

## ERROR BOUNDS FOR MONOTONE APPROXIMATION SCHEMES FOR PARABOLIC HAMILTON-JACOBI-BELLMAN EQUATIONS

GUY BARLES AND ESPEN R. JAKOBSEN

ABSTRACT. We obtain nonsymmetric upper and lower bounds on the rate of convergence of general monotone approximation/numerical schemes for parabolic Hamilton-Jacobi-Bellman equations by introducing a new notion of consistency. Our results are robust and general – they improve and extend earlier results by Krylov, Barles, and Jakobsen. We apply our general results to various schemes including Crank–Nicholson type finite difference schemes, splitting methods, and the classical approximation by piecewise constant controls. In the first two cases our results are new, and in the last two cases the results are obtained by a new method which we develop here.

### 1. INTRODUCTION

In this article, we are interested in the rate of convergence of general monotone approximation/numerical schemes for time-dependent Hamilton-Jacobi-Bellman (HJB) equations.

The HJB equations we consider are written in the following form:

$$(1.1) \quad u_t + F(t, x, u, Du, D^2u) = 0 \quad \text{in } Q_T := (0, T] \times \mathbb{R}^N,$$

$$(1.2) \quad u(0, x) = u_0(x) \quad \text{in } \mathbb{R}^N,$$

where

$$F(t, x, r, p, X) = \sup_{\alpha \in \mathcal{A}} \{ \mathcal{L}^\alpha(t, x, r, p, X) \}$$

with

$$\mathcal{L}^\alpha(t, x, r, p, X) := -\text{tr}[a^\alpha(t, x)X] - b^\alpha(t, x)p - c^\alpha(t, x)r - f^\alpha(t, x).$$

The coefficients  $a^\alpha$ ,  $b^\alpha$ ,  $c^\alpha$ ,  $f^\alpha$  and the initial data  $u_0$  take values respectively in  $\mathbb{S}^N$ , the space of  $N \times N$  symmetric matrices,  $\mathbb{R}^N$ ,  $\mathbb{R}$ ,  $\mathbb{R}$ , and  $\mathbb{R}$ . Under suitable assumptions (see **(A1)** in Section 2), the initial value problem (1.1)-(1.2) has a unique, bounded, Hölder continuous, viscosity solution  $u$  which is the value function of a finite horizon, optimal stochastic control problem.

---

Received by the editor June 24, 2005 and, in revised form, June 29, 2006.

2000 *Mathematics Subject Classification*. Primary 65M15, 65M06, 35K60, 35K70, 49L25.

*Key words and phrases*. Hamilton-Jacobi-Bellman equations, switching system, viscosity solution, approximation schemes, finite difference methods, splitting methods, convergence rate, error bound.

Jakobsen was supported by the Research Council of Norway, grant no. 151608/432.

©2007 American Mathematical Society  
Reverts to public domain 28 years from publication

We consider approximation/numerical schemes for (1.1)–(1.2) written in the following abstract way:

$$(1.3) \quad \begin{aligned} S(h, t, x, u_h(t, x), [u_h]_{t,x}) &= 0 && \text{in } \mathcal{G}_h^+ := \mathcal{G}_h \setminus \{t = 0\}, \\ u_h(0, x) &= u_{h,0}(x) && \text{in } \mathcal{G}_h^0 := \mathcal{G}_h \cap \{t = 0\}, \end{aligned}$$

where  $S$  is, loosely speaking, a consistent, monotone and uniformly continuous approximation of the equation (1.1) defined on a grid/mesh  $\mathcal{G}_h \subset \overline{Q}_T$ . The approximation parameter  $h$  can be multi-dimensional, e.g.  $h$  could be  $(\Delta t, \Delta x)$ ,  $\Delta t$ ,  $\Delta x$  denoting time and space discretization parameters, where  $\Delta x$  can itself be multi-dimensional. The approximate solution is  $u_h : \mathcal{G}_h \rightarrow \mathbb{R}$ , and  $[u_h]_{t,x}$  is a function defined from  $u_h$  representing, typically, the value of  $u_h$  at points other than  $(t, x)$ . We assume that the total scheme including the initial value is well-defined on some appropriate subset of the space of bounded continuous functions on  $\mathcal{G}_h$ . The grid  $\mathcal{G}_h$  will vary from application to application, and when  $\mathcal{G}_h \subset \overline{Q}_T$  is discrete (as is not always the case!), any function on  $\mathcal{G}_h$  is automatically continuous.

The abstract notation  $S$  was introduced by Barles and Souganidis [3] to display clearly the monotonicity of the scheme. One of the main assumptions is that  $S$  is nondecreasing in  $u_h$  and nonincreasing in  $[u_h]_{t,x}$  with the classical ordering of functions. The typical approximation schemes we have in mind are various finite differences schemes (see, e.g., Kushner and Dupuis [12] and Bonnans and Zidani [5]), (operator) splitting methods (see, e.g., Tourin [27] and references therein), and control schemes based on the dynamic programming principle (see, e.g., Camilli and Falcone [6]). However, we will not discuss control schemes in this paper, since better results can be obtained using the different approach of [1].

The aim of this paper is to obtain estimates on the rate of the convergence of  $u_h$  to  $u$ . To obtain such results, one faces the double difficulty of having to deal with both *fully nonlinear* equations and *nonsmooth* solutions. Since these equations may also be degenerate, the (viscosity) solutions are expected to be no more than Hölder continuous in general.

Despite these difficulties, in the 1980s, Crandall and Lions [9] provided the first optimal rates of convergence for first-order equations. We refer to Souganidis [26] for more general results in this direction. For technical reasons, the problem turns out to be more difficult for second-order equations, and the question remained open for a long time.

The breakthrough came in 1997 and 2000 with Krylov's papers [19, 20], and by now there exist several papers based on and extending his ideas, e.g. [1, 2, 10, 17, 21, 22]. One of the main ideas of Krylov is a method named by himself "shaking the coefficients". Combined with a standard mollification argument, it allows one to get smooth *subsolutions* of the equation which approximate the solution. Then classical arguments involving consistency and monotonicity of the scheme yield a *one-sided bound* on the error. This method uses in a crucial way the convexity of the equation in  $u$ ,  $Du$ , and  $D^2u$ .

It is much more difficult to obtain the other bound, and essentially there are two main approaches. The first one consists of interchanging the role of the scheme and the equation. By applying the above explained ideas, one gets a sequence of appropriate smooth subsolutions of the scheme and concludes by consistency and the comparison principle for the equation. This idea was used in different articles; see [1, 10, 17, 19, 22]. Here, the key difficulty is to obtain a "continuous dependence"

result for the scheme. Even though it is now standard to prove that the solutions of the HJB equation with “shaken coefficients” remain close to the solution of the original equation, such types of results are not known for numerical schemes in general. We mention here the nice paper of Krylov [22] where such kinds of results are obtained by a tricky Bernstein type of argument. However, these results along with the corresponding error bounds, only hold for equations and schemes with special structures.

The second approach consists of considering some approximation of the equation or the associated control problem and obtaining the other bound either by probabilistic arguments (as Krylov first did using piecewise constant controls, [21, 20]) or by building a sequence of appropriate “smooth supersolutions” of the equation (see [2] where, as in the present paper, approximations by switching are considered).

The first approach leads to better error bounds than the second one, but it seems to work only for very specific schemes and with restrictions on the equations. The second approach yields error bounds in “the general case” but at the expense of lower rates.

In this paper we use the second approach by extending and improving the arguments of [2]. Compared with the various results of Krylov, we obtain better rates in most cases, our results apply to more general schemes, and we use a simpler, purely analytical approach. In fact our method is robust in the sense that it applies to “general” schemes without any particular form and under rather natural assumptions. However, as mentioned before, in certain situations the first approach yields better rates [22].

The results in [2] apply to stationary HJB equations set in the entire space  $\mathbb{R}^N$ . In this paper we extend these results to initial value problems for time-dependent HJB equations. The latter case is much more interesting in view of applications, and slightly more difficult from a mathematical point of view. We introduce a new and very general monotonicity assumption for the scheme, its generality allowing us to handle more general time-discretizations of HJB equations than have appeared in the literature so far. We relax the assumptions on the controls compared with [2], the new assumptions being more natural, and we also present several technical improvements and simplifications in the proofs.

However, the most important difference between the two papers lies in our opinion in the formulation of the consistency requirements of the main (abstract) result and in the new applications we are able to handle. Here we introduce a new, more general formulation of consistency that emphasizes more the nonsymmetrical feature of the upper and lower bounds and their proofs. It is a kind of a recipe on how to obtain error bounds in different situations, one which we feel is easier to apply to new problems and gives better insight into how the error bounds are produced. Other important contributions of this paper are the new applications we consider: (i) Finite difference methods (FDMs) using the  $\theta$ -method for time discretization (Crank–Nicholson type schemes), (ii) semidiscrete splitting methods, and (iii) approximation by piecewise constant controls. In the first two cases error bounds have not appeared before, and in the last two cases the results are obtained by a new method based on semigroup techniques which we develop here.

The results for finite difference approximations can be compared with the ones obtained by Krylov in [20, 21]. As in [2], we get the rate  $1/5$  for monotone FDMs while the corresponding result in [21] is  $1/21$ . Of course, in special situations the

rate can be improved to  $1/2$ , which is the maximal rate under our assumptions. We refer to [22] for the most general results in that direction, and to [11] for the optimality of the rate  $1/2$ . The results for semidiscrete splitting methods are new, while the result we get for the control approximation is  $1/10$ , which is worse than the  $1/6$  obtained by Krylov in [21]. It would be interesting to understand why Krylov is doing better than we are here but not in the other cases.

We conclude this introduction by explaining the notation we will use throughout this paper. By  $|\cdot|$  we mean the standard Euclidean norm in any  $\mathbb{R}^p$  type space (including the space of  $N \times P$  matrices). In particular, if  $X \in \mathbb{S}^N$ , then  $|X|^2 = \text{tr}(XX^T)$ , where  $X^T$  denotes the transpose of  $X$ .

If  $w$  is a bounded function from some set  $Q' \subset \overline{Q}_\infty$  into either  $\mathbb{R}$ ,  $\mathbb{R}^M$ , or the space of  $N \times P$  matrices, we set

$$|w|_0 = \sup_{(t,y) \in Q'} |w(t,y)|.$$

Furthermore, for  $\delta \in (0, 1]$ , we set

$$[w]_\delta = \sup_{(t,x) \neq (s,y)} \frac{|w(t,x) - w(s,y)|}{(|x - y| + |t - s|^{1/2})^\delta} \quad \text{and} \quad |w|_\delta = |w|_0 + [w]_\delta.$$

Let  $C_b(Q')$  and  $\mathcal{C}^{0,\delta}(Q')$ ,  $\delta \in (0, 1]$ , denote respectively the space of bounded continuous functions on  $Q'$  and the subset of  $C_b(Q')$  in which the norm  $|\cdot|_\delta$  is finite. Note in particular the choices  $Q' = Q_T$  and  $Q' = \mathbb{R}^N$ . In the following we always suppress the domain  $Q'$  when writing norms.

We denote by  $\leq$  the component by component ordering in  $\mathbb{R}^M$  and the ordering in the sense of positive semi-definite matrices in  $\mathbb{S}^N$ . For the rest of this paper we let  $\rho$  denote the same, fixed, positive smooth function with support in  $\{0 < t < 1\} \times \{|x| < 1\}$  and mass 1. From this function  $\rho$ , we define the sequence of mollifiers  $\{\rho_\varepsilon\}_{\varepsilon > 0}$  as follows:

$$\rho_\varepsilon(t, x) = \frac{1}{\varepsilon^{N+2}} \rho\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right) \quad \text{in} \quad \overline{Q}_\infty.$$

The rest of this paper is organized as follows: In the next section we present results on the so-called switching approximation for the problem (1.1)–(1.2). As in [2], these results are crucial to obtaining the general results on the rate of convergence of approximation/numerical schemes and are of an independent interest. Section 3 is devoted to stating and proving the main result on the rate of convergence. Finally we present applications to classical finite difference schemes, the splitting method, and classical approximation by piecewise constant controls.

## 2. CONVERGENCE RATE FOR A SWITCHING SYSTEM

In this section, we obtain the rate of convergence for a certain switching system approximations to the HJB equation (1.1). Such approximations have been studied in [13, 7], and a viscosity solutions theory of switching systems can be found in [28, 16, 15]. We consider the following type of switching systems:

$$(2.1) \quad \begin{aligned} F_i(t, x, v, \partial_t v_i, Dv_i, D^2v_i) &= 0 && \text{in} \quad Q_T, \quad i \in \mathcal{I} := \{1, \dots, M\}, \\ v(0, x) &= v_0(x) && \text{in} \quad \mathbb{R}^N, \end{aligned}$$

where the solution  $v = (v_1, \dots, v_M)$  is in  $\mathbb{R}^M$ , and for  $i \in \mathcal{I}$ ,  $(t, x) \in Q_T$ ,  $r = (r_1, \dots, r_M) \in \mathbb{R}^M$ ,  $p_t \in \mathbb{R}$ ,  $p_x \in \mathbb{R}^N$ , and  $X \in \mathcal{S}^N$ ,  $F_i$  is given by

$$F_i(t, x, r, p_t, p_x, X) = \max \left\{ p_t + \sup_{\alpha \in \mathcal{A}_i} \mathcal{L}^\alpha(t, x, r_i, p_x, X); r_i - \mathcal{M}_i r \right\},$$

where the  $\mathcal{A}_i$ 's are subsets of  $\mathcal{A}$ ,  $\mathcal{L}^\alpha$  is defined below (1.1), and for  $k > 0$ ,

$$\mathcal{M}_i r = \min_{j \neq i} \{r_j + k\}.$$

Finally, for the initial data, we are interested here in the case when  $v_0 = (u_0, \dots, u_0)$ .

Under suitable assumptions on the data (see (A1) below), we have existence and uniqueness of a solution  $v$  of this system. Moreover, it is not so difficult to see that, as  $k \rightarrow 0$ , every component of  $v$  converges locally uniformly to the solution of the following HJB equation:

$$(2.2) \quad \begin{aligned} u_t + \sup_{\alpha \in \tilde{\mathcal{A}}} \mathcal{L}^\alpha(x, u, Du, D^2u) &= 0 && \text{in } Q_T, \\ u(0, x) &= u_0(x) && \text{in } \mathbb{R}^N, \end{aligned}$$

where  $\tilde{\mathcal{A}} = \bigcup_i \mathcal{A}_i$ .

The objective of this section is to obtain an error bound for this convergence. For the sake of simplicity, we restrict ourselves to the situation where the solutions are in  $\mathcal{C}^{0,1}(Q_T)$ , i.e. when they are bounded, Lipschitz continuous in  $x$ , and Hölder 1/2 in  $t$ . This type of regularity is natural in this context. However, it is not difficult to adapt our approach to more general situations, and we give results in this direction in Section 6.

We will use the following assumption.

**(A1).** For any  $\alpha \in \mathcal{A}$ ,  $a^\alpha = \frac{1}{2} \sigma^\alpha \sigma^{\alpha T}$  for some  $N \times P$  matrix  $\sigma^\alpha$ . Moreover, there is a constant  $K$  independent of  $\alpha$  such that

$$|u_0|_1 + |\sigma^\alpha|_1 + |b^\alpha|_1 + |c^\alpha|_1 + |f^\alpha|_1 \leq K.$$

Assumption (A1) ensures the well-posedness of all the equations and systems of equations we consider in this paper; we refer the reader to the Appendix for a (partial) proof of this claim. In the present situation, we have the following well-posedness and regularity result.

**Proposition 2.1.** *Assume (A1). Then there exist unique solutions  $v$  and  $u$  of (2.1) and (2.2) respectively, satisfying*

$$|v|_1 + |u|_1 \leq C,$$

where the constant  $C$  only depends on  $T$  and  $K$  appearing in (A1).

Furthermore, if  $w_1$  and  $w_2$  are sub- and supersolutions of (2.1) or (2.2) satisfying  $w_1(0, \cdot) \leq w_2(0, \cdot)$ , then  $w_1 \leq w_2$ .

*Remark 2.1.* The functions  $\sigma^\alpha, b^\alpha, c^\alpha, f^\alpha$  are a priori only defined for times  $t \in [0, T]$ . But they can easily be extended to times  $[-r, T + r]$  for any  $r \in \mathbb{R}^+$  in such a way that (A1) still holds. In view of Proposition 2.1 we can then solve our initial value problems (2.1) and (2.2) either up to time  $T + r$  and even, by using a translation in time, on time intervals of the form  $[-r, T + r]$ . We will use this fact several times below.

