

# A Probabilistic Numerical Method for Fully Nonlinear Parabolic PDEs \*

Arash Fahim<sup>†</sup>      Nizar Touzi<sup>‡</sup>      Xavier Warin<sup>§</sup>

March 9, 2009

## Abstract

We consider the probabilistic numerical scheme for fully nonlinear PDEs suggested in [10], and show that it can be introduced naturally as a combination of Monte Carlo and finite differences scheme without appealing to the theory of backward stochastic differential equations. Our first main result provides the convergence of the discrete-time approximation and derives a bound on the discretization error in terms of the time step. An explicit implementable scheme requires to approximate the conditional expectation operators involved in the discretization. This induces a further Monte Carlo error. Our second main result is to prove the convergence of the latter approximation scheme, and to derive an upper bound on the approximation error. Numerical experiments are performed for the approximation of the solution of the mean curvature flow equation in dimensions two and three, and for two and five-dimensional (plus time) fully-nonlinear Hamilton-Jacobi-Bellman equations arising in the theory of portfolio optimization in financial mathematics.

**Key words:** Viscosity Solutions, monotone schemes, Monte Carlo approximation, second order backward stochastic differential equations.

**AMS 2000 subject classifications:** 65C05, 49L25.

## 1 Introduction

We consider the probabilistic numerical scheme for the approximation of the solution of a fully-nonlinear parabolic Cauchy problem suggested in [10]. In the latter paper, a representation of the solution of the PDE is derived in terms of the newly introduced notion of second order backward stochastic differential equations, assuming that the fully-nonlinear

---

\*We are grateful to Mete Soner, Bruno Bouchard, Denis Talay, and Frédéric Bonnans for fruitful comments and suggestions. This research is part of the Chair *Financial Risks* of the *Risk Foundation* sponsored by Société Générale, the Chair *Derivatives of the Future* sponsored by the Fédération Bancaire Française, the Chair *Finance and Sustainable Development* sponsored by EDF and Calyon.

<sup>†</sup>CMAP, Ecole Polytechnique Paris & Sharif University of Technology, Tehran, arash.fahim@polytechnique.edu.

<sup>‡</sup>CMAP, Ecole Polytechnique Paris, nizar.touzi@polytechnique.edu.

<sup>§</sup>EDF R&D, Paris, xavier.warin@edf.fr.

<sup>¶</sup>The authors want to thank Bruno Bouchard for his good suggestions during this work.

parabolic Cauchy problem has a smooth solution. Then, similarly to the case of backward stochastic differential equations which are connected to semi-linear PDEs, this representation suggests a backward probabilistic numerical scheme.

The representation result of [10] can be viewed as an extension of the Feynman-Kac representation result, for the linear case, which is widely used in order to approach the numerical approximation problem from the probabilistic viewpoint, and to take advantage of the high dimensional properties of Monte Carlo methods. Previously, the theory of backward stochastic differential equations provided an extension of these approximation methods to the semilinear case. See for instance Chevance [11], El Karoui, Peng and Quenez [15], Bally and Pagès [2], Bouchard and Touzi [9] and Zhang [28]. In particular, the latter papers provide the convergence of the “natural” discrete-time approximation of the value function and its partial space gradient with the same  $L^2$  error of order  $\sqrt{h}$ , where  $h$  is the length of time step. The discretization involves the computation of conditional expectations, which need to be further approximated in order to result into an implementable scheme. We refer to [2], [9] and [17] for an complete asymptotic analysis of the approximation, including the regression error.

In this paper, we observe that the backward probabilistic scheme of [10] can be introduced naturally without appealing to the notion of backward stochastic differential equation. This is shown in Section 2 where the scheme is decomposed into three steps:

- (i) The Monte Carlo step consists in isolating the linear generator of some underlying diffusion process, so as to split the PDE into this linear part and a remaining nonlinear one.
- (ii) Evaluating the PDE along the underlying diffusion process, we obtain a natural discrete-time approximation by using finite differences approximation in the remaining nonlinear part of the equation.
- (iii) Finally, the backward discrete-time approximation obtained by the above steps (i)-(ii) involves the conditional expectation operator which is not computable in explicit form. An implementable probabilistic numerical scheme therefore requires to replace such conditional expectations by a convenient approximation, and induces a further Monte Carlo type of error.

In the present paper, we do not require the fully nonlinear PDE to have a smooth solution, and we only assume that it satisfies a comparison result in the sense of viscosity solutions. Our main objective is to establish the convergence of this approximation towards the unique viscosity solution of the fully-nonlinear PDE, and to provide an asymptotic analysis of the approximation error.

Our main results are the following. We first prove the convergence of the discrete-time approximation for general nonlinear PDEs, and we provide bounds on the corresponding approximation error for a class of Hamilton-Jacobi-Bellman PDEs. Then, we consider the implementable scheme involving the Monte Carlo error, and we similarly prove a convergence result for general nonlinear PDEs, and we provide bounds on the error of approximation for Hamilton-Jacobi-Bellman PDEs. We observe that our convergence results place some restrictions on the choice of the diffusion of the underlying diffusion process. First, a uniform ellipticity condition is needed; we believe that this technical condition can be re-

laxed in some future work. More importantly, the diffusion coefficient is needed to dominate the partial gradient of the remaining nonlinearity with respect to its Hessian component. Although we have no theoretical result that this condition is necessary, our numerical experiments show that the violation of this condition leads to a serious mis-performance of the method, see Figure 5.

Our proofs rely on the monotonic scheme method developed by Barles and Souganidis [7] in the theory of viscosity solutions, and the recent method of shaking coefficients of Krylov [21], [22] and [23] and Barles and Jakobsen [6], [5] and [4]. The use of the latter type of methods in the context of a stochastic scheme seems to be new. Notice however, that our results are of a different nature than the classical error analysis results in the theory of backward stochastic differential equations, as we only study the convergence of the approximation of the value function, and no information is available for its gradient or Hessian with respect to the space variable.

A related work in the context of Hamilton-Jacobi-Bellman nonlinear PDEs, is by Bonnans and Zidani [8] who introduced a finite difference scheme which satisfies the crucial monotonicity condition of Barles and Souganidis [7] so as to ensure its convergence. Their main idea is to discretize both time and space, approximate the underlying controlled forward diffusion for each fixed control by a controlled local Markov chain on the grid, approximate the derivatives in certain directions which are found by solving some further optimization problem, and optimize over the control. Beyond the curse of dimensionality problem which is encountered by finite differences schemes, we believe that our method is much simpler as the monotonicity is satisfied without any need to treat separately the linear structures for each fixed control, and without any further investigation of some direction of discretization for the finite differences.

We finally observe a connection with the recent work of Kohn and Serfaty [20] who provide a deterministic game theoretic interpretation for fully nonlinear parabolic problems. The game is time limited and consists of two players. At each time step, one tries to maximize her gain and the other to minimize it by imposing a penalty term to her gain. The nonlinearity of the fully nonlinear PDE appears in the penalty. Also, although the nonlinear penalty does not need to be elliptic, a parabolic nonlinearity appears in the limiting PDE. This approach is very similar to the representation of [10] where such a parabolic envelope appears in the PDE, and where the Brownian motion plays the role of the nature playing against the player.

The paper is organized as follows. In Section 2, we provide a natural presentation of the scheme without appealing to the theory of backward stochastic differential equations. Section 3 is dedicated to the asymptotic analysis of the discrete-time approximation, and contains our first main convergence result and the corresponding error estimate. In Section 4, we introduce the implementable backward scheme, and we further investigate the induced Monte Carlo error. We again prove convergence and we provide bounds on the approximation error. Finally, Section 5 contains some numerical results for the mean curvature flow equation on the plane and space, and for a five-dimensional Hamilton-Jacobi-Bellman equation arising in the problem of portfolio optimization in financial mathematics.

**Notations** For scalars  $a, b \in \mathbb{R}$ , we write  $a \wedge b := \min\{a, b\}$ ,  $a \vee b := \max\{a, b\}$ , and  $a^+ := \max\{a, 0\}$ . By  $\mathbb{M}(n, d)$ , we denote the collection of all  $n \times d$  matrices with real entries. The collection of all symmetric matrices of size  $d$  is denoted  $\mathbb{S}_d$ , and its subset of nonnegative symmetric matrices is denoted by  $\mathbb{S}_d^+$ . For a matrix  $A \in \mathbb{M}(n, d)$ , we denote by  $A^T$  its transpose. For  $A, B \in \mathbb{M}(n, d)$ , we denote  $A \cdot B := \text{Tr}[A^T B]$ . In particular, for  $d = 1$ ,  $A$  and  $B$  are vectors of  $\mathbb{R}^n$  and  $A \cdot B$  reduces to the Euclidean scalar product.

