}\\
&\leq \sum_{k=0}^{\frac{T}{h_i}-1}\big\langle\!\big\langle h_i\mathsf{F}_i^\alpha, v_i(s_i^k,\cdot)\big\rangle\!\big\rangle + \tfrac12\langle\!\langle v_i(T,\cdot),v_i(T,\cdot)\rangle\!\rangle + h_i\|v_i(T,\cdot)\|^2_{H^1(\Omega)}\\
&\lesssim T\,\|f_i^\alpha\|_{L^1(\Omega)}\,\|v_i\|_{L^\infty([0,T]\times\Omega)} + h_i\|v_i(T,\cdot)\|^2_{H^1(\Omega)}.
\end{aligned}$$

Thus, with the L^∞ control established in the previous section and (7.1), it is apparent that the v_i are bounded in $L^2([0,T],H^1(\Omega))$ provided that $h_i v_i(T,\cdot) = h_i \mathcal{I}_i v_T$ are bounded in $H^1(\Omega)$; this condition holds for instance if $v(T,\cdot) \in W^{1,\infty}(\Omega)$. The first convergence result for the gradient is therefore that, owing to the Banach–Alaoglu theorem, $v_i \rightharpoonup v$ weakly in $L^2([0,T],H^1(\Omega))$, using $L^\infty((0,T)\times\Omega)$ convergence to pass from $L^2([0,T],H^1(\Omega))$ weak convergence of subsequences to $L^2([0,T],H^1(\Omega))$ weak convergence of the whole sequence.

The question arises under which circumstances the convergence in the gradient is also strong. We demonstrate this under Assumption 7.1. Let Λ_0 be the level set $\{(t,x) \in (0,T)\times\Omega : v(t,x) = 0\}$. For a smooth v the boundary of Λ_0 is always a $d-1$ dimensional set if 0 is a regular value.

Assumption 7.1. The value function v belongs to the space W, and the d-dimensional Lebesgue measure of the boundary of Λ_0 vanishes: $\mathrm{vol}(\partial\Lambda_0) = 0$. There is an α such that the coefficients $\bar{a}_i^\alpha$ and $\bar{\bar{a}}_i^\alpha$ belong to $W^{1,\infty}(\Omega)$ and (7.1) is satisfied.

Let us suppose momentarily that there are approximations $Q_i v \in W_i$ to v such that $Q_i v \le v_i$ for all $i \in \mathbb{N}$ and, as $i \to \infty$,

(7.2)
$$\|v - Q_i v\|_{L^2([0,T],H^1(\Omega))} + h_i \|(v - Q_i v)(T,\cdot)\|_{H^1(\Omega)} + \|(v - Q_i v)(T,\cdot)\|_{L^2(\Omega)} \to 0,$$

as well as

$$\text{(7.3)} \quad \sum_{k=0}^{\frac{T}{h_i}-1} \left\langle\!\left\langle \big(h_i \mathsf{E}_i^\alpha - \mathsf{Id}\big) Q_i v(s_i^{k+1},\cdot) + \big(h_i \mathsf{I}_i^\alpha + \mathsf{Id}\big) Q_i v(s_i^k,\cdot), (v_i - Q_i v)(s_i^k,\cdot) \right\rangle\!\right\rangle \to 0.$$

We will construct such $Q_i v$ below. With $\xi^k = v_i(s_i^k,\cdot) - Q_i v(s_i^k,\cdot)$,

(7.4)
$$\begin{aligned}
&|v_i - Q_i v|^2_{L^2([0,T],H^1(\Omega))} \\
&\overset{(7.1)}{\lesssim} \sum_{k=0}^{\frac{T}{h_i}-1} \left\langle\!\left\langle \big(h_i \mathsf{E}_i^\alpha - \mathsf{Id}\big)\xi^{k+1} + \big(h_i \mathsf{I}_i^\alpha + \mathsf{Id}\big)\xi^k, \xi^k \right\rangle\!\right\rangle \\
&\qquad + \tfrac{1}{2}\langle\!\langle \xi^{T/h_i}, \xi^{T/h_i} \rangle\!\rangle + h_i \|\xi^{T/h_i}\|^2_{H^1(\Omega)} \\
&= \sum_{k=0}^{\frac{T}{h_i}-1} \left\langle\!\left\langle \big(h_i \mathsf{E}_i^\alpha - \mathsf{Id}\big) v_i(s_i^{k+1},\cdot) + \big(h_i \mathsf{I}_i^\alpha + \mathsf{Id}\big) v_i(s_i^k,\cdot), \xi^k \right\rangle\!\right\rangle + \tfrac{1}{2}\langle\!\langle \xi^{T/h_i}, \xi^{T/h_i} \rangle\!\rangle \\
&\qquad - \sum_{k=0}^{\frac{T}{h_i}-1} \left\langle\!\left\langle \big(h_i \mathsf{E}_i^\alpha - \mathsf{Id}\big) Q_i v(s_i^{k+1},\cdot) + \big(h_i \mathsf{I}_i^\alpha + \mathsf{Id}\big) Q_i v(s_i^k,\cdot), \xi^k \right\rangle\!\right\rangle + h_i \|\xi^{T/h_i}\|^2_{H^1(\Omega)} \\
&\overset{(*)}{\le} \sum_{k=0}^{\frac{T}{h_i}-1} \left\langle\!\left\langle h_i \mathsf{F}_i^\alpha, \xi^k \right\rangle\!\right\rangle - \sum_{k=0}^{\frac{T}{h_i}-1} \left\langle\!\left\langle \big(h_i \mathsf{E}_i^\alpha - \mathsf{Id}\big) Q_i v(s_i^{k+1},\cdot) + \big(h_i \mathsf{I}_i^\alpha + \mathsf{Id}\big) Q_i v(s_i^k,\cdot), \xi^k \right\rangle\!\right\rangle \\
&\qquad + \tfrac{1}{2}\langle\!\langle \xi^{T/h_i}, \xi^{T/h_i} \rangle\!\rangle + h_i \|\xi^{T/h_i}\|^2_{H^1(\Omega)},
\end{aligned}$$