In order to obtain the rate of convergence for the switching approximation, we use a regularization procedure introduced by Krylov [20, 1]. This procedure requires the following auxiliary system:

$$(2.3) \quad \begin{aligned} F_i^\varepsilon(t, x, v^\varepsilon, \partial_t v_i^\varepsilon, Dv_i^\varepsilon, D^2v_i^\varepsilon) &= 0 && \text{in } Q_{T+\varepsilon^2}, \quad i \in \mathcal{I}, \\ v^\varepsilon(0, x) &= v_0(x) && \text{in } \mathbb{R}^N, \end{aligned}$$

where  $v^\varepsilon = (v_1^\varepsilon, \dots, v_M^\varepsilon)$ ,

$$\begin{aligned} F_i^\varepsilon(t, x, r, p_t, p_x, M) \\ = \max \left\{ p_t + \sup_{\substack{\alpha \in \mathcal{A}_i \\ 0 \leq s \leq \varepsilon^2, |e| \leq \varepsilon}} \mathcal{L}^\alpha(t + s, x + e, r_i, p_x, X); r_i - M_i r \right\}, \end{aligned}$$

and  $\mathcal{L}$  and  $\mathcal{M}$  are defined below (1.2) and (2.1) respectively. Note that we use here the extension mentioned in Remark 2.1.

By Theorems A.1 and A.3 in the Appendix, we have the following result:

**Proposition 2.2.** *Assume (A1). Then there exists a unique solution  $v^\varepsilon : \overline{Q}_{T+\varepsilon^2} \rightarrow \mathbb{R}$  of (2.3) satisfying*

$$|v^\varepsilon|_1 + \frac{1}{\varepsilon} |v^\varepsilon - v|_0 \leq C,$$

where  $v$  solves (2.1) and the constant  $C$  only depends on  $T$  and  $K$  from (A1).

Furthermore, if  $w_1$  and  $w_2$  are sub- and supersolutions of (2.3) satisfying  $w_1(0, \cdot) \leq w_2(0, \cdot)$ , then  $w_1 \leq w_2$ .

We are now in a position to state and prove the main result of this section.

**Theorem 2.3.** *Assume (A1) and  $v_0 = (u_0, \dots, u_0)$ . If  $u$  and  $v$  are the solutions of (2.2) and (2.1) respectively, then for  $k$  small enough,*

$$0 \leq v_i - u \leq Ck^{1/3} \quad \text{in } Q_T, \quad i \in \mathcal{I},$$

where  $C$  only depends on  $T$  and  $K$  from (A1).

*Proof.* Since  $w = (u, \dots, u)$  is a subsolution of (2.1), comparison for (2.1) (Proposition 2.1) yields  $u \leq v_i$  for  $i \in \mathcal{I}$ .

To get the other bound, we use an argument suggested by P.-L. Lions [23] together with the regularization procedure of Krylov [20]. Consider first system (2.3). It follows that, for every  $0 \leq s \leq \varepsilon^2, |e| \leq \varepsilon$ ,

$$\partial_t v_i^\varepsilon + \sup_{\alpha \in \mathcal{A}_i} \mathcal{L}^\alpha(t + s, x + e, v_i^\varepsilon(t, x), Dv_i^\varepsilon, D^2v_i^\varepsilon) \leq 0 \quad \text{in } Q_{T+\varepsilon^2}, \quad i \in \mathcal{I}.$$

After a change of variables, we see that for every  $0 \leq s \leq \varepsilon^2, |e| \leq \varepsilon, v^\varepsilon(t - s, x - e)$  is a subsolution of the following system of uncoupled equations:

$$(2.4) \quad \partial_t w_i + \sup_{\alpha \in \mathcal{A}_i} \mathcal{L}^\alpha(t, x, w_i, Dw_i, D^2w_i) = 0 \quad \text{in } Q_T^\varepsilon, \quad i \in \mathcal{I},$$

where  $Q_T^\varepsilon := (\varepsilon^2, T) \times \mathbb{R}^N$ . Define  $v_\varepsilon := v^\varepsilon * \rho_\varepsilon$ , where  $\{\rho_\varepsilon\}_\varepsilon$  is the sequence of mollifiers defined at the end of the introduction. A Riemann-sum approximation shows that  $v_\varepsilon(t, x)$  can be viewed as the limit of convex combinations of  $v^\varepsilon(t - s, x - e)$ 's for  $0 < s < \varepsilon^2$  and  $|e| < \varepsilon$ . Since the  $v^\varepsilon(t - s, x - e)$ 's are subsolutions of the convex equation (2.4), so are the convex combinations. By the stability result for viscosity subsolutions we can now conclude that  $v_\varepsilon$  is itself a subsolution of (2.4). We refer to the Appendix in [1] for more details.

On the other hand, since  $v^\varepsilon$  is a continuous subsolution of (2.3), we have

$$v_i^\varepsilon \leq \min_{j \neq i} v_j^\varepsilon + k \quad \text{in } Q_{T+\varepsilon^2}, \quad i \in \mathcal{I}.$$

It follows that  $\max_i v_i^\varepsilon(t, x) - \min_i v_i^\varepsilon(t, x) \leq k$  in  $Q_{T+\varepsilon^2}$ , and hence

$$|v_i^\varepsilon - v_j^\varepsilon|_0 \leq k, \quad i, j \in \mathcal{I}.$$

Then, by the definition and properties of  $v_\varepsilon$ , we have

$$|\partial_t v_{\varepsilon i} - \partial_t v_{\varepsilon j}|_0 \leq C \frac{k}{\varepsilon^2}, \quad |D^n v_{\varepsilon i} - D^n v_{\varepsilon j}|_0 \leq C \frac{k}{\varepsilon^n}, \quad n \in \mathbb{N}, \quad i, j \in \mathcal{I},$$

where  $C$  depends only on  $\rho$  and the uniform bounds on  $v_{\varepsilon i}$  and  $Dv_{\varepsilon i}$ , i.e. on  $T$  and  $K$  given in (A1). Furthermore, from these bounds, we see that for  $\varepsilon < 1$ ,

$$\left| \partial_t v_{\varepsilon j} + \sup_{\alpha \in \mathcal{A}_i} \mathcal{L}^\alpha[v_{\varepsilon j}] - \partial_t v_{\varepsilon i} - \sup_{\alpha \in \mathcal{A}_i} \mathcal{L}^\alpha[v_{\varepsilon i}] \right| \leq C \frac{k}{\varepsilon^2} \quad \text{in } Q_T^\varepsilon, \quad i, j \in \mathcal{I}.$$

Here, as above,  $C$  only depends on  $\rho$ ,  $T$  and  $K$ . Since  $v_\varepsilon$  is a subsolution of (2.4), this means that

$$\partial_t v_{\varepsilon i} + \sup_{\alpha \in \mathcal{A}} \mathcal{L}^\alpha(x, v_{\varepsilon i}, Dv_{\varepsilon i}, D^2 v_{\varepsilon i}) \leq C \frac{k}{\varepsilon^2} \quad \text{in } Q_T^\varepsilon, \quad i \in \mathcal{I}.$$

From assumption (A1) and the structure of the equation, we see that  $v_{\varepsilon i} - te^{Kt} C \frac{k}{\varepsilon^2}$  is a subsolution of equation (2.2) restricted to  $Q_T^\varepsilon$ .

Comparison for (2.2) restricted to  $Q_T^\varepsilon$  (Proposition 2.1) yields

$$v_{\varepsilon i} - u \leq e^{Kt} \left( |v_{\varepsilon i}(\varepsilon^2, \cdot) - u(\varepsilon^2, \cdot)|_0 + Ct \frac{k}{\varepsilon^2} \right) \quad \text{in } Q_T^\varepsilon, \quad i \in \mathcal{I}.$$

Regularity of  $u$  and  $v_i$  (Proposition 2.1) implies that

$$|u(t, \cdot) - v_i(t, \cdot)|_0 \leq ([u]_1 + [v_i]_1)\varepsilon \quad \text{in } [0, \varepsilon^2].$$

Hence by Proposition 2.2, regularity of  $u$  and  $v_i^\varepsilon$ , and properties of mollifiers, we have

$$v_i - u \leq v_i - v_{\varepsilon i} + v_{\varepsilon i} - u \leq C\left(\varepsilon + \frac{k}{\varepsilon^2}\right) \quad \text{in } Q_T^\varepsilon, \quad i \in \mathcal{I}.$$

Minimizing w.r.t.  $\varepsilon$  now yields the result. □

### 3. CONVERGENCE RATE FOR THE HJB EQUATION

In this section we derive our main result, an error bound for the convergence of the solution of the scheme (1.3) to the solution of the HJB equation (1.1)–(1.2). This result extends and improves earlier results by Krylov [19, 20], Barles and Jakobsen [1, 2, 17].

Throughout this section, we assume that (A1) holds and we recall that, by Proposition 2.1, there exists a unique  $C^{0,1}$ -solution  $u$  of (1.1) satisfying  $|u|_1 \leq C$ , where the constant  $C$  only depends on  $T$  and  $K$  from (A1). In Section 6, we will weaken this assumption and give results for  $C^{0,\beta}$  solutions,  $\beta \in (0, 1)$ . In addition, to get a lower bound on the error, we need the following technical assumption that allows us to approximate the HJB equation (1.1) by another HJB equation where the supremum (maximum) is over a finite set:

**(A2).** The control set  $\mathcal{A}$  is a separable metric space, and the coefficients  $\sigma^\alpha, b^\alpha, c^\alpha, f^\alpha$  are continuous in  $\alpha$  for all  $x, t$ .

This assumption is used by Krylov [20, 21], and it is more natural and general than the one used in [2] (Assumption (A3) page 8).

Next we introduce the assumptions for the scheme (1.3).

**(S1) (Monotonicity).** There exist  $\lambda, \mu \geq 0, h_0 > 0$  such that if  $|h| \leq h_0, u \leq v$  are functions in  $C_b(\mathcal{G}_h)$ , and  $\phi(t) = e^{\mu t}(a + bt) + c$  for  $a, b, c \geq 0$ , then

$$S(h, t, x, r + \phi(t), [u + \phi]_{t,x}) \geq S(h, t, x, r, [v]_{t,x}) + b/2 - \lambda c \quad \text{in } \mathcal{G}_h^+.$$

**(S2) (Regularity).** For every  $h$  and  $\phi \in C_b(\mathcal{G}_h)$ , the function  $(t, x) \mapsto S(h, t, x, \phi(t, x), [\phi]_{t,x})$  is bounded and continuous in  $\mathcal{G}_h^+$  and the function  $r \mapsto S(h, t, x, r, [\phi]_{t,x})$  is uniformly continuous for bounded  $r$ , uniformly in  $(t, x) \in \mathcal{G}_h^+$ .

Assumptions (S1) and (S2) imply a comparison result for the scheme (1.3); see Lemma 3.2 below. Assumption (S1) is very general and allows one to handle more general time discretizations than in previous papers; see Remark 3.4 below.

*Remark 3.1.* In (S1) and (S2) (and in Theorem 3.1 below) we may replace  $C_b(\mathcal{G}_h)$  by any subset of this space as long as (1.3) is well-defined for functions in this set. E.g. in Section 4 we replace it by  $C_b(\mathbb{R}^N)$  and in Section 5 by  $C(\{0, 1, \dots, n_T\}; \mathcal{C}^{0,1}(\mathbb{R}^N))$  (since  $\mathcal{G}_h = \{0, 1, \dots, n_T\} \times \mathbb{R}^N$  there).

Let us now state the key consistency conditions.

**(S3)(i) (Subconsistency).** There exists a function  $E_1(\tilde{K}, h, \varepsilon)$  such that for any sequence  $\{\phi_\varepsilon\}_{\varepsilon>0}$  of smooth functions satisfying

$$|\partial_t^{\beta_0} D^{\beta'} \phi_\varepsilon(x, t)| \leq \tilde{K} \varepsilon^{1-2\beta_0-|\beta'|} \quad \text{in } \overline{Q}_T, \quad \text{for any } \beta_0 \in \mathbb{N}, \beta' = (\beta'_i)_i \in \mathbb{N}^N,$$

where  $|\beta'| = \sum_{i=1}^N \beta'_i$ , the following inequality holds:

$$S(h, t, x, \phi_\varepsilon(t, x), [\phi_\varepsilon]_{t,x}) \leq \phi_{\varepsilon t} + F(t, x, \phi, D\phi_\varepsilon, D^2\phi_\varepsilon) + E_1(\tilde{K}, h, \varepsilon) \quad \text{in } \mathcal{G}_h^+.$$

**(S3)(ii) (Superconsistency).** There exists a function  $E_2(\tilde{K}, h, \varepsilon)$  such that for any sequence  $\{\phi_\varepsilon\}_\varepsilon$  of smooth functions satisfying

$$|\partial_t^{\beta_0} D^{\beta'} \phi_\varepsilon(x, t)| \leq \tilde{K} \varepsilon^{1-2\beta_0-|\beta'|} \quad \text{in } \overline{Q}_T, \quad \text{for any } \beta_0 \in \mathbb{N}, \beta' \in \mathbb{N}^N,$$

the following inequality holds:

$$S(h, t, x, \phi_\varepsilon(t, x), [\phi_\varepsilon]_{t,x}) \geq \phi_{\varepsilon t} + F(t, x, \phi, D\phi_\varepsilon, D^2\phi_\varepsilon) - E_2(\tilde{K}, h, \varepsilon) \quad \text{in } \mathcal{G}_h^+.$$

The typical  $\phi_\varepsilon$  we have in mind in (S3) is of the form  $\chi_\varepsilon * \rho_\varepsilon$ , where  $(\chi_\varepsilon)_\varepsilon$  is a sequence of uniformly bounded functions in  $\mathcal{C}^{0,1}$  and  $\rho_\varepsilon$  is the mollifier defined at the end of the introduction. This function satisfies the derivative bounds of (S3).

The main result in this paper is the following:

**Theorem 3.1.** *Assume (A1), (S1), (S2) and that (1.3) has a unique solution  $u_h$  in  $C_b(\mathcal{G}_h)$ . Let  $u$  denote the solution of (1.1)–(1.2), and let  $h$  be sufficiently small.*

(a) **(Upper bound)** *If (S3)(i) holds, then there exists a constant  $C$  depending only on  $\mu, K$  in (S1), (A1) such that*

$$u - u_h \leq e^{\mu t} |(u_0 - u_{0,h})^+|_0 + C \min_{\varepsilon>0} (\varepsilon + E_1(\tilde{K}, h, \varepsilon)) \quad \text{in } \mathcal{G}_h,$$

for  $h$  small enough and  $\tilde{K} = |u|_1$ .



(b) **(Lower bound)** If (S3)(ii) and (A3) hold, then there exists a constant  $C$  depending only on  $\mu, K$  in (S1), (A1) such that

$$u - u_h \geq -e^{\mu t} |(u_0 - u_{0,h})^-|_0 - C \min_{\varepsilon > 0} \left( \varepsilon^{1/3} + E_2(\tilde{K}, h, \varepsilon) \right) \quad \text{in } \mathcal{G}_h,$$

for  $h$  small enough and  $\tilde{K} = |u|_1$ .

The motivation for this new formulation of the upper and lower bounds is three-fold: (i) in some applications,  $E_1 \neq E_2$  and therefore it is natural to have such asymmetry in the consistency requirement (see Section 5), (ii) from the proof it can be seen that the upper bound (a) is proven independently of the lower bound (b), and most importantly, (iii) the new formulation describes completely how the bounds are obtained from the consistency requirements. The good  $h$ -dependence and the bad  $\varepsilon$ -dependence of  $E_1$  and  $E_2$  are combined in the minimization process to give the final bounds; see Remark 3.2 below.

Since the minimum is achieved for  $\varepsilon \ll 1$ , the upper bound is in general much better than the lower bound (in particular in cases where  $E_1 = E_2$ ).

Finally note that the existence of a  $u_h$  in (some appropriate subset of)  $C_b(\mathcal{G}_h)$  must be proved for each particular scheme  $S$ . We refer to [19, 20, 1, 17] for examples of such arguments.

*Remark 3.2.* In the case of a finite difference method with a time step  $\Delta t$  and maximal mesh size in space  $\Delta x$ , a standard formulation of the consistency requirement would be

**(S3')**. There exist finite sets  $I \subset \mathbb{N} \times \mathbb{N}_0^N, \bar{I} \subset \mathbb{N}_0 \times \mathbb{N}^N$  and constants  $K_c \geq 0, k_\beta, \bar{k}_{\bar{\beta}}$  for  $\beta = (\beta_0, \beta') \in I, \bar{\beta} = (\bar{\beta}_0, \bar{\beta}') \in \bar{I}$  such that for every  $h = (\Delta t, \Delta x) > 0, (t, x) \in \mathcal{G}_h^+$ , and smooth functions  $\phi$ :

$$\begin{aligned} & \left| \phi_t + F(t, x, \phi, D\phi, D^2\phi) - S(h, t, x, \phi(t, x), [\phi]_{t,x}) \right| \\ & \leq K_c \sum_{\beta \in I} |\partial_t^{\beta_0} D^{\beta'} \phi|_0 \Delta t^{k_\beta} + K_c \sum_{\bar{\beta} \in \bar{I}} |\partial_t^{\bar{\beta}_0} D^{\bar{\beta}'} \phi|_0 \Delta x^{\bar{k}_{\bar{\beta}}}. \end{aligned}$$

The corresponding version of (S3) is obtained by plugging  $\phi_\varepsilon$  into (S3') and using the estimates on its derivatives. The result is

$$\begin{aligned} E_1(\tilde{K}, h, \varepsilon) &= E_2(\tilde{K}, h, \varepsilon) \\ &= \tilde{K} K_c \sum_{\beta \in I} \varepsilon^{1-2\beta_0-|\beta'|} \Delta t^{k_\beta} + \tilde{K} K_c \sum_{\bar{\beta} \in \bar{I}} \varepsilon^{1-2\bar{\beta}_0-|\bar{\beta}'|} \Delta x^{\bar{k}_{\bar{\beta}}}. \end{aligned}$$

From this formula we see that the dependence in the small parameter  $\varepsilon$  is bad since all the exponents of  $\varepsilon$  are negative, while the dependence on  $\Delta t, \Delta x$  is good since their exponents are positive.