For a function  $u$  from  $[0, T] \times \mathbb{R}^d$  to  $\mathbb{R}$ , we say that  $u$  has  $q$ -polynomial growth (resp.  $\alpha$ -exponential growth) if

$$\sup_{t \leq T, x \in \mathbb{R}^d} \frac{|u(t, x)|}{1 + |x|^q} < \infty, \quad (\text{resp.} \quad \sup_{t \leq T, x \in \mathbb{R}^d} e^{-\alpha|x|} |u(t, x)| < \infty).$$

For a suitably smooth function  $\varphi$  on  $Q_T := (0, T] \times \mathbb{R}^d$ , we define

$$|\varphi|_\infty := \sup_{(t, x) \in Q_T} |\varphi(t, x)| \quad \text{and} \quad |\varphi|_1 := |\varphi|_\infty + \sup_{Q_T \times Q_T} \frac{|\varphi(t, x) - \varphi(t', x')|}{|x - x'| + |t - t'|^{\frac{1}{2}}}.$$

Finally, we denote the  $\mathbb{L}^p$ -norm of a r.v.  $R$  by  $\|R\|_p := (\mathbb{E}[|R|^p])^{1/p}$ .

## 2 Discretization

Let  $\mu$  and  $\sigma$  be two maps from  $\mathbb{R}_+ \times \mathbb{R}^d$  to  $\mathbb{M}(d, d)$  and  $\mathbb{R}^d$ . With  $a := \sigma\sigma^T$  be the map from  $\mathbb{R}_+ \times \mathbb{R}^d$  to  $\mathbb{S}_d^+$ , and define the linear operator:

$$\mathcal{L}^X \varphi := \frac{\partial \varphi}{\partial t} + \mu \cdot D\varphi + \frac{1}{2} a \cdot D^2 \varphi.$$

Given a map

$$F : (t, x, r, p, \gamma) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \longmapsto F(x, r, p, \gamma) \in \mathbb{R}$$

we consider the Cauchy problem:

$$-\mathcal{L}^X v - F(\cdot, v, Dv, D^2v) = 0, \quad \text{on } [0, T] \times \mathbb{R}^d, \quad (2.1)$$

$$v(T, \cdot) = g, \quad \text{on } \mathbb{R}^d. \quad (2.2)$$

Under some conditions, a stochastic representation for the solution of this problem was provided in [10] by means of the newly introduced notion of second order backward stochastic differential equations. As an important implication, such a stochastic representation suggests a probabilistic numerical scheme for the above Cauchy problem.

The chief goal of this section is to obtain the probabilistic numerical scheme suggested in [10] by a direct manipulation of (2.1)-(2.2) without appealing to the notion of backward stochastic differential equations.

To do this, we consider an  $\mathbb{R}^d$ -valued Brownian motion  $W$  on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ , where the filtration  $\mathbb{F} = \{\mathcal{F}_t, t \in [0, T]\}$  satisfies the usual completeness conditions, and  $\mathcal{F}_0$  is trivial.

For a positive integer  $n$ , let  $h := T/n$ ,  $t_i = ih$ ,  $i = 0, \dots, n$ , and consider the one step ahead Euler discretization

$$\hat{X}_h^{t,x} := x + \mu(t, x)h + \sigma(t, x)(W_{t+h} - W_t), \quad (2.3)$$

of the diffusion  $X$  corresponding to the linear operator  $\mathcal{L}^X$ . Our analysis does not require any existence and uniqueness result for the underlying diffusion  $X$ . However, the subsequent formal discussion assumes it in order to provides a natural justification of our numerical scheme.

Assuming that the PDE (2.1) has a classical solution, it follows from Itô's formula that

$$\mathbb{E}_{t_i, x} [v(t_{i+1}, X_{t_{i+1}})] = v(t_i, x) + \mathbb{E}_{t_i, x} \left[ \int_{t_i}^{t_{i+1}} \mathcal{L}^X v(t, X_t) dt \right]$$

where we ignored the difficulties related to local martingale part, and  $\mathbb{E}_{t_i, x} := \mathbb{E}[\cdot | X_{t_i} = x]$  denotes the expectation operator conditional on  $\{X_{t_i} = x\}$ . Since  $v$  solves the PDE (2.1), this provides

$$\mathbb{E}_{t_i, x} [v(t_{i+1}, X_{t_{i+1}})] = v(t_i, x) - \mathbb{E}_{t_i, x} \left[ \int_{t_i}^{t_{i+1}} F(\cdot, v, Dv, D^2v)(t, X_t) dt \right].$$

By approximating the Riemann integral, and replacing the process  $X$  by its Euler discretization, this suggest the following approximation of the value function  $v$

$$v^h(T, \cdot) := g \quad \text{and} \quad v^h(t_i, x) := \mathbf{T}_h[v^h](t_i, x), \quad (2.4)$$

where we denoted for a function  $\psi : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$  with exponential growth:

$$\mathbf{T}_h[\psi](t, x) := \mathbb{E} \left[ \psi(t+h, \hat{X}_h^{t,x}) \right] + hF(\cdot, \mathcal{D}_h\psi)(t, x), \quad (2.5)$$

$$\mathcal{D}_h^k \psi(t, x) := \mathbb{E} [D^k \psi(t+h, \hat{X}_h^{t,x})], \quad k = 0, 1, 2, \quad \mathcal{D}_h \psi := (\mathcal{D}_h^0 \psi, \mathcal{D}_h^1 \psi, \mathcal{D}_h^2 \psi), \quad (2.6)$$

and  $D^k$  is the  $k$ -th order partial differential operator with respect to the space variable  $x$ . The differentiations in the above scheme are to be understood in the sense of distributions. This algorithm is well-defined whenever  $g$  has exponential growth and  $F$  is a Lipschitz map. To see this, observe that any function with exponential growth has weak gradient and Hessian because the Gaussian kernel is a Schwartz function, and the exponential growth is inherited at each time step from the Lipschitz property of  $F$ .

At this stage, the above backward algorithm presents the serious drawback of involving the gradient  $Dv^h(t_{i+1}, \cdot)$  and the Hessian  $D^2v^h(t_{i+1}, \cdot)$  in order to compute  $v^h(t_i, \cdot)$ . The following result avoids this difficulty by an easy integration by parts argument.

**Lemma 2.1** *For every function  $\varphi : Q_T \rightarrow \mathbb{R}$  with exponential growth, we have:*

$$\mathcal{D}_h \varphi(t_i, x) = \mathbb{E} [\varphi(t_{i+1}, \hat{X}_h^{t_i, x}) H_h(t_i, x)],$$

where  $H_h = (H_0^h, H_1^h, H_2^h)$  and

$$H_0^h = 1, \quad H_1^h = (\sigma^\top)^{-1} \frac{W_h}{h}, \quad H_2^h = (\sigma^\top)^{-1} \frac{W_h W_h^\top - h \mathbf{I}_d}{h^2} \sigma^{-1}. \quad (2.7)$$

**Proof.** The main ingredient is the following easy observation. Let  $G$  be a one dimensional Gaussian random variable with unit variance. Then, for any function  $f : \mathbb{R} \rightarrow \mathbb{R}$  with exponential growth, we have:

$$\mathbb{E}[f(G)H^k(G)] = \mathbb{E}[f^{(k)}(G)], \quad (2.8)$$

where  $f^{(k)}$  is the  $k$ -th order derivative of  $f$  in the sense of distributions, and  $H^k$  is the one-dimensional Hermite polynomial of degree  $k$ .