using in $(*)$ the numerical scheme and that, due to the assumptions on the Q_i, the sign of $v_i - Q_i v$ is known. Since

$$\begin{aligned}\sum_{k=0}^{\frac{T}{h_i}-1} \langle\!\langle h_i \mathsf{F}_i^\alpha, \xi^k \rangle\!\rangle &\le \|f_i^\alpha\|_{L^2} \sum_{k=0}^{\frac{T}{h_i}-1} h_i \big(\|v_i(s_i^k,\cdot) - v(s_i^k,\cdot)\|_{L^2} + \|v(s_i^k,\cdot) - Q_i v(s_i^k,\cdot)\|_{L^2} \big) \\ &\lesssim \|f_i^\alpha\|_{L^2(\Omega)} \big(\|v_i - v\|_{L^2((0,T)\times\Omega)} + \|v - Q_i v\|_{L^2((0,T)\times\Omega)} \big),\end{aligned}$$

the first term in (7.4) vanishes as $i \to \infty$. The second term vanishes due to (7.3). For the remaining terms, we recall $\xi^{T/h_i} = (v - Q_i v)(T,\cdot) \to 0$ by (7.2). Hence $|v_i - v|_{L^2([0,T],H^1(\Omega))} \to 0$ as $i \to \infty$.

THEOREM 7.1. *If there is an $\alpha \in A$ such that Assumption 7.1 holds, then the numerical solutions converge to the exact solution strongly in $L^2([0,T],H^1(\Omega))$.*

Proof. It remains to be shown that suitable Q_i can be constructed, given Assumption 7.1. Denoting the nodal interpolant on $[0,T]\times\overline{\Omega}$ by $\mathcal{I}_i$ we define

$$Q_i : W \to W_i,\ w \mapsto \mathcal{I}_i \max\{w - \|v - v_i\|_{L^\infty((0,T)\times\Omega)}, 0\}. \tag{7.5}$$

Observe that for $Q_i v$ the max operator in (7.5) switches between the first and second argument in the vicinity of $\partial\Lambda_0$ for i sufficiently large. Furthermore, $Q_i v \in W_i$ satisfies homogeneous boundary conditions and $Q_i v \le v_i$, and by the mean value theorem,

$$\begin{aligned}\|Q_i v\|_{W^{1,\infty}((0,T)\times\Omega)} &\le \|v\|_{W^{1,\infty}((0,T)\times\Omega)}, \\ \|Q_i v(T,\cdot)\|_{W^{1,\infty}(\Omega)} &\le \|v(T,\cdot)\|_{W^{1,\infty}(\Omega)}.\end{aligned} \tag{7.6}$$

Note also that for all nodes y_i^ℓ and time levels s_i^k

$$\begin{aligned}0 &\le (v_i - Q_i v)\,(s_i^k, y_i^\ell) = \min\big\{(v_i - v)\,(s_i^k, y_i^\ell) + \|v_i - v\|_{L^\infty((0,T)\times\Omega)}\,, v_i(s_i^k, y_i^\ell)\big\} \\ &\le 2\,\|v_i - v\|_{L^\infty((0,T)\times\Omega)}\,.\end{aligned} \tag{7.7}$$

Consider the set Γ_i of points which is not affected by the cutoff below 0 in (7.5) in the sense that

$$\Gamma_i := \big\{(t,x) \in (0,T)\times\Omega : \inf_{j\ge i} Q_j v(t,x) > 0 \ \text{ or } (t,x)\ \in \Lambda_0\big\}.$$

The set Γ_i' contains the points which are at least one element's length away from the boundary of $\Gamma_i \setminus \partial\Lambda_0$:

$$\Gamma_i' := \big\{(t,x) \in \Gamma_i : \{(s,y) \in (0,T)\times\Omega : \|(t,x)-(s,y)\| < \sup_{j\ge i} h_j + \Delta x_j\} \subset \Gamma_i \setminus \partial\Lambda_0\big\}.$$

Notice that Γ_i and Γ_i' are hierarchical families. Since $\|v - v_i\|_{L^\infty((0,T)\times\Omega)} \to 0$ and $h_i + \Delta x_i \to 0$ as $i \to \infty$ it follows that

$$\bigcup_{i\in\mathbb{N}} \Gamma_i' = \big((0,T)\times\Omega\big) \setminus \partial\Lambda_0.$$

Crucially, $(\partial_t Q_j v)|_{\Gamma_i'} = (\partial_t \mathcal{I}_j v)|_{\Gamma_i'}$ and $(\nabla Q_j v)|_{\Gamma_i'} = (\nabla \mathcal{I}_j v)|_{\Gamma_i'}$ for $j \ge i$.