*Remark 3.3.* Assumption (S1) contains two different kinds of information. First, by taking  $\phi \equiv 0$  it implies that the scheme is nondecreasing with respect to the  $[u]$  argument. Second, by taking  $u \equiv v$  it indicates that a parabolic equation, i.e., an equation with a  $u_t$  term, is being approximated. Both these points play a crucial role in the proof of the comparison principle for (1.3) (Lemma 3.2 below).

To better understand that assumption (S1) implies parabolicity of the scheme, consider the following more restrictive assumption:

**(S1') (Monotonicity).** There exists  $\lambda \geq 0$ ,  $\bar{K} > 0$  such that if  $u \leq v$ ,  $u, v \in C_b(\mathcal{G}_h)$ , and  $\phi : [0, T] \rightarrow \mathbb{R}$  is smooth, then

$$\begin{aligned} & S(h, t, x, r + \phi(t), [u + \phi]_{t,x}) \\ & \geq S(h, t, x, r, [v]_{t,x}) + \phi'(t) - \bar{K}\Delta t|\phi''|_0 - \lambda\phi^+(t) \quad \text{in } \mathcal{G}_h^+. \end{aligned}$$

Here  $h = (\Delta t, h')$  where  $h'$  represents a small parameter related to, e.g., the space discretization. It is easy to see that (S1') implies (S1), e.g. with the same value for  $\lambda$  and the following values of  $\mu$  and  $h_0$ :

$$\mu = \lambda + 1 \quad \text{and} \quad h_0^{-1} = 2\bar{K}e^{(\lambda+1)T}(\lambda + 1)(2 + (\lambda + 1)T).$$

*Remark 3.4.* Assumptions (S1') and (S1) are more general than the corresponding assumptions used in earlier papers by Krylov and ourselves [20, 17]. They are satisfied for all *monotone* finite differences in time approximations of (1.1), e.g. *monotone* Runge-Kutta methods and *monotone* multi-step methods, both explicit and implicit methods. Note that whereas many Runge-Kutta methods lead to monotone schemes for (1.1) (possibly under a CFL condition), it seems that the most commonly used multi-step methods (Adams-Bashforth, BDS) do not. We refer to [25] for a multi-step method that yields a monotone approximation of (1.1).

Using the generality of (S1), we provide the first error bounds for Crank–Nicholson type schemes for (1.1) in Section 4.

In the proof of Theorem 3.1 we need the following comparison result for bounded continuous sub- and supersolutions of (1.3).

**Lemma 3.2.** *Assume (S1), (S2), and that  $u, v \in C_b(\mathcal{G}_h)$  satisfy*

$$\begin{aligned} S(h, t, x, u(t, x), [u]_{t,x}) &\leq g_1 \quad \text{in } \mathcal{G}_h^+, \\ S(h, t, x, v(t, x), [v]_{t,x}) &\geq g_2 \quad \text{in } \mathcal{G}_h^+, \end{aligned}$$

where  $g_1, g_2 \in C_b(\mathcal{G}_h)$ . Then

$$u - v \leq e^{\mu t} |(u(0, \cdot) - v(0, \cdot))^+|_0 + 2te^{\mu t} |(g_1 - g_2)^+|_0,$$

where  $\lambda$  and  $\mu$  are given by (S1).

*Proof.* 1. First note that it suffices to prove the lemma in the case

$$(3.1) \quad u(0, x) - v(0, x) \leq 0 \quad \text{in } \mathcal{G}_h^0,$$

$$(3.2) \quad g_1(t, x) - g_2(t, x) \leq 0 \quad \text{in } \mathcal{G}_h.$$

The general case follows from this result after noting that by (S1),

$$w = v + e^{\mu t} (|(u(0, \cdot) - v(0, \cdot))^+|_0 + 2t|(g_1 - g_2)^+|_0)$$

satisfies  $S(h, t, x, w(t, x), [w]_{t,x}) \geq g_1$  in  $\mathcal{G}_h^+$  and  $u(0, x) - w(0, x) \leq 0$  in  $\mathcal{G}_h^0$ .

2. We assume that (3.1) and (3.2) hold and, for  $b \geq 0$ , we set  $\psi_b(t) = e^{\mu t} 2bt$ , where  $\mu$  is given by (S1) and

$$M(b) = \sup_{\mathcal{G}_h} \{u - v - \psi_b\}.$$

We have to prove that  $M(0) \leq 0$  and we argue by contradiction assuming  $M(0) > 0$ .

3. Consider  $b \geq 0$  for which  $M(b) > 0$  and take a sequence  $\{(t_n, x_n)\}_n \subset \mathcal{G}_h$  such that

$$\delta_n := M(b) - (u - v - \psi_b)(t_n, x_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Since  $M(b) > 0$  and (3.1) holds,  $t_n > 0$  for all sufficiently large  $n$ . For such  $n$ ,

$$\begin{aligned} g_1 &\geq S(h, t_n, x_n, u, [u]_{t_n, x_n}) && (u \text{ subsolution}) \\ &\geq S(h, t_n, x_n, v + \psi_b + M(b) - \delta_n, [v + \psi_b + M(b)]_{t_n, x_n}) && (S1), \phi \equiv 0 \\ &\geq \omega(\delta_n) \\ &\quad + S(h, t_n, x_n, v + \psi_b + M(b), [v + \psi_b + M(b)]_{t_n, x_n}) && (S2) \\ &\geq \omega(\delta_n) + b - \lambda M(b) + S(h, t_n, x_n, v, [v]_{t_n, x_n}) && (S1), \phi = \psi + M \\ &\geq \omega(\delta_n) + b - \lambda M(b) + g_2, && (v \text{ supersolution}) \end{aligned}$$

where we have dropped the dependence in  $t_n, x_n$  of  $u, v$  and  $\psi_b$  to simplify the notation. Recalling (3.2) and sending  $n \rightarrow \infty$  lead to

$$b - \lambda M(b) \leq 0 .$$

4. Since  $M(b) \leq M(0)$ , the above inequality yields a contradiction for  $b$  large, so for such  $b$ ,  $M(b) \leq 0$ . On the other hand, since  $M(b)$  is a continuous function of  $b$  and  $M(0) > 0$ , there exists a minimal solution  $\bar{b} > 0$  of  $M(\bar{b}) = 0$ . For  $\delta > 0$  satisfying  $\bar{b} - \delta > 0$ , we have  $M(\bar{b} - \delta) > 0$  and  $M(\bar{b} - \delta) \rightarrow 0$  as  $\delta \rightarrow 0$ . But, by 3 we have

$$\bar{b} - \delta \leq \lambda M(\bar{b} - \delta),$$

which is a contradiction for  $\delta$  small enough since  $\bar{b} > 0$ . □

**Proof of Theorem 3.1 (a).** We only sketch it since it relies on the regularization procedure of Krylov used in Section 2. More details can be found in [19, 20, 1, 17]. The main steps are:

1. Introduce the solution  $u^\varepsilon$  of

$$\begin{aligned} u_t^\varepsilon + \sup_{0 \leq s \leq \varepsilon^2, |e| \leq \varepsilon} F(t + s, x + e, u^\varepsilon(t, x), Du^\varepsilon, D^2u^\varepsilon) &= 0 && \text{in } Q_{T+\varepsilon^2}, \\ u^\varepsilon(x, 0) &= u_0(x) && \text{in } \mathbb{R}^N. \end{aligned}$$

Essentially as a consequence of Proposition 2.1, it follows that  $u^\varepsilon$  belongs to  $\mathcal{C}^{0,1}(Q_T)$  with a uniform  $\mathcal{C}^{0,1}(Q_T)$ -bound  $\bar{K}$ .

2. By analogous arguments to the ones used in Section 2, it is easy to see that  $u_\varepsilon := u^\varepsilon * \rho_\varepsilon$  is a subsolution of (1.1). By combining regularity and continuous dependence results (Theorem A.3 in the Appendix), we also have  $|u_\varepsilon - u|_0 \leq C\varepsilon$ , where  $C$  only depends on  $T$  and  $K$  in (A1).

3. Plugging  $u_\varepsilon$  into the scheme and using (S3)(i) and the uniform estimates on  $u^\varepsilon$  we get

$$S(h, t, x, u_\varepsilon(t, x), [u_\varepsilon]_{t, x}) \leq E_1(\bar{K}, h, \varepsilon) \text{ in } \mathcal{G}_h^+,$$

where  $\bar{K}$  is the above-mentioned uniform  $\mathcal{C}^{0,1}$  estimate on  $u^\varepsilon$ ; see step 1.

4. Use Lemma 3.2 to compare  $u_\varepsilon$  and  $u_h$  and conclude by using the control we have on  $u - u_\varepsilon$  and by taking the minimum in  $\varepsilon$ .

**Proof of Theorem 3.1 (b).** Here, unfortunately, we cannot follow the scheme of the proof of (a) since we do not know how to obtain a sequence of approximate, global, smooth supersolutions of (1.1). Instead we build approximate supersolutions which are smooth at the “right points”. First we consider the case when  $\mathcal{A}$  is finite using a switching system approach similar to the approach in [2]. Then we do the general case using our new Assumption (A2), and finally we prove a key lemma used in the first part of the proof.

**1. The case  $\mathcal{A} = \{\alpha_1, \dots, \alpha_M\}$ .** We build the “almost smooth” supersolutions out of the solutions of the following switching system approximation of (1.1):

$$(3.3) \quad \begin{aligned} F_i^\varepsilon(t, x, v^\varepsilon, \partial_t v_i^\varepsilon, Dv_i^\varepsilon, D^2v_i^\varepsilon) &= 0 && \text{in } Q_{T+2\varepsilon^2}, \quad i \in \mathcal{I} := \{1, \dots, M\}, \\ v^\varepsilon(0, x) &= v_0(x) && \text{in } \mathbb{R}^N, \end{aligned}$$

where  $v^\varepsilon = (v_1^\varepsilon, \dots, v_M^\varepsilon)$ ,  $v_0 = (u_0, \dots, u_0)$ ,

$$(3.4) \quad \begin{aligned} &F_i^\varepsilon(t, x, r, p_t, p_x, X) \\ &= \max \left\{ p_t + \min_{0 \leq s \leq \varepsilon^2, |e| \leq \varepsilon} \mathcal{L}^{\alpha_i}(t + s - \varepsilon^2, x + e, r_i, p_x, X); r_i - \mathcal{M}_i r \right\}, \end{aligned}$$

and  $\mathcal{L}$  and  $\mathcal{M}$  are defined below (1.1) and (2.1) respectively. When  $k$  and  $\varepsilon$  are small, the solution of this system is expected to be close to the solution of (1.1) (remember that  $\mathcal{A} = \{\alpha_1, \dots, \alpha_M\}$ !). In fact we have the following result.

**Lemma 3.3.** *Assume (A1). There exists a unique solution  $v^\varepsilon$  of (3.3) satisfying*

$$|v^\varepsilon|_1 \leq \bar{K}, \quad |v_i^\varepsilon - v_j^\varepsilon|_0 \leq k, \quad \text{and, for } k \text{ small, } \max_{i \in \mathcal{I}} |u - v_i^\varepsilon|_0 \leq C(\varepsilon + k^{1/3}),$$

where  $u$  solves (1.1) and (1.2),  $i, j \in \mathcal{I}$ , and  $\bar{K}, C$  only depend on  $T$  and  $K$  from (A1).

*Proof.* Theorems A.1 and A.3 in the Appendix yield the existence and uniqueness of a solution  $v^\varepsilon : \bar{Q}_{T+2\varepsilon^2} \rightarrow \mathbb{R}$  of (3.3) satisfying  $|v^\varepsilon|_1 + \frac{1}{\varepsilon}|v^\varepsilon - v^0|_0 \leq C$ . Here  $v^0$  is the solution of (3.3) with  $\varepsilon = 0$ , i.e. of (2.3) in Section 2. Theorem 2.3 then implies that  $|u - v_i^0|_0 \leq Ck^{1/3}$ . Finally, an argument given in the proof of Theorem 2.3 in Section 2 shows that  $0 \leq \max_i v_i^\varepsilon - \min_i v_i^\varepsilon \leq k$  in  $Q_{T+\varepsilon^2}$ . The constants  $C$  depend only on  $T$  and  $K$  from (A1).  $\square$

In order to simplify the arguments in this proof (to have the simplest possible formulation of Lemma 3.5 below), we have defined the solutions of the equation with “shaken coefficients” (3.3) in a slightly larger domain than  $Q_T$ .

The (almost) smooth supersolutions of (1.1) we are looking for are built out of  $v^\varepsilon$  by mollification. But to have initial data at  $t = 0$  after mollification of  $v^\varepsilon$ , we have to shift  $v^\varepsilon$  in time and consider

$$\bar{v}_i^\varepsilon(t, x) := v_i^\varepsilon(t - \varepsilon^2, x).$$

The function  $\bar{v}^\varepsilon$  is defined on  $Q_T^\varepsilon := (-\varepsilon^2, T + \varepsilon^2] \times \mathbb{R}^N$ , and by Lemma 3.3 it satisfies  $|\bar{v}^\varepsilon|_1 \leq \bar{K}$  and hence

$$(3.5) \quad |\bar{v}_i^\varepsilon - v_i^\varepsilon|_0 \leq \bar{K}\varepsilon \quad \text{for } i \in \mathcal{I}.$$

We mollify  $\bar{v}^\varepsilon$  and define

$$v_{\varepsilon,i} := \rho_\varepsilon * \bar{v}_i^\varepsilon \quad \text{for } i \in \mathcal{I},$$

where  $\rho_\varepsilon$  is the mollifier defined at the end of the introduction. Note that  $v_\varepsilon$  is defined on  $\bar{Q}_{T+\varepsilon^2}$ , and in view of Lemma 3.3, inequality (3.5), and properties of mollifiers it satisfies  $|v_\varepsilon|_1 \leq \bar{K}$ , and for  $k$  small,

$$(3.6) \quad |v_{\varepsilon,i} - v_{\varepsilon,j}| \leq C(k + \varepsilon) \text{ in } Q_{T+\varepsilon^2}, \quad \max_{i \in \mathcal{I}} |u - v_{\varepsilon,i}| \leq C(\varepsilon + k^{1/3}) \text{ in } Q_T,$$

where  $i, j \in \mathcal{I}$  and the constants  $C$  only depend on  $T$  and  $K$  from (A1).

Now we are in a position to define our “almost smooth” supersolutions  $w$ ,

$$w := \min_{i \in \mathcal{I}} v_{\varepsilon,i} \quad \text{for } i \in \mathcal{I}.$$

The way we have built  $v_\varepsilon$ , it turns out that  $w$  will be a supersolution of (1.1) in  $Q_T$  when  $\varepsilon$  is small compared with  $k$ . This is a consequence of the following lemma.

**Lemma 3.4.** *Assume (A1) and  $\varepsilon \leq (8 \sup_i [v_i^\varepsilon]_1)^{-1}k$ . For every  $(t, x) \in Q_T$ , if  $j := \operatorname{argmin}_{i \in \mathcal{I}} v_{\varepsilon_i}(t, x)$ , then*

$$\partial_t v_{\varepsilon_j}(t, x) + \mathcal{L}^{\alpha_j}(t, x, v_{\varepsilon_j}(t, x), Dv_{\varepsilon_j}(t, x), D^2v_{\varepsilon_j}(t, x)) \geq 0.$$

The proof of this lemma will be given at the end of this section. The key consequence is the next lemma, which shows that  $w$  is an approximate supersolution of the scheme (1.3). This result is the cornerstone of the proof of the lower bound, and its proof should explain the name “almost smooth” supersolution for  $w$ .

**Lemma 3.5.** *Assume (A1) and  $\varepsilon \leq (8 \sup_i [v_i^\varepsilon]_1)^{-1}k$ . Then the function  $w := \min_{i \in \mathcal{I}} v_{\varepsilon_i}$  is an approximate supersolution of the scheme (1.3) in the sense that*

$$S(h, t, x, w(t, x), [w]_{t,x}) \geq -E_2(\bar{K}, h, \varepsilon) \quad \text{in } \mathcal{G}_h^+,$$

where  $\bar{K}$  comes from Lemma 3.3.