**1** Now, let  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  be a function with exponential growth. Then, by direct conditioning, it follows from (2.8) that

$$\mathbb{E} \left[ \varphi(\hat{X}_h^{t,x}) W_h^i \right] = h \sum_{j=1}^d \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_j}(\hat{X}_h^{t,x}) \sigma_{ji}(t,x) \right],$$

and therefore:

$$\mathbb{E} \left[ \varphi(\hat{X}_h^{t,x}) H_1^h(t,x) \right] = \sigma(t,x)^T \mathbb{E} \left[ \nabla \varphi(\hat{X}_h^{t,x}) \right].$$

**2** For  $i \neq j$ , it follows from (2.8) that

$$\begin{aligned} \mathbb{E} \left[ \varphi(\hat{X}_h^{t,x}) W_h^i W_h^j \right] &= h \sum_{k=1}^d \mathbb{E} \left[ \frac{\partial \varphi}{\partial x_k}(\hat{X}_h^{t,x}) W_h^j \sigma_{ki}(t,x) \right] \\ &= h^2 \sum_{k,l=1}^d \mathbb{E} \left[ \frac{\partial^2 \varphi}{\partial x_k \partial x_l}(\hat{X}_h^{t,x}) \sigma_{lj}(t,x) \sigma_{ki}(t,x) \right], \end{aligned}$$

and for  $j = i$ :

$$\mathbb{E} \left[ \varphi(\hat{X}_h^{t,x}) ((W_h^i)^2 - h) \right] = h^2 \sum_{k,l=1}^d \mathbb{E} \left[ \frac{\partial^2 \varphi}{\partial x_k \partial x_l}(\hat{X}_h^{t,x}) \sigma_{li}(t,x) \sigma_{ki}(t,x) \right].$$

This provides:

$$\mathbb{E} \left[ \varphi(\hat{X}_h^{t,x}) H_2^h(t,x) \right] = \sigma(t,x)^T \mathbb{E} \left[ \nabla^2 \varphi(\hat{X}_h^{t,x}) \sigma(t,x) \right].$$

□

In view of Lemma 2.1 the iteration which computes  $v^h(t_i, \cdot)$  out of  $v^h(t_{i+1}, \cdot)$  in (2.4)-(2.5) does not involve the gradient and the Hessian of the latter function.

**Remark 2.1** Clearly, one can proceed to different choices for the integration by parts in Lemma 2.1. One such possibility leads to the representation of  $\mathcal{D}_2^h \varphi$  as:

$$\mathcal{D}_2^h \varphi(t,x) = \mathbb{E} \left[ \varphi(\hat{X}_h^{t,x}) (\sigma^T)^{-1} \frac{W_{h/2}}{(h/2)} \frac{W_{h/2}^T}{(h/2)} \sigma^{-1} \right].$$

This representation shows that the backward scheme (2.4) is very similar to the probabilistic numerical algorithm suggested in [10].

Observe that the choice of the drift and the diffusion coefficients  $\mu$  and  $\sigma$  in the nonlinear PDE (2.1) is arbitrary. So far, it has been only used in order to define the underlying diffusion  $X$ . Our convergence result will however place some restrictions on the choice of the diffusion coefficient, see Remark 3.2.

Once the linear operator  $\mathcal{L}^X$  is chosen in the nonlinear PDE, the above algorithm handles the remaining nonlinearity by the classical finite differences approximation. This connection with finite differences is motivated by the following formal interpretation of Lemma 2.1, where for ease of presentation, we set  $d = 1$ ,  $\mu \equiv 0$ , and  $\sigma(x) \equiv 1$ :

- Consider the binomial random walk approximation of the Brownian motion  $\hat{W}_{t_k} := \sum_{j=1}^k w_j$ ,  $t_k := kh$ ,  $k \geq 1$ , where  $\{w_j, j \geq 1\}$  are independent random variables distributed as  $\frac{1}{2} \left( \delta_{\sqrt{h}} + \delta_{-\sqrt{h}} \right)$ . Then, this induces the following approximation:

$$\mathcal{D}_h^1 \psi(t, x) := \mathbb{E} \left[ \psi(t+h, X_h^{t,x}) H_1^h \right] \approx \frac{\psi(t, x + \sqrt{h}) - \psi(t, x - \sqrt{h})}{2\sqrt{h}},$$

which is the centered finite differences approximation of the gradient.

- Similarly, consider the trinomial random walk approximation  $\hat{W}_{t_k} := \sum_{j=1}^k w_j$ ,  $t_k := kh$ ,  $k \geq 1$ , where  $\{w_j, j \geq 1\}$  are independent random variables distributed as  $\frac{1}{6} \left( \delta_{\{\sqrt{3h}\}} + 4\delta_{\{0\}} + \delta_{\{-\sqrt{3h}\}} \right)$ , so that  $\mathbb{E}[w_j^n] = \mathbb{E}[W_h^n]$  for all integers  $n \leq 4$ . Then, this induces the following approximation:

$$\mathcal{D}_h^2 \psi(t, x) := \mathbb{E} \left[ \psi(t+h, X_h^{t,x}) H_2^h \right] \approx \frac{\psi(t, x + \sqrt{3h}) - 2\psi(t, x) + \psi(t, x - \sqrt{3h})}{3h},$$

which is the centered finite differences approximation of the Hessian.

In view of the above interpretation, the numerical scheme studied in this paper can be viewed as a mixed Monte Carlo–Finite Differences algorithm. The Monte Carlo component of the scheme consists in the choice of an underlying diffusion process  $X$ . The finite differences component of the scheme consists in approximating the remaining nonlinearity by means of the integration-by-parts formula of Lemma 2.1.

### 3 Asymptotics of the discrete-time approximation

#### 3.1 The main results

Our first main convergence results follow the general methodology of Barles and Souganidis [7], and requires that the nonlinear PDE (2.1) satisfies a comparison result in the sense of viscosity solutions.

**Definition 3.1** *We say that (2.1) has strong comparison for bounded functions if for any bounded upper semicontinuous viscosity supersolution  $\bar{v}$  and any bounded lower semicontinuous subsolution  $\underline{v}$  on  $[0, T] \times \mathbb{R}^d$ , satisfying*

- *either  $\bar{v}(T, \cdot) \geq \underline{v}(T, \cdot)$ ,*
  - *or the viscosity property of  $\bar{v}$  and  $\underline{v}$  holds true on  $[0, T] \times \mathbb{R}^d$ ,*
- we have  $\bar{v} \geq \underline{v}$ .*

The strong comparison principle is an important notion in the theory of viscosity solutions which allows to handle situations where the boundary condition  $g$  is not compatible with the equation. We recall that, if  $v$  is a continuous viscosity solution of (2.1)-(2.2), then  $v$  is the unique continuous viscosity solution of (2.1) on  $[0, T) \times \mathbb{R}^d$  which is in addition satisfies in the viscosity sense:

$$\begin{aligned} \max \{ -\mathcal{L}^X v - F(\cdot, v, Dv, D^2v), v - g \} &\geq 0 \text{ on } \{T\} \times \mathbb{R}^d \\ \min \{ -\mathcal{L}^X v - F(\cdot, v, Dv, D^2v), v - g \} &\leq 0 \text{ on } \{T\} \times \mathbb{R}^d. \end{aligned} \quad (3.1)$$

**Remark 3.1** We could have imposed some technical assumption on  $F$  and  $g$  which ensure that the above strong comparison assumption is satisfied. For more discussion, see Remark 4.8(2) in [10]. Moreover, notice that in some cases the strong comparison hypothesis in theorem 3.1 could be replaced by weak comparison hypothesis.

In the sequel, we denote by  $F_r$ ,  $F_p$  and  $F_\gamma$  the partial gradients of  $F$  with respect to  $r$ ,  $p$  and  $\gamma$ , respectively. We recall that any Lipschitz function is differentiable a.e.

**Assumption F** *The nonlinearity  $F$  is Lipschitz-continuous with respect to  $(r, p, \gamma)$  uniformly in  $(t, x)$  and  $|F(t, x, 0, 0, 0)|_\infty < \infty$ . Moreover,  $F$  is uniformly elliptic and dominated by the diffusion of the linear operator  $\mathcal{L}^X$ , i.e.*

$$\varepsilon I_d \leq \nabla_\gamma F \leq a \text{ on } \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \text{ for some } \varepsilon > 0. \quad (3.2)$$

**Remark 3.2** The above Condition (3.2) places some restrictions on the choice of the linear operator  $\mathcal{L}^X$  in the nonlinear PDE (2.1). First,  $F$  is required to be uniformly elliptic, implying an upper bound on the choice of the diffusion matrix  $\sigma$ . Since  $\sigma \sigma^T \in \mathcal{S}_d^+$ , this implies in particular that our main results do not apply to general degenerate nonlinear parabolic PDEs. Second, the diffusion of the linear operator  $\sigma$  is required to dominate the nonlinearity  $F$  which places implicitly a lower bound on the choice of the diffusion  $\sigma$ .