For each $\varepsilon > 0$ there are $i, j \in \mathbb{N}$ such that $\mathrm{vol}((0,T)\times\Omega\setminus\Gamma_i') \le \varepsilon^2$ and $\|Q_k v - v\|_{H^1(\Gamma_i')} \le \varepsilon$ for all $k \ge j$. Therefore, by (7.6),

$$\begin{aligned}\|Q_k v - v\|_{H^1((0,T)\times\Omega)} &\lesssim \|Q_k v - v\|_{H^1(\Gamma_i')} + \sqrt{\mathrm{vol}((0,T)\times\Omega\setminus\Gamma_i')}\,\|v\|_{W^{1,\infty}((0,T)\times\Omega)} \\ &\le \varepsilon(1 + \|v\|_{W^{1,\infty}((0,T)\times\Omega)}),\end{aligned}$$

giving strong convergence in $H^1((0,T)\times\Omega)$, meaning convergence in the spatial gradient and the time derivative. Hence we proved that the first term in (7.2) vanishes. The second term goes to 0 because $h_i \to 0$ and since $\|(v-Q_iv)(T,\cdot)\|_{H^1(\Omega)}$ is bounded by (7.6). The last term in (7.2) vanishes in the limit, owing to inequality (7.7) and Theorem 6.2.

The terms connected to the time derivative in (7.3) vanish in the limit as

$$\sum_{k=0}^{\frac{T}{h_i}-1} \left\langle\!\!\left\langle Q_iv(s_i^{k+1},\cdot) - Q_iv(s_i^k,\cdot), \xi^k \right\rangle\!\!\right\rangle = \sum_{k=0}^{\frac{T}{h_i}-1} h_i \left\langle\!\!\left\langle (\partial_t Q_i v)|_{(s_i^k,s_i^{k+1})}, \xi^k \right\rangle\!\!\right\rangle \tag{7.8}$$

$$\lesssim \|\partial_t v\|_{L^\infty((0,T)\times\Omega)}\, \|\xi^k\|_{L^2((0,T)\times\Omega)}, \tag{7.9}$$

using the uniform convergence in ξ^k. Recall that

$$\begin{aligned}\langle\!\langle \mathsf{I}_i^\alpha Q_iv(s_i^k,\cdot), \xi^k \rangle\!\rangle = \sum_\ell (v_i - Q_iv)(s_i^k, y_i^\ell)\Big(&\bar{\bar{a}}_i^\alpha(y_i^\ell)\langle \nabla Q_iv(s_i^k,\cdot), \nabla\phi_i^\ell\rangle \\ &+ \langle \bar{\bar{b}}_i^\alpha \cdot \nabla Q_iv(s_i^k,\cdot) + \bar{\bar{c}}_i^\alpha\, Q_iv(s_i^k,\cdot), \phi_i^\ell\rangle\Big).\end{aligned}$$

The lower-order terms vanish due to the uniform convergence of $v_i - Q_iv$ to 0 and the bound

$$\sup_i \|\bar{\bar{b}}_i^\alpha \cdot \nabla Q_iv(s_i^k,\cdot) + \bar{\bar{c}}_i^\alpha\, Q_iv(s_i^k,\cdot)\|_{L^\infty(\Omega)} < \infty.$$

We note for the second-order term that

$$\sum_\ell (v_i - Q_iv)(s_i^k, y_i^\ell)\bar{\bar{a}}_i^\alpha(y_i^\ell)\langle \nabla Q_iv(s_i^k,\cdot), \nabla\phi_i^\ell\rangle = \langle \nabla Q_iv(s_i^k,\cdot), \nabla\mathcal{I}_i(\bar{\bar{a}}_i^\alpha(v_i - Q_iv))(s_i^k,\cdot)\rangle,$$

so that in (7.3) the implicit part of the second-order term becomes

$$\begin{aligned}&\sum_{k=0}^{\frac{T}{h_i}-1} h_i\, \langle \nabla Q_iv(s_i^k,\cdot), \nabla\mathcal{I}_i(\bar{\bar{a}}_i^\alpha(v_i - Q_iv))(s_i^k,\cdot)\rangle \\ &\qquad = \int_0^T \langle \mathcal{J}_i\nabla Q_iv, \mathcal{J}_i\nabla\mathcal{I}_i(\bar{\bar{a}}_i^\alpha(v_i - Q_iv))\rangle\,\mathrm{d}t,\end{aligned} \tag{7.10}$$

where $\mathcal{J}_i$ maps any $w : [0,T] \to L^2(\Omega;\mathbb{R}^d)$ onto the step function $(\mathcal{J}_iw)|_{[s_i^k,s_i^{k+1})} \equiv w(s_i^k,\cdot)$. Note that $\mathcal{J}_i\nabla Q_iv$ converges strongly in $L^2((0,T)\times\Omega;\mathbb{R}^d)$. At a time $s_i^k \in [0,T)$ the bound

$$\begin{aligned}\|\nabla\mathcal{I}_i(\bar{\bar{a}}_i^\alpha(v_i - Q_iv))\|_{L^2(\Omega;\mathbb{R}^d)} &\lesssim \|\nabla\mathcal{I}_i(\bar{\bar{a}}_i^\alpha v_i)\|_{L^2(\Omega;\mathbb{R}^d)} + \|\bar{\bar{a}}_i^\alpha Q_iv\|_{W^{1,\infty}(\Omega)} \\ &\lesssim \|\bar{\bar{a}}_i^\alpha\|_{W^{1,\infty}(\Omega)} \cdot \big(\|v_i\|_{H^1(\Omega)} + \|v\|_{W^{1,\infty}(\Omega)}\big)\end{aligned}$$

follows from an inverse estimate and

$$\begin{aligned}\sum_{K\in\mathcal{T}_i} \|\nabla\mathcal{I}_i(\bar{\bar{a}}_i^\alpha v_i)\|^2_{L^2(K;\mathbb{R}^d)} &\lesssim \sum_{K\in\mathcal{T}_i} \Delta x_K^d\, \|\nabla\mathcal{I}_i(\bar{\bar{a}}_i^\alpha v_i)\|^2_{L^\infty(K;\mathbb{R}^d)} \\ &\lesssim \sum_{K\in\mathcal{T}_i} \|\bar{\bar{a}}_i^\alpha\|^2_{W^{1,\infty}(K)}\, \big(\Delta x_K^d\, \|v_i\|^2_{W^{1,\infty}(K)}\big),\end{aligned}$$