*Proof.* Let  $(t, x) \in Q_T$  and  $j$  be as in Lemma 3.4. We see that  $w(t, x) = v_{\varepsilon_j}(t, x)$  and  $w \leq v_{\varepsilon_j}$  in  $\mathcal{G}_h$ , and hence the monotonicity of the scheme (cf. (S1)) implies that

$$S(h, t, x, w(t, x), [w]_{t,x}) \geq S(h, t, x, v_{\varepsilon_j}(t, x), [v_{\varepsilon_j}]_{t,x}).$$

But then, by (S3)(ii),

$$\begin{aligned} & S(h, t, x, w(t, x), [w]_{t,x}) \\ & \geq \partial_t v_{\varepsilon_j}(t, x) + \mathcal{L}^{\alpha_j}(t, x, v_{\varepsilon_j}(t, x), Dv_{\varepsilon_j}(t, x), D^2v_{\varepsilon_j}(t, x)) - E_2(\bar{K}, h, \varepsilon), \end{aligned}$$

and the proof is complete after an application of Lemma 3.4. □

It is now straightforward to derive the lower bound; we simply choose  $k = 8 \sup_i [v_i^\varepsilon]_1 \varepsilon$  and use Lemma 3.2 to compare  $u_h$  and  $w$ . This yields

$$u_h - w \leq e^{\mu t} |(u_{h,0} - w(0, \cdot))^+|_0 + 2te^{\mu t} E_2(\bar{K}, h, \varepsilon) \quad \text{in } \mathcal{G}_h.$$

But by inequalities (3.6), for small enough  $\varepsilon$  (and hence small enough  $k$ ) we have

$$|w - u|_0 \leq C(\varepsilon + k + k^{1/3}),$$

and therefore

$$u_h - u \leq e^{\mu t} |(u_{h,0} - u_0)^+|_0 + 2te^{\mu t} E_2(\bar{K}, h, \varepsilon) + C(\varepsilon + k + k^{1/3}) \quad \text{in } \mathcal{G}_h,$$

for some constant  $C$ . In view of our choice of  $k$ , we get the lower bound by minimizing w.r.t  $\varepsilon$ . Note that here  $\varepsilon$  will be an increasing function of  $h$  and therefore the requirement that  $\varepsilon$  is small implies that  $h$  must be small as well. This concludes the proof in the case when  $\mathcal{A}$  is finite.

**2. The case of general  $\mathcal{A}$ .** By (A2)  $\mathcal{A}$  is a separable metric space and hence has a countable dense subset  $\mathcal{A}_\infty$ . Furthermore, by (A2) the coefficients are continuous in  $\alpha$  and therefore

$$\sup_{\mathcal{A}} \mathcal{L}^\alpha(t, x, r, p, X) = \sup_{\mathcal{A}_\infty} \mathcal{L}^\alpha(t, x, r, p, X).$$

In other words, under assumption (A2) we may replace  $\mathcal{A}$  by  $\mathcal{A}_\infty$  in equation (1.1).

Now take a sequence of sets  $\{\mathcal{A}_M\}_{M=1}^\infty \subset \mathcal{A}_\infty$  such that

$$\mathcal{A}_M \subset \mathcal{A}_{M+1} \text{ for } M \in \mathbb{N} \quad \text{and} \quad \bigcup_{M=1}^\infty \mathcal{A}_M = \mathcal{A}_\infty.$$

Let  $u$  denote the solution of (1.1) and (1.2) and let  $u^M$  be the solution of (1.1) and (1.2) when  $\mathcal{A}$  is replaced by  $\mathcal{A}_M$ . By Proposition 2.1 we have the following  $\mathcal{C}^{0,1}$  bounds:

$$|u|_1 + |u^M|_1 \leq \bar{K},$$

where  $\bar{K}$  only depends on  $T$  and  $K$  from (A1) and not on  $M$ . Arzela–Ascoli’s theorem then implies that a subsequence of  $\{u^M\}_M$ , also denoted  $\{u^M\}_M$ , converges locally uniformly. It is an easy exercise to prove that the limit function solves (1.1) and (1.2), and hence by uniqueness is equal to  $u$ .

By part 1 of this proof we have that

$$u_h^M - u^M \leq \text{lower bound},$$

where  $u_h^M$  solves a scheme (1.3) related to equation (1.1) when  $\mathcal{A}$  is replaced by  $\mathcal{A}_M$ . In fact we do not need (or want!) to introduce  $u_h^M$  at all. To see this note that the “almost smooth” supersolution  $w$  of part 1 is a supersolution of (1.1) also for general  $\mathcal{A}$ . Now by re-examining the proof of Lemma 3.5 we see that we can replace  $u_h^M$  by  $u_h$  in that lemma, and repeating the rest of part 1 then leads to

$$u_h - u^M \leq \text{lower bound}.$$

The lower bound on  $u - u_h$  can then be obtained by sending  $M \rightarrow \infty$  after noting that the “lower bound” mentioned above (see part 1) does not depend on  $M$ .

**3. The proof of Lemma 3.4.** Fix an arbitrary point  $(t, x) \in Q_T$  and set  $j := \operatorname{argmin}_{i \in \mathcal{I}} v_{\varepsilon i}(t, x)$ . By definition of  $\mathcal{M}$  and  $j$  we have

$$v_{\varepsilon j}(t, x) - \mathcal{M}_j v_\varepsilon(t, x) = \max_{i \neq j} \{v_{\varepsilon j}(t, x) - v_{\varepsilon i}(t, x) - k\} \leq -k,$$

and by Hölder continuity of  $\bar{v}^\varepsilon$  (Lemma 3.3) and properties of mollifiers we see that

$$\bar{v}_j^\varepsilon(t, x) - \mathcal{M}_j \bar{v}^\varepsilon(t, x) \leq -k + 2 \max_i [v_i^\varepsilon]_1 2\varepsilon.$$

Using again Hölder continuity of  $\bar{v}^\varepsilon$ , for any  $(\bar{t}, \bar{x}) \in Q_T^\varepsilon$ , we get

$$\bar{v}_j^\varepsilon(\bar{t}, \bar{x}) - \mathcal{M}_j \bar{v}^\varepsilon(\bar{t}, \bar{x}) \leq -k + 2 \max_i [v_i^\varepsilon]_1 (2\varepsilon + |x - \bar{x}| + |t - \bar{t}|^{1/2}).$$

Now we conclude that if  $|x - \bar{x}| < \varepsilon$ ,  $|t - \bar{t}| < \varepsilon^2$ , and  $\varepsilon \leq (8 \max_i [v_i^\varepsilon]_1)^{-1} k$ , then

$$\bar{v}_j^\varepsilon(\bar{t}, \bar{x}) - \mathcal{M}_j \bar{v}^\varepsilon(\bar{t}, \bar{x}) < 0,$$

and by equation (3.3) and the definition of  $\bar{v}^\varepsilon$ ,  $\bar{v}^\varepsilon(t, x) = v^\varepsilon(t - \varepsilon^2, x)$ ,

$$\partial_t \bar{v}_j^\varepsilon(\bar{t}, \bar{x}) + \inf_{0 \leq s \leq \varepsilon^2, |e| \leq \varepsilon} \mathcal{L}^{\alpha_j}(\bar{t} + s, \bar{x} + e, \bar{v}_j^\varepsilon(\bar{t}, \bar{x}), D\bar{v}_j^\varepsilon(\bar{t}, \bar{x}), D^2\bar{v}_j^\varepsilon(\bar{t}, \bar{x})) = 0.$$

After a change of variables, we see that for every  $0 \leq s < \varepsilon^2$ ,  $|e| < \varepsilon$ ,

$$(3.7) \quad \begin{aligned} & \partial_t \bar{v}_j^\varepsilon(t - s, x - e)(t, x) \\ & + \mathcal{L}^{\alpha_j}(t, x, \bar{v}_j^\varepsilon(t - s, x - e), D\bar{v}_j^\varepsilon(t - s, x - e), D^2\bar{v}_j^\varepsilon(t - s, x - e)) \geq 0. \end{aligned}$$

In other words, for every  $0 \leq s < \varepsilon^2$  and  $|e| < \varepsilon$ ,  $\bar{v}_j^\varepsilon(t - s, x - e)$  is a (viscosity) supersolution at  $(t, x)$  of

$$(3.8) \quad \chi_t + \mathcal{L}^{\alpha_j}(t, x, \chi, D\chi, D^2\chi) = 0.$$

By mollifying (3.7) w.r.t.  $(s, e)$ , we see that  $v_{\varepsilon_j}$  is also a (viscosity) supersolution of (3.8) at  $(t, x)$  and hence a supersolution of the HJB equation (1.1) at  $(t, x)$ . This is correct since  $v_{\varepsilon_j}$  can be viewed as the limit of convex combinations of supersolutions  $\bar{v}_j^\varepsilon(t - s, x - e)$  of the linear and hence concave equation (3.8); we refer to the proof of Theorem 2.3 and to the Appendix in [1] for the details. We conclude the proof of Lemma 3.4 by noting that since  $v_{\varepsilon_j}$  is smooth, it is in fact a classical supersolution of (1.1) at  $(t, x)$ .

This completes the proof of Theorem 3.1 (b).

4. MONOTONE FINITE DIFFERENCE METHODS

In this section, we apply our main result to finite difference approximations of (1.1) based on the  $\vartheta$ -method of approximation in time and two different approximations in space: one proposed by Kushner [12] which is monotone when  $a$  is diagonally dominant and a (more) general approach based on directional second derivatives proposed by Bonnans and Zidani [5], but see also Dong and Krylov [10]. For simplicity we take  $h = (\Delta t, \Delta x)$  and consider the uniform grid

$$\mathcal{G}_h = \Delta t\{0, 1, \dots, n_T\} \times \Delta x\mathbb{Z}^N.$$

4.1. **Discretization in space.** To explain the methods we first write equation (1.1) as

$$u_t + \sup_{\alpha \in \mathcal{A}} \left\{ -L^\alpha u - c^\alpha(t, x)u - f^\alpha(t, x) \right\} = 0 \quad \text{in } Q_T,$$

where

$$L^\alpha \phi(t, x) = \text{tr}[a^\alpha(t, x)D^2\phi(t, x)] + b^\alpha(t, x)D\phi(t, x).$$

To obtain a discretization in space we approximate  $L$  by a finite difference operator  $L_h$ , which we will take to be of the form

$$(4.1) \quad L_h^\alpha \phi(t, x) = \sum_{\beta \in \mathcal{S}} C_h^\alpha(t, x, \beta)(\phi(t, x + \beta\Delta x) - \phi(t, x)),$$

for  $(t, x) \in \mathcal{G}_h$ , where the *stencil*  $\mathcal{S}$  is a finite subset of  $\mathbb{Z}^N \setminus \{0\}$ , and where

$$(4.2) \quad C_h^\alpha(t, x, \beta) \geq 0 \quad \text{for all } \beta \in \mathcal{S}, (t, x) \in \mathcal{G}_h^+, h = (\Delta x, \Delta t) > 0, \alpha \in \mathcal{A}.$$

The last assumption says that the difference approximation is of *positive* type. This is a sufficient assumption for monotonicity in the stationary case.

(i) The approximation of Kushner. We denote by  $\{e_i\}_{i=1}^N$  the standard basis in  $\mathbb{R}^N$  and define

$$(4.3) \quad L_h^\alpha \phi = \sum_{i=1}^N \left[ \frac{a_{ii}^\alpha}{2} \Delta_{ii} + \sum_{j \neq i} \left( \frac{a_{ij}^{\alpha+}}{2} \Delta_{ij}^+ - \frac{a_{ij}^{\alpha-}}{2} \Delta_{ij}^- \right) + b_i^{\alpha+} \delta_i^+ - b_i^{\alpha-} \delta_i^- \right] \phi,$$

where  $b^+ = \max\{b, 0\}$ ,  $b^- = (-b)^+$  ( $b = b^+ - b^-$ ), and

$$\begin{aligned} \delta_i^\pm w(x) &= \pm \frac{1}{\Delta x} \{w(x \pm e_i \Delta x) - w(x)\}, \\ \Delta_{ii} w(x) &= \frac{1}{\Delta x^2} \{w(x + e_i \Delta x) - 2w(x) + w(x - e_i \Delta x)\}, \\ \Delta_{ij}^+ w(x) &= \frac{1}{2\Delta x^2} \{2w(x) + w(x + e_i \Delta x + e_j \Delta x) + w(x - e_i \Delta x - e_j \Delta x)\} \\ &\quad - \frac{1}{2\Delta x^2} \{w(x + e_i \Delta x) + w(x - e_i \Delta x) + w(x + e_j \Delta x) + w(x - e_j \Delta x)\}, \\ \Delta_{ij}^- w(x) &= -\frac{1}{2\Delta x^2} \{2w(x) + w(x + e_i \Delta x - e_j \Delta x) + w(x - e_i \Delta x + e_j \Delta x)\} \\ &\quad + \frac{1}{2\Delta x^2} \{w(x + e_i \Delta x) + w(x - e_i \Delta x) + w(x + e_j \Delta x) + w(x - e_j \Delta x)\}. \end{aligned}$$

The stencil is  $\mathcal{S} = \{\pm e_i, \pm(e_i \pm e_j) : i, j = 1, \dots, N\}$ , and it is easy to see that the coefficients in (4.1) are

$$\begin{aligned} C_h^\alpha(t, x, \pm e_i) &= \frac{a_{ii}^\alpha(x)}{2\Delta x^2} - \sum_{j \neq i} \frac{|a_{ij}^\alpha(x)|}{4\Delta x^2} + \frac{b_i^{\alpha \pm}(x)}{\Delta x}, \\ C_h^\alpha(t, x, e_i h \pm e_j h) &= \frac{a_{ij}^{\alpha \pm}(x)}{2\Delta x^2}, \quad i \neq j, \\ C_h^\alpha(t, x, -e_i h \pm e_j h) &= \frac{a_{ij}^{\alpha \mp}(x)}{2\Delta x^2}, \quad i \neq j. \end{aligned}$$

The approximation is of positive type (4.2) if and only if  $a$  is diagonally dominant, i.e.,

$$(4.4) \quad a_{ii}^\alpha(t, x) - \sum_{j \neq i} |a_{ij}^\alpha(t, x)| \geq 0 \quad \text{in } Q_T, \quad \alpha \in \mathcal{A}, \quad i = 1, \dots, N.$$

(ii) The approximation of Bonnans and Zidani.

We assume that there is a (finite) stencil  $\bar{\mathcal{S}} \subset \mathbb{Z}^N \setminus \{0\}$  and a set of positive coefficients  $\{\bar{a}_\beta : \beta \in \bar{\mathcal{S}}\} \subset \mathbb{R}_+$  such that

$$(4.5) \quad a^\alpha(t, x) = \sum_{\beta \in \bar{\mathcal{S}}} \bar{a}_\beta^\alpha(t, x) \beta^T \beta \quad \text{in } Q_T, \quad \alpha \in \mathcal{A}.$$

Under assumption (4.5) we may rewrite the operator  $L$  using second-order directional derivatives  $D_\beta^2 = \text{tr}[\beta \beta^T D^2] = (\beta \cdot D)^2$ ,

$$L^\alpha \phi(t, x) = \sum_{\beta \in \bar{\mathcal{S}}} \bar{a}_\beta^\alpha(t, x) D_\beta^2 \phi(t, x) + b^\alpha(t, x) D \phi(t, x).$$

The approximation of Bonnans and Zidani is given by

$$(4.6) \quad L_h^\alpha \phi = \sum_{\beta \in \bar{\mathcal{S}}} \bar{a}_\beta^\alpha \Delta_\beta \phi + \sum_{i=1}^N \left[ b_i^{\alpha+} \delta_i^+ - b_i^{\alpha-} \delta_i^- \right] \phi,$$

where  $\Delta_\beta$  is an approximation of  $D_\beta^2$  given by

$$\Delta_\beta w(x) = \frac{1}{|\beta|^2 \Delta x^2} \{w(x + \beta \Delta x) - 2w(x) + w(x - \beta \Delta x)\}.$$



In this case, the stencil is  $\mathcal{S} = \pm\bar{\mathcal{S}} \cup \{\pm e_i : i = 1, \dots, N\}$  and the coefficients corresponding to (4.1) are given by

$$C_h^\alpha(t, x, \pm e_i) = \frac{b_i^{\alpha\pm}(x)}{\Delta x}, \quad i = 1, \dots, N,$$

$$C_h^\alpha(t, x, \pm\beta) = \frac{\bar{a}_\beta^\alpha(t, x)}{|\beta|^2 \Delta x^2}, \quad \beta \in \bar{\mathcal{S}},$$

and the sum of the two whenever  $\beta = e_i$ . Under assumption (4.5), which is more general than (4.4) (see below), this approximation is always of positive type.