**Example 3.1** Let us consider the nonlinear PDE in the one-dimensional case  $-\frac{\partial v}{\partial t} - \frac{1}{2}(a^2 v_{xx}^+ - b^2 v_{xx}^-)$  where  $0 < b < a$  are given constants. Then if we restrict the choice of the diffusion to be constant, it follows from Condition **F** that  $\frac{1}{2}a^2 \leq \sigma^2 \leq b^2$ , which implies that  $a^2 \leq 2b^2$ . If the parameters  $a$  and  $b$  do not satisfy the latter condition, then the diffusion  $\sigma$  has to be chosen to be state and time dependent.

**Theorem 3.1 (Convergence)** *Let Assumption **F** hold true, and assume that the fully nonlinear PDE (2.1) has strong comparison for bounded functions. Then for every bounded function  $g$ , there exists a bounded function  $v$  so that*

$$v^h \longrightarrow v \quad \text{locally uniformly.}$$

*In addition,  $v$  is the unique bounded viscosity solution of the relaxed boundary problem (2.1)-(2.2).*

**Remark 3.3** Assume that the coefficients  $\mu$  and  $\sigma$  are bounded. Then, The restriction to bounded terminal data  $g$  in the above Theorem 3.1 can be relaxed by an immediate change of variable. Let  $g$  be a function with  $\alpha$ -exponential growth for some  $\alpha > 0$ . Fix some  $M > 0$ , and let  $\rho$  be an arbitrary smooth positive function with:

$$\rho(x) = e^{\alpha|x|} \quad \text{for } |x| \geq M,$$

so that both  $\rho(x)^{-1}\nabla\rho(x)$  and  $\rho(x)^{-1}\nabla^2\rho(x)$  are bounded. Let

$$u(t, x) := \rho(x)^{-1}v(t, x) \quad \text{for } (t, x) \in [0, T] \times \mathbb{R}^d.$$

Then, the nonlinear PDE problem (2.1)-(2.2) satisfied by  $v$  converts into the following nonlinear PDE for  $u$ :

$$-\mathcal{L}^X u - \tilde{F}(\cdot, u, Du, D^2u) = 0, \quad \text{on } [0, T] \times \mathbb{R}^d \quad \text{and} \quad v(T, \cdot) = \tilde{g} := \rho^{-1}g \quad \text{on } \mathbb{R}^d \quad (3.3)$$

where

$$\begin{aligned} \tilde{F}(t, x, r, p, \gamma) &:= r\mu(x) \cdot \rho^{-1}\nabla\rho + \frac{1}{2}\text{Tr} [a(x) (r\rho^{-1}\nabla^2\rho + 2p\rho^{-1}\nabla\rho^T)] \\ &\quad + \rho^{-1}F(t, x, r\rho, r\nabla\rho + p\rho, r\nabla^2\rho + 2p\nabla\rho^T + \rho\gamma). \end{aligned}$$

Recall that the coefficients  $\mu$  and  $\sigma$  are assumed to be bounded. Then, it is easy to see that  $\tilde{F}$  satisfies the same conditions as  $F$ . Since  $\tilde{g}$  is bounded, the convergence Theorem 3.1 applies to the nonlinear PDE (3.3).  $\square$

**Remark 3.4** The strong comparison assumption can be relaxed under some conditions to the standard comparison assumption. For example if  $F$  is bounded or  $\sigma$  is constant and  $F(x, r, p, \gamma)$  is between two linear functions in  $\gamma$ , then one can show that the subsolution  $v_*$  and supersolution  $v^*$ , introduced in [7], will satisfy the final condition:

$$v_*(T, \cdot) = v^*(T, \cdot) = g.$$

One can also check that, if  $g$  is  $C^2$  and  $F$  is Lipschitz continuous, standard comparison is enough to achieve the results of this paper.

**Remark 3.5** Theorem 3.1 states that the right hand-side inequality of (3.2) (i.e. diffusion must dominate the nonlinearity in  $\gamma$ ) is sufficient for the convergence of the Monte Carlo–Finite Differences scheme. We do not know whether this condition is necessary:

- Subsection 3.4 suggests that this condition is not sharp in the simple linear case,
- however, our numerical experiments of Section 5 reveal that the method may have a poor performance in the absence of this condition, see Figure 5.

We next provide bounds on the rate of convergence of the Monte Carlo–Finite Differences scheme in the context of nonlinear PDEs of the Hamilton-Jacobi-Bellman type in the same context as [6]. The following assumptions are stronger than Assumption **F** and imply that the nonlinear PDE (2.1) satisfies a strong comparison result for bounded functions.

**Assumption HJB** *The nonlinearity  $F$  satisfies (3.2), and is of the Hamilton-Jacobi-Bellman type:*

$$\begin{aligned} F(t, x, r, p, \gamma) &= \inf_{\alpha \in \mathcal{A}} \{ \mathcal{L}^\alpha(t, x, r, p, \gamma) \} \\ \mathcal{L}^\alpha(t, x, r, p, \gamma) &:= \frac{1}{2} \text{Tr}[\sigma^\alpha \sigma^{\alpha \text{T}}(t, x) \gamma] + b^\alpha(t, x) p + c^\alpha(t, x) r + f^\alpha(t, x) \end{aligned}$$

where the functions  $\mu$ ,  $\sigma$ ,  $\sigma^\alpha$ ,  $b^\alpha$ ,  $c^\alpha$  and  $f^\alpha$  satisfy:

$$|\mu|_\infty + |\sigma|_\infty + \sup_{\alpha \in \mathcal{A}} (|\sigma^\alpha|_1 + |b^\alpha|_1 + |c^\alpha|_1 + |f^\alpha|_1) < \infty.$$

**Assumption HJB+** *The nonlinearity  $F$  satisfies HJB, and for any  $\delta > 0$ , there exists a finite set  $\{\alpha_i\}_{i=1}^{M_\delta}$  such that for any  $\alpha \in \mathcal{A}$ :*

$$\inf_{1 \leq i \leq M_\delta} |\sigma^\alpha - \sigma^{\alpha_i}|_\infty + |b^\alpha - b^{\alpha_i}|_\infty + |c^\alpha - c^{\alpha_i}|_\infty + |f^\alpha - f^{\alpha_i}|_\infty \leq \delta.$$

**Theorem 3.2 (Rate of Convergence)** *Assume that the final condition  $g$  is bounded Lipschitz-continuous. Then, there is a constant  $C > 0$  such that:*

- (i) *under Assumption HJB, we have  $v - v^h \leq Ch^{1/4}$ ,*
- (ii) *under the stronger condition HJB+, we have  $-Ch^{1/10} \leq v - v^h \leq Ch^{1/4}$ .*

The above bounds can be improved in some specific examples. See Subsection 3.4 for the linear case where the rate of convergence is improved to  $\sqrt{h}$ .

We also observe that, in the PDE Finite Differences literature, the rate of convergence is usually stated in terms of the discretization in the space variable  $|\Delta x|$ . In our context of stochastic differential equation, notice that  $|\Delta x|$  is of the order of  $h^{1/2}$ . Therefore, the above upper and lower bounds on the rate of convergence corresponds to the classical rate  $|\Delta x|^{1/2}$  and  $|\Delta x|^{1/5}$ , respectively.

### 3.2 Proof of the convergence result

We now provide the proof Theorem 3.1 by building on Theorem 2.1 and Remark 2.1 of Barles and Souganidis [7]. In this subsection, we show that our scheme is consistent, monotone and stable in the sense of [7] as required by their Theorem 2.1.

**Lemma 3.1** *Let  $\varphi$  be a smooth function with bounded derivatives. Then for all  $(t, x) \in [0, T] \times \mathbb{R}^d$ :*

$$\lim_{\substack{(t', x') \rightarrow (t, x) \\ (h, c) \rightarrow (0, 0) \\ t' + h \leq T}} \frac{[c + \varphi](t', x') - \mathbf{T}_h[c + \varphi](t', x')}{h} = -(\mathcal{L}^X \varphi + F(\cdot, \varphi, D\varphi, D^2\varphi))(t, x).$$

The proof is a straightforward application of Itô's formula, and is omitted.