where Δx_K denotes the diameter of the element K of the mesh $\mathcal{T}_i$. The convergence

$$\lim_{i\to\infty}\int_0^T \langle w, \mathcal{J}_i \nabla \mathcal{I}_i(\bar{\bar{a}}_i^\alpha(v_i - Q_i v))\rangle \mathrm{d}t = -\lim_{i\to\infty}\int_0^T \langle \nabla\cdot w, \mathcal{J}_I \mathcal{I}_i(\bar{\bar{a}}_i^\alpha(v_i - Q_i v))\rangle \mathrm{d}t = 0$$

with test functions w in the dense subset $C_0^1((0,T)\times\Omega;\mathbb{R}^d)$ gives weak convergence of $\nabla\mathcal{I}_i(\bar{\bar{a}}_i^\alpha(v_i - Q_i v))$ in $L^2((0,T)\times\Omega;\mathbb{R}^d)$; see [37, p. 121]. Combining weak and strong convergence [38, Prop. 21.23], it is ensured that (7.10) converges to 0 as $i\to\infty$. A similar argument guarantees that $\sum_k h_i \langle\!\langle \mathsf{E}_i^\alpha Q_i v(s_i^{k+1},\cdot),\xi^k\rangle\!\rangle$ vanishes in the limit. Therefore we proved (7.3). ∎

The regularity of the exact value function v is, for instance, discussed in section IV.8 and IV.9 of [19]. Another item of Assumption 7.1, namely, the justification of (7.1), is examined in the following example.

Example 3. (a) Suppose that a^α is positive and constant and for all smooth w,

$$L^\alpha w = I^\alpha w = -a^\alpha \Delta w + b^\alpha\cdot\nabla w + c^\alpha w, \qquad E^\alpha w = 0,$$

and to obtain semidefiniteness in the lower-order terms, $c^\alpha - \frac{1}{2}\nabla\cdot b^\alpha \geq 0$. Then, for $w\in W_i$,

$$\begin{aligned}|w|^2_{L^2([0,T],H^1(\Omega))} &\lesssim \sum_{k=0}^{\frac{T}{h_i}-1} h_i\big\langle \nabla w(s_i^k,\cdot),\nabla w(s_i^k,\cdot)\big\rangle + h_i\|w(T,\cdot)\|^2_{H^1(\Omega)}\\ &\lesssim \sum_{k=0}^{\frac{T}{h_i}-1} h_i\big\langle\!\big\langle \mathsf{I}_i^\alpha w(s_i^k,\cdot), w(s_i^k,\cdot)\big\rangle\!\big\rangle + h_i\|w(T,\cdot)\|^2_{H^1(\Omega)}.\end{aligned}$$

(b) Suppose that $a^\alpha\in W^{2,\infty}(\Omega)$ is nonconstant, positive, and uniformly bounded from below and that $c^\alpha - \frac{1}{2}(\nabla\cdot b^\alpha + \Delta a^\alpha)\geq 0$, noting for $w\in H^1(\Omega)$

$$\langle L^\alpha w, w\rangle = \langle a^\alpha\nabla w,\nabla w\rangle + \langle (c^\alpha - \tfrac{1}{2}(\nabla\cdot b^\alpha + \Delta a^\alpha))w, w\rangle. \tag{7.11}$$

Again choosing a fully implicit scheme with $L^\alpha = I^\alpha$, the highest-order term in $\langle\!\langle \mathsf{I}_i^\alpha w, w\rangle\!\rangle$ is at time s_i^k

$$\sum_\ell w(s_i^k, y_i^\ell) a^\alpha(s_i^k, y_i^\ell)\langle\nabla w(s_i^k,\cdot),\nabla\phi_i^\ell\rangle = \langle \nabla w(s_i^k,\cdot),\nabla\mathcal{I}_i(a^\alpha(s_i^k,\cdot)w(s_i^k,\cdot))\rangle.$$

According to Theorem 2.1 in [14] there is a constant $C = C(\|a^\alpha\|_{W^{2,\infty}(\Omega)})$ such that for i sufficiently large

$$\begin{aligned}\langle\nabla w,\nabla\mathcal{I}_i(a^\alpha w)\rangle - \langle\nabla w,\nabla a^\alpha w\rangle &\leq \|\nabla w\|_{L^2(\Omega;\mathbb{R}^d)}\cdot\|\mathcal{I}_i(a^\alpha w) - a^\alpha w\|_{H^1(\Omega)}\\ &\leq C\,\Delta x_i\,\|w\|^2_{H^1(\Omega)},\end{aligned}$$

using that the η appearing in the proof in [14] is defined in terms of nodal interpolation. Therefore for large i the difference between I_i^α and L^α is small, making the positivity of (7.11) available. More precisely, from Poincaré's inequality there is some C such that for $C\Delta x_i < \frac{1}{2}\inf_\Omega a^\alpha$ one has $|w|^2_{H^1(\Omega)}\lesssim\langle\!\langle \mathsf{I}_i^\alpha w, w\rangle\!\rangle$ for $w\in V_i^0$, implying (7.1).

8. Example: The method of artificial diffusion. The purpose of this section is to provide a way of constructing the operators E_i^α and I_i^α in order to satisfy Assumptions 2.1 and 2.2. This approach, called the method of artificial diffusion, is based on the fact that for strictly acute meshes, the discrete Laplacian is monotone. Further details on the method of artificial diffusion and monotone finite element schemes may, for example, be found in [7], [11], [35], and [36].