For both approximations there is a constant  $C > 0$ , independent of  $\Delta x$ , such that, for every  $\phi \in C^4(\mathbb{R}^N)$  and  $(t, x) \in \mathcal{G}_h^+$ ,

$$(4.7) \quad |L^\alpha \phi(t, x) - L_h^\alpha \phi(t, x)| \leq C(|b^\alpha|_0 |D^2 \phi|_0 \Delta x + |a^\alpha|_0 |D^4 \phi|_0 \Delta x^2).$$

**4.2. The fully discrete scheme.** To obtain a fully discrete scheme, we apply the  $\vartheta$ -method,  $\vartheta \in [0, 1]$ , to discretize the time derivative. The result is the following scheme:

$$(4.8) \quad \begin{aligned} u(t, x) &= u(t - \Delta t, x) \\ &\quad - (1 - \vartheta) \Delta t \sup_{\alpha \in \mathcal{A}} \{-L_h^\alpha u - c^\alpha u - f^\alpha\}(t - \Delta t, x) \\ &\quad - \vartheta \Delta t \sup_{\alpha \in \mathcal{A}} \{-L_h^\alpha u - c^\alpha u - f^\alpha\}(t, x) \quad \text{in } \mathcal{G}_h^+. \end{aligned}$$

The cases  $\vartheta = 0$  and  $\vartheta = 1$  correspond to the forward and backward Euler time-discretizations, respectively, while for  $\vartheta = 1/2$  the scheme is a generalization of the second order in time Crank–Nicholson scheme. Note that the scheme is implicit except for the value  $\vartheta = 0$ . We may write (4.8) in the form (1.3) by setting

$$\begin{aligned} S(h, t, x, r, [u]_{t,x}) &= \sup_{\alpha \in \mathcal{A}} \left\{ \left[ \frac{1}{\Delta t} - \vartheta c^\alpha + \vartheta \sum_{\beta \in \mathcal{S}} C_h^\alpha(t, x, \beta) \right] r \right. \\ &\quad - \left[ \frac{1}{\Delta t} + (1 - \vartheta) c^\alpha - (1 - \vartheta) \sum_{\beta \in \mathcal{S}} C_h^\alpha(t, x, \beta) \right] [u]_{t,x}(-\Delta t, 0) \\ &\quad \left. - \sum_{\beta \in \mathcal{S}} C_h^\alpha(t, x, \beta) \left[ \vartheta [u]_{t,x}(0, \beta \Delta x) + (1 - \vartheta) [u]_{t,x}(-\Delta t, \beta \Delta x) \right] \right\}, \end{aligned}$$

where  $[u]_{t,x}(s, y) = u(t + s, x + y)$ . Under assumption (4.2) the scheme (4.8) is monotone (i.e., satisfies (S1) or (S1')) provided the following CFL conditions hold:

$$(4.9) \quad \Delta t (1 - \vartheta) \left( -c^\alpha(t, x) + \sum_{\beta \in \mathcal{S}} C_h^\alpha(t, x, \beta) \right) \leq 1,$$

$$(4.10) \quad \Delta t \vartheta \left( c^\alpha(t, x) - \sum_{\beta \in \mathcal{S}} C_h^\alpha(t, x, \beta) \right) \leq 1.$$

Furthermore, in view of (A1) and (4.7), Taylor expansion in (4.8) yields the following consistency result for smooth functions  $\phi$  and  $(t, x) \in \mathcal{G}_h^+$ :

$$\begin{aligned} &|\phi_t + F(t, x, \phi, D\phi, D^2\phi) - S(h, t, x, \phi, [\phi]_{t,x})| \\ &\leq C(\Delta t |\phi_{tt}|_0 + \Delta x |D^2 \phi|_0 + \Delta x^2 |D^4 \phi|_0 + (1 - \vartheta) \Delta t (|D\phi_t|_0 + |D^2 \phi_t|_0)). \end{aligned}$$

The  $(1 - \vartheta)\Delta t$ -term is a nonstandard term coming from the fact that we need the equation and the scheme to be satisfied in the *same* point; see assumption (S3). The necessity of this assumption follows from the proof of Theorem 3.1.

We have seen that if (4.2) and (4.7) hold along with the CFL conditions (4.9) and (4.10), then the scheme (4.8) satisfies assumptions (S1)–(S3) in Section 3. Theorem 3.1 therefore yields the following error bound:

**Theorem 4.1.** *Assume (A1), (A2), (4.2), (4.7), (4.9), (4.10) hold. If  $u_h \in C_b(\mathcal{G}_h)$  is the solution of (4.8) and  $u$  is the solution of (1.1), then there is  $C > 0$  such that in  $\mathcal{G}_h$ ,*

$$-e^{\mu t} |(u_0 - u_{0,h})^-|_0 - C|h|^{\frac{1}{5}} \leq u - u_h \leq e^{\mu t} |(u_0 - u_{0,h})^+|_0 + C|h|^{\frac{1}{2}},$$

where  $|h| := \sqrt{\Delta x^2 + \Delta t}$ .

*Remark 4.1.* Except when  $\vartheta = 1$ , the CFL condition (4.9) essentially implies that  $\Delta t \leq C\Delta x^2$ . Therefore  $\Delta t$  and  $\Delta x^2$  play essentially the same role. Also note that the CFL condition (4.10) is satisfied if, e.g.,  $\Delta t \leq (\sup_{\alpha} |(c^{\alpha})^+|_0)^{-1}$ .

*Remark 4.2.* Even though the above consistency relationship is not quite the “standard” one, it gives the correct asymptotic behavior of our scheme. First of all note that the new term, the  $(1 - \vartheta)$ -term, behaves just like the  $\Delta t$  and  $\Delta x^2$  terms. To see this, we note that according to (S3) we only need the above relation when  $\phi$  is replaced by  $\phi_{\varepsilon}$  defined in (S3). But for  $\phi_{\varepsilon}$  we have  $|\phi_{\varepsilon,tt}|_0 \approx |D^4\phi_{\varepsilon}|_0 \approx |D^2\phi_{\varepsilon,t}|_0 \approx \tilde{K}\varepsilon^{-3}$ . By the CFL conditions (4.9) and (4.10) we have essentially that  $\Delta x^2 \approx \Delta t$ , so

$$\Delta t |\phi_{\varepsilon,tt}|_0 \approx \Delta x^2 |D^4\phi_{\varepsilon}|_0 \approx \Delta t |D^2\phi_{\varepsilon,t}|_0 \approx \tilde{K}\Delta x^2\varepsilon^{-3}.$$

Next note that for  $\vartheta = 1/2$  (the Crank–Nicholson case) the scheme is formally second order in time. However this is no longer the case for the monotone version. It is only first order in time due to the CFL condition, which implies that  $\Delta x^2 \|D^4\phi\| = C\Delta t \|D^4\phi\|$ .

*Proof.* In this case,

$$\begin{aligned} E_1(\bar{K}, h, \varepsilon) &= E_2(\bar{K}, h, \varepsilon) \\ &= C(\Delta t\varepsilon^{-3} + \Delta x\varepsilon^{-1} + \Delta x^2\varepsilon^{-3} + (1 - \vartheta)\Delta t(\varepsilon^{-2} + \varepsilon^{-3})). \end{aligned}$$

So we have to minimize w.r.t.  $\varepsilon$  the following functions:

$$\begin{aligned} &\varepsilon + C(\Delta t\varepsilon^{-3} + \Delta x\varepsilon^{-1} + \Delta x^2\varepsilon^{-3}), \\ &\varepsilon^{1/3} + C(\Delta t\varepsilon^{-3} + \Delta x\varepsilon^{-1} + \Delta x^2\varepsilon^{-3}). \end{aligned}$$

By minimizing separately in  $\Delta t$  and  $\Delta x$ , one finds that  $\varepsilon$  has to be like  $\Delta t^{1/4}$  and  $\Delta x^{1/2}$  in the first case, and that  $\varepsilon^{1/3}$  has to be like  $\Delta t^{1/10}$  and  $\Delta x^{1/5}$  in the second case. The result now follows by taking  $\varepsilon = \max(\Delta t^{1/4}, \Delta x^{1/2})$  in the first case and  $\varepsilon^{1/3} = \max(\Delta t^{1/10}, \Delta x^{1/5})$  in the second case.  $\square$

**4.3. Remarks.** For approximations of nonlinear equations monotonicity is a key property since it ensures (along with consistency) that the approximate solutions converge to the *correct* generalized solution of the problem (the viscosity solution in our case). This is not the case for nonmonotone methods, at least not in any generality.

However, the monotonicity requirement poses certain problems. Monotone schemes are low-order schemes, and maybe more importantly, it is not always possible to find consistent monotone approximations for a given problem. To see the last point we note that in general the second derivative coefficient matrix  $a$  is only positive semidefinite, while the monotone schemes of Kushner and Bonnans/Zidani require the stronger assumptions (4.4) and (4.5), respectively. In fact, in Dong and Krylov [10] it was proved that if an operator  $L$  admits an approximation  $L_h$  of the form (4.1) which is of positive type, then  $a$  has to satisfy (4.5) (at least if  $a$  is bounded).

This is a problem in real applications, e.g. in finance, and this problem was the motivation behind the approximation of Bonnans and Zidani. First of all we note that their condition (4.5) is more general than (4.4) because any symmetric  $N \times N$  matrix  $a$  can be decomposed as

$$a = \sum_{i=1}^N \sum_{j \neq i} (a_{ii} - |a_{ij}|) e_i e_i^T + \frac{a_{ij}^+}{2} (e_i + e_j)(e_i + e_j)^T + \frac{a_{ij}^-}{2} (e_i - e_j)(e_i - e_j)^T,$$

where the coefficients are nonnegative if and only if  $a$  is diagonally dominant. More importantly, it turns out that any symmetric positive semidefinite matrix can be approximated by a sequence of matrices satisfying (4.5). In Bonnans, Ottenwaelter, and Zidani [4], this was proved in the case of symmetric  $2 \times 2$  matrices along with an explicit error bound and an algorithm for computing the approximate matrices. Because of continuous dependence results for the equations, convergence of the coefficients immediately implies convergence of the solutions of the corresponding equations. Hence the Bonnans/Zidani approximation yields a way of approximating general problems where  $a$  is only positive semidefinite.

### 5. SEMIGROUP APPROXIMATIONS AND SPLITTING METHODS

In this section, we consider various approximations of semigroups obtained by a semidiscretization in time. In order to simplify the presentation we start by specializing Theorem 3.1 to the semigroup setting. To be precise we consider one-step-in-time approximations of (1.1) given by

$$(5.1) \quad \begin{aligned} u_h(t_n, x) &= S_h(t_n, t_{n-1})u_h(t_{n-1}, x) && \text{in } \mathbb{R}^N, \\ u_h(0, x) &= u_{h,0}(x) && \text{in } \mathbb{R}^N, \end{aligned}$$

where  $t_0 = 0 < t_1 < \dots < t_n < \dots < t_{n_T} = T$ ,  $h := \max_n(t_{n+1} - t_n)$ , and the approximation semigroup  $S_h$  satisfies the following sub- and superconsistency requirements: There exist a constant  $K_c$ , a subset  $I$  of  $\mathbb{N} \times \mathbb{N}^N$ , and constants  $\gamma_\beta, \delta_\beta$  for  $\beta \in I$  such that for any smooth functions  $\phi$ ,

$$(5.2) \quad \begin{aligned} &\frac{1}{\Delta t} \left[ S_h(t_n, t_{n-1}) - 1 \right] \phi(t_{n-1}, x) - F(t, x, \phi, D\phi, D^2\phi)_{t=t_n} \\ &\leq K_c \sum_{\beta \in I} |\partial_t^{\beta_0} D^{\beta'} \phi|_0^{\gamma_\beta} \Delta t^{\delta_\beta}, \end{aligned}$$

where  $\beta = (\beta_0, \beta') \in I$  for  $\beta_0 \in \mathbb{N}$  and  $\beta' \in \mathbb{N}^N$ , and in a similar way,

$$(5.3) \quad \begin{aligned} & \frac{1}{\Delta t} \left[ S_h(t_n, t_{n-1}) - 1 \right] \phi(t_{n-1}, x) - F(t, x, \phi, D\phi, D^2\phi)_{t=t_n} \\ & \geq -\bar{K}_c \sum_{\beta \in \bar{I}} |\partial_t^{\beta_0} D^{\beta'} \phi|_0^{\bar{\gamma}_\beta} \Delta t^{\bar{\delta}_\beta}, \end{aligned}$$

with corresponding data  $\bar{K}_c, \bar{I}, \bar{\gamma}_\beta, \bar{\delta}_\beta$ . We say that the semigroup is *monotone* if

$$\phi \leq \psi \quad \Rightarrow \quad S_h(t_n, t_{n-1})\phi \leq S_h(t_n, t_{n-1})\psi, \quad n = 1, \dots, n_T,$$

for all continuous bounded functions  $\phi, \psi$  for which  $S_h(t)\phi$  and  $S_h(t)\psi$  are well defined.

We have the following corollary to Theorem 3.1.

**Proposition 5.1.** *Assume (A1), (A2), and that  $S_h$  is a monotone semigroup satisfying (5.2) and (5.3) and which is defined on a subset of  $C_b(\mathbb{R}^N)$ . If  $u$  is the solution of (1.1) and  $u_h$  is the solution of (5.1), then*

$$-C(|u_0 - u_{h,0}|_0 + \Delta t^{\frac{1}{10} \wedge r_1}) \leq u - u_h \leq C(|u_0 - u_{h,0}|_0 + \Delta t^{\frac{1}{4} \wedge r_2})$$

in  $\mathbb{R}^N$ , where

$$\begin{aligned} r_1 &:= \min_{\beta \in I} \left\{ \frac{\delta_\beta}{3(2\beta_0 + |\beta'| - 1)\gamma_\beta + 1} \right\}, \\ r_2 &:= \min_{\beta \in \bar{I}} \left\{ \frac{\bar{\delta}_\beta}{(2\beta_0 + |\beta'| - 1)\bar{\gamma}_\beta + 1} \right\}, \end{aligned}$$

where  $|\beta'|$  denotes the sum of the components of  $\beta'$ .

*Proof.* We define

$$S(h, t_n, x, u_h, [u_h]_{t_n, x}) = \frac{1}{\Delta t} \left( u_h(t_n, x) - [u_h]_{t_n, x} \right),$$

where

$$[u_h]_{t_n, x} = S_h(t_{n-1}, t_n)u_h(t_{n-1}, x).$$

To apply Theorem 3.1, we just have to check that (S1)–(S3) hold, and this is clear for (S1) and (S2) (see Remark 3.1). For (S3)(i), note that by (5.2) we have

$$\begin{aligned} & \phi_t + F(t_n, x, \phi, D\phi, D^2\phi) - S(h, t_n, x, \phi, [\phi]_{t_n, x}) \\ & \leq \frac{1}{2} |\partial_t^2 \phi|_0 \Delta t + K_c \sum_{\beta \in I} |\partial_t^{\beta_0} D^{\beta'} \phi|_0^{\gamma_\beta} \Delta t^{\delta_\beta}, \end{aligned}$$

which leads to

$$E_1(\bar{K}, h, \varepsilon) = \frac{1}{2} \bar{K} \varepsilon^{1-4} \Delta t + K_c \sum_{\beta \in I} (\bar{K} \varepsilon^{1-2\beta_0-|\beta'|})^{\gamma_\beta} \Delta t^{\delta_\beta}.$$

The upper bound now follows by optimizing with respect to  $\varepsilon$  as in the proof of Theorem 4.1. In a similar way we may use (5.3) to define  $E_2$  and then prove the lower bound. □

*Remark 5.1.* In view of the consistency requirements (5.2) and (5.3), for schemes such as (5.1) it is natural to think that only the  $x$ -variable is really playing a role and that one can get results on the rate of convergence by using this special “semigroup type” structure. More specifically, one might think that a different proof, using a mollification of the solution with respect to the space variable only, can

produce the estimates in an easier and maybe better way. We tried this strategy, but we could not avoid using the short time expansion of the solution of the HJB equation associated with smooth initial data (the short time expansion of the semigroup), and this leads to worse rates, even in cases where  $F$  is smooth. One way of understanding this (without justifying it completely) consists of looking at our estimates for the  $\phi_{tt}$ -term (cf. (S3)(i) and (ii)). The present approach leads to an estimate of order  $\varepsilon^{-3}$ , while if we use the short time expansion, we are led to a worse estimate of order  $\varepsilon^{-4}$ . We refer the reader to Subsection 5.1 and in particular to Lemma 5.6 below, where short time expansions for semigroups are obtained and used to study the rate of convergence for splitting problems.

**5.1. Semidiscrete splitting.** We consider an equation of the form

$$(5.4) \quad u_t + F_1(D^2u) + F_2(D^2u) = 0 \quad \text{in } Q_T,$$

where

$$F_j(X) = \sup_{\alpha \in \mathcal{A}} \{-\text{tr}[a_j^\alpha X] - f_j^\alpha\}, \quad j = 1, 2,$$

and  $a_j^\alpha \geq 0$  are matrices and  $f_j^\alpha$  real numbers. We assume that they are both uniformly bounded in  $\alpha$  and are independent of  $(t, x)$ . It follows that  $F_1$  and  $F_2$  are Lipschitz continuous and that (A1) is satisfied.

