

Some non monotone schemes for Hamilton-Jacobi-Bellman equations

Xavier Warin , EDF R&D & FiME

February 2015

Abstract

We extend the theory of Barles Jakobsen [2] for a class of “nearly” monotone schemes to solve Hamilton Jacobi Bellman equations. We show that the monotonicity of the schemes can be relaxed still leading to the convergence to the viscosity solution of the equation. We give some examples of such numerical schemes and show that the bounds obtained by the framework developed are not tight. At last we test the schemes.

Introduction

We are interested in the following HJB equation arising in infinite horizon, discounted, stochastic control problems

$$F(x, u, \mathbf{D}u, \mathcal{D}^2u) = 0 \text{ in } \mathbf{R}^N, \quad (1)$$

with

$$\begin{aligned} F(x, t, p, X) &= \sup_{\alpha \in \mathcal{A}} \mathcal{L}^\alpha(x, t, p, X), \\ \mathcal{L}^\alpha(x, t, p, X) &= -tr[a^\alpha(x)X] - b^\alpha(x)p + c^\alpha(x)t - f^\alpha(x). \end{aligned} \quad (2)$$

where a, b, c, f are at least continuous functions on $\mathbf{R}^N \times \mathcal{A}$ with values in $S(N)$ the space of symmetric $N \times N$ matrices, \mathbf{R}^N , \mathbf{R} and \mathbf{R} respectively. The space of controls \mathcal{A} is supposed to be a compact metric space.

Supposing h is an approximating parameter, we consider an approximation S of F such that the approximate function u_h satisfies:

$$S(h, x, u_h(x), [u_h]_x) = 0, x \in \mathbf{R}^N, \quad (3)$$

where $S(h, x, r, [t]_x)$ is defined for $x \in \mathbf{R}^N$, $r \in \mathbf{R}$, t a function defined on \mathbf{R}^N , and $[t]_x$ is a function defined at x from t . This notation was introduced by [3] to prove that a scheme S which is non decreasing in r and non increasing in $[t]_x$ is monotone. When the scheme S is a monotone, uniformly continuous and a consistent approximation of F and when a discrete bounded solution u_h can

be found for (3), then u_h converges to the viscosity solution of the problem (1) [1, 2].

When the scheme is non monotone, no general theory is available : [5] showed for example that some non monotone schemes may not converge, [6] gave some examples of non monotone schemes converging toward a false solution.

Methods to solve HJB equations includes Finite Difference methods and Semi Lagrangian methods. Classical Finite Difference method often can be interpreted as a Markov Chain [4] leading to monotone schemes. When a^α is not diagonally dominant, the requirement about monotonicity of the scheme leads to Finite Difference scheme such as in [7] using ideas independently developed in [8]. As an alternative to Finite Difference schemes, Semi Lagrangian schemes of low order based on the original work in [10] have been developed in [9, 1]. Monotonicity of the schemes is desirable because a convenient framework is available. Nevertheless, it is to notice that some high order non monotone schemes have been developed and proved convergent for some first order Hamilton-Jacobi equations for examples in [11] (with some filtered schemes to satisfy some given bounds) or in [12] using spectral methods on a periodic domain. In the case of first order Hamilton-Jacobi-Bellman equation, some non monotone explicit schemes were developed because of the inefficiency of monotone scheme for discontinuous initial data and proved to be convergent in [13].

In this paper we will relax too the constraint on the monotonicity of the scheme such that it can converge to the right solution. We will suppose that this scheme is a perturbation of a monotone scheme \hat{S} and we have in mind the schemes based on interpolation method (Semi Lagrangian scheme or Finite Difference scheme with carefully chosen directions). This idea is not new : [3] already have emphasized the fact that monotonicity could be relaxed. As for Finite Difference schemes, for first order Hamilton-Jacobi, following the ideas in [14], some potentially high order scheme have been developed in [15] by blending two schemes: one of high order potentially instable and one monotone of low order that will be used near singularity of the solution. As for Semi Lagrangian, results for high order schemes interpolators are given by the Italian School for first order Hamilton-Jacobi in [16]. The interest in nearly monotone scheme is driven by the fact that monotone scheme such as Semi-Lagrangian schemes with linear interpolators are converging numerically very slowly. Then it seems natural to try to use schemes that are potentially fast convergent (with potentially a high order of consistency) while being sure that they converge towards the true solution even if we cannot prove a higher rate of convergence than in the monotone case. Our goal is to develop a framework that could be used to develop new schemes and easily prove that they are convergent.

Specifically we treat the second order Hamilton Jacobi Bellman stationary problem with nearly monotone schemes. Because in some case the resolution of (3) can be impossible we will try to relax the equality requirement and only assume we can find a function u_h such that on a given grid X that may depend on h

$$|S(h, x, u_h(x), [u_h]_x)| \leq \epsilon(h), x \in X, \quad (4)$$

where $\epsilon(h)$ is a continuous function of h with $\epsilon(0) = 0$.

If existing, the solution of such a scheme is not unique, so we consider a constructed sequence of solution u_h of (4) and get bounds proving the convergence of u_h towards the right solution.

In a second part of the article, we detail some schemes that can be cast into this framework and derived some convergence properties.

First we show that Semi Lagrangian methods with high order interpolation and truncation can be cast in this framework. In a previous work [19], it has been proven that an interpolation of high degree with truncation could be used to estimate the solution of time dependent Hamilton-Jacobi-Bellman equations. A numerical study has been achieved comparing different types of Lagrange interpolators, Bernstein approximations, and Cubic spline interpolators. The conclusion was that Gauss Lobatto Legendre interpolators were the most interesting given the accuracy obtained compared to the CPU time used. In the present article, we only focus on Gauss Lobatto Legendre interpolators to study the stationary Hamilton-Jacobi-Bellman equations in our framework. Using a direct estimation, we besides prove that the result obtained by the framework is not optimal. In fact, with the direct estimation, we get back the convergence result in $O(h^{\frac{1}{4}})$ previously obtained in [9, 1].

We then develop a Finite Difference approach with interpolation. In order to use the developed framework, we have to suppose that the diffusion coefficient is independent of the space. Once again the method can be cast in the framework developed and the convergence rate obtained is not optimal. The best rate we found is in $O(h^{\frac{1}{2}})$ which is the rate found in [1] with the same assumptions. As for the general case where the diffusion depends on the space, the same rate of convergence is reached using the Bonnans/Krylov discretization [7, 8] as shown in [8].

In the sequel the constant C may vary between lines.

1 Main result

We define the norm denoted $||$ as follows: for any integer $m \geq 1$ and $z = (z_i)_i \in \mathbf{R}^m$, we set $|z|^2 = \sum_{i=1}^m z_i^2$. For a matrix $M \in \mathbf{R}^{n_1 \times n_2}$, $|M|^2 = \text{tr}[M^t M]$ with M^t the transpose of M .

If $f : \mathbf{R}^N \rightarrow \mathbf{R}^M$ we define the semi-norms:

$$|f|_0 = \sup_{x \in \mathbf{R}^N} |f(x)|, [f]_1 = \sup_{\substack{x, y \in \mathbf{R}^N \\ x \neq y}} \frac{|f(x) - f(y)|}{|x - y|},$$

and

$$|f|_1 = |f|_0 + [f]_1.$$

$C^{0,1}(\mathbf{R}^N)$ stands for the set of functions $f : \mathbf{R}^n \rightarrow \mathbf{R}$ with finite norm $|f|_1$, $C_b(\mathbf{R}^N)$ the set with finite norm $|f|_0$. In the sequel we make the following classical assumptions

Assumption (A1). For any $\alpha \in \mathcal{A}$, $a^\alpha = \frac{1}{2}\sigma^\alpha\sigma^{\alpha t}$ for some $N \times P$ matrix σ^α . Furthermore, there exists λ, K independent of α such that:

$$c^\alpha \geq \lambda > 0, \text{ and } |\sigma^\alpha|_1 + |b^\alpha|_1 + |f^\alpha|_1 \leq K. \quad (5)$$

Assumption (A2). The constant λ in (A1) satisfies $\lambda > \sup_\alpha \frac{1}{2}[\sigma^\alpha]_1^2 + [b^\alpha]_1$.

We just recall the well-posedness and regularity result given in [1] with demonstrations in the references therein.

Proposition 1.1. Assume (A1): There exists a unique viscosity u solution of (1) in $C_b(\mathbf{R}^N)$. If w_1 and w_2 are in $C_b(\mathbf{R}^N)$ and are sub- and supersolution of (1) respectively, then $w_1 \leq w_2$ in \mathbf{R}^N .

Assume (A1), (A2): There exists a unique bounded viscosity u solution of (1) in $C^{0,1}(\mathbf{R}^N)$.

Remark 1.2. Assumptions (A1) can be given for more general Hölder spaces, and regularity of the solution is then given in [2].