**Lemma 3.2** *Let  $\varphi, \psi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  be two functions with exponential growth. Then:*

$$\varphi \leq \psi \implies \mathbf{T}_h[\varphi](t, x) \leq \mathbf{T}_h[\psi](t, x) + Ch \mathbb{E}[(\psi - \varphi)(t + h, \hat{X}_h^{t,x})] \text{ for some } C > 0.$$

**Proof.** By Lemma 2.1 the operator  $\mathbf{T}_h$  can be written as:

$$\mathbf{T}_h[\psi](t, x) = \mathbb{E} \left[ \psi(\hat{X}_h^{t,x}) \right] + hF \left( t, x, \mathbb{E}[\psi(\hat{X}_h^{t,x})H_h(t, x)] \right).$$

Let  $f := \psi - \varphi \geq 0$  where  $\varphi$  and  $\psi$  are as in the statement of the lemma. Let  $F_\tau$  denote the partial gradient with respect to  $\tau = (r, p, \gamma)$ . By the mean value Theorem:

$$\begin{aligned} \mathbf{T}_h[\psi](t, x) - \mathbf{T}_h[\varphi](t, x) &= \mathbb{E} \left[ f(\hat{X}_h^{t,x}) \right] + hF_\tau(\theta) \cdot \mathcal{D}_h f(\hat{X}_h^{t,x}) \\ &= \mathbb{E} \left[ f(\hat{X}_h^{t,x}) (1 + hF_\tau(\theta) \cdot H_h(t, x)) \right], \end{aligned}$$

for some  $\theta = (t, x, \bar{r}, \bar{p}, \bar{\gamma})$ . By the definition of  $H_h(t, x)$ :

$$\mathbf{T}_h[\psi] - \mathbf{T}_h[\varphi] = \mathbb{E} \left[ f(\hat{X}_h^{t,x}) \left( 1 + hF_r + F_p \cdot \sigma^{\text{T}-1} W_h + h^{-1} F_\gamma \cdot \sigma^{\text{T}-1} (W_h W_h^{\text{T}} - hI) \sigma^{-1} \right) \right],$$

where the dependence on  $\theta$  and  $x$  has been omitted for notational simplicity. Since  $F_\gamma \leq a$  by (3.2) of Assumption **F**, we have  $1 - a^{-1} \cdot F_\gamma \geq 0$  and therefore:

$$\begin{aligned} \mathbf{T}_h[\psi] - \mathbf{T}_h[\varphi] &\geq \mathbb{E} \left[ f(\hat{X}_h^{t,x}) \left( hF_r + F_p \cdot \sigma^{\text{T}-1} W_h + h^{-1} F_\gamma \cdot \sigma^{\text{T}-1} W_h W_h^{\text{T}} \sigma^{-1} \right) \right] \\ &= \mathbb{E} \left[ f(\hat{X}_h^{t,x}) \left( hF_r + |A_h|^2 - \frac{h}{4} F_p^{\text{T}} (F_\gamma)^{-1} F_p \right) \right] \end{aligned}$$

where

$$A_h := \frac{1}{\sqrt{h}} F_\gamma^{1/2} \sigma^{\text{T}-1} W_h + \frac{\sqrt{h}}{2} F_\gamma^{-1/2} F_p. \quad (3.4)$$

Hence:

$$\mathbf{T}_h[\psi] - \mathbf{T}_h[\varphi] \geq \mathbb{E} \left[ f(\hat{X}_h^{t,x}) \left( hF_r - \frac{h}{4} F_p^{\text{T}} F_\gamma^{-1} F_p \right) \right] \geq -Ch \mathbb{E} \left[ f(\hat{X}_h^{t,x}) \right]$$

for some constant  $C > 0$ , where the last inequality follows from the Lipschitz property of  $F$  and the fact that  $|F_\gamma^{-1}|_\infty < \infty$  by (3.2).  $\square$

The following observation will be used in the proof of Theorem 3.2 below.

**Remark 3.6** The monotonicity result of the previous Lemma 3.2 is slightly different from that required in [7]. However, as it is observed in Remark 2.1 in [7], their convergence theorem holds under this approximate monotonicity. From the previous proof, we observe that if the function  $F$  satisfies the condition:

$$F_r - \frac{1}{4} F_p^{\text{T}} F_\gamma^{-1} F_p \geq 0, \quad (3.5)$$

then, the standard monotonicity condition

$$\varphi \leq \psi \implies \mathbf{T}_h[\varphi](t, x) \leq \mathbf{T}_h[\psi](t, x) \quad (3.6)$$

holds. Using the parabolic feature of the equation, we may introduce a new function  $u(t, x) := e^{\theta(T-t)}v(t, x)$  which solves a nonlinear PDE satisfying (3.5). Indeed, direct calculation shows that the PDE inherited by  $u$  is:

$$-\mathcal{L}^X u - \bar{F}(\cdot, u, Du, D^2u) = 0, \quad \text{on } [0, T) \times \mathbb{R}^d \quad (3.7)$$

$$u(T, x) = g(x), \quad \text{on } \mathbb{R}^d, \quad (3.8)$$

where  $\bar{F}(t, x, r, p, \gamma) = e^{\theta(T-t)}F(t, x, e^{-\theta(T-t)}r, e^{-\theta(T-t)}p, e^{-\theta(T-t)}\gamma) + \theta r$ . Then, it is easily seen that  $\bar{F}$  satisfies the same conditions as  $F$  together with (3.5) for sufficiently large  $\theta$ .

**Lemma 3.3** *Let  $\varphi, \psi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  be two  $\mathbb{L}^\infty$ -bounded functions. Then there exists a constant  $C > 0$  such that*

$$|\mathbf{T}_h[\varphi] - \mathbf{T}_h[\psi]|_\infty \leq |\varphi - \psi|_\infty(1 + Ch)$$

*In particular, if  $g$  is  $\mathbb{L}^\infty$ -bounded, the family  $(v^h)_h$  defined in (2.4) is  $\mathbb{L}^\infty$ -bounded, uniformly in  $h$ .*

**Proof.** Let  $f := \varphi - \psi$ . Then, arguing as in the previous proof,

$$\mathbf{T}_h[\varphi] - \mathbf{T}_h[\psi] = \mathbb{E} \left[ f(\hat{X}_h) \left( 1 - a^{-1} \cdot F_\gamma + |A_h|^2 + hF_r - \frac{h}{4} F_p^\top F_\gamma^{-1} F_p \right) \right]$$

where  $A_h$  is given by (3.4). Since  $1 - \text{Tr}[a^{-1}F_\gamma] \geq 0$  and  $|F_\gamma^{-1}|_\infty < \infty$  by (3.2) of Assumption **F**, it follows from the Lipschitz property of  $F$  that

$$\begin{aligned} |\mathbf{T}_h[\varphi] - \mathbf{T}_h[\psi]|_\infty &\leq |f|_\infty (1 - a^{-1} \cdot F_\gamma + \mathbb{E}[|A_h|^2] + Ch) \\ &= |f|_\infty \left( 1 + \frac{h}{4} F_p^\top F_\gamma^{-1} F_p + Ch \right) \leq |f|_\infty(1 + Ch). \end{aligned}$$

To prove that the family  $(v^h)_h$  is bounded, we proceed by backward induction. By the assumption of the lemma  $v^h(T, \cdot) = g$  is  $\mathbb{L}^\infty$ -bounded. We next fix some  $i < n$  and we assume that  $|v^h(t_j, \cdot)|_\infty \leq C_j$  for every  $i + 1 \leq j \leq n - 1$ . Proceeding as in the proof of Lemma 3.2 with  $\varphi \equiv v^h(t_{i+1}, \cdot)$  and  $\psi \equiv 0$ , we see that

$$\left| v^h(t_i, \cdot) \right|_\infty \leq h |F(t, x, 0, 0, 0)| + C_{i+1}(1 + Ch).$$

Since  $F(t, x, 0, 0, 0)$  is bounded by Assumption **F**, it follows from the discrete Gronwall inequality that  $|v^h(t_i, \cdot)|_\infty \leq Ce^{CT}$  for some constant  $C$  independent of  $h$ .  $\square$

### 3.3 Derivation of the rate of convergence

The proof of Theorem 3.2 is based on Barles and Jakobsen [6], which uses the Krylov method of shaking coefficients [21].