Let  $S$  denote the semigroup of (5.4); i.e.,  $S(\Delta t)\phi$  is the solution at time  $t = \Delta t$  of (5.4) with initial value  $\phi$ . Similarly, let  $S_1$  and  $S_2$  denote the semigroups associated with the equations  $u_t + F_1(D^2u) = 0$  and  $u_t + F_2(D^2u) = 0$ .

We can define a semidiscrete splitting method by taking (5.1) with  $t_n := n\Delta t$  and

$$(5.5) \quad S_h(t_{n-1}, t_n) = S_1(\Delta t)S_2(\Delta t).$$

Under the current assumptions all these semigroups map  $W^{1,\infty}(\mathbb{R}^N)$  into itself, they are monotone, and they satisfy the following comparison principle:

$$\bar{S}(t)\phi_1 - \bar{S}(t)\phi_2 \leq |(\phi_1 - \phi_2)^+|_0,$$

for  $\phi_1, \phi_2 \in W^{1,\infty}(\mathbb{R}^N)$  and where  $\bar{S}$  denotes one of the semigroups above.

As soon as we know the consistency relation for this scheme, we can find an error bound using Theorem 5.1. However, contrary to the case of finite difference schemes in the previous section, here the precise form of the consistency requirement is not well known. We are going to provide such results under different assumptions on  $F_1, F_2$ . Our first result is the following:

**Lemma 5.2.** *Under the above assumptions, if in addition  $|DF_1| \in W^{1,\infty}(\mathbb{S}^N)$  and  $|DF_2| \in W^{3,\infty}(\mathbb{S}^N)$ , then*

$$\begin{aligned} & -C(\Delta t|D^2\phi_t|_0 + \Delta t^2|D^3\phi|_0^4) - h.o.t. \\ & \leq \frac{1}{\Delta t}[S_h(t) - 1]\phi(t_{n-1}, x) + F_1(D^2\phi(t_n, x)) + F_2(D^2\phi(t_n, x)) \\ & \leq C(\Delta t|D^2\phi_t|_0 + \Delta t|D^3\phi|_0^2) + h.o.t. \end{aligned}$$

for all smooth functions  $\phi$ , where “h.o.t.” stands for “higher order terms”.

*Remark 5.2.* The upper and lower bounds are different due to convexity of the equation.

*Remark 5.3.* To simplify the presentation we have only stated the principal error terms in Lemma 5.2, i.e., the terms determining the rate in Corollary 5.3 below. To see which terms are principal, one must look also at the proof of Corollary 5.3. The “h.o.t.” category contains the terms that are not principal, both high and low order terms, and maybe a better name would be the “less important terms”.

A direct consequence of Proposition 5.1 is the following result:

**Corollary 5.3.** *Let  $u_h$  denote the solution of (5.1) where  $S_h$  is defined in (5.5) and  $u_{h,0} = u_0$ , and let  $u$  be the solution of (5.4) with initial value  $u_0$ . Under the assumptions of Lemma 5.2 we have*

$$-C\Delta t^{\frac{1}{13}} \leq u - u_h \leq C\Delta t^{\frac{2}{9}} \quad \text{in } \Delta t\{0, 1, 2, \dots, n_T\} \times \mathbb{R}^N.$$

Next, we give the result when  $F_1$  and  $F_2$  are assumed to be only Lipschitz continuous (which is the natural regularity assumption here). In this case the consistency relation is:

**Lemma 5.4.** *Under the above assumptions, if  $F_1$  and  $F_2$  are only Lipschitz continuous, we have*

$$\begin{aligned} & \left| \frac{1}{\Delta t} [S_h(t) - 1] \phi(t_{n-1}, x) + F_1(D^2 \phi(t_n, x)) + F_2(D^2 \phi(t_n, x)) \right| \\ & \leq C\Delta t |D^2 \phi_t|_0 + C\Delta t^{\frac{1}{2}} |D^3 \phi|_0 + h.o.t. \end{aligned}$$

for all smooth functions  $\phi$ .

Again as a direct consequence of Proposition 5.1 we have the following error bound:

**Corollary 5.5.** *Under the assumptions of Corollary 5.3 but where  $F_1$  and  $F_2$  are only assumed to be Lipschitz continuous, we have*

$$-C\Delta t^{\frac{1}{14}} \leq u - u_h \leq C\Delta t^{\frac{1}{6}} \quad \text{in } \Delta t\{0, 1, 2, \dots, n_T\} \times \mathbb{R}^N.$$

*Remark 5.4.* We see a slight reduction of the rates in the Lipschitz case but not as important as one might have guessed. For first-order equations these methods lead to the same rates in the smooth and Lipschitz cases.

*Remark 5.5.* If we change operators  $S_1, S_2$  so that  $S_1(t)\phi$  and  $S_2(t)\phi$  denote the viscosity solutions of

$$\begin{aligned} u(x) &= \phi(x) - tF_1(D^2 u(x)) \quad \text{in } \mathbb{R}^N, \\ u(x) &= \phi(x) - tF_2(D^2 u(x)) \quad \text{in } \mathbb{R}^N, \end{aligned}$$

respectively, then the statements of Corollaries 5.3 and 5.5 still hold.

In the proofs of Lemmas 5.2 and 5.4 we will use the following lemma:

**Lemma 5.6.** *Let  $\bar{S}$  be the semigroup associated to the equation*

$$u_t + \bar{F}(D^2 u) = 0,$$

where  $\bar{F}$  is Lipschitz, convex, and nonincreasing. Define  $\bar{F}_\delta$  by

$$\bar{F}_\delta = \bar{F} * \bar{\rho}_\delta,$$

where  $\bar{\rho}_\delta(X) = \delta^{-N^2} \bar{\rho}(X/\delta)$  and  $\bar{\rho}$  is a smooth function on  $\mathbb{S}(N)$  with mass one and support in  $B(0, 1)$ . Then for any smooth function  $\phi$ ,

$$\bar{S}(t)\phi - \phi + t\bar{F}_\delta(D^2\phi) \leq t\delta|D\bar{F}|_0 + \frac{1}{2}t^2|D\bar{F}|_0|D\bar{F}_\delta|_0|D^4\phi|_0$$

and

$$\bar{S}(t)\phi - \phi + t\bar{F}_\delta(D^2\phi) \geq -\frac{1}{2}t^2|D\bar{F}|_0(|D^2\bar{F}_\delta|_0|D^3\phi|_0^2 + |D\bar{F}_\delta|_0|D^4\phi|_0).$$

The proof of this result will be given after the proofs of Lemmas 5.2 and 5.4.

**Proofs of Lemmas 5.2 and 5.4.** In order to treat the two results at the same time, we mollify  $F_1$  and  $F_2$  and consider  $F_{1,\delta}$  and  $F_{2,\delta}$  (see Lemma 5.6 for the definitions). By Lemma 5.6 we have the following (small time) expansions:

$$(5.6) \quad S_j(t)\phi - \phi + tF_{j,\delta}(D^2\phi) \leq t\delta|DF_j|_0 + \frac{1}{2}t^2|DF_j|_0|DF_{j,\delta}|_0|D^4\phi|_0,$$

$$(5.7) \quad \begin{aligned} &S_j(t)\phi - \phi + tF_{j,\delta}(D^2\phi) \\ &\geq -\frac{1}{2}t^2|DF_j|_0(|D^2F_{j,\delta}|_0|D^3\phi|_0^2 + |DF_{j,\delta}|_0|D^4\phi|_0), \end{aligned}$$

for smooth functions  $\phi$  and  $j = 1, 2$ .

Now we want to find an (small time) expansion for  $S_h$ . We write

$$\begin{aligned} &S_h(t)\phi - \phi + t(F_1 + F_2)(D^2\phi) \\ &= [S_1(t)S_2(t)\phi - S_1(t)(\phi - tF_{2,\delta}(D^2\phi))] \\ &\quad + [S_1(t)(\phi - tF_{2,\delta}(D^2\phi)) - (\phi - tF_{2,\delta}(D^2\phi)) + tF_{1,\delta}(D^2(\phi - tF_{2,\delta}(D^2\phi)))] \\ &\quad + t[F_{1,\delta}(D^2\phi) - F_{1,\delta}(D^2(\phi - tF_{2,\delta}(D^2\phi)))] \\ &\quad + t[(F_1 + F_2)(D^2\phi) - (F_{1,\delta} + F_{2,\delta})(D^2\phi)]. \end{aligned}$$

In view of the Lipschitz regularity and convexity of  $F_1$  and  $F_2$ , the last term on the right-hand side is between  $-Ct\delta$  (Lipschitz regularity) and 0 (convexity), while the other terms can be estimated using the comparison principle for  $S_1$ , small time expansions for  $S_1$  and  $S_2$ , and Lipschitz regularity of  $F_1$  and  $F_2$ . Consider the second term on the right-hand side. By (5.6) and (5.7),

$$t\delta|DF_1|_0 + \frac{1}{2}t^2|DF_1|_0|DF_{1,\delta}|_0|D^4\{\phi - tF_{2,\delta}(D^2\phi)\}|_0$$

is an upper bound on this term, while

$$\frac{1}{2}t^2|DF_1|_0 \left[ |D^2F_{1,\delta}|_0|D^3\{\phi - tF_{2,\delta}(D^2\phi)\}|_0^2 + |DF_{1,\delta}|_0|D^4\{\phi - tF_{2,\delta}(D^2\phi)\}|_0 \right]$$

is a lower bound. Expanding these expressions, keeping only the “worst terms” and bearing in mind the Lipschitz regularity of  $F_1$  and  $F_2$ , lead to the following upper and lower bounds respectively:

$$\begin{aligned} &C(t\delta + t^2|D^4\phi|_0 + t^3|D^4F_{2,\delta}|_0|D^3\phi|_0^4) \quad \text{and} \\ &C|D^2F_{1,\delta}|_0 \left( t^2|D^3\phi|_0^2 + t^4|D^3F_{2,\delta}|_0^2|D^3\phi|_0^6 \right). \end{aligned}$$

One can check that the “worst terms” coming from the first and third term above are either of the same type or better than these.

To conclude the proofs of the upper bounds in Lemmas 5.2 and 5.4, we note that

$$\begin{aligned} & \frac{1}{\Delta t} [S_h(t) - 1] \phi(t_{n-1}, x) + F_1(D^2 \phi(t_n, x)) + F_2(D^2 \phi(t_n, x)) \\ & \leq \Delta t (|DF_1|_0 + |DF_2|_0) |D^2 \phi_t|_0 + \left[ \frac{1}{\Delta t} [S_h(t) - 1] \phi + F_1(D^2 \phi) + F_2(D^2 \phi) \right]_{(t_{n-1}, x)}. \end{aligned}$$

In view of the above estimates, the right-hand side can be upper bounded by terms that are no worse than

$$(5.8) \quad C \left[ \Delta t |D^2 \phi_t|_0 + \delta + \Delta t |D^4 \phi|_0 + \Delta t^2 |D^4 F_{2,\delta}|_0 |D^3 \phi|_0^4 \right].$$

This proves the upper bound in Lemma 5.2 after sending  $\delta \rightarrow 0$  while keeping in mind that in this case,

$$|D^n F_2^\delta|_0 \leq |D^n F_2|_0 < \infty \quad \text{and} \quad |D^m F_1^\delta|_0 \leq |D^m F_1|_0 < \infty$$

for  $n = 1, \dots, 4$  and  $m = 1, 2$ . The upper bound in Lemma 5.4 follows from minimizing (5.8) w.r.t.  $\delta$  since in this case  $|D^n F_j|_0 \leq C \delta^{1-\delta}$ ,  $n \in \mathbb{N}$ ,  $j = 1, 2$ .

The upper bounds follow in a similar way. We conclude the proof simply by giving the expression corresponding to (5.8):

$$C \left[ \Delta t |D^2 \phi_t|_0 + \delta + |D^2 F_{1,\delta}|_0 \left( \Delta t |D^3 \phi|^2 + \Delta t^3 |D^3 F_{2,\delta}|^2 |D^3 \phi|_0^6 \right) \right].$$

*Proof of Lemma 5.6.* Let

$$w = \phi - t \bar{F}_\delta(D^2 \phi),$$

and observe that

$$(5.9) \quad w_t + \bar{F}(D^2 w) = -\bar{F}_\delta(D^2 \phi) + \bar{F}(D^2 \phi) - \bar{F}(D^2 \phi) + \bar{F}(D^2 w).$$

Since  $\bar{F}$  is convex, it is easy to see that  $\bar{F}_\delta(X) \geq \bar{F}(X)$ , and hence

$$(5.10) \quad -|D\bar{F}|_0 \delta \leq -\bar{F}_\delta(D^2 \phi) + \bar{F}(D^2 \phi) \leq 0.$$

The second difference,  $-\bar{F}(D^2 \phi) + \bar{F}(D^2 w)$ , can be written as

$$\begin{aligned} & \int_0^1 \frac{d}{ds} \{ \bar{F}(sD^2 w + (1-s)D^2 \phi) \} ds \\ (5.11) \quad & = \sum_{i,j} \partial_i \partial_j (w - \phi) \int_0^1 (\partial_{X_{ij}} \bar{F})(sD^2 w + (1-s)D^2 \phi) ds \\ & = -t \sum_{i,j} \partial_i \partial_j \{ \bar{F}_\delta(D^2 \phi) \} \int_0^1 (\partial_{X_{ij}} \bar{F})(sD^2 w + (1-s)D^2 \phi) ds. \end{aligned}$$

We expand  $\partial_i \partial_j \{ \bar{F}_\delta(D^2 \phi) \}$  and get

$$\begin{aligned} & \sum_{k,l,m,n} (\partial_{X_{kl}} \partial_{X_{mn}} \bar{F}_\delta)(D^2 \phi) (\partial_i \partial_k \partial_l \phi) (\partial_j \partial_m \partial_n \phi) \\ & + \sum_{k,l} (\partial_{X_{kl}} \bar{F}_\delta)(D^2 \phi) (\partial_i \partial_j \partial_k \partial_l \phi). \end{aligned}$$

We call the first term  $M[\phi]_{ij}$ .

Since  $\partial_{X_{kl}} \partial_{X_{mn}} \bar{F}_\delta = \partial_{X_{mn}} \partial_{X_{kl}} \bar{F}_\delta$ , it follows that  $M$  is symmetric,

$$M[\phi]_{ij} = M[\phi]_{ji}.$$



Moreover, since  $\bar{F}$  is convex,  $M$  is positive semidefinite: For every  $\xi \in \mathbb{R}^N$ ,

$$\begin{aligned} & \sum_i M[\phi]_{ij} \xi_i \xi_j \\ &= \sum_{k,l,m,n} (\partial_{X_{kl}} \partial_{X_{mn}} \bar{F}_\delta)(D^2\phi) (\partial_k \partial_l (\sum_i \xi_i \partial_i \phi)) (\partial_m \partial_n (\sum_j \xi_j \partial_j \phi)) \\ &= \sum_{k,l,m,n} (\partial_{X_{kl}} \partial_{X_{mn}} \bar{F}_\delta)(D^2\phi) Y_{kl} Y_{mn} \geq 0, \end{aligned}$$

where  $Y_{ij} = \partial_i \partial_j (\sum_k \xi_k \partial_k \phi)$  and where the inequality follows by convexity of  $\bar{F}$ .

By the spectral theorem there exists  $e^k \in \mathbb{R}^N$  and  $\lambda_k \in \mathbb{R}$  for  $k = 1, \dots, N$  (depending on  $\phi$ ) such that

$$M[\phi] = \sum_k \lambda_k e^k \otimes e^k.$$

Furthermore, since  $M$  is positive semidefinite,  $\lambda_i \geq 0$  for  $i = 1, \dots, N$ . Therefore we have

$$\begin{aligned} & \sum_{i,j} M[\phi]_{ij} \int_0^1 (\partial_{X_{ij}} \bar{F})(sD^2w + (1-s)D^2\phi) ds \\ &= \sum_k \lambda_k \int_0^1 \sum_{i,j} e_i^k e_j^k (\partial_{X_{ij}} \bar{F})(sD^2w + (1-s)D^2\phi) ds \leq 0, \end{aligned}$$

where the inequality follows from the fact that  $\bar{F}$  is nonincreasing. We conclude that

$$\bar{F}(D^2w) - \bar{F}(D^2\phi) \geq -t|D\bar{F}|_0 |D\bar{F}_\delta|_0 |D^4\phi|_0,$$

and hence by (5.9)–(5.11) we get

$$w_t + \bar{F}(D^2w) \geq -|D\bar{F}|_0 \delta - t|D\bar{F}|_0 |D\bar{F}_\delta|_0 |D^4\phi|_0.$$

The first part of the lemma now follows from the comparison principle.