Here we add some new definitions that will be helpful in the sequel. First we introduce the notion of ϵ monotone scheme stating that the scheme S is “nearly” monotone

Definition 1.3. An $\epsilon(p, K)$ **monotone scheme** S is a scheme such that there exists $\bar{\lambda}$ satisfying:

- for every $h > 0$, $x \in \mathbf{R}^N$, $r \in \mathbf{R}$, for every function $w \in C^{0,1}(\mathbf{R}^N)$, $v \in C_b(\mathbf{R}^N)$ such that $v \geq w$:

$$S(h, x, r, [v]_x) \leq S(h, x, r, [w]_x) + K|w|_1 h^p, \quad (6)$$

- for every $h > 0$, $x \in \mathbf{R}^N$, $r \in \mathbf{R}$, for every function $w \in C_b(\mathbf{R}^N)$, $v \in C^{0,1}(\mathbf{R}^N)$ such that $v \geq w$:

$$S(h, x, r, [v]_x) \leq S(h, x, r, [w]_x) + K|v|_1 h^p, \quad (7)$$

- for every $h > 0$, $x \in \mathbf{R}^N$, $r \in \mathbf{R}$, $m \geq 0$, for every function $u \in C_b(\mathbf{R}^N)$:

$$S(h, x, r + m, [u + m]_x) \geq S(h, x, r, [u]_x) + \bar{\lambda}m. \quad (8)$$

Because it is sometimes difficult or impossible to prove that a discrete scheme has a solution, we relax the notion of solution as we did in (4):

Definition 1.4. An $\epsilon(c)$ **solution** u of scheme S is a continuous function which satisfies:

$$|S(h, x, u(x), [u]_x)| < c.$$

In the same spirit we relax the notion of subsolution and supersolution.

Definition 1.5. An $\epsilon(c)$ **subsolution (supersolution)** u of scheme S is a continuous function which satisfies:

$$S(h, x, u(x), [u]_x) < c \quad (S(h, x, u(x), [u]_x) > -c).$$

On the scheme we make the following assumptions:

Assumption (S1). The scheme S is $\epsilon(p, K)$ monotone.

Assumption (S2). (Regularity of S scheme) For every $h > 0$ and $\phi \in C_b(\mathbf{R}^N)$, $x \rightarrow S(h, x, \phi(x), [\phi]_x)$ is bounded and continuous in \mathbf{R}^N and the function $r \rightarrow S(h, x, r, [\phi]_x)$ is uniformly continuous for bounded r , uniformly in $x \in \mathbf{R}^N$.

Assumption (S3). (Consistency) There exists a set of integers $(k_i)_{i=1, m}$, and a constant K such that for every $h \geq 0$, $x \in \mathbf{R}^N$ and a smooth function ϕ

$$|F(x, \phi(x), \mathcal{D}\phi(x), \mathcal{D}^2\phi) - S(h, x, \phi(x), [\phi]_x)| \leq K \sum_{i=1}^m h^{k_i} |\mathcal{D}^i \phi|_0.$$

We add an assumption on the existence of a solution of the discretized scheme with sufficient regularity which has to be checked for each scheme:

Assumption (S4). We suppose there exists C and r independent of h such that for each h we can find an $\epsilon(Ch^r)$ solution u_h of scheme S .

Remark 1.6. Assumptions (S2) and (S3) are classical. Assumption (S1) is a relaxation of the constraint on the monotonicity of the scheme.

An ϵ monotone scheme doesn't ensure the existence of a discrete comparison result but we give here a relaxation of this result:

Lemma 1.7. Assume (S1). Let $v \in C_b(\mathbf{R}^N)$ and $u \in C^{0,1}(\mathbf{R}^N)$. If u is a subsolution of (4) and v is an $\epsilon(C)$ supersolution of (4) then

$$u \leq v + \frac{1}{\lambda}(Kh^p|u|_1 + C).$$

Proof. Mimicking Lemma 2.3 in [1], we assume $m := \sup_{\mathbf{R}^N}(u-v) > \frac{1}{\lambda}(Kh^p|u|_1 + C)$. Let $\{x_n\}_n$ be a sequence such that $u(x_n) - v(x_n) := \delta_n \rightarrow m$. For n large enough, $\delta_n > \frac{1}{\lambda}(Kh^p|u|_1 + C)$. Using the subsolution definition, assumption (S1) :

$$\begin{aligned} 0 &\geq S(h, x_n, u(x_n), [u]_{x_n}) - S(h, x_n, v(x_n), [v]_{x_n}) - C, \\ 0 &\geq S(h, x_n, v(x_n) + \delta_n, [v + m]_{x_n}) - S(h, x_n, v(x_n), [v]_{x_n}) - Kh^p|u|_{1,0} - C, \\ 0 &\geq \bar{\lambda}\delta_n + w(m - \delta_n) - Kh^p|u|_1 - C, \end{aligned}$$

where $w(t) \rightarrow 0$ when $t^+ \rightarrow 0$ by assumption (S2). Letting $n \rightarrow \infty$, we get

$$m \leq \frac{1}{\lambda} (Kh^p |u|_1 + C),$$

which gives the contradiction. \square

We now give the lower and upper bound for the error given by the scheme. The lower bound will be given by the classical Krylov method of shaking coefficients [18], while the upper bound will be given by the use of a switching system as in [2, 20] that gives a supersolution of the problem. To use the theory developed in [2], we need to add a last assumption:

Assumption (A3). *For every $\delta > 0$, there are $M \in \mathbf{N}$ and $\{\alpha_i\}_{i=1}^M \subset \mathcal{A}$, such that for any $\alpha \in \mathcal{A}$*

$$\inf_{1 \leq i \leq M} (|\sigma^\alpha - \sigma^{\alpha_i}|_0 + |b^\alpha - b^{\alpha_i}|_0 + |c^\alpha - c^{\alpha_i}|_0 + |f^\alpha - f^{\alpha_i}|_0) < \delta.$$

We introduce the following switching system:

$$F_i^\epsilon(x, v^\epsilon, \mathcal{D}v_i^\epsilon, \mathcal{D}^2v_i^\epsilon) = 0 \text{ in } \mathbf{R}^N, i \in \mathcal{I} := \{1, \dots, M\}, \quad (9)$$

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

$$F_i^\epsilon(x, r, p, X) = \max \left\{ \min_{|e| \leq \epsilon} \mathcal{L}^{\alpha_i}(x + e, r_i, p, X); r_i - \mathcal{M}_i r \right\},$$

and \mathcal{L}^α given by (2),

$$\mathcal{M}_i r = \min_{j \neq i} \{r_j + k\}. \quad (10)$$

We give two lemmas proved in [2]:

Lemma 1.8. *Assume (A1) and (A2).*

- *There exists a unique solution v^ϵ of (9) satisfying $|v^\epsilon|_1 \leq C$ where C depends only on λ, K from (A1).*
- *Assume in addition (A3), then for any $\delta > 0$, there are $M \in \mathbf{N}$ and $\{\alpha_i\}_{i=1}^M \subset \mathcal{A}$ such that the solution v_ϵ of (9) satisfies*

$$\max_i |u - v_i^\epsilon|_0 \leq C(\epsilon + k^{\frac{1}{3}} + \delta),$$

where C depends on λ, K from (A1).

A function u can be regularized by

$$\rho_\epsilon * u(x) := \int_{\mathbf{R}^N} u(x - e) \rho_\epsilon(e) de, \quad (11)$$

where ρ_ϵ is the mollifier sequence such that

$$\rho_\epsilon(x) = \frac{1}{\epsilon^N} \rho\left(\frac{x}{\epsilon}\right) \text{ where } \rho \in C^\infty(\mathbf{R}^N), \int_{\mathbf{R}^N} \rho = 1, \text{ and } \text{supp}(\rho) = \bar{B}(0, 1).$$

Lemma 1.9. Assume (A1) and (A2) and define $v_{\epsilon,i} := \rho_\epsilon * v_i^\epsilon$ for $i \in \mathcal{I}$.

- There is a constant C depending only on λ, K from (A1), such that

$$|v_{\epsilon,j} - v_i^\epsilon|_0 \leq C(k + \epsilon) \text{ for } i, j \in \mathcal{I}.$$

- Assume in addition that $\epsilon \leq (4 \sup_i [v_i^\epsilon]_1)^{-1} k$. For every $x \in \mathbf{R}^N$, if $j := \operatorname{argmin}_{i \in \mathcal{I}} v_{\epsilon,i}(x)$, then

$$\mathcal{L}^{\alpha_j}(x, v_{\epsilon,j}(x), \mathcal{D}v_{\epsilon,j}(x), \mathcal{D}^2 v_{\epsilon,j}(x)) \geq 0. \quad (12)$$

We can now give the main result of the paper using the two previous lemmas for an upper bound of the ϵ solution.

Theorem 1.10. Assume (A1), (A2), (S1), (S2), (S3), and (S4). We have the following bounds for u_h a sequence of $\epsilon(\tilde{C}h^r)$ solutions:

$$u - u_h \leq \hat{C}(h^{\min(p,r)} + h^{\min_{i=1,m} \frac{k_i}{i}}),$$

where \hat{C} depends on \tilde{C} .

Besides assume (A3) then there exists \hat{C} depending on \tilde{C} such that :

$$u_h - u \leq \hat{C}(h^{\min(p,r)} + h^{\min_{i=1,m} \frac{k_i}{3i-2}}).$$

Proof. For the lower bound, we follow the Krylov demonstration [18] as done in [1]. First we introduce the solution u^ϵ

$$\max_{|e| \leq \epsilon} [F(x + e, u^\epsilon, \mathcal{D}u^\epsilon, \mathcal{D}^2 u^\epsilon)] = 0 \text{ in } \mathbf{R}^N.$$

The existence a solution u^ϵ in $C^{0,1}(\mathbf{R}^N)$ satisfying $|u^\epsilon| \leq C$ and $|u^\epsilon - u|_0 \leq C\epsilon$ is given by Lemma 2.6 in [1]. Noting that for each $e \leq \epsilon$, $u^\epsilon(\cdot - e)$ satisfies for each function $\phi = \psi(\cdot - e) \in C^2(\mathbf{R}^N)$ and each y where $u^\epsilon(y - e) - \phi(y)$ is maximal

$$\begin{aligned} F(y, u^\epsilon(y - e), \mathcal{D}\phi(y), \mathcal{D}^2\phi(y)) &= F(y, u^\epsilon(y - e), \mathcal{D}\psi(y - e), \mathcal{D}^2\psi(y - e)), \\ &\leq \sup_{|e| \leq \epsilon} F(y, u^\epsilon(y - e), \mathcal{D}\psi(y - e), \mathcal{D}^2\psi(y - e)), \\ &\leq 0, \end{aligned}$$

so $u^\epsilon(\cdot - e)$ is a subsolution of (1).