Let $\mathcal{T}_i$ be the mesh corresponding to the finite element space V_i. Given a function $g : \Omega \to \mathbb{R}^d$ and an element K of $\mathcal{T}_i$, we denote

$$|g|_K := \Big(\sum_{j=1}^{d} \|g_j\|^2_{L^\infty(K)}\Big)^{\frac{1}{2}}.$$

If g is elementwise constant, then $|g|_K$ is simply the Euclidean norm of g on K. Let Δx_K denote the diameter of K. We assume that the meshes $\mathcal{T}_i$ are strictly acute [7] in the sense that there exists $\vartheta \in (0, \pi/2)$ such that for all $i \in \mathbb{N}$

$$\nabla\phi_i^\ell \cdot \nabla\phi_i^l\big|_K \le -\sin(\vartheta)\, |\nabla\phi_i^\ell|_K\, |\nabla\phi_i^l|_K \qquad \forall \ell, l \le \dim V_i,\ \ell \ne l,\ \forall K \in \mathcal{T}_i. \tag{8.1}$$

We choose a splitting of the form $a^\alpha = \tilde{a}_i^\alpha + \tilde{\tilde{a}}_i^\alpha$, $b^\alpha = \bar{b}_i^\alpha + \bar{\bar{b}}_i^\alpha$, $c^\alpha = \bar{c}_i^\alpha + \bar{\bar{c}}_i^\alpha$, and $f^\alpha = f_i^\alpha$, where all terms are in $C(\overline{\Omega})$, $\tilde{a}_i^\alpha$ and $\tilde{\tilde{a}}_i^\alpha$ are nonnegative, and all $\bar{c}_i^\alpha$ and $\bar{\bar{c}}_i^\alpha$ are nonnegative and satisfy inequality (2.5). Choose nonnegative $\bar{\nu}_i^{\alpha,\ell}$ and $\bar{\bar{\nu}}_i^{\alpha,\ell}$ such that for all K that have y_i^ℓ as vertex

$$\big(|\bar{b}_i^\alpha|_K + \Delta x_K \|\bar{c}_i^\alpha\|_{L^\infty(K)}\big) \le \bar{\nu}_i^{\alpha,\ell}\, \sin(\vartheta)\, |\nabla\hat{\phi}_i^\ell|_K \operatorname{vol}(K), \tag{8.2a}$$

$$\big(|\bar{\bar{b}}_i^\alpha|_K + \Delta x_K \|\bar{\bar{c}}_i^\alpha\|_{L^\infty(K)}\big) \le \bar{\bar{\nu}}_i^{\alpha,\ell}\, \sin(\vartheta)\, |\nabla\hat{\phi}_i^\ell|_K \operatorname{vol}(K). \tag{8.2b}$$

Choose $\bar{a}_i^\alpha$ and $\bar{\bar{a}}_i^\alpha$ both in $C(\overline{\Omega})$ such that $\bar{a}_i^\alpha(y_i^\ell) \ge \max\{\tilde{a}_i^\alpha(y_i^\ell), \bar{\nu}_i^{\alpha,\ell}\}$ and $\bar{\bar{a}}_i^\alpha(y_i^\ell) \ge \max\{\tilde{\tilde{a}}_i^\alpha(y_i^\ell), \bar{\bar{\nu}}_i^{\alpha,\ell}\}$. Now suppose that $w \in V_i$ has a nonpositive local minimum at an interior node y_i^ℓ. By extending the arguments of [7], we show that

$$(\mathsf{E}_i^\alpha w)_\ell \le 0, \qquad (\mathsf{I}_i^\alpha w)_\ell \le 0. \tag{8.3}$$

We illustrate the proof of (8.3) for the implicit term. From the strict acuteness condition on the mesh, it can be shown that on the restriction to K [7, Lem. 3.1]

$$\nabla w \cdot \nabla\phi_i^\ell = \cos\big(\angle(\nabla w, \nabla\phi_i^\ell)\big)\, |\nabla w|_K\, |\nabla\phi_i^\ell|_K \le -\sin(\vartheta) |\nabla w|_K\, |\nabla\phi_i^\ell|_K.$$

Using $\bar{\bar{c}}_i^\alpha \ge 0$, $w(y_i^\ell) \le 0$ and $\|\hat{\phi}_i^\ell\|_{L^1(\Omega)} = 1$,

$$\begin{aligned}\langle \bar{\bar{c}}_i^\alpha\, w, \hat{\phi}_i^\ell\rangle &= \int_\Omega \bar{\bar{c}}_i^\alpha(x)\, \big(w(y_i^\ell) + \nabla w(x)\cdot(x - y_i^\ell)\big)\, \hat{\phi}_i^\ell(x)\,\mathrm{d}x \\ &\le \int_\Omega \bar{\bar{c}}_i^\alpha(x)\, \nabla w(x)\cdot(x - y_i^\ell)\, \hat{\phi}_i^\ell(x)\,\mathrm{d}x \le \sum_K \|\bar{\bar{c}}_i^\alpha\|_{L^\infty(K)}\, |\nabla w|_K\, \Delta x_K.\end{aligned}$$