### 3.3.1 Comparison result for the scheme

Because  $F$  does not satisfy the standard monotonicity condition (3.6) of Barles and Souganidis [7], we need to introduce the nonlinearity  $\bar{F}$  of Remark 3.8 so that  $\bar{F}$  satisfies (3.5). Let  $u^h$  be the family of functions defined by

$$u^h(T, \cdot) = g \text{ and } u^h(t_i, x) = \bar{\mathbf{T}}_h[u^h](t_i, x), \quad (3.9)$$

where for a function  $\psi$  from  $[0, T] \times \mathbb{R}^d$  to  $\mathbb{R}$  with exponential growth:

$$\bar{\mathbf{T}}_h[\psi](t, x) := \mathbb{E} \left[ \psi(t+h, \hat{X}_h^{t,x}) \right] + h\bar{F}(\cdot, \mathcal{D}_h\psi)(t, x),$$

and set

$$\bar{v}^h(t_i, x) := e^{-\theta(T-t_i)} u^h(t_i, x), \quad i = 0, \dots, n. \quad (3.10)$$

The following result shows that the difference  $v^h - \bar{v}^h$  is of higher order, and thus reduces the error estimate problem to the analysis of the difference  $\bar{v}^h - v$ .

**Lemma 3.4** *Under Assumption  $\mathbf{F}$ , we have*

$$\limsup_{h \searrow 0} h^{-1} |(v^h - \bar{v}^h)(t, \cdot)|_\infty < \infty.$$

**Proof.** By definition of  $\bar{F}$ , we directly calculate that:

$$\bar{v}^h(t, x) = e^{-\theta h} (1 + h\theta) \mathbb{E}[\bar{v}^h(t+h, \hat{X}_h^{t,x})] + hF(t+h, x, \mathcal{D}_h \bar{v}^h(t, x)).$$

Since  $1 + h\theta = e^{\theta h} + O(h^2)$ , this shows that  $\bar{v}^h(t, x) = \mathbf{T}_h[\bar{v}^h](t, x) + O(h^2)$ . By lemma 3.3, we conclude that:

$$|(\bar{v}^h - v^h)(t, \cdot)|_\infty \leq (1 + Ch) |(\bar{v}^h - v^h)(t+h, \cdot)|_\infty + O(h^2),$$

which shows by the Gronwall inequality that  $|(\bar{v}^h - v^h)(t, \cdot)|_\infty \leq O(h)$  for all  $t \leq T - h$ .  $\square$

By Remark 3.8, the operator  $\bar{\mathbf{T}}_h$  satisfies the standard monotonicity condition (3.6):

$$\varphi \leq \psi \implies \bar{\mathbf{T}}_h[\varphi] \leq \bar{\mathbf{T}}_h[\psi]. \quad (3.11)$$

The key-ingredient for the derivation of the error estimate is the following comparison result for the scheme.

**Proposition 3.1** *Let Assumption  $\mathbf{F}$  holds true, and consider two arbitrary bounded functions  $\varphi$  and  $\psi$  satisfying:*

$$h^{-1} (\varphi - \bar{\mathbf{T}}_h[\varphi]) \leq g_1 \text{ and } h^{-1} (\psi - \bar{\mathbf{T}}_h[\psi]) \geq g_2 \quad (3.12)$$

for some bounded functions  $g_1$  and  $g_2$ . Then, for every  $i = 0, \dots, n$ :

$$(\varphi - \psi)(t_i, x) \leq e^{\beta(T-t_i)} |(\varphi - \psi)^+(T, \cdot)|_\infty + (T-h)e^{\beta(T-t_i)} |(g_1 - g_2)^+|_\infty \quad (3.13)$$

for some parameter  $\beta > |F_r|_\infty$ .

To prove this comparison result, we need the following strengthening of the monotonicity condition:

**Lemma 3.5** *Let Assumption **F** hold true, let  $\lambda := |F_r|_\infty$ , and choose some parameter  $\beta > |F_r|_\infty$ . Then, for every  $a \in \mathbb{R}$ ,  $b \in \mathbb{R}_+$  and  $c \in \mathbb{R}$ , and every bounded functions  $\varphi \leq \psi$ , the function  $\delta(t) := e^{\beta(T-t)}(a + b(T-t)) + c$  satisfies:*

$$\overline{\mathbf{T}}_h[\varphi + \delta](t, x) \leq \overline{\mathbf{T}}_h[\psi](t, x) + \delta(t) - h(b - \lambda c), \quad t \leq T - h, \quad x \in \mathbb{R}^d.$$

**Proof.** Because  $\delta$  does not depend on  $x$ , we have  $\mathcal{D}_h[\varphi + \delta] = \mathcal{D}_h\varphi + \delta(t+h)e_1$ , where  $e_1 := (1, 0, 0)$ . Then, it follows from the regularity of  $\overline{F}$  that there exist some  $\xi$  such that:

$$\overline{F}(t+h, x, \mathcal{D}_h[\varphi + \delta](t, x)) = \overline{F}(t+h, x, \mathcal{D}_h\varphi(t, x)) + \delta(t+h)\overline{F}_r(t+h, x, \xi e_1 + \mathcal{D}_h\varphi(t, x)),$$

and

$$\begin{aligned} \overline{\mathbf{T}}_h[\varphi + \delta](t, x) &= \delta(t+h) + \mathbb{E}[\varphi(t+h, \hat{X}_h^{t,x})] + h\overline{F}(t+h, x, \mathcal{D}_h\varphi(t, x)) \\ &\quad + h\delta(t+h)\overline{F}_r(t+h, x, \xi e_1 + \mathcal{D}_h\varphi(t, x)) \\ &= \overline{\mathbf{T}}_h[\varphi](t, x) + \delta(t+h) \{1 + h\overline{F}_r(t+h, x, \xi e_1 + \mathcal{D}_h\varphi(t, x))\} \\ &\leq \overline{\mathbf{T}}_h[\varphi](t, x) + (1 + h|\overline{F}_r|_\infty) \delta(t+h). \end{aligned}$$

Since  $\overline{\mathbf{T}}_h$  satisfies the standard monotonicity condition (3.11), this provides:

$$\overline{\mathbf{T}}_h[\varphi + \delta](t, x) \leq \overline{\mathbf{T}}_h[\psi](t, x) + \delta(t) + \zeta(t), \quad \text{where } \zeta(t) := (1 + h|\overline{F}_r|_\infty) \delta(t+h) - \delta(t).$$

It remains to prove that  $\zeta(t) \leq h(-b + \lambda c)$  for a suitable choice of  $\lambda$  and  $\beta$ . We now choose  $\beta > |\overline{F}_r|_\infty$  and  $\lambda = |\overline{F}_r|_\infty$ . From the smoothness of  $\delta$ , we have  $\delta(t+h) - \delta(t) = h\delta'(\bar{t})$  for some  $\bar{t} \in [t, t+h]$ . Then, since  $\delta$  is decreasing in  $t$ , we see that

$$h^{-1}\zeta(t) = \delta'(\bar{t}) + \lambda\delta(t+h) \leq \delta'(\bar{t}) + \lambda\delta(\bar{t}) \leq -be^{\beta(T-\bar{t})} + \lambda c,$$

and the required estimate follows from the restriction  $b \geq 0$ .  $\square$

**Proof of Proposition 3.1.** We may refer directly to the similar result of [6]. However in our context, we give the following simpler proof. Observe that we may assume without loss of generality that

$$\varphi(T, \cdot) \leq \psi(T, \cdot) \quad \text{and} \quad g_1 \leq g_2. \quad (3.14)$$

Indeed, one can otherwise consider the function

$$\bar{\psi} := \psi + e^{\beta(T-t)}(a + b(T-t)) \quad \text{where } a = |(\varphi - \psi)^+(T, \cdot)|_\infty, \quad b = |(g_1 - g_2)^+|_\infty,$$

and  $\beta$  is the parameter defined in the previous Lemma 3.5, so that  $\bar{\psi}(T, \cdot) \geq \varphi(T, \cdot)$  and, by Lemma (3.5),  $\bar{\psi}(t, x) - \overline{\mathbf{T}}_h[\bar{\psi}](t, x) \geq h(g_1 \vee g_2)$ . Hence (3.14) holds true for  $\varphi$  and  $\bar{\psi}$ .