Then $u^\epsilon(\cdot - e)$ is regularized by

$$u_\epsilon(x) := \int_{\mathbf{R}^N} u^\epsilon(x - e) \rho_\epsilon(e) de.$$

The regularized function u_ϵ is a subsolution of problem (1) as given by Lemma 2.7 in [1]. First use the relation for $m > 0$,

$$F(x, t + m, p, X) \geq F(x, t, p, X) + \lambda m,$$

and the consistency property (S3) to get

$$\begin{aligned} F(y, u_\epsilon(y), \mathcal{D}u_\epsilon(y), \mathcal{D}^2u_\epsilon(y)) &\geq F(y, u_\epsilon(y) - \frac{K}{\lambda} \sum_{i=1}^m h^{k_i} |\mathcal{D}^i u_\epsilon|_0, \mathcal{D}u_\epsilon(y), \mathcal{D}^2u_\epsilon(y)) + K \sum_{i=1}^m h^{k_i} |\mathcal{D}^i u_\epsilon|_0, \\ &\geq S(h, y, u_\epsilon(y) - \frac{K}{\lambda} \sum_{i=1}^m h^{k_i} |\mathcal{D}^i u_\epsilon|_0, [u_\epsilon - \frac{K}{\lambda} \sum_{i=1}^m h^{k_i} |\mathcal{D}^i u_\epsilon|_0]_y). \end{aligned}$$

Then $u_\epsilon - \frac{K}{\lambda} \sum_{i=1}^m |\mathcal{D}^i u_\epsilon|_0$ is a subsolution of equation (4). Using lemma 1.7, and assumption (S4), we get that there exists C such that

$$\begin{aligned} u_\epsilon - u_h &\leq C(|u_\epsilon|_1 h^p + \sum_{i=1}^m |\mathcal{D}^i u_\epsilon|_0 h^{k_i} + h^r), \\ &\leq C(h^{\min(p,r)} + \sum_{i=1}^m \epsilon^{1-i} h^{k_i}). \end{aligned}$$

In the last line we have used that because u^ϵ is bounded uniformly in $C^{0,1}(\mathbf{R}^N)$, u_ϵ is regular and $|\mathcal{D}^n u_\epsilon|_0 \leq C\epsilon^{1-n}$ for $n \geq 1$.

At last using the mollifier properties, the uniform boundedness of u^ϵ in $C^{0,1}$ gives us that $|u_\epsilon - u^\epsilon| \leq C\epsilon$. Besides $|u^\epsilon - u|_0 \leq C\epsilon$ (Lemma 2.6 in [1]) so there exists \hat{C} depending on \tilde{C} such that

$$\begin{aligned} u - u_h &\leq |u - u^\epsilon|_0 + |u^\epsilon - u_\epsilon|_0 + u_\epsilon - u_h, \\ &\leq C(h^{\min(p,r)} + \sum_{i=1}^m \epsilon^{1-i} h^{k_i} + \epsilon). \end{aligned}$$

Choosing $\epsilon = h^{\min_{i=1,m} \frac{k_i}{i}}$ we get the lower bound in the theorem.

For the upper bound we follow the Barles Jakobsen demonstration that we shorten except for points different from the initial proof. We fix a $\delta > 0$, and pick up the corresponding $\{\alpha_i\}_{i \in \mathcal{I}}$ according to (A3). The corresponding solution v^ϵ of (9) exists according to lemma 1.8 and is regularized as in lemma 1.9. We note

$$m := \sup_{y \in \mathbf{R}^N} \{u_h(y) - w(y)\},$$

where $w := \min_{i \in \mathcal{I}} v_{\epsilon,i}$. We approximate m by

$$m_\kappa := \sup_{y \in \mathbf{R}^N} \{u_h(y) - w(y) - \kappa\phi(y)\}, \quad (13)$$

where $\phi(y) = (1 + |y|^2)^{\frac{1}{2}}$. Because of the boundedness and continuity of u_h and w , the maximum is attained at a point x . Because of the definition of x ,

$$m_\kappa := \sup_{y \in \mathbf{R}^N} \{u_h(y) - v_{\epsilon,i}(y) - \kappa\phi(y)\},$$

where $i = \operatorname{argmin}_{i \in \mathcal{I}} v_{\epsilon, i}(x)$.

Taking $\epsilon = (4 \sup_i [v_i^\epsilon]_1)^{-1} k$, from lemma 1.9, the definition of ϕ and (A1) we get

$$\sup_{\alpha \in \mathcal{A}} \mathcal{L}^\alpha(x, (v_{\epsilon, i} + \kappa\phi)(x), \mathcal{D}(v_{\epsilon, i} + \kappa\phi)(x), \mathcal{D}^2(v_{\epsilon, i} + \kappa\phi)(x)) \geq -C\kappa.$$

Then using (S3):

$$-C\kappa \leq S(h, (v_{\epsilon, i} + \kappa\phi)(x), [v_{\epsilon, i} + \kappa\phi]_x) + K \sum_{j=1}^m h^{k_j} |\mathcal{D}^j(v_{\epsilon, i} + \kappa\phi)|_0.$$

Using the properties of the mollified $v_{\epsilon, i}$, and the definition of ϕ :

$$-K \sum_{j=1}^m h^{k_j} \epsilon^{1-j} + \mathcal{O}(\kappa) \leq S(h, (v_{\epsilon, i} + \kappa\phi)(x), [v_{\epsilon, i} + \kappa\phi]_x). \quad (14)$$

Then we use the ϵ monotony property (8), the definition of m_κ , the fact that $v_{\epsilon, i}$ is bounded uniformly by the properties of mollifiers and lemma 1.8 to get

$$\begin{aligned} S(h, (v_{\epsilon, i} + \kappa\phi)(x), [v_{\epsilon, i} + \kappa\phi]_x) &\leq S(h, u_h(x) - m_\kappa, [u_h - m_\kappa]_x) + Kh^p |v_{\epsilon, i} + \kappa\phi|_1, \\ &\leq -\lambda m_\kappa + S(h, u_h(x), [u_h]_x) + Ch^p(1 + \kappa), \\ &\leq \tilde{C}(1 + \kappa)h^{\min(r, p)} - \lambda m_\kappa. \end{aligned} \quad (15)$$

Using (14) and (15) we get

$$\lambda m_\kappa \leq \tilde{C}(1 + \kappa)h^{\min(r, p)} + K \sum_{j=1}^m h^{k_j} \epsilon^{1-j} + \mathcal{O}(\kappa).$$

An estimate of m is obtained by letting κ goes to 0. Then for any $y \in \mathbf{R}^N$,

$$\begin{aligned} u_h(y) - u(y) &\leq u_h(y) - v_{\epsilon, i}(y) + v_{\epsilon, i}(y) - u(y), \\ &\leq m + v_{\epsilon, i}(y) - u(y). \end{aligned}$$

Using the lemma 1.8 and 1.9 we get

$$u_h(y) - u(y) \leq \hat{C}(h^{\min(r, p)} + \sum_{i=1}^m h^{k_i} \epsilon^{1-i} + \epsilon + k + k^{\frac{1}{3}} + \delta),$$

with \hat{C} depending on \tilde{C} and uniform in y . The conclusion is obtained by taking $\epsilon = h^{\min_{i=1, m} \frac{3k_i}{3i-2}}$, remembering that $k = \epsilon \sup_i [v_i^\epsilon]_1$ and letting δ going to 0. \square

2 Some numerical schemes

In this section we take the notations and we will follow arguments similar to [1, 2]. We give the notations used for the discretization and interpolation used by the scheme. A thorough study of interpolation method for time dependent HJB equation can be found in [19]. All the schemes defined in [19] can be used : it includes truncated Lagrangian interpolators, classical cubic spline truncated interpolators, the monotone cubic spline first defined in [17]. In the sequel we focus on the truncated Gauss Lobatto Legendre interpolators which are the most effective according to [19].

A spatial discretization Δx of the problem being given, in the sequel a mesh $\hat{M}_{\bar{i}}$ corresponds to the hyper-cube $[i_1\Delta x, (i_1 + 1)\Delta x] \times \dots \times [i_N\Delta x, (i_N + 1)\Delta x]$ with $\bar{i} = (i_1, \dots, i_N) \in \mathbf{Z}^N$. For Gauss Lobatto Legendre quadrature grid $(\xi_i)_{i=1, \dots, M+1} \in [-1, 1]$, and for a mesh \bar{i} , the point $y_{\bar{i}, \tilde{j}}$ with $\tilde{j} = (j_1, \dots, j_N) \in [1, M + 1]^N$ will have the coordinate $(\Delta x(i_1 + 0.5(1 + \xi_{j_1})), \dots, \Delta x(i_N + 0.5(1 + \xi_{j_N})))$. We denote $X_{\Delta x, M} := (y_{\bar{i}, \tilde{j}})_{\bar{i}, \tilde{j}}$ the set of all the grids points on the whole domain.