Consequently,

$$\begin{aligned}(\mathsf{I}_i^\alpha w)_\ell &= \bar{\bar{a}}_i^\alpha(y_i^\ell)\langle \nabla w, \nabla\hat{\phi}_i^\ell\rangle + \langle \bar{\bar{b}}_i^\alpha\cdot\nabla w + \bar{\bar{c}}_i^\alpha\, w, \hat{\phi}_i^\ell\rangle \\ &\le \sum_K -\bar{\bar{a}}_i^\alpha(y_i^\ell)\sin(\vartheta)|\nabla w|_K\, |\nabla\hat{\phi}_i^\ell|_K \operatorname{vol}(K) \\ &\qquad + |\bar{\bar{b}}_i^\alpha|_K\, |\nabla w|_K + \|\bar{\bar{c}}_i^\alpha\|_{L^\infty(K)}\, |\nabla w|_K\, \Delta x_K \\ &\le \sum_K |\nabla w|_K \big(\big(|\bar{\bar{b}}_i^\alpha|_K + \Delta x_K \|\bar{\bar{c}}_i^\alpha\|_{L^\infty(K)}\big) - \bar{\bar{\nu}}_i^{\alpha,\ell}\, \sin(\vartheta)\, |\nabla\hat{\phi}_i^\ell|_K \operatorname{vol}(K)\big) \le 0.\end{aligned}$$

The proof of $(\mathsf{E}_i^\alpha w)_\ell \le 0$ is analogous. As hat functions ϕ_i^l attain a nonpositive minimum at all y_i^l, where $l \neq \ell$, all off-diagonal entries of E_i^α are nonpositive. Hence with a suitable time step restriction the $h_i\mathsf{E}_i^\alpha - \mathsf{Id}$ are monotone, which ensures that Assumption 2.2 is satisfied. We make the time step restriction more precise below.

The scaling of the terms in (8.2) with respect to Δx_K leads to Assumption 2.1. Due to shape-regularity, all elements K on a patch are of comparable size, giving $\|\phi_i^\ell\|_{L^1(\Omega)} \le C\,\mathrm{vol}(K)$ for all $K \subset \operatorname{supp}\phi_i^\ell$ with a constant C which is independent of i and ℓ. Hence in (8.2), we see that

$$\mathrm{vol}(K)\;|\nabla\hat{\phi}_i^\ell|_K \ge \frac{\mathrm{vol}(K)}{\Delta x_K\,\|\phi_i^\ell\|_{L^1(\Omega)}} \ge \frac{1}{C\Delta x_K}.$$

Thus, if $\bar{\nu}_i^{\alpha,\ell}$ and $\bar{\bar{\nu}}_i^{\alpha,\ell}$ are chosen optimally, then for $K \subset \operatorname{supp}\phi_i^\ell$

$$\left.\begin{aligned} \bar{\nu}_i^{\alpha,\ell} &= \mathsf{O}\big(\sup\nolimits_K\big\{|\bar{b}_i^\alpha|_K\Delta x_K + \|\bar{c}_i^\alpha\|_{L^\infty(K)}\Delta x_K^2\big\}\big)\\ \bar{\bar{\nu}}_i^{\alpha,\ell} &= \mathsf{O}\big(\sup\nolimits_K\big\{|\bar{\bar{b}}_i^\alpha|_K\Delta x_K + \|\bar{\bar{c}}_i^\alpha\|_{L^\infty(K)}\Delta x_K^2\big\}\big)\end{aligned}\right\}. \tag{8.4}$$

With (8.4) in mind we return to the time step restriction for semi-implicit and explicit methods. The nonpositivity of the diagonal terms of $h_i\mathsf{E}_i^\alpha - \mathsf{Id}$ expands to

$$\begin{aligned} 1 &\ge h_i\Big(\bar{a}_i^\alpha(y_i^\ell)\langle\nabla\phi_i^\ell,\nabla\hat{\phi}_i^\ell\rangle + \langle\bar{b}_i^\alpha\cdot\nabla\phi_i^\ell + \bar{c}_i^\alpha\,\phi_i^\ell,\hat{\phi}_i^\ell\rangle\Big)\\ &= h_i\Big(\mathsf{O}\big(\bar{a}_i^\alpha\,\Delta x_K^{-2}\big) + \mathsf{O}\big(|\bar{b}_i^\alpha|_K\,\Delta x_K^{-1}\big) + \mathsf{O}\big(\bar{c}_i^\alpha\big)\Big).\end{aligned}$$

Therefore the time step restriction imposed by L^α is $h_i \lesssim \inf_K(\Delta x_K^2/\bar{a}_i^\alpha(y_i^\ell))$, $y_i^\ell \in \overline{K}$, if there is a nonzero $\tilde{a}_i^\alpha$ and i is large. It is $h_i \lesssim \inf_K(\Delta x_K/|\bar{b}_i^\alpha(y_i^\ell)|_K)$ if all $\bar{a}_i^\alpha = 0$, $i \in \mathbb{N}$, and there are nonzero $\bar{b}_i^\alpha$, and is $\mathsf{O}(1)$ if all $\bar{a}_i^\alpha$ and $\bar{b}_i^\alpha$ vanish. There is no restriction if also all $\bar{c}_i^\alpha$ are zero.

9. Numerical experiment. Consider the HJB equation (2.2) with the following data. Let $\Omega = (-1,1)^2$, $T = 1$, and $A = [0,4]$. Let the final time data be $v_T = (1-x^2)(1-y^2)$. Let the operators L^α be defined by

$$L^\alpha v = -\big(\alpha + |x|^2/2\big)\Delta v + x\,v_x,$$

and let $f^\alpha = \alpha^2 f_2 + \alpha f_1 + f_0$ be chosen such that the exact solution of (2.2) is