We now prove the required result by induction. First  $\varphi(T, \cdot) \leq \psi(T, \cdot)$  by (3.14). We next assume that  $\varphi(t+h, \cdot) \leq \psi(t+h, \cdot)$  for some  $t+h \leq T$ . Since  $\overline{\mathbf{T}}_h$  satisfies the standard monotonicity condition (3.11), it follows from (3.14) that

$$\overline{\mathbf{T}}_h[\varphi](t, x) \leq \overline{\mathbf{T}}_h[\psi](t, x).$$

On the other hand, under (3.14), the hypothesis of the lemma implies:

$$\varphi(t, x) - \overline{\mathbf{T}}_h[\varphi](t, x) \leq \psi(t, x) - \overline{\mathbf{T}}_h[\psi](t, x).$$

Then  $\varphi(t, \cdot) \leq \psi(t, \cdot)$ . □

### 3.3.2 Proof of Theorem 3.2 (i)

Under the conditions of Assumption HJB on the coefficients, we may build a bounded subsolution  $\underline{v}^\varepsilon$  of the nonlinear PDE, by the method of shaking the coefficients, which is Lipschitz in  $x$ ,  $1/2$ -Hölder continuous in  $t$ , and approximates uniformly the solution  $v$ :

$$v - \varepsilon \leq \underline{v}^\varepsilon \leq v.$$

Let  $\rho(t, x)$  be a  $C^\infty$  positive function supported in  $\{(t, x) : t \in [0, 1], |x| \leq 1\}$  with unit mass, and define

$$\underline{w}^\varepsilon(t, x) := v^\varepsilon * \rho^\varepsilon \quad \text{where} \quad \rho^\varepsilon(t, x) := \frac{1}{\varepsilon^{d+2}} \rho\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right) \quad (3.15)$$

so that, from the convexity of the operator  $F$ ,

$$\underline{w}^\varepsilon \text{ is a subsolution of (2.1), } |\underline{w}^\varepsilon - v| \leq 2\varepsilon. \quad (3.16)$$

Moreover, since  $v^\varepsilon$  is Lipschitz in  $x$ , and  $1/2$ -Hölder continuous in  $t$ ,

$$\underline{w}^\varepsilon \text{ is } C^\infty, \text{ and } \left| \partial_t^{\beta_0} D^\beta \underline{w}^\varepsilon \right| \leq C \varepsilon^{1-2\beta_0-|\beta|_1} \text{ for any } (\beta_0, \beta) \in \mathbb{N} \times \mathbb{N}^d \setminus \{0\}, \quad (3.17)$$

where  $|\beta|_1 := \sum_{i=1}^d \beta_i$ , and  $C > 0$  is some constant. As a consequence of the consistency result of Lemma 3.1 above, we know that

$$\mathcal{R}_h[\underline{w}^\varepsilon](t, x) := \frac{\underline{w}^\varepsilon(t, x) - \mathbf{T}_h[\underline{w}^\varepsilon](t, x)}{h} + \mathcal{L}^X \underline{w}^\varepsilon(t, x) + F(\cdot, \underline{w}^\varepsilon, D\underline{w}^\varepsilon, D^2\underline{w}^\varepsilon)(t, x)$$

converges to 0 as  $h \rightarrow 0$ . The next key-ingredient is to estimate the rate of convergence of  $\mathcal{R}_h[\underline{w}^\varepsilon]$  to zero:

**Lemma 3.6** *For a family  $\{\varphi_\varepsilon\}_{0 < \varepsilon < 1}$  of smooth functions satisfying (3.17), we have:*

$$|\mathcal{R}_h[\varphi_\varepsilon]|_\infty \leq R(h, \varepsilon) := C h \varepsilon^{-3} \text{ for some constant } C > 0.$$

The proof of this result is reported at the end of this section. From the previous estimate together with the subsolution property of  $w^\varepsilon$ , we see that  $\underline{w}^\varepsilon \leq \overline{\mathbf{T}}_h[\underline{w}^\varepsilon] + Ch^2\varepsilon^{-3}$ . Then, it follows from Proposition 3.1 that

$$\underline{w}^\varepsilon - \overline{v}^h \leq C|(\underline{w}^\varepsilon - \overline{v}^h)(T, \cdot)|_\infty + Ch\varepsilon^{-3} \leq C(\varepsilon + h\varepsilon^{-3}). \quad (3.18)$$

We now use (3.16) and (3.18) to conclude that

$$v - \overline{v}^h \leq v - \underline{w}^\varepsilon + \underline{w}^\varepsilon - \overline{v}^h \leq C(\varepsilon + h\varepsilon^{-3}).$$

Minimizing the right hand-side estimate over the choice of  $\varepsilon > 0$ , this implies the upper bound on the error  $v - v^h$ :

$$v - \overline{v}^h \leq Ch^{1/4}. \quad (3.19)$$

### 3.3.3 Proof of Theorem 3.2 (ii)

The results of the previous section, together with the reinforced assumption HJB+, allow to apply the switching system method of Barles and Jakobsen [6] which provides the lower bound on the error:

$$v - \bar{v}^h \geq -\inf_{\varepsilon > 0} \{C\varepsilon^{1/3} + R(h, \varepsilon)\} = -C'h^{1/10},$$

for some constants  $C, C' > 0$ . The required rate of convergence follows again from Lemma 3.4 which states that the difference  $v^h - \bar{v}^h$  is dominated by the above rate of convergence.

**Proof of Lemma 3.6** Notice that the evolution of the Euler approximation  $\hat{X}_h^{t,x}$  between  $t$  and  $t+h$  is driven by a constant drift  $\mu(t, x)$  and a constant diffusion  $\sigma(t, x)$ . Since  $D\varphi_\varepsilon$  is bounded, it follows from Itô's formula that:

$$\frac{1}{h} \left[ \mathbb{E}\varphi_\varepsilon(t+h, \hat{X}_h^x) - \varphi_\varepsilon(t, x) \right] - \mathcal{L}^X \varphi_\varepsilon(t, x) = \frac{1}{h} \mathbb{E} \int_t^{t+h} \left( \mathcal{L}^{\hat{X}^{t,x}} \varphi_\varepsilon(u, \hat{X}_u^x) - \mathcal{L}^X \varphi_\varepsilon(t, x) \right) du,$$

where  $\mathcal{L}^{\hat{X}^{t,x}}$  is the Dynkin operator associated to the Euler scheme:

$$\mathcal{L}^{\hat{X}^{t,x}} \varphi(t', x') = \partial_t \varphi(t', x') + \mu(t, x) D\varphi(t', x') + \frac{1}{2} \text{Tr} [a(t, x) D^2 \varphi(t', x')].$$

Applying again Itô's formula, and using the fact that  $\mathcal{L}^{\hat{X}^{t,x}} D\varphi_\varepsilon$  is bounded, leads to

$$\frac{1}{h} \left[ \mathbb{E}\varphi_\varepsilon(t+h, \hat{X}_h^x) - \varphi_\varepsilon(t, x) \right] - \mathcal{L}^X \varphi_\varepsilon(t, x) = \frac{1}{h} \mathbb{E} \int_t^{t+h} \int_t^u \mathcal{L}^{\hat{X}^{t,x}} \mathcal{L}^{\hat{X}^{t,x}} \varphi_\varepsilon(s, \hat{X}_s^x) ds du.$$

Using the boundedness of the coefficients  $\mu$  and  $\sigma$ , it follows from (3.17) that for  $\varepsilon \in (0, 1)$ :

$$\left| \frac{\mathbb{E}\varphi_\varepsilon(t+h, \hat{X}_h^x) - \varphi_\varepsilon(t, x)}{h} - \mathcal{L}^X \varphi_\varepsilon(t, x) \right| \leq R_0(h, \varepsilon) := C h \varepsilon^{-3}.$$

Step 2 This implies that

$$\begin{aligned} |\mathcal{R}_h[\varphi_\varepsilon](t, x)| &\leq \left| \frac{\mathbb{E}\varphi_\varepsilon(t+h, \hat{X}_h^{t,x}) - \varphi_\varepsilon(t, x)}{h} - \mathcal{L}^X \varphi_\varepsilon(t, x) \right| \\ &\quad + |F(x, \varphi_\varepsilon(t, x), D\varphi_\varepsilon(t, x), D^2\varphi_\varepsilon(t, x)) - F(\cdot, \mathcal{D}_h[\varphi_\varepsilon](t, x))| \\ &\leq R_0(h, \varepsilon) + C \sum_{k=0}^2 \left| \mathbb{E} D^k \varphi_\varepsilon(t+h, \hat{X}_h^{t,x}) - D^k \varphi_\varepsilon(t, x) \right| \end{aligned} \quad (3.20)$$

by the Lipschitz continuity of the nonlinearity  $F$ .

By a similar calculation as in Step 1, we see that:

$$|\mathbb{E} D^i \varphi_\varepsilon(t+h, \hat{X}_h^{t,x}) - D^i \varphi_\varepsilon(t, x)| \leq C h \varepsilon^{-1-i}, \quad i = 0, 1, 2,$$

which, together with (3.20), provides the required result.  $\square$

**Remark 3.7** Let  $\mu$  and  $\sigma$  be Lipschitz in  $x$  and  $1/2$ -Hölder-continuous in  $t$ . Then under Assumption **F**, the function  $\bar{v}^h$  is Lipschitz in  $x$  and  $1/2$ -Hölder-continuous in  $t$ , uniformly in  $h$ .