We notice that each mesh $\hat{M}_{\bar{i}}$ has a constant volume Δx^N , so we have the following relation for all $x \in \mathbf{R}^N$:

$$\min_{\bar{i}, \tilde{j}} |x - y_{\bar{i}, \tilde{j}}| \leq C\Delta x. \quad (16)$$

We introduce $I_{\Delta x, M}$ the Lagrange interpolator associated to the Gauss Lobatto Legendre quadrature. We recall that in one dimension, the Gauss Lobatto Legendre Lagrange interpolator I_M on $[-1, 1]$ is given by (see for example [21, 22, 23]):

$$\begin{aligned} I_M(f) &= \sum_{k=1}^{M+1} \tilde{f}_k L'_k(x), \\ \tilde{f}_k &= \frac{1}{\gamma_k} \sum_{i=1}^{M+1} \rho_i f(\eta_i) L'_k(\eta_i), \\ \gamma_k &= \sum_{i=1}^{M+1} L'_k(\eta_i)^2 \rho_i, \end{aligned}$$

where the functions L_N satisfy the recurrence

$$\begin{aligned} (N + 1)L_{N+1}(x) &= (2N + 1)xL_N(x) - NL_{N-1}(x), \\ L_0 &= 1, \quad L_1 = x, \end{aligned}$$

$\eta_1 = -1, \eta_{M+1} = 1$, the η_i ($i = 2, \dots, M$) are the zeros of L'_M and the eigenvalues

of the matrix P

$$P = \begin{pmatrix} 0 & \gamma_1 & \dots & 0 & 0 \\ \gamma_1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \gamma_{M-2} \\ 0 & 0 & \dots & \gamma_{M-2} & 0 \end{pmatrix},$$

$$\gamma_n = \frac{1}{2} \sqrt{\frac{n(n+2)}{(n+\frac{1}{2})(n+\frac{3}{2})}}, 1 \leq n \leq M-2,$$

and the weights satisfies

$$\rho_i = \frac{2}{(M+1)ML_M^2(\eta_i)}, 1 \leq i \leq M+1.$$

The interpolator $I_{\Delta x, M}$ on a mesh $\hat{M}_{\bar{i}}$ is obtained by first rescaling I_M and by tensorization.

On a mesh $\hat{M}_{\bar{i}}$ and for a point x in this mesh, we note $\underline{v}_{\bar{i}} = \min_{\bar{j}} v(y_{\bar{i}, \bar{j}})$, $\bar{v}_{\bar{i}} = \max_{\bar{j}} v(y_{\bar{i}, \bar{j}})$. We introduce the following truncated operator:

$$\hat{I}_{\Delta x, M}(v) = \underline{v}_{\bar{i}} \vee I_{\Delta x, M}(v) \wedge \bar{v}_{\bar{i}},$$

where \wedge denotes the minimum and \vee the maximum.

We first give some properties associated to the truncated interpolation operator.

Lemma 2.1. *The interpolator $\hat{I}_{\Delta x, M}$ satisfies*

$$(\hat{I}_{\Delta x, M} f)(x) = \sum_{\bar{j}} (w_{\bar{i}, \bar{j}}(f))(x) f(y_{\bar{i}, \bar{j}})$$

$$\sum_{\bar{j}} (w_{\bar{i}, \bar{j}}^h(f))(x) = 1,$$

and the positive weights $w_{\bar{i}, \bar{j}}^h(f)$ are functions of f .

Proof. Because of the truncation for each point x of a mesh $\hat{M}_{\bar{i}}$, we have

$$(\hat{I}_{\Delta x, M}(f))(x) = \underline{w}_{\bar{i}}^h(f)(x) \underline{f}_{\bar{i}} + \bar{w}_{\bar{i}}^h(f)(x) \bar{f}_{\bar{i}}.$$

If $\underline{f}_{\bar{i}} \leq \hat{I}_{\Delta x, M}(f)(x) \leq \bar{f}_{\bar{i}}$ then

$$\underline{w}_{\bar{i}}^h(f)(x) = \frac{\hat{I}_{\Delta x, M}(f)(x) - \bar{f}_{\bar{i}}}{\underline{f}_{\bar{i}} - \bar{f}_{\bar{i}}},$$

$$\bar{w}_{\bar{i}}^h(f)(x) = 1 - \underline{w}_{\bar{i}}^h(f)(x),$$
(17)

If $f_{\bar{i}} > \hat{I}_{\Delta x, M}(f)(x)$,

$$\begin{aligned}\bar{w}_i^h(f)(x) &= 0, \\ \underline{w}_i^h(f)(x) &= 1.\end{aligned}$$

Otherwise

$$\begin{aligned}\underline{w}_i^h(f)(x) &= 0, \\ \bar{w}_i^h(f)(x) &= 1.\end{aligned}$$

The weights associated to non extremal points are taken equal to 0. \square
Then we add a result for the interpolation error :

Lemma 2.2. • For each K -Lipschitz bounded function f :

$$|\hat{I}_{\Delta x, M}(f) - f|_0 \leq K(\Delta x).$$

- Suppose $M \geq 2$, for each f \mathbf{R} value function defined on \mathbf{R}^N and twice differentiable, there exists C such that:

$$|\hat{I}_{\Delta x, M}(f) - f|_0 \leq C\Delta x^2 |D^2 f|_0.$$

- For $m \in \mathbf{R}^N$,

$$\hat{I}_{\Delta x, M}(f + m)(x) = \hat{I}_{\Delta x, M}(f) + m.$$

Proof. First because of the truncation, continuity of $\hat{I}_{\Delta x, M}(f)$, for each $x \in \mathbf{R}^N$, $x \in \hat{M}_{\bar{i}}$, there exists $\tilde{x} \in \hat{M}_{\bar{i}}$ such that $\hat{I}_{\Delta x, M}(f)(x) = f(\tilde{x})$. We then use the Lipschitz property of f and (16) to get the result.

When no truncation is achieved, we have a rate of convergence in $O(\Delta x^{M+1})$.
When the truncation is achieved, for example $\hat{I}_{\Delta x, M}(f)(x) = \bar{f}_{\bar{i}}$, then

$$I_{\Delta x, 1} \leq \hat{I}_{\Delta x, M}(f)(x) \leq I_{\Delta x, M}(f)(x),$$

where $I_{\Delta x, 1}$ correspond to the linear interpolator and then

$$|\hat{I}_{\Delta x, M}(f)(x) - f(x)| \leq \max(|I_{\Delta x, M}(f)(x) - f(x)|, |I_{\Delta x, 1}(f)(x) - f(x)|),$$

so the rate of convergence remains at least equal to 2.

The third point is easily check by noticing that $I_{\Delta x, M}$ is a Lagrange interpolator so linear, that $I_{\Delta x, M}(m) = m$ and that the truncation operator tr satisfies $tr(f + m) = tr(f) + m$. \square

Remark 2.3. Some effective interpolation methods such as ENO, WENO can be used for interpolation [24, 25, 26] while solving Hamilton Jacobi equations but they are not proved convergent. We will show that the previously defined interpolator ensures convergence for Semi Lagrangian schemes and for some Finite Difference schemes but at a rate not better than linear interpolator. The interest of such interpolators will be checked numerically on some examples in the last section. Besides, they are easy to implement independently of the dimension of the problem.

2.1 A Camilli Falcone style scheme

The first scheme we study is a modification of Camilli Falcone scheme [9] where the linear interpolator $I_{\Delta x,1}$ is replaced by a potentially high order interpolator $\hat{I}_{\Delta x,M}$ with $M > 1$. We begin by defining the monotone operator [1, 9] \hat{S} which is the Lagrangian scheme without interpolation. First for any bounded continuous function ϕ , $x, y, z \in \mathbf{R}^N$, we set $[\phi]_x(z) = \phi(x+z)$ and

$$\hat{S}(h, y, t, [\phi]_x) = \sup_{\alpha \in \mathcal{A}} \left\{ -\frac{1}{h} (G(h, \alpha, y, [\phi]_x) - t) + c^\alpha(y)t - f^\alpha(y) \right\},$$

$$G(h, \alpha, y, [\phi]_x) = \frac{1 - hc^\alpha(y)}{2P} \sum_{i=1}^P \left([\phi]_x(hb^\alpha(y) + \sqrt{hP}\sigma_i^\alpha(y)) + [\phi]_x(hb^\alpha(y) - \sqrt{hP}\sigma_i^\alpha(y)) \right),$$

where σ_i^α is the i -th column of σ .