$$v(x,y,t) = t\,v_T(x,y) + (1-t)\sin(\pi x)\cos(\pi y/2).$$

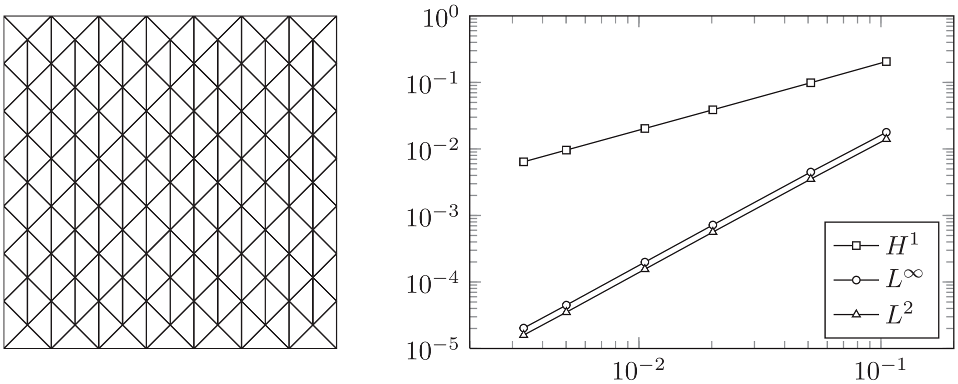
Note that the operator L^α is degenerate at the origin for $\alpha = 0$.

The problem is discretized as follows. In order to illustrate the fact that strictly acute meshes are sufficient but not always necessary, we use a sequence of non-strictly acute Delaunay triangulations of the type depicted in Figure 9.1(a). The operators L^α are split into explicit and implicit parts defined by

$$(\mathsf{I}_i^\alpha v)_\ell = \big(\max(\alpha - \nu_i^\ell, 0) + |x|^2/2\big)\langle\nabla v,\nabla\hat{\phi}_i^\ell\rangle, \tag{9.1a}$$

$$(\mathsf{E}_i^\alpha v)_\ell = \nu_i^\ell\langle\nabla v,\nabla\hat{\phi}_i^\ell\rangle + \langle x\,v_x,\hat{\phi}_i^\ell\rangle, \tag{9.1b}$$

where ν_i^ℓ is the smallest value such that the off-diagonal entries of the ℓth row of E_i^α are nonpositive. It can be checked that for meshes similar to that of Figure 9.1(a), the above choice fulfills the requirements of Assumptions 2.1 and 2.2, and it can also be



(a) (b)

FIG. 9.1. (a) *depicts the type of mesh used;* (b) *plots Δx_i on the abscissa against the error at $t = 0$ on the ordinate when using a time step $h_i = \mathsf{O}(\Delta x_i^2)$.*

TABLE 9.1
(a) *shows that a stability is achieved with a time step $h_i = \mathsf{O}(\Delta x_i)$, whereas* (b) *demonstrates that a time step $h_i = \mathsf{O}(\Delta x_i^2)$ yields optimal convergence rates in H^1 and L^2 norms.*

(a) Errors and convergence rates obtained by using a time step $h_i = \mathsf{O}(\Delta x_i)$.

Δx_i	DoF	$\|v(0,\cdot) - v_i(0,\cdot)\|_{H^1}$	Rate	$\|v(0,\cdot) - v_i(0,\cdot)\|_{L^2}$	Rate
0.1053	685	2.590e-1	1.02	4.643e-2	1.08
0.0513	2965	1.246e-1	1.00	2.141e-2	1.02
0.0202	19405	4.899e-2	1.00	8.308e-3	0.99
0.0106	71065	2.570e-2	1.00	4.369e-3	0.99
0.0050	317605	1.220e-2	1.00	2.086e-3	0.99
0.0033	716405	8.136e-3	1.00	1.395e-3	0.99
0.0025	1275205	6.103e-3		1.049e-3	

(b) Errors and convergence rates obtained by using a time step $h_i = \mathsf{O}(\Delta x_i^2)$.

Δx_i	DoF	$\|v(0,\cdot) - v_i(0,\cdot)\|_{H^1}$	Rate	$\|v(0,\cdot) - v_i(0,\cdot)\|_{L^2}$	Rate
0.1053	685	2.059e-1	1.02	1.404e-2	1.92
0.0513	2965	9.887e-2	1.00	3.520e-3	1.97
0.0202	19405	3.878e-2	1.00	5.622e-4	1.99
0.0106	71065	2.030e-2	1.00	1.554e-4	1.99
0.0050	317605	9.614e-3	1.00	3.508e-5	1.95
0.0033	716405	6.404e-3		1.588e-5	

seen that a time step $h_i = \mathsf{O}(\Delta x_i)$ is sufficient for stability. Moreover, Assumption 4.1 holds for the current choice of meshes. It is also found that ν_i^ℓ may be taken to be 0 everywhere except at the nodes closest to the origin, in which case $\nu_i^\ell = \mathsf{O}(\Delta x_i^2)$.

The numerical solutions were obtained on a sequence of meshes with mesh sizes ranging from 0.10 to 0.0025 with corresponding number of degrees of freedom ranging from 685 to 1.275×10^6. Table 9.1 presents the results of two sets of computations. The first set of results show that the discretization of (9.1) and a time step size $h_i = \mathsf{O}(\Delta x_i)$ lead to stability, whereas the second set of results, also plotted in Figure 9.1(b), shows that, similarly to linear parabolic problems, a time step size $h_i = \mathsf{O}(\Delta x_i^2)$ gives optimal convergence rates in the $H^1(\Omega)$ and $L^2(\Omega)$ norms, evaluated at the initial time $t = 0$. The number of time steps ranges in Table 9.1(a) from 20 to 956 and in 9.1(b) from 91 to 89701.

The results shown here are further supported by [22], which presents further investigations into the application of locally adapted artificial diffusion, the treatment of first-order problems, the performance of Algorithm 1 for solving the discrete equations, and the use of unstructured meshes for problems on complicated geometries.