The second part of the lemma follows from (5.9)–(5.11) and the comparison principle after noting that this time, due to its sign, the  $D^2\bar{F}_\delta$  term will be part of the error expression.  $\square$

**5.2. Piecewise constant controls.** Here we study approximations by piecewise constant controls. Such approximations have been studied, e.g., in [24, 21] (see also the references therein). We consider the following simplified version of equation (1.1):

$$(5.12) \quad u_t + \max_i \{-L^i u - f^i(x)\} = 0 \quad \text{in } Q_T,$$

where

$$L^i \phi = \text{tr}[\sigma^i(x) \sigma^{iT}(x) D^2\phi] + b^i(x) D\phi + c^i(x) \phi,$$

and  $\sigma, b, c$  and  $f$  satisfy assumption (A1) when  $\alpha$  is replaced by  $i$ . Note that the coefficients are independent of time. We approximate (5.12) in the following way:

$$(5.13) \quad u^{n+1}(x) = \min_i S_i(\Delta t) u^n(x) \quad \text{in } \{0, 1, \dots, n_T\} \times \mathbb{R}^N,$$

where  $S_i(t)\phi(x)$  denotes the solution at  $(t, x)$  of the linear equation

$$(5.14) \quad u_t - L^i u - f^i(x) = 0$$

with initial data  $\phi$  at time  $t = 0$ . As usual,  $u^n$  is expected to be an approximation of  $u(t_n, x)$ ,  $t_n := n\Delta t$ , and we are looking for a bound on the approximation error.

Under assumption (A1) the comparison principle holds for the linear equations (5.14); hence  $S_i$  and  $\min_i S_i$  are monotone. Furthermore, we have the following consistency relation:

**Lemma 5.7.** *If (A1) holds, then for any smooth function  $\phi$  we have*

$$\begin{aligned} & \left| \frac{1}{\Delta t} [\min_i S_i(\Delta t) - 1] \phi(t_{n-1}, x) + \Delta t \max_i \{-L^i \phi(t_n, x) - f^i(x)\} \right| \\ & \leq C\Delta t |D^2 \phi|_0 + C\Delta t \left( \sum_{n=0}^4 |D^n \phi|_0 + 1 \right) + C\Delta t^{1/2} \left( \sum_{n=0}^2 |D^n \phi|_0 + 1 \right). \end{aligned}$$

We have the following error bound:

**Proposition 5.8.** *Assume (A1). Let  $u_h$  denote the solution of (5.1) corresponding to*

$$S_h(t_{n-1}, t_n) = \min_i S_i(\Delta t)$$

*and  $u_{h,0} = u_0$ , and let  $u$  be the solution of (5.12) with initial value  $u_0$ . Then*

$$-C\Delta t^{\frac{1}{10}} \leq u - u_h \leq 0 \quad \text{in} \quad \Delta t \{0, 1, 2, \dots, n_T\} \times \mathbb{R}^N.$$

*Proof.* We first observe that  $u_h \geq u$  in  $Q_T$ . This can be easily seen from the control interpretation of  $u_h$  (which we have not provided!) or from the comparison principle since  $u_h$  is a supersolution of (5.12) (solutions of (5.14) are supersolutions of (5.12) and so is the min of such solutions). The other bound follows from Lemma 5.7 and Proposition 5.1. □

*Remark 5.6.* Assuming more regularity on the coefficients does not lead to any improvement of the bound. The principal contribution to the error comes from the  $|D^4 \phi|_0$ -term, and this term does not depend on the regularity of the coefficients (only on the  $L^\infty$  norm of  $\sigma$ ).

*Remark 5.7.* In [21] Krylov obtains a better rate, namely 1/6. His approach is different from ours; he works on the dynamic programming principle directly using control techniques.

*Proof of Lemma 5.7.* Let  $\sigma_\delta^i = \sigma^i * \rho_\delta$ , and define similarly  $b_\delta^i, c_\delta^i$ , and  $f_\delta^i$ , and let  $L_\delta^i$  be the operator  $L^i$  corresponding to  $\sigma_\delta^i, b_\delta^i$ , and  $c_\delta^i$ . Observe that

$$\begin{aligned} & \left| \min_i S_i(t) \phi - \phi + t \max_i \{-L_\delta^i \phi - f_\delta^i(x)\} \right| \\ & = \left| \min_i (S_i(t) \phi - \phi) - t \min_i \{L_\delta^i \phi + f_\delta^i(x)\} \right| \\ & \leq \max_i |S_i(t) \phi - \phi + t(-L_\delta^i \phi - f_\delta^i(x))|. \end{aligned}$$

Next, define

$$w^\pm = \phi - t(-L_\delta^i \phi - f_\delta^i(x)) \pm \frac{1}{2} t^2 |L^i L_\delta^i \phi - L^i f_\delta^i|_0 \pm t |(L_\delta^i - L^i) \phi - (f_\delta^i - f^i)|_0,$$

and observe that  $w^+$  is a supersolution of (5.14) while  $w^-$  is a subsolution. By the comparison principle and properties of mollifiers we get

$$\begin{aligned} & |S_i(t)\phi - \phi + t(-L_\delta^i\phi - f_\delta^i(x))| \\ & \leq \frac{1}{2}t^2|L^iL_\delta^i\phi - L^if_\delta^i|_0 + t\delta C \left( \sum_{n=0}^2 |D^n\phi|_0 + 1 \right). \end{aligned}$$

Furthermore, by properties of mollifiers and the Lipschitz regularity of the coefficients we see that

$$|L^iL_\delta^i\phi + L_if_i|_0 \leq C \left( \sum_{n=0}^4 |D^n\phi|_0 + \delta^{-1} \sum_{n=0}^2 |D^n\phi|_0 + \delta^{-1} + 1 \right).$$

By combining the above estimates we get

$$\begin{aligned} & |\min_i S_i(t)\phi - \phi + t \max\{-L^i\phi - f^i(x)\}| \\ & \leq Ct^2 \left( \sum_{n=0}^4 |D^n\phi|_0 + \delta^{-1} \sum_{n=0}^2 |D^n\phi|_0 + \delta^{-1} + 1 \right) + t\delta C \left( \sum_{n=0}^2 |D^n\phi|_0 + 1 \right), \end{aligned}$$

and the result follows by similar arguments as were given in the proofs of Lemmas 5.2 and 5.4 after optimizing w.r.t.  $\delta$ .  $\square$

### 6. REMARKS ON THE HÖLDER CONTINUOUS CASE

In this section we give an extension of the main result, Theorem 3.1, to the case when solutions of (1.1) no longer belong to the space  $C^{0,1}$  but rather belong to the bigger space  $C^\beta$  for some  $\beta \in (0, 1)$ .

In the time-dependent case,  $C^\beta$  regularity of the solution is observed typically when assumption (A1) is relaxed in the following way:

**(A1')**. For any  $\alpha \in \mathcal{A}$ ,  $a^\alpha = \frac{1}{2}\sigma^\alpha\sigma^{\alpha T}$  for some  $N \times P$  matrix  $\sigma^\alpha$ . Moreover, there is a constant  $K$  independent of  $\alpha$  such that

$$|u_0|_\beta + |\sigma^\alpha|_1 + |b^\alpha|_1 + |c^\alpha|_\beta + |f^\alpha|_\beta \leq K.$$

In other words  $u_0, c^\alpha, f^\alpha$  now belongs to  $C^\beta$ .

**Lemma 6.1.** *If (A1') holds, then there exists a unique solution  $u \in C^{0,\beta}(\overline{Q}_T)$  of (1.1) and (1.2).*

This standard result is proved, e.g., in [18]. We claim that under (A1'), we have the same regularity (the same  $\beta$ ) for all equations considered in this paper. We skip the proof of this claim. In the rest of this section, the solutions of the different equations belong to  $C^{0,\beta}(\overline{Q}_T)$  with the same fixed  $\beta \in (0, 1]$ .

Lower than  $C^{0,1}$  regularity of solutions implies lower convergence rates than those obtained in Sections 2–5. We will now state the Hölder versions of some of these results without proofs. The proofs are not much different from the proofs given above, and moreover, the Hölder case was extensively studied in [1]. We start by the convergence rate for the switching system approximation of Section 2.

**Proposition 6.2.** *Assume (A1'). If  $\bar{u}$  and  $v$  are the solutions of (2.2) and (2.1) in  $C^{0,\beta}(\overline{Q}_T)$ , then for  $k$  small enough,*

$$0 \leq v_i - \bar{u} \leq Ck^{\frac{\beta}{2+\beta}} \quad \text{in } \overline{Q}_T, \quad i \in \mathcal{I},$$

where  $C$  only depends on  $T$  and  $K$  from (A1').

In order to state a  $\mathcal{C}^\beta$  version of Theorem 3.1 we need to modify assumption (S3). The requirement on  $\phi_\varepsilon$  should be changed to

$$|\partial_t^{\beta_0} D^{\beta'} \phi_\varepsilon(x, t)| \leq \tilde{K} \varepsilon^{\beta - 2\beta_0 - |\beta'|} \quad \text{in } \overline{Q}_T, \quad \text{for any } \beta_0 \in \mathbb{N}, \beta' \in \mathbb{N}^N.$$

We will denote the modified assumption by (S3'). Now we state the  $\mathcal{C}^\beta$  version of our main result, Theorem 3.1.

**Theorem 6.3.** *Assume (A1'), (S1), (S2) and that (1.3) has a unique solution  $u_h \in C_b(\mathcal{G}_h)$ . Let  $u$  denote the solution of (1.1)–(1.2), and let  $h$  be sufficiently small.*

(a) **(Upper bound)** *If (S3')(i) holds, then there exists a constant  $C$  depending only on  $\mu, K$  in (S1), (A1') such that*

$$u - u_h \leq e^{\mu t} |(u_0 - u_{0,h})^+|_0 + C \min_{\varepsilon > 0} \left( \varepsilon^\beta + E_1(\tilde{K}, h, \varepsilon) \right) \quad \text{in } \mathcal{G}_h,$$

where  $\tilde{K} = |u|_1$ .

(b) **(Lower bound)** *If (S3')(ii) and (A2) hold, then there exists a constant  $C$  depending only on  $\mu, K$  in (S1), (A1') such that*

$$u - u_h \geq -e^{\mu t} |(u_0 - u_{0,h})^-|_0 - C \min_{\varepsilon > 0} \left( \varepsilon^{\frac{\beta^2}{2+\beta}} + E_2(\tilde{K}, h, \varepsilon) \right) \quad \text{in } \mathcal{G}_h,$$

where  $\tilde{K} = |u|_1$ .

*Remark 6.1.* For the FDMs described in Section 4 we get an upper rate of  $\frac{\beta}{2}$  and a lower rate of  $\frac{2\beta^2}{4(2+\beta)-2\beta}$  in the  $\mathcal{C}^\beta$  case. Compare with Theorem 4.1.

APPENDIX A. WELL-POSEDNESS, REGULARITY, AND CONTINUOUS DEPENDENCE FOR SWITCHING SYSTEMS

In this section we give well-posedness, regularity, and continuous dependence results for solutions of a very general switching system that has as special cases the scalar HJB equations (1.1), and the switching systems (2.1), (2.3), (3.3).

We consider the following system:

$$(A.1) \quad F_i(x, u, \partial_t u_i, Du_i, D^2 u_i) = 0 \quad \text{in } Q_T, \quad i \in \mathcal{I} := \{1, \dots, M\},$$

with

$$F_i(t, x, r, p_t, p_x, X) = \max \left\{ p_t + \sup_{\alpha \in \mathcal{A}} \inf_{\beta \in \mathcal{B}} \mathcal{L}_i^{\alpha, \beta}(x, r_i, p_x, X); r_i - \mathcal{M}_i r \right\},$$

$$\mathcal{L}_i^{\alpha, \beta}(t, x, s, q, X) = -\text{tr}[a_i^{\alpha, \beta}(t, x)X] - b_i^{\alpha, \beta}(t, x)q - c_i^{\alpha, \beta}(t, x)s - f_i^{\alpha, \beta}(t, x),$$

where  $\mathcal{M}$  is defined below (2.1),  $\mathcal{A}, \mathcal{B}$  are compact metric spaces,  $r$  is a vector,  $r = (r_1, \dots, r_M)$ , and  $k > 0$  is a constant (the switching cost). See [13, 7, 28, 16, 15] for more information about such systems.

We make the following assumption:

(A). For any  $\alpha, \beta, i, a_i^{\alpha, \beta} = \frac{1}{2} \sigma_i^{\alpha, \beta} \sigma_i^{\alpha, \beta T}$  for some  $N \times P$  matrix  $\sigma_i^{\alpha, \beta}$ . Furthermore, there is a constant  $C$  independent of  $i, \alpha, \beta, t$ , such that

$$|\sigma_i^{\alpha, \beta}(t, \cdot)|_1 + |b_i^{\alpha, \beta}(t, \cdot)|_1 + |c_i^{\alpha, \beta}(t, \cdot)|_1 + |f_i^{\alpha, \beta}(t, \cdot)|_1 \leq \bar{C}.$$

We start by comparison, existence, uniqueness, and  $L^\infty$  bounds on the solution and its gradient. Before stating the results, we define  $USC(\bar{Q}_T; \mathbb{R}^M)$  and  $LSC(\bar{Q}_T; \mathbb{R}^M)$  to be the spaces of upper and lower semicontinuous functions from  $\bar{Q}_T$  into  $\mathbb{R}^M$  respectively.

**Theorem A.1.** *Assume (A) holds.*

(i) *If  $u \in USC(\bar{Q}_T; \mathbb{R}^M)$  is a subsolution of (A.1) bounded above and  $v \in LSC(\bar{Q}_T; \mathbb{R}^M)$  is a supersolution of (A.1) bounded below, then  $u \leq v$  in  $\bar{Q}_T$ .*

(ii) *There exists a unique bounded continuous solution  $u$  of (A.1).*

(iii) *The solution  $u$  of (A.1) belongs to  $C^{0,1}(\bar{Q}_T)$  and satisfies for all  $t, s \in [0, T]$ ,*

$$e^{-\lambda t} \max_i |u_i(t, \cdot)|_0 \leq \max_i |u_{0,i}|_0 + t \sup_{i, \alpha, \beta} |f_i^{\alpha, \beta}|_0,$$

where  $\lambda := \sup_{i, \alpha, \beta} |c_i^{\alpha, \beta+}|_0$ ,

$$e^{\lambda_0 t} \max_i [u_i(t, \cdot)]_1 \leq \max_i [u_{0,i}]_1 + t \sup_{i, \alpha, \beta, s} \left\{ |u^i|_0 [c_i^{\alpha, \beta}(s, \cdot)]_1 + [f_i^{\alpha, \beta}(s, \cdot)]_1 \right\},$$

where  $\lambda_0 := \sup_{i, \alpha, \beta, s} \{ |c_i^{\alpha, \beta+}(s, \cdot)|_0 + [\sigma_i^{\alpha, \beta}(s, \cdot)]_1^2 + [b_i^{\alpha, \beta}(s, \cdot)]_1 \}$ , and

$$\max_i |u_i(t, x) - u_i(s, x)| \leq C |t - s|^{1/2},$$

where  $C \leq 8M\bar{C} + \sqrt{T}\bar{C}(2M + 1)$  and  $M := \sup_{i,t} |u_i(t, \cdot)|_1$ .

Before giving the proof we state a key technical lemma.

**Lemma A.2.** *Let  $u \in USC(\bar{Q}_T; \mathbb{R}^M)$  be a bounded above subsolution of (A.1) and  $\bar{u} \in LSC(\bar{Q}_T; \mathbb{R}^M)$  be a bounded below supersolution of another equation (A.1), where the functions  $\mathcal{L}_i^{\alpha, \beta}$  are replaced by functions  $\bar{\mathcal{L}}_i^{\alpha, \beta}$  satisfying the same assumptions. Let  $\phi : [0, T] \times \mathbb{R}^{2N} \rightarrow \mathbb{R}$  be a smooth function bounded from below. We denote by*

$$\psi_i(t, x, y) = u_i(t, x) - \bar{u}_i(t, y) - \phi(t, x, y),$$

and  $M = \sup_{i,t,x,y} \psi_i(t, x, y)$ . *If there exists a maximum point for  $M$ , i.e., a point  $(i', t_0, x_0, y_0)$  such that  $\psi_{i'}(t_0, x_0, y_0) = M$ , then there exists  $i_0 \in \mathcal{I}$  such that  $(i_0, t_0, x_0, y_0)$  is also a maximum point for  $M$ , and, in addition  $\bar{u}_{i_0}(t_0, y_0) < \mathcal{M}_{i_0} \bar{u}(t_0, y_0)$ .*

Loosely speaking this lemma means that whenever we do doubling of variables for systems of the type (A.1), we can ignore the  $u_i - \mathcal{M}_i u$  parts of the equations and we can proceed as if working with scalar equations. We skip the proof since it is similar to the proof given in [2] for the stationary case.