The semi discretized scheme S is defined as follows for y a quadrature point.

$$S(h, y, t, [\phi]_x) = \hat{S}(h, y, t, [\hat{I}_{\Delta x,M}\phi]_x).$$

So the discretized problem leads to find U function on $X_{\Delta x,M}$ such that

$$|\hat{S}(h, y, U(y), [\hat{I}_{\Delta x,M}U]_y)| \leq \epsilon(h, \Delta x), \text{ for } y \in X_{\Delta x,M}. \quad (18)$$

We first recall some results on the solution associated to the semi discretized scheme that can be found in [9, 1]

Proposition 2.4. *Assume that (A1), (A2) hold. Then there exists a unique bounded function v_h uniformly in $C^{0,1}(\mathbf{R}^N)$ satisfying*

$$\hat{S}(h, x, v_h(x), [v_h]_x) = 0 \text{ for } x \in \mathbf{R}^N. \quad (19)$$

We next prove that the scheme S satisfies the first assumptions of the article :

Proposition 2.5. *Assume (A1) hold and that $\Delta x = h^q$. Then the scheme (18) satisfies assumptions (S1), (S2), (S3) with $k_2 = \min(2q - 1, 1)$, $k_4 = 1$, $p = q - 1$.*

Proof. First assumption (S2) follows easily from (A1). (S3) follows easily using lemma 2.2: for v regular the consistency error is bounded by

$$\epsilon(h, \Delta x) = C(h|\mathcal{D}^4 v|_0 + h|\mathcal{D}^2 v|_0 + \frac{\Delta x^2}{h}|\mathcal{D}^2 v|_0). \quad (20)$$

Using (A1), and lemma 2.2, for $m > 0$:

$$G(h, \alpha, y, [\hat{I}_{\Delta x,M}(\phi_x + m)]_x) = G(h, \alpha, y, [\hat{I}_{\Delta x,M}(\phi)]_x) + (1 - hc^\alpha(y))m,$$

so

$$S(h, y, t + m, [I_{\Delta x,M}(\phi + m)]_x) \geq S(h, y, t, [I_{\Delta x,M}(\phi)]_x) + 2\lambda m,$$

and property (8) is checked.

Suppose $v \in C_b(\mathbf{R}^N)$, $w \in C^{0,1}(\mathbf{R}^N)$, $v \geq w$. If $x \in \hat{M}_{\bar{i}}$ is such that $\hat{I}_{\Delta x, M}(v)(x) \leq \hat{I}_{\Delta x, M}(w)(x)$ let's introduce $v(x_{\bar{i}, \bar{i}}) = \min_{\bar{j}} v(x_{\bar{i}, \bar{j}})$. It satisfies $v(x_{\bar{i}, \bar{i}}) \leq \hat{I}_{\Delta x, M}(v)(x)$ so using lemma 2.2

$$\begin{aligned} \hat{I}_{\Delta x, M}(v)(x) - \hat{I}_{\Delta x, M}(w)(x) &\geq v(x_{\bar{i}, \bar{i}}) - \hat{I}_{\Delta x, M}(w)(x), \\ &\geq v(x_{\bar{i}, \bar{i}}) - w(x_{\bar{i}, \bar{i}}) - |w|_1 \Delta x \sqrt{N}, \geq -|w|_1 \Delta x \sqrt{N}. \end{aligned}$$

So for all $x \in \mathbf{R}^N$

$$\hat{I}_{\Delta x, M}(v)(x) \geq \hat{I}_{\Delta x, M}(w)(x) - |w|_1 \Delta x \sqrt{N}, \quad (21)$$

and

$$G(h, \alpha, y, [\hat{I}_{\Delta x, M}(v)]_x) \geq G(h, \alpha, y, [\hat{I}_{\Delta x, M}(w)]_x) - (1 + h|c^\alpha(y)|_0) |w|_1 \Delta x \sqrt{N}.$$

So

$$S(h, y, t, [\hat{I}_{\Delta x, M}(v)]_x) \leq S(h, y, t, [\hat{I}_{\Delta x, M}(w)]_x) + (1 + h|c^\alpha(y)|_0) |w|_1 \frac{\sqrt{N} \Delta x}{h}.$$

Similarly if $v \in C^{0,1}(\mathbf{R}^N)$, $w \in C_b(\mathbf{R}^N)$, $v \geq w$, noting $w(x_{\bar{i}, \bar{i}}) = \max_{\bar{j}} w(x_{\bar{i}, \bar{j}})$, and using lemma 2.2

$$\begin{aligned} \hat{I}_{\Delta x, M}(v)(x) - \hat{I}_{\Delta x, M}(w)(x) &\geq \hat{I}_{\Delta x, M}(v)(x) - w(x_{\bar{i}, \bar{i}}), \\ &\geq v(x_{\bar{i}, \bar{i}}) - w(x_{\bar{i}, \bar{i}}) - |v|_1 \Delta x \sqrt{N} \geq -|v|_1 \Delta x \sqrt{N}, \end{aligned}$$

so

$$\hat{I}_{\Delta x, M}(v)(x) \geq \hat{I}_{\Delta x, M}(w)(x) - |v|_1 \Delta x \sqrt{N}, \quad (22)$$

and

$$S(h, y, t, [\hat{I}_{\Delta x, M}(v)]_x) \leq S(h, y, t, [\hat{I}_{\Delta x, M}(w)]_x) + (1 + h|c^\alpha(y)|_0) |v|_1 \frac{\sqrt{N} \Delta x}{h},$$

so that the (6) and (7) properties are checked. \square

We need to prove that we can construct an approximate solution of the discretized problem. We introduce the operator T defined for U a function on $X_{\Delta x, M}$:

$$(T_{h, \Delta x} U)(x) = \inf_{\alpha \in \mathcal{A}} \{(1 - hc^\alpha(x))(\Pi_{\Delta x, h}(U))(x) + hf^\alpha(x)\} \text{ for } x \in X_{\Delta x, M},$$

where the operator $\Pi_{\Delta x, h}$ is

$$(\Pi_{\Delta x, h} U)(x) = \frac{1}{2P} \sum_{i=1}^{2P} \left((\hat{I}_{\Delta x, M} U)(x + hb^\alpha(x) + \sqrt{Ph} \sigma_i^\alpha(x)) + (\hat{I}_{\Delta x, M} U)(x + hb^\alpha(x) - \sqrt{Ph} \sigma_i^\alpha(x)) \right).$$

We then recursively define $T_{h, \Delta x}^s$ for $s \in \mathbf{N}$ and $s \geq 2$ by

$$(T_{h, \Delta x}^s U)(x) = (T_{h, \Delta x}(T_{h, \Delta x}^{s-1} U))(x). \quad (23)$$

Proposition 2.6. *Assume (A1), (A2) hold. Suppose that $\Delta x = h^q$ with $q > 2$. There exists $s \in \mathbf{N}$ depending on h and C independent of h such that $u_h = \hat{I}_{\Delta x, M}(T_{h, \Delta x}^s \mathbf{0})$ is an $\epsilon(C h^{q-2})$ solution of scheme S .*

Proof. Note v_h the unique solution given of scheme (19).

For $x \in X_{\Delta x, M}$, U a function on $X_{\Delta x, M}$, using $|\inf \dots - \inf \dots| \leq \sup |\dots - \dots|$

$$\begin{aligned} |(T_{h, \Delta x} U)(x) - v_h(x)| &\leq (1 - \lambda h) \sum_{i=1}^{2P} \left(\sup_{\alpha \in \mathcal{A}} |(\hat{I}_{\Delta x, M} U)(x + hb^\alpha(x) + \sqrt{hP}\sigma_i^\alpha(x)) - v_h(x + hb^\alpha(x) + \sqrt{hP}\sigma_i^\alpha(x))| \right. \\ &\quad \left. + |(\hat{I}_{\Delta x, M} U)(x + hb^\alpha(x) - \sqrt{hP}\sigma_i^\alpha(x)) - v_h(x + hb^\alpha(x) - \sqrt{hP}\sigma_i^\alpha(x))| \right), \end{aligned}$$

then notice that for $x \in \hat{M}_{\bar{i}}$

$$\begin{aligned} |\hat{I}_{\Delta x, M}(U)(x) - v_h(x)| &= \left| \sum_{\bar{j}} w_{\bar{i}, \bar{j}}(U)(x) (U(x_{\bar{i}, \bar{j}}) - v_h(x)) \right| \\ &\leq \sum_{j, l} w_{\bar{i}, \bar{j}}(U)(x) |U(x_{\bar{i}, \bar{j}}) - v_h(x)|, \\ &\leq \sup_{\bar{j}} |U(x_{\bar{i}, \bar{j}}) - v_h(x)|, \\ &\leq \sup_{\bar{j}} |U(x_{\bar{i}, \bar{j}}) - v_h(x_{\bar{i}, \bar{j}})| + \Delta x \sqrt{N} |v_h|_1, \\ &\leq |U - v_h|_0 + C \Delta x, \end{aligned} \tag{25}$$

where we have use the uniform boundedness of v_h in $C^{0,1}(\mathbf{R}^N)$. Gathering (24) and (25) we get

$$|(T_{h, \Delta x} U)(x) - v_h(x)| \leq (1 - \lambda h) (|U - v_h|_0 + C \Delta x).$$

Iterating we find that

$$|(T_{h, \Delta x}^k \mathbf{0})(x) - v_h(x)| \leq (1 - \lambda h)^k |v_h|_0 + C \frac{\Delta x}{h}. \tag{26}$$