We only report the proof that  $\bar{v}^h$  is Lipschitz in  $x$ , the  $1/2$ -Hölder continuity follows the same line of argument. We also report the following calculation in the one-dimensional case  $d = 1$  in order to simplify the presentation. For fixed  $t \in [0, T - h]$ , we argue as in the proof of Lemma 3.2 to see that for  $x, x' \in \mathbb{R}^d$  with  $x \neq x'$ :

$$\begin{aligned} |\bar{v}^h(t, x) - \bar{v}^h(t, x')| &\leq \left| \mathbb{E} \left[ \left( \bar{v}^h(t+h, \hat{X}_h^{t,x}) - \bar{v}^h(t+h, \hat{X}_h^{t,x'}) \right) \right. \right. \\ &\quad \left. \left. (1 - \sigma^{-2}(x)F_\gamma + h^{-1}F_\gamma\sigma^{-2}(x)W_h^2 + Ch) \right. \right. \\ &\quad \left. \left. + F_\gamma D\bar{v}^h(t+h, \hat{X}_h^{t,x'}) (\sigma^{-1}(x) - \sigma^{-1}(x')) W_h \right] \right|, \end{aligned}$$

for some positive constant  $C$ . Assuming that  $v^h(t+h, \cdot)$  is Lipschitz with constant  $L_{t+h}$ , this shows that

$$\limsup_{|x-x'| \rightarrow 0} \frac{|\bar{v}^h(t, x) - \bar{v}^h(t, x')|}{|x-x'|} \leq \left| \mathbb{E} \left[ D\bar{v}^h(t+h, \hat{X}_h^{t,x}) \left\{ (1+ah+b\sqrt{h}N)(1-\alpha+\alpha N^2+Ch) - \beta b\sqrt{h}N \right\} \right] \right|,$$

where  $a := \mu'(x)$ ,  $b := \sigma'(x)$ , and  $\beta := \sigma^{-2}(x)F_\gamma$  are all bounded and deterministic, and  $N$  is a standard normal distribution. Let  $\tilde{\mathbb{P}}$  be the probability measure equivalent to  $\mathbb{P}$  with density  $Z := (1 - \alpha + \alpha N^2 + Ch)/(1 + Ch)$ . Then

$$\begin{aligned} \limsup_{|x-x'| \rightarrow 0} \frac{|\bar{v}^h(t, x) - \bar{v}^h(t, x')|}{|x-x'|} &\leq L_{t+h} \mathbb{E}^{\tilde{\mathbb{P}}} \left[ \left| (1+ah+b\sqrt{h}N)(1+Ch) - \beta b\sqrt{h}N Z^{-1} \right| \right] \\ &\leq L_{t+h} \left( \mathbb{E}^{\tilde{\mathbb{P}}} \left[ \left| (1+ah+b\sqrt{h}N)(1+Ch) - \beta b\sqrt{h}N Z^{-1} \right|^2 \right] \right)^{1/2} \\ &= L_{t+h} \left( \mathbb{E} \left[ Z \left| (1+ah+b\sqrt{h}N)(1+Ch) - \beta b\sqrt{h}N Z^{-1} \right|^2 \right] \right)^{1/2}. \end{aligned}$$

We next develop the calculation of the square term inside the expectation, and collect the terms which are multiplied by the same power of  $h$ . The crucial observation is that the coefficient  $\sqrt{h}$  is multiplied by  $(1+Ch)bNZ - 2\beta bN$  which has zero mean. Hence

$$\limsup_{|x-x'| \rightarrow 0} \frac{|\bar{v}^h(t, x) - \bar{v}^h(t, x')|}{|x-x'|} \leq L_{t+h} \sqrt{1 + C'h}$$

for some constant  $C'$ . Since the scheme consists of  $T/h$  steps, the latter estimate shows that  $\bar{v}^h(t, \cdot)$  is Lipschitz with constant bounded by:

$$L_T \sqrt{1 + C'h}^{T/h} \sim L_T e^{TC'/2}.$$

□

### 3.4 The rate of convergence in the linear case

In this subsection, we specialize the discussion to the linear one-dimensional case

$$F(\gamma) = c\gamma, \quad (3.21)$$

for some  $c > 0$ . The multi-dimensional case  $d > 1$  can be handled similarly. Assuming that  $g$  is bounded, the linear PDE (2.1)-(2.2) has a unique bounded solution

$$v(t, x) = \mathbb{E} \left[ g \left( x + \sqrt{1 + 2c} W_{T-t} \right) \right] \quad \text{for } (t, x) \in [0, T] \times \mathbb{R}^d. \quad (3.22)$$

We also observe that this solution  $v$  is  $C^\infty([0, T] \times \mathbb{R})$  with

$$D^k v(t, x) = \mathbb{E} \left[ g^{(k)} \left( x + \sqrt{1 + 2c} W_{T-t} \right) \right], \quad t < T, \quad x \in \mathbb{R}. \quad (3.23)$$

This shows in particular that  $v$  has bounded derivatives of any order, whenever the terminal data  $g$  is  $C^\infty$  and has bounded derivatives of any order.

Of course, one can use the classical Monte Carlo estimate to produce an approximation of the function  $v$  of (3.25). The objective of this section is to analyze the error of the numerical scheme outlined in the previous sections. Namely:

$$v^h(T, \cdot) = g, \quad v^h(t_{i-1}, x) = \mathbb{E} \left[ v^h(t_i, x + W_h) \right] + ch \mathbb{E} \left[ v^h(t_i, x + W_h) H_2^h \right], \quad i \leq n. \quad (3.24)$$

Here,  $\sigma = 1$  and  $\mu = 0$  are used to write the above scheme.

**Proposition 3.2** *Consider the linear  $F$  of (3.24), and assume that  $D^{(2k+1)}v$  is bounded for every  $k \geq 0$ . Then*

$$\limsup_{h \rightarrow 0} h^{-1/2} |v^h - v|_\infty < \infty.$$

**Proof.** Since  $v$  has bounded first derivative with respect to  $x$ , it follows from Itô's formula that:

$$v(t, x) = \mathbb{E} [v(t+h, x + W_h)] + c \mathbb{E} \left[ \int_0^h \Delta v(t+s, v + W_s) ds \right],$$

Then, in view of Lemma 2.1, the error  $u := v - v^h$  satisfies  $u(t_n, X_{t_n}) = 0$  and for  $i \leq n-1$ :

$$\begin{aligned} u(t_i, X_{t_i}) &= \mathbb{E}_i [u(t_{i+1}, X_{t_{i+1}})] + ch \mathbb{E}_i [\Delta u(t_{i+1}, X_{t_{i+1}})] \\ &\quad + c \mathbb{E}_i \int_0^h [\Delta v(ih+s, X_{ih+s}) - \Delta v((i+1)h, X_{(i+1)h})] ds, \end{aligned} \quad (3.25)$$

where  $\mathbb{E}_i := \mathbb{E}[\cdot | \mathcal{F}_{t_i}]$  is the expectation operator conditional on  $\mathcal{F}_{t_i}$ .

*Step 1* Set

$$a_i^k := \mathbb{E} [\Delta^k u(t_i, X_{t_i})], \quad b_i^k := \mathbb{E} \int_0^h [\Delta^k v(t_{i-1} + s, X_{t_{i-1} + s}) - \Delta^k v(t_i, X_{t_i})] ds,$$

and we introduce the matrices

$$A := \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & -1 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}, \quad B := \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix},$$

and we observe that (3.28) implies that the vectors  $a^k := (a_1^k, \dots, a_n^k)^\top$  and  $b^k := (b_1^k, \dots, b_n^k)^\top$  satisfy  $Aa^k = chBa^{k+1} + cBb^k$  for all  $k \geq 0$ , and therefore:

$$a^k = chA^{-1}Ba^{k+1} + cA^{-1}Bb^k \quad \text{where} \quad A^{-1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \quad (3.26)$$

By direct calculation, we see that the powers  $(A^{-1}B)^k$  are given by:

$$(A^{-1}B)_{i,j}^k = \mathbf{1}_{\{j \geq i+k\}} \binom{j-i-1}{k-1} \quad \text{for all } k \geq 1 \text{ and } i, j = 1, \dots, n.$