*Proof of Theorem A.1.* Comparison, uniqueness, and existence is proved in [16] for the stationary Dirichlet problem for (1.1) on a bounded domain under similar assumptions on the data. To extend the comparison result to a time-dependent problem in an unbounded domain, we only need to modify the test function used in [16] in the standard way. (See also the arguments given below.) Comparison implies uniqueness, and existence follows from Perron's method. This last argument

is similar to the argument given in [16], but easier since we have no boundary conditions other than the initial condition.

Let

$$w(t) := e^{\lambda t} \left\{ \max_i |u_{0,i}|_0 + t \sup_{i,\alpha,\beta} |f_i^{\alpha,\beta}|_0 \right\}.$$

Then the bound on  $|u|_0$  follows from the comparison principle after checking that  $w$  ( $-w$ ) is a supersolution (subsolution) of (A.1).

To get the bound on the gradient of  $u$ , consider

$$m := \sup_{i,t,x,y \in \mathbb{R}^N} \{u_i(t,x) - u_i(t,y) - \bar{w}(t)|x - y|\},$$

where

$$\bar{w}(t) := e^{\lambda_0 t} \left\{ \max_i [u_{0,i}]_1 + t \sup_{i,\alpha,\beta,s} \left\{ |u_i|_0 [c_i^{\alpha,\beta}(s, \cdot)]_1 + [f_i^{\alpha,\beta}(s, \cdot)]_1 \right\} \right\}.$$

We are done if we can prove that  $m \leq 0$ . Assume this is not the case,  $m > 0$ , and for simplicity that this maximum is attained in  $\bar{t}, \bar{x}, \bar{y}$ . Then there exists a  $k > 0$  such that

$$u_i(\bar{t}, \bar{x}) - u_i(\bar{t}, \bar{y}) - \bar{w}(\bar{t})|\bar{x} - \bar{y}| - \bar{t}e^{\lambda_0 \bar{t}}k > 0, \quad i \in \mathcal{I}.$$

Let  $\psi_i(t, x, y) := u_i(t, x) - u_i(t, y) - \bar{w}(t)|x - y| - te^{\lambda_0 t}k$ . Then  $\psi$  also has a maximum  $M > 0$  at some point  $(\tilde{i}, \tilde{t}, \tilde{x}, \tilde{y})$ . Since  $M > 0$ ,  $\tilde{x} \neq \tilde{y}$  and  $\tilde{t} > 0$ . Therefore  $\bar{w}(t)|x - y| + te^{\lambda_0 t}k$  is a smooth function at  $(\tilde{t}, \tilde{x}, \tilde{y})$  and a standard argument using the viscosity sub- and supersolution inequalities for (A.1) at  $(\tilde{t}, \tilde{x}, \tilde{y})$  and Lemma A.2 leads to  $k \leq 0$ . See the proof of Theorem A.3 for a similar argument. This is a contradiction and hence  $m \leq 0$ .

In the general case when the maximum  $m$  need not be attained at some finite point, we must modify the test function in the standard way. We skip the details.

To get the time regularity result, assume that  $s < t$  and let  $u^\varepsilon$  be the solution of (A.1) in  $t \in (s, T]$  starting from  $u(s, \cdot) * \rho_\varepsilon(x) =: u_0^\varepsilon(x)$ . By the comparison principle,

$$|u - u^\varepsilon| \leq \sup_{r \in [s, T]} [u(r, \cdot)]_1 \varepsilon \quad \text{in } [s, T] \times \mathbb{R}^N,$$

and easy computations show that

$$w_\varepsilon^\pm(t, x) = e^{\lambda t} \left\{ u_0^\varepsilon(x) \pm (t - s)C_\varepsilon \right\}$$

are the subsolution ( $w^-$ ) and the supersolution ( $w^+$ ) of (A.1) if

$$C_\varepsilon = \bar{C}^2 |D^2 u_0^\varepsilon|_0 + \bar{C} (|Du_0^\varepsilon|_0 + |u_0^\varepsilon|_0 + 1)$$

and  $\bar{C}$  is given by (A). Another application of the comparison principle then yields

$$w_\varepsilon^- \leq u^\varepsilon \leq w_\varepsilon^+ \quad \text{in } [s, T] \times \mathbb{R}^N.$$

The result now follows from

$$\begin{aligned} & |u(t, x) - u(s, x)| \\ & \leq |u(t, x) - u^\varepsilon(t, x)| + |u^\varepsilon(t, x) - u_0^\varepsilon(x)| + |u_0^\varepsilon(x) - u(s, x)| \\ & \leq ([u(t, \cdot)]_1 + [u(s, \cdot)]_1)\varepsilon + |t - s|C_\varepsilon, \end{aligned}$$

and a minimization in  $\varepsilon$  after noting that  $C_\varepsilon \leq C(\varepsilon^{-1} + 1)$ . □

We proceed to obtain continuous dependence on the coefficients following the approach of [18].

**Theorem A.3.** *Let  $u$  and  $\bar{u}$  be solutions of (A.1) with coefficients  $\sigma, b, c, f$  and  $\bar{\sigma}, \bar{b}, \bar{c}, \bar{f}$ , respectively. If both sets of coefficients satisfy (A1), and  $|u|_0 + |\bar{u}|_0 + [u(t, \cdot)]_1 + [\bar{u}(t, \cdot)]_1 \leq M < \infty$  for  $t \in [0, T]$ , then*

$$e^{-\lambda t} \max_i |u_i(t, \cdot) - \bar{u}_i(t, \cdot)|_0 \leq \max_i |u_i(0, \cdot) - \bar{u}_i(0, \cdot)|_0 + t^{1/2} K \sup_{i, \alpha, \beta} |\sigma - \bar{\sigma}|_0 + t \sup_{i, \alpha, \beta} \left\{ 2M|b - \bar{b}|_0 + M|c - \bar{c}|_0 + |f - \bar{f}|_0 \right\},$$

where  $\lambda := \sup_{i, \alpha, \beta} |c^-|_0$  and

$$K^2 \leq 8M^2 + 8MT \sup_{i, \alpha, \beta} \left\{ 2M[\sigma]_1^2 \wedge [\bar{\sigma}]_1^2 + 2M[b]_1 \wedge [\bar{b}]_1 + M[c]_1 \vee [\bar{c}]_1 + [f]_1 \wedge [\bar{f}]_1 \right\}.$$

*Proof.* We only indicate the proof in the case  $\lambda = 0$ . Define

$$\begin{aligned} \psi^i(t, x, y) &:= u_i(t, x) - \bar{u}_i(t, y) - \frac{1}{\delta} |x - y|^2 - \varepsilon(|x|^2 + |y|^2), \\ m &:= \sup_{i, t, x, y} \psi^i(t, x, y) - \sup_{i, x, y} (\psi^i(0, x, y))^+, \\ \bar{m} &:= \sup_{i, t, x, y} \left\{ \psi^i(t, x, y) - \frac{\sigma m t}{T} \right\}, \end{aligned}$$

where  $\sigma \in (0, 1)$ . We assume  $m > 0$  since otherwise we are done. We will now derive an upper bound on  $m$ . To do this we consider  $\bar{m}$ . By the assumptions this supremum is attained at some point  $(i_0, t_0, x_0, y_0)$ . Since  $m > 0$  it follows that  $\bar{m} > 0$  and  $t_0 > 0$ , and by Lemma A.2, the index  $i_0$  may be chosen so that  $\bar{u}_{i_0}(t_0, y_0) < \mathcal{M}_{i_0} \bar{u}(t_0, y_0)$ . With this in mind, the maximum principle for semicontinuous functions [8] and the definition of viscosity solutions imply the following inequality:

$$p_t - \bar{p}_t + \sup_{\alpha} \inf_{\beta} \mathcal{L}_{i_0}^{\alpha, \beta}(t_0, x_0, u_{i_0}, p_x, X) - \sup_{\alpha} \inf_{\beta} \bar{\mathcal{L}}_{i_0}^{\alpha, \beta}(t_0, y_0, \bar{u}_{i_0}, p_y, Y) \leq 0,$$

where  $(p_t, p_x, X) \in \bar{\mathcal{P}}^{2,+} u_{i_0}(x_0)$  and  $(\bar{p}_t, p_y, Y) \in \bar{\mathcal{P}}^{2,-} \bar{u}_{i_0}(y_0)$  (see [8] for the notation). Furthermore  $p_t - \bar{p}_t = \frac{\sigma m}{T}$ ,  $p_x = \frac{2}{\delta}(x_0 - y_0) + 2\varepsilon x_0$ ,  $p_y = \frac{2}{\delta}(x_0 - y_0) - 2\varepsilon y_0$ , and

$$\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \leq \frac{2}{\delta} \begin{pmatrix} I & -I \\ -I & I \end{pmatrix} + 2\varepsilon \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + \mathcal{O}(\kappa),$$

for some  $\kappa > 0$ . In the end we will fix  $\sigma, \delta$ , and  $\varepsilon$  and send  $\kappa \rightarrow 0$ , so we simply ignore the  $\mathcal{O}(\kappa)$ -term in the following. The first inequality implies

$$\begin{aligned} \frac{\sigma m}{T} &\leq \sup_{i, \alpha, \beta} \left\{ -\text{tr}[\bar{a}(t_0, y_0)Y] + \text{tr}[a(t_0, x_0)X] + \bar{b}(t_0, y_0)p_x - b(t_0, x_0)p_y \right. \\ &\quad \left. + \bar{c}(t_0, y_0)\bar{u}(t_0, y_0) - c(t_0, x_0)u(t_0, x_0) + \bar{f}(t_0, y_0) + f(t_0, x_0) \right\}. \end{aligned}$$

Note that Lipschitz regularity of the solutions and a standard argument yield

$$|x_0 - y_0| \leq \delta M.$$

So using Ishii's trick on the second-order terms [14, pp. 33, 34], and a few other manipulations, we get

$$\begin{aligned} \frac{\sigma m}{T} \leq \sup_{i,\alpha,\beta} \left\{ \frac{2}{\delta} |\sigma(t_0, x_0) - \bar{\sigma}(t_0, y_0)|^2 + 2M |b(t_0, x_0) - \bar{b}(t_0, y_0)| \right. \\ \left. + C\varepsilon(1 + |x_0|^2 + |y_0|^2) \right. \\ \left. + M |c(t_0, x_0) - \bar{c}(t_0, y_0)| + |f(t_0, x_0) - \bar{f}(t_0, y_0)| \right\}. \end{aligned}$$

Some more work leads to an estimate for  $m$  depending on  $T$ ,  $\sigma$ ,  $\delta$ , and  $\varepsilon$ , and using the definition of  $m$  and estimates on  $\sup_{i,x,y} \psi_i(0, x, y)$ , we obtain a similar upper bound for  $u - \bar{u}$ . We finish the proof of the upper bound on  $u - \bar{u}$  by sending  $\sigma \rightarrow 1$ , minimizing this expression w.r.t.  $\delta$ , sending  $\varepsilon \rightarrow 0$ , and noting that the result still holds if we replace  $T$  by any  $t \in [0, T]$ . The lower bound follows in a similar fashion.  $\square$

#### REFERENCES

- [1] G. Barles and E. R. Jakobsen. On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations. *M2AN Math. Model. Numer. Anal.* 36(1):33–54, 2002. MR1916291 (2003h:65142)
- [2] G. Barles and E. R. Jakobsen. Error bounds for monotone approximation schemes for Hamilton-Jacobi-Bellman equations. *SIAM J. Numer. Anal.* 43(2):540–558, 2005. MR2177879
- [3] G. Barles and P. E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Anal.* 4(3):271–283, 1991. MR1115933 (92d:35137)
- [4] F. Bonnans, E. Ottenwaelter, and H. Zidani. A fast algorithm for the two dimensional HJB equation of stochastic control. *M2AN Math. Model. Numer. Anal.* 38(4):723–735, 2004. MR2087732 (2005e:93165)
- [5] F. Bonnans and H. Zidani. Consistency of generalized finite difference schemes for the stochastic HJB equation. *SIAM J. Numer. Anal.* 41(3):1008–1021, 2003. MR2005192 (2004i:49061)
- [6] F. Camilli and M. Falcone. An approximation scheme for the optimal control of diffusion processes. *RAIRO Modél. Math. Anal. Numér.* 29(1): 97–122, 1995. MR1326802 (96a:49033)
- [7] I. Capuzzo-Dolcetta and L. C. Evans. Optimal switching for ordinary differential equations. *SIAM J. Control Optim.* 22(1):143–161, 1984. MR728678 (85b:49043)
- [8] 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(1):1–67, 1992. MR1118699 (92j:35050)
- [9] M. G. Crandall and P.-L. Lions. Two approximations of solutions of Hamilton-Jacobi equations. *Math. Comp.* 43(167):1–19, 1984. MR744921 (86j:65121)
- [10] H. Dong and N. V. Krylov. On the rate of convergence of finite-difference approximations for Bellman equations with constant coefficients. *St. Petersburg Math. J.* 17(2): 295–313, 2006. MR2159586 (2006f:49050)
- [11] H. Dong and N. V. Krylov. On the Rate of Convergence of Finite-difference Approximations for Parabolic Equations with  $C^1$  and  $C^2$  Coefficients. *Electron. J. Differential Equations* 2005(102): 1–25, 2005. MR2162263 (2006i:35008)
- [12] H. J. Kushner and P. Dupuis. *Numerical methods for stochastic control problems in continuous time*. Springer-Verlag, New York, 2001. MR1800098 (2001g:93002)
- [13] L. C. Evans and A. Friedman. Optimal stochastic switching and the Dirichlet problem for the Bellman equation. *Trans. Amer. Math. Soc.* 253:365–389, 1979. MR536953 (80f:93091)
- [14] H. Ishii. On uniqueness and existence of viscosity solutions of fully nonlinear second-order elliptic PDEs. *Comm. Pure Appl. Math.*, 42(1):15–45, 1989. MR973743 (89m:35070)
- [15] H. Ishii and S. Koike. Viscosity solutions for monotone systems of second-order elliptic PDEs. *Comm. Partial Differential Equations* 16(6-7):1095–1128, 1991. MR1116855 (92h:35066)
- [16] H. Ishii and S. Koike. Viscosity solutions of a system of nonlinear second-order elliptic PDEs arising in switching games. *Funkcial. Ekvac.*, 34:143–155, 1991. MR1116886 (92h:35067)



- [17] E. R. Jakobsen. On the rate of convergence of approximation schemes for Bellman equations associated with optimal stopping time problems. *Math. Models Methods Appl. Sci. (M3AS)* 13(5):613-644, 2003. MR1978929 (2004g:49054)
- [18] E. R. Jakobsen and K. H. Karlsen. Continuous dependence estimates for viscosity solutions of fully nonlinear degenerate parabolic equations. *J. Differential Equations* 183:497-525, 2002. MR1919788 (2003i:35135)
- [19] N. V. Krylov. On the rate of convergence of finite-difference approximations for Bellman's equations. *St. Petersburg Math. J.*, 9(3):639-650, 1997. MR1466804 (98h:49033)
- [20] N. V. Krylov. On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients. *Probab. Theory Relat. Fields*, 117:1-16, 2000. MR1759507 (2001j:65134)
- [21] N. V. Krylov. Approximating value functions for controlled degenerate diffusion processes by using piece-wise constant policies. *Electron. J. Probab.* 4(2), 1999. MR1668597 (2000b:49056)
- [22] N. V. Krylov. On the rate of convergence of finite-difference approximations for Bellman equations with Lipschitz coefficients. *Appl. Math. Optim.* 52(2): 365-399, 2005. MR2174020 (2006k:65219)
- [23] P.-L. Lions. Personal communication.
- [24] P.-L. Lions and B. Mercier. Approximation numérique des équations de Hamilton-Jacobi-Bellman. *RAIRO Anal. Numér.* 14(4):369-393, 1980. MR596541 (82b:65055)
- [25] C.-W. Shu. Total-variation-diminishing time discretizations. *SIAM J. Sci. Statist. Comput.* 9(6):1073-1084, 1988. MR963855 (90a:65196)
- [26] P. E. Souganidis. Approximation schemes for viscosity solutions of Hamilton-Jacobi equations. *J. Differential Equations* 59(1):1-43, 1985. MR803085 (86k:35028)
- [27] A. Tourin. Splitting methods for Hamilton-Jacobi equations. *Numer. Methods Partial Differential Equations*, 22(2): 381-396, 2006. MR2201439 (2006j:65254)
- [28] N. Yamada. Viscosity solutions for a system of elliptic inequalities with bilateral obstacles. *Funkcial. Ekvac.* 30(2-3):417-425, 1987. MR927191 (88m:35061)

LABORATOIRE DE MATHÉMATIQUES ET PHYSIQUE THÉORIQUE, UNIVERSITY OF TOURS, 37200 TOURS, FRANCE

*E-mail address:* `barles@lmpt.univ-tours.fr`

DEPARTMENT OF MATHEMATICAL SCIENCES, NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY, 7491 TRONDHEIM, NORWAY

*E-mail address:* `erj@math.ntnu.no`