Taking $k = \min(i \in \mathbf{N} \text{ such that } i \geq \frac{(q-1)\log(h)}{\log(1-\lambda h)})$, using the fact that $\Delta x = h^q$, we get that

$$|(T_{h, \Delta x}^k \mathbf{0})(x) - v_h(x)| \leq C h^{q-1}. \tag{27}$$

Let's prove that $u_h = \hat{I}_{\Delta x, N}(T_{h, \Delta x}^k \mathbf{0})$ is an $\epsilon(h^{q-2})$ solution of the scheme S . As in (25)

$$\begin{aligned} |\hat{I}_{\Delta x, M}(T_{h, \Delta x}^k \mathbf{0})(x) - v_h(x)| &\leq |T_{h, \Delta x}^k \mathbf{0} - v_h|_0 + C \Delta x, \\ &\leq C h^{q-1}, \end{aligned}$$

where we have used (27), so

$$|u_h(x) - v_h(x)| \leq C h^{q-1}. \tag{28}$$

Then using (A1), the fact that $u_h = \hat{I}_{h,\Delta x} u_h$, and (28)

$$\begin{aligned} |S(h, x, u_h(x), [u_h]_x)| &= |\hat{S}(h, x, u_h(x), [u_h]_x) - \hat{S}(h, x, v_h(x), [v_h]_x)|, \\ &\leq \sup_{\alpha \in \mathcal{A}} \left| \frac{1}{h} [(G(h, \alpha, x, [u_h]_x) - u_h(x)) - (G(h, \alpha, x, [v_h]_x) - v_h(x))] \right|, \\ &\leq ch^{q-2}. \end{aligned}$$

□

Proposition 2.7. *Assume (A1), (A2), (A3) hold. Suppose u_h has been constructed as in proposition 2.6, the previous developed framework gives us that we can find q such that*

$$|u - u_h| < Ch^{\frac{1}{10}}.$$

Proposition 2.8. *Assume (A1), (A2) hold. Taking $q \geq \frac{5}{4}$, u_h being constructed as in proposition 2.6, we have*

$$|u - u_h| < Ch^{\frac{1}{4}}.$$

Proof. Because we have a bound on $|v_h - u|_0$ in $O(h^{\frac{1}{4}})$ (see [1]), a direct use of (28) shows

$$\begin{aligned} |u - u_h| &\leq |u - v_h|_0 + |v_h - u_h|_0, \\ &\leq c(h^{\frac{1}{4}} + h^{q-1}), \end{aligned} \tag{29}$$

giving the result. □

Remark 2.9. *Propositions 2.7 and 2.8 are generalization of the results of theorem 5.1 and 6.1 in [27] in the near monotone case.*

Remark 2.10. *The bound given by the framework is not tight because we introduced a switching system in our approach and not in the latter proposition.*

2.2 Finite Difference scheme

In this part we suppose that a^α is independent of x . The matrix a^α can be written $P^\alpha D^\alpha (P^\alpha)^t$ where P^α is a unitary matrix with i -th columns ξ_i^α and D^α is a diagonal $D_{i,j}^\alpha = \delta_{i,j} d_i^\alpha \geq 0$. We note $(b^\alpha)^+(x)$ the vector such that $(b^\alpha)_i^+(x) = \max(0, b^{\alpha i}(x))$ and $(b^\alpha)_i^-(x) = \max(0, -b^{\alpha i}(x))$. The operator \mathcal{L}^α

can be discretized for a regular function u using two parameters h and \hat{h} by

$$\begin{aligned} \mathcal{L}^\alpha(x, u, \mathcal{D}u, \mathcal{D}^2u) &\simeq \sum_{i=1}^N \frac{d_i^\alpha}{2} \frac{2u(x) - u(x - h\xi_i^\alpha) - u(x + h\xi_i^\alpha)}{h^2} \\ &\quad - \sum_{i=1}^N (b^\alpha)_i^+(x) \frac{u(x + \hat{h}e_i) - u(x)}{\hat{h}} + \sum_{i=1}^N (b^\alpha)_i^-(x) \frac{u(x) - u(x - \hat{h}e_i)}{\hat{h}} \\ &\quad + c^\alpha(x)u(x) - f(x), \end{aligned}$$

where $(e_i)_{i=1, N}$ is the canonical basis in \mathbf{R}^N . We suppose that the equation has been normalized such that

$$\sup_\alpha \sum_i d_i^\alpha + |b_i^\alpha(x)| \leq 1.$$

This is always possible because of (A1) and noting that $\sum_i d_i^\alpha = \text{tr}[a^\alpha]$. For $z \in \mathcal{Q}^\alpha = \{0, \{\xi_i^\alpha\}_{i=1, N}, \{e_i\}_{i=1, N}\}$ we define the transition probability (see [4]) for $\frac{h^2}{\hat{h}} \leq 1$

$$\begin{aligned} p^\alpha(x, x) &= 1 - \sum_i \left\{ d_i^\alpha + |b_i^\alpha(x)| \frac{h^2}{\hat{h}} \right\}, \\ p^\alpha(x, x \pm \hat{h}e_i) &= (b^\alpha)_i^\pm(x) \frac{h^2}{\hat{h}}, \\ p^\alpha(x, x \pm h\xi_i^\alpha) &= \frac{d_i^\alpha}{2}. \end{aligned}$$

For any bounded continuous function ϕ , $x, y, z \in \mathbf{R}^N$, we set $[\phi]_x(z) = \phi(x + z)$ and define the operator \hat{S}

$$\hat{S}(h, y, t, [\phi]_x) = \sup_{\alpha \in \mathcal{A}} \left\{ -\frac{1}{h^2} \left[\sum_{z \in \mathcal{Q}^\alpha} p^\alpha(y, y + z) [\phi]_x(z) - t \right] + c^\alpha(y)t - f^\alpha(y) \right\}.$$

First we are interested in getting a solution v_h of

$$\hat{S}(h, x, v_h(x), [v_h]_x) = 0, x \in \mathbf{R}^N, \quad (30)$$

in $C^{0,1}(\mathbf{R}^N)$ and get a convergence bound of $|u - v_h|$.

We introduce another assumption stronger than assumption (A2):

Assumption (A4). *The constant λ in (A1) satisfies $\lambda > \max\left(\sup_\alpha \frac{1}{2}[\sigma^\alpha]_1^2 + [b^\alpha]_1, 2\sqrt{N} \sup_\alpha [b^\alpha]_1\right)$.*

By using [1] we get the following proposition

Proposition 2.11. *Under assumptions (A1), (A4) the operator \hat{S} is monotone, consistent, and there exists a unique solution v_h bounded uniformly in $C^{0,1}(\mathbf{R}^N)$ of equation (30) and u solution of (1) satisfies*

$$|u - v_h| \leq Ch^{\frac{1}{2}}. \quad (31)$$

Remark 2.12. We imposed that the a^α is independent of x to get a solution v_h in $C^{0,1}$ which is necessary for us in order to satisfy (S1). Without this assumption we can't prove that v_h is in $C^{0,1}$: only L_∞ bounds on the error are available (see [2]). As an alternative it would have been possible to use Bonnans/Krylov discretization [8, 7] with Krylov convergence results [8] with bounds for v_h in $C^{0,1}(\mathbf{R}^N)$.

Remark 2.13. It is possible to handle certain situations where a^α is not constant in x using Remark 4.9 in [1].

We introduce the operator

$$S(h, y, t, [\phi]_x) = \hat{S}(h, y, t, [\hat{I}_{\Delta x, M} \phi]_x). \quad (32)$$

We give the convergence results obtained with the Finite Difference scheme :

Proposition 2.14. Assume (A1) hold and that $\Delta x = h^q$, $\hat{h} = h$. Then the scheme (32) satisfies assumptions (S1), (S2), (S3) with $k_2 = \min(1, 2q - 2)$, $k_4 = 2$, $p = q - 2$

Proof. First assumption (S2) follows easily from (A1). (S3) follows easily using lemma 2.2: for v regular the consistency error is bounded by

$$\epsilon(h, \Delta x) = C(h^2 |\mathcal{D}^4 v|_0 + \hat{h} |\mathcal{D}^2 v|_0 + \Delta x^2 (\frac{1}{h^2} + \frac{1}{h}) |\mathcal{D}^2 v|_0). \quad (33)$$

Using the fact that $\sum_{z \in \mathcal{Q}^\alpha} p^\alpha(y, y + z) = 1$ with positive weights,

$$S(h, y, t + m, [I_{\Delta x, M}(\phi + m)]_x) = S(h, y, t, [I_{\Delta x, M}(\phi)]_x),$$

and property (8) is checked.

Suppose $v \in C_b(\mathbf{R}^N)$, $w \in C^{0,1}(\mathbf{R}^N)$, $v \geq w$, using estimate (21), the fact that $\sum_{z \in \mathcal{Q}^\alpha} p^\alpha(y, y + z) = 1$ with positive weights,

$$S(h, y, t, [\hat{I}_{\Delta x, M}(v)]_x) \leq S(h, y, t, [\hat{I}_{\Delta x, M}(w)]_x) + |w|_1 \frac{\sqrt{N} \Delta x}{h^2}.$$

Similarly if $v \in C^{0,1}(\mathbf{R}^N)$, $w \in C_b(\mathbf{R}^N)$, $v \geq w$ using estimate (22),

$$S(h, y, t, [\hat{I}_{\Delta x, M}(v)]_x) \leq S(h, y, t, [\hat{I}_{\Delta x, M}(w)]_x) + |v|_1 \frac{\sqrt{N} \Delta x}{h^2},$$

so that the (6) and (7) properties are checked. \square

We then have to check that we can find an ϵ solution u_h satisfying (4). We introduce the operator defined for U a function on $X_{\Delta x, M}$: for $x \in X_{\Delta x, M}$

$$(T_{h, \Delta x} U)(x) = \inf_{\alpha \in \mathcal{A}} \left\{ \frac{1}{1 + h^2 c^\alpha(x)} \sum_{z \in \mathcal{Q}} p^\alpha(x, x + z) \hat{I}_{\Delta x, M}(U)(x + z) + h^2 f^\alpha(x) \right\},$$

and $T_{h, \Delta x}^s$ operator is still defined by equation (23).