REFERENCES

[1] G. Barles and E.R. Jakobsen, *Error bounds for monotone approximation schemes for parabolic Hamilton–Jacobi–Bellman equations*, Math. Comp., 76 (2007), pp. 1861–1893.

[2] G. Barles and B. Perthame, *Exit time problems in optimal control and vanishing viscosity method*, SIAM J. Control Optim., 26 (1988), pp. 1133–1148.

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

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

[5] K. Böhmer, *Numerical Methods for Nonlinear Elliptic Differential Equations*, Numer. Math. Sci. Comput., Oxford University Press, Oxford, UK, 2010.

[6] S. Brenner, T. Gudi, M. Neilan, and L.-Y. Sung, $\mathcal{C}^0$ *penalty methods for the fully nonlinear Monge-Ampère equation*, Math. Comp., 80 (2011), pp. 1979–1995.

[7] E. Burman and A. Ern, *Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection-diffusion-reaction equation*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 3833–3855.

[8] E. Burman and A. Ern, *Stabilized Galerkin approximation of convection-diffusion-reaction equations*, Math. Comp., 74 (2005), pp. 1637–1652.

[9] F. Camilli and M. Falcone, *An approximation scheme for the optimal control of diffusion processes*, RAIRO Modél. Math. Anal. Numér., 29 (1995), pp. 97–122.

[10] F. Camilli and E.R. Jakobsen, *A finite element like scheme for integro-partial differential Hamilton-Jacobi-Bellman equations*, SIAM J. Numer. Anal., 47 (2009), pp. 2407–2431.

[11] R. Codina, *A discontinuity capturing crosswind dissipation for the finite element solution of the convection-diffusion equation*, Comput. Methods Appl. Mech. Engrg., 110 (1993), pp. 325–342.

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

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

[14] A. Demlow, J. Guzmán, and A.H. Schatz, *Local energy estimates for the finite element method on sharply varying grids*, Math. Comp., 80 (2011), pp. 1–9.

[15] A. Demlow, D. Leykekhman, A.H. Schatz, and L.B. Wahlbin, *Best approximation property in the* W^1_∞ *norm on graded meshes*, Math. Comp., 81 (2012), pp. 743–764.

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

[17] X. Feng, R. Glowinsky, and M. Neilan, *Recent developments in numerical methods for fully nonlinear second order partial differential equations*, SIAM Rev., accepted.

[18] X. Feng and M. Neilan, *The Vanishing Moment Method for Fully Nonlinear Second Order Partial Differential Equations: Formulation, Theory, and Numerical Analysis*, arXiv:1109.1183, 2011.

[19] W. Fleming and H. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed., Stoch. Model. Appl. Probab. 25, Springer, New York, 2006.

[20] R. Horn and C. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.

[21] R. Howard, *Dynamic Programming and Markov Processes*, The Technology Press of MIT, Cambridge, MA, 1960.

[22] M. Jensen and I. Smears, *Finite element methods with artificial diffusion for Hamilton–Jacobi–Bellman equations*, in Numerical Mathematics and Advanced Applications 2011, Springer, New York, 2013.

[23] M. Kocan, *Approximation of viscosity solutions of elliptic partial differential equations on minimal grids*, Numer. Math., 72 (1995), pp. 73–92.

[24] N.V. Krylov, *On the rate of convergence of finite–difference approximations for Bellman's equations*, Algebra i Analiz, 9 (1997), pp. 245–256.

[25] N.V. Krylov, *On the rate of convergence of finite–difference approximations for Bellman's equations with variable coefficients*, Probab. Theory Related Fields, 117 (2000), pp. 1–16.

[26] H.J. Kushner, *Numerical methods for stochastic control problems in continuous time*, SIAM J. Control Optim., 28 (1990), pp. 999–1048.

[27] H.J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Appl. Math. (N.Y.) 24, Springer-Verlag, New York, 2001.

[28] O. Lakkis and T. Pryer, *A Finite Element Method for Fully Nonlinear Elliptic Problems*, arXiv:1103.2970, 2011.

[29] P.-L. Lions and B. Mercier, *Approximation numérique des équations de Hamilton–Jacobi–Bellman*, RAIRO Anal. Numér., 14 (1980), pp. 369–393.

[30] T.S. Motzkin and W. Wasow, *On the approximation of linear elliptic differential equations by difference equations with positive coefficients*, J. Math. Phys., 31 (1953), pp. 253–259.

[31] A. M. Oberman, *Convergent difference schemes for degenerate elliptic and parabolic equations*, SIAM J. Numer. Anal., 44 (2006), pp. 879–895.

[32] O.A. Oleinik and E.V. Radkevič, *Second Order Equations with Nonnegative Characteristic Form*, AMS, Providence, RI, 1973.

[33] J.-P. Quadrat, *Sur L'identification et le Contrôle de Systèmes Dynamiques Stochastiques*, Thèse d'Etat, Paris 9, 1981.

[34] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, 2nd ed., Springer Ser. Comput. Math. 25, Springer-Verlag, Berlin, 2006.

[35] J. Wang and R. Zhang, *Maximum principles for $P1$-conforming finite element approximations of quasi-linear second order elliptic equations*, SIAM' J. Numer. Anal., 50 (2012), pp. 626–642.

[36] J. Xu and L. Zikatanov, *A monotone finite element scheme for convection-diffusion equations*, Math. Comp., 68 (1999), pp. 1429–1446.

[37] K. Yosida, *Functional Analysis*, 6th ed., Springer, New York, 1980.

[38] E. Zeidler, *Nonlinear Functional Analysis and Its Applications* II, Springer, New York, 1990.