Proposition 2.15. *Assume (A1), (A4) hold. Suppose that $\Delta x = h^q$ with $q > 2$, $\hat{h} = h$. There exists $s \in \mathbf{N}$ depending on h and C independent of h such that $u_h = \hat{I}_{h,\Delta x}(T_{h,\Delta x}^s 0)$ is an $\epsilon(C h^{q-4})$ solution of scheme S .*

Proof. The proof is similar to the one of proposition 2.6. We first prove that

$$|(T_{h,\Delta x}^k 0)(x) - v_h(x)| \leq \frac{1}{(1 + \lambda h^2)^k} |v_h|_0 + C \frac{\Delta x}{h^2}. \quad (34)$$

Taking $k = \min(i \in \mathbf{N} \text{ such that } i \geq -\frac{(q-2)\log(h)}{\log(1+\lambda h^2)})$, using the fact that $\Delta x = h^q$, we get that

$$|(T_{h,\Delta x}^k 0)(x) - v_h(x)| \leq C h^{q-2}. \quad (35)$$

Let's prove that $u_h = \hat{I}_{\Delta x, N}(T_{h,\Delta x}^k 0)$ is an $\epsilon(h^{q-4})$ solution of the scheme S . As in (25),

$$\begin{aligned} |\hat{I}_{\Delta x, M}(T_{h,\Delta x}^k 0)(x) - v_h(x)| &\leq |T_{h,\Delta x}^k 0 - v_h|_0 + C \Delta x \\ &\leq C h^{q-2}, \end{aligned}$$

where we have used (35), so

$$|u_h(x) - v_h(x)| \leq C h^{q-2}. \quad (36)$$

Then using (A1), (36), the relation $|\sup \dots - \sup \dots| \leq \sup |\dots - \dots|$, the fact that the probabilities are between 0 and 1 with sum equal to 1 and the fact that $u_h = \hat{I}_{h,\Delta x} u_h$:

$$\begin{aligned} |S(h, x, u_h(x), [u_h]_x)| &= |\hat{S}(h, x, u_h(x), [u_h]_x) - \hat{S}(h, x, v_h(x), [v_h]_x)| \\ &\leq \frac{1}{h^2} |u_h - v_h|_0 \\ &\leq c h^{q-4}. \end{aligned} \quad (37)$$

□

Proposition 2.16. *Assume (A1), (A4), (A3) hold. Constructing an ϵ solution of (4) with the u_h given by proposition 2.15, we get*

$$|u - u_h|_0 \leq h^{\frac{1}{5}}. \quad (38)$$

with q above 21/5.

Proof. The rate of convergence and the value $q = \frac{21}{5}$ is due to a direct use of theorem 1.10. □

Proposition 2.17. *Assume (A1), (A4) hold. Using u_h given by proposition 2.15, we get*

$$|u - u_h|_0 \leq h^{\frac{1}{2}}. \quad (39)$$

Proof. This is a direct use of (36) with the rate of convergence $|u - v_h|_0 \leq h^{\frac{1}{2}}$ of proposition 2.11 taking $q = \frac{5}{2}$. □

3 Numerical tests

The theoretical bounds obtained in the previous section are not better than the ones obtained with linear interpolation. The interest of this approximation relies on the fact that where the solution is smooth we expect that the solution won't be truncated and that the consistency error will be far better than the theoretical one. All case treated are two dimensional cases. For the first three examples, we only give results for the Semi-Lagrangian scheme because the Finite Difference scheme developed coincides with the classical Finite Difference. The domain linked to the resolution of equation (1) will be noted Q , $\mathbb{1}$ is the diagonal unitary matrix and $\mathbf{1}$ is the vector with 1 components. For all Semi-Lagrangian Schemes we choose $h = 0.0002$. We discretized the one dimensional space of controls \mathcal{A} with 2000 controls. The software is parallelized with 48 cores as explained in [19]. The interpolation used are either linear (2 points per mesh in each direction, monotone scheme), or quadratic (3 points per mesh in each direction), or Cubic (4 points per mesh in each direction). On all the cases and all the tests the fixed point iteration is converging but it can be very slow especially for Finite Differences. The maximum number of iterations is taken equal to 100000 and no acceleration was used. The criterion for stopping between iteration i and $i + 1$ was $|u^{i+1} - u^i| \leq 10^{-7}$. The different schemes are stable and numerically convergent. The empirical rate of convergence is not regular so not reported. In the table Err indicate the sup of the error for a given discretization given by NbM (number of nodes in each direction), ItN gives the number of fixed point iteration used and Time the time needed to compute in seconds.

3.1 A first regular test case

The solution of this test case is given by

$$u(x, y) = \sin(\pi x) \sin(\pi y).$$

The coefficients are given by

$$c_a(t, x) = C, \quad \sigma_a(t, x) = \sigma a \mathbf{1}, \quad b^\alpha = b \mathbf{1},$$

and the function f^α is given by

$$f^\alpha(x, y) = (C + \pi^2 \sigma^2 \mathbf{1}_{u(x, y) > 0}) u(x, y) - b\pi(\cos(\pi x) \sin(\pi y) + \sin(\pi x) \cos(\pi y)).$$

The numerical coefficients are

$$Q = [0, 2]^2, \quad b = 0.3, \quad c = 0.55, \quad \sigma = 1, \quad A = [0, 1],$$

and the boundary condition is a Dirichlet one with 0 value. Results are given in table 1 and this first regular case clearly indicates that the quadratic approximation is by far the most efficient interpolation : even on this regular case the use of a cubic interpolator doesn't decrease the error with a CPU time multiplied at least three-fold. In fact it is rapidly converging to the h discretized operator so that the interpolation error becomes negligible for a number of mesh equal to 80.

Table 1: Test case 1

NbM	LINEAR			QUADRATIC			CUBIC		
	Err	ItN	Time	Err	ItN	Time	Err	ItN	Time
10	1.169	25900	138	0.051	29483	919	0.183	45108	5182
20	1.028	26555	468	0.0065	29398	3619	0.0083	29486	10770
40	0.758	27211	1879	0.0012	29485	12633	0.0011	29706	43371
80	0.243	28076	6923	0.0003	29621	50988	0.0003	29788	175163
160	0.103	28620	27762	0.0003	29748	19517			
320	0.018	29282	108000						

3.2 A second regular problem

The solution is here again

$$u(x, y) = \sin(\pi x) \sin(\pi y),$$

$$c_a(t, x) = C, \quad \sigma_a(t, x) = \sigma a \mathbb{1}, \quad b^\alpha = b(a, \sqrt{1 - a^2}), \quad a \in [\underline{a}, \bar{a}].$$

Noting

$$\tilde{a} = \frac{\sin(\pi y) \cos(\pi x)}{\sqrt{\sin(\pi x)^2 \cos(\pi y)^2 + \cos(\pi x)^2 \sin(\pi y)^2}},$$

$$\phi(a) = a \sin(\pi y) \cos(\pi x) + \sqrt{1 - a^2} \sin(\pi x) \cos(\pi y),$$

for $b \leq 0$, the function f^α is here given by

$$f^\alpha(x, y) = (C + \pi^2 \sigma^2) u(x, y) - b \pi K,$$

where K is the maximum of $\phi(\underline{a})$, $\phi(\bar{a})$ and $\phi(\tilde{a})$ conditionally to $\underline{a} \leq \tilde{a} \leq \bar{a}$. We take $\underline{a} = -1$, $\bar{a} = 1$, $\sigma = 1$, $C = 0.6$, $b = -1$, $Q = [0, \frac{1}{2}]$. As boundary condition we take the Dirichlet value given by u .

Results obtained in table 2 still show the superiority of the quadratic interpolation for this regular problem with exactly the same conclusions.

Table 2: Test case 2

NbM	LINEAR			QUADRATIC			CUBIC		
	Err	ItN	Time	Err	ItN	Time	Err	ItN	Time
8	0.1195	493	2	0.0115	1304	33	0.012	1571	99
16	0.0632	903	13	0.0022	1318	100	0.0022	1589	390
32	0.0191	1197	50	0.0003	1319	405	0.0003	1590	1574
64	0.0062	1278	207						

3.3 A non regular problem

We keep the same notations as in the subsection 3.2. We introduce the function

$$v(x, y) = \sin(\pi y) \begin{cases} \sin(\pi x) & \text{pour } -1 \leq x \leq 0, \\ \sin(\frac{1}{2}\pi x) & \text{pour } 1 \geq x \geq 0. \end{cases}$$

Introducing

$$\begin{aligned} \tilde{a} &= \frac{\frac{1}{2} \sin(\pi y) \cos(\frac{1}{2}\pi x)}{\sqrt{\sin(0.5\pi x)^2 \cos(\pi y)^2 + \frac{1}{4} \cos(0.5\pi x)^2 \sin(\pi y)^2}}, \\ \hat{a} &= -\tilde{a}, \\ \hat{\phi}(a) &= \frac{1}{2} a \sin(\pi y) \cos(\frac{1}{2}\pi x) + \sqrt{1 - a^2} \sin(\frac{1}{2}\pi x) \cos(\pi y), \end{aligned}$$

for $b \leq 0$, the function f is then

$$f^\alpha = \begin{cases} (C + \pi^2 \sigma^2) v(x, u) - b\pi K & \text{for } -1 \leq x \leq 0, \\ (C + \pi^2 \sigma^2 \frac{5}{8}) \sin(\frac{1}{2}\pi x) \cos(\pi x) - b\pi \hat{K} & \text{for } 0 \leq x \leq 1, \end{cases}$$

where \hat{K} is the maximum between $\hat{\phi}(\underline{a})$, $\hat{\phi}(\bar{a})$ and $\hat{\phi}(\tilde{a})$ conditionally to $\underline{a} \leq \tilde{a} \leq \bar{a}$, $\hat{\phi}(\hat{a})$ conditionally to $\underline{a} \leq \hat{a} \leq \bar{a}$. We take the values $\underline{a} = -1$, $\bar{a} = 1$, $\sigma = 1$, $C = 0.6$, $b = -1$, $Q = [-1, 1]^2$ and the boundary conditions are given by the values of v . This test case is interesting because the continuous function v is a regular solution for $x < 0$ and for $x > 0$ but it turns out that it is not the viscosity solution u of the problem. The reference solution (an estimation of u) is numerically calculated with a quadratic interpolation with 128 meshes per directions. All the methods seems to converge towards the same solution but here again the superiority of the quadratic interpolation is obvious.

Table 3: Test case 3

NbM	LINEAR			QUADRATIC			CUBIC		
	Err	ItN	Time	Err	ItN	Time	Err	ItN	Time
8	0.970	1810	6	0.10	9406	239	0.097	8877	536
16	0.909	3256	44	0.0141	10546	777	0.0132	10677	2553
32	0.767	5551	214	0.00363	10942	3277	0.0030	10916	10469
64	0.436	8259	1313	0.00187	10984	12107	0.0004	10980	41952
128	0.140	8854	5348						
256	0.0478	10386	26056						

3.4 A last problem for degeneracy of the diffusion operator

The last test case will allow us to test the Finite Difference method with a diffusion operator which is degenerated : The solution is here again

$$u(x, y) = \sin(\pi x) \sin(\pi y).$$

The coefficients are given by

$$c_a(t, x) = C, \quad \sigma_a(t, x) = \sigma \left(\frac{1}{a} \right), \quad b^\alpha = b \left(\frac{1}{a} \right),$$

and the function f^α is given by

$$\begin{aligned} f^\alpha(x, y) = & \sup_{\hat{a} \in [\underline{a}, \bar{a}]} ((C + \pi^2 \sigma^2 (1 + \hat{a}^2))u(x, u) \\ & - (\sigma^2 \pi^2 (\cos(\pi x) \cos(\pi y) + \pi b \sin(\pi x) \cos(\pi y)))\hat{a} \\ & - b\pi \cos(\pi x) \sin(\pi y)). \end{aligned} \quad (40)$$

The values taken are $\underline{a} = -1$, $\bar{a} = 1$, $\sigma = 1$, $C = 0.7$, $b = 0.5$, $Q = [-1, 1]^2$. The results clearly indicate that the Finite Difference method proposed is not competitive with the Semi Lagrangian scheme. The convergence of the fixed point iteration for a step $h = 0.01$ is not achieved with Finite Difference with 100000 iterations (the error given between the last two iterations is around 2.10^{-7}). Once again, the Semi Lagrangian scheme with quadratic interpolation is the most effective method.

Table 4: Test case 4

Semi Lagrangian									
NbM	LINEAR			QUADRATIC			CUBIC		
	Err	ItN	Time	Err	ItN	Time	Err	ItN	Time
8	0.958	1298	2	0.126	12879	176	0.117	17817	564
16	0.838	3448	20	0.0227	14066	576	0.0147	18854	2347
32	0.639	6841	653	0.0024	14326	2314	0.0015	19088	9430
64	0.154	13004	4825	0.0018	14429	8499			
Finite Difference									
$h = 0.01$									
NbM	LINEAR			QUADRATIC			CUBIC		
	Err	ItN	Time	Err	ItN	Time	Err	ItN	Time
8	0.9598	8258	38	0.200	10^5	1649	0.139	10^5	7168
16	0.9276	15240	257	0.064	10^5	4986	0.0345	10^5	28590
32	0.8517	26815	1307	0.023	10^5	19450	0.0129	10^5	113792
64	0.69840	46830	8787	0.008	10^5	145509			
128	0.4599	79658	49851						
256	0.119	50000	198970						

Special thank to the two anonymous referees for their corrections and suggestions.

References

- [1] G. BARLES AND E. R. JAKOBSEN, *On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations*, M2AN, vol 36, N1,(2002), pp 33–54
- [2] G. BARLES AND E. R. JAKOBSEN, *Error bounds for monotone approximation schemes for Hamilton-Jacobi-Bellman equations*, SIAM J. Numer. Anal., 43(2), (2005), pp 540-558
- [3] G. BARLES AND P. E. SOUGANIDIS, *Convergence of Approximation Schemes for Fully Non-linear Second Order Equation* , Asymptotic Anal., 4, (1991), pp. 271–283
- [4] H. J. KUSHNER AND P. DUPUIS, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer Science & Business Media, (2001)
- [5] A. M. OBERMAN, *Convergent difference schemes for degenerate elliptic and parabolic equations: HamiltonJacobi equations and free boundary problems*, SIAM J. Numer. Anal, 44, pp 879–895,(2006)
- [6] D.M. POLLEY, F.A. FORSYTH, K.R. VETZAL, *Numerical Convergence Properties of Option Pricing PDEs with Uncertain Volatility* , IMA Journal of Numerical Analysis, 23, pp 24–267, (2003).
- [7] J. F. BONNANS AND H. ZIDANI, *A fast algorithm for the two dimensional HJB equation of stochastic control*, ESAIM:M2AN, 38-4 , (2004), pp. 723–735
- [8] N. V. KRYLOV, *The rate of convergence of Finite-Difference approximations for Bellman equations with Lipschitz coefficients*, Appl Math Optim, 52,(2005), pp. 365–399
- [9] F. CAMILLI AND M. FALCONE , *An approximation scheme for the optimal control of diffusion processes*, Modélisation Mathématique et Analyse Numérique 29.1,(1995), pp. 97–122
- [10] J. L. MENALDI, *Some estimates for Finite Difference approximations*, SIAM J. Control and Optimization, Vol 27, 3, (1989), pp. 579–607942
- [11] P. L. LIONS, P. E. SOUGANIDIS, *Convergence of MUSCL and filtered schemes for scalar conservation laws and Hamilton-Jacobi equations* , Numer. Math., 69, (1995), pp. 441–470
- [12] O. LEPSKY, *Spectral viscosity approximations to HamiltonJacobi solutions*, SIAM J. Numer. Anal., 38(5), (2000), pp; 1439-1453
- [13] O. BOKANOWSKI, N. MEGDICH AND H. ZIDANI, *Convergence of a non-monotone scheme for Hamilton-Jacobi-Bellman equations with discontinuous initial data*, Numer. Math., 115(1), (2010), pp. 1–44

- [14] R. ABGRALL, *Construction of simple, stable and convergent high order schemes for steady first order Hamilton Jacobi equations*, SIAM J. Sci. Compu, 31, Vol 4, (2009), pp. 2419–2446
- [15] A. OBERMAN, T. SALVADOR , *Filtered schemes for Hamilton-Jacobi equations: a simple construction of convergent accurate difference schemes*, arXiv:1411.3205v1 [math.NA] ,(2014)
- [16] M. FALCONE, R. FERRETTI , *Semi-Lagrangian Approximation Schemes for Linear and Hamilton-Jacobi Equations*, Society for Industrial and Applied Mathematics, (2014)
- [17] K. DEBRABANT AND E. R. JAKOBSEN., *Semi-Lagrangian schemes for linear and fully non-linear diffusion equations*, Math. Comp, no. 283 (2013), pp. 1433–1462
- [18] N. V. KRYLOV, *On The Rate Of Convergence Of Finite-Difference Approximations For Bellman’s Equations With Variable Coefficients*, Probab. Theory Relat. fields,117,(2000), pp. 1-16
- [19] X. WARIN, *Some non monotone schemes for time dependent Hamilton-Jacobi-Bellman equations*, (2014)
- [20] L. C. EVANS, A. FRIEDMAN, *Optimal stochastic switching and the Dirichlet problem for the Bellman equation*, Trans. Amer. Math. Soc. 253:365-389, (1979)
- [21] K. ATKINSON, *An introduction to numerical analysis* ,Willey, Singapore, (1988)
- [22] A. QUARTERONI , R. SACCO , F. SALERI, *Mthodes numriques*, Springer, (2007)
- [23] M. AZAIEZ, M. DAUGE, Y. MADAY , *Methodes Spectrales et des Eléments Spectraux* , master course Nantes, (1993)
- [24] G. S. JIANG, D. PENG, *Weighted ENO schemes for Hamilton-Jacobi equations*, SIAM Journal on Scientific Computing, vol. 21, pp. 2126-2143, (2000)
- [25] S. OSHER, C. W. SHU, *High-order essentially nonoscillatory schemes for Hamilton-Jacobi equations* , SIAM Journal on Numerical Analysis, v28, pp.907-922, (1991)
- [26] C. W. SHU, *High order numerical methods for time dependent Hamilton-Jacobi equations*, Mathematics and computation in imaging science and information processing, pp. 47-91, (2007)
- [27] K. DEBRABANT, E. R. JAKOBSEN, *Semi-Lagrangian schemes for linear and fully non-linear Hamilton-Jacobi-Bellman equations*, arXiv:1403.1217 [math.NA], (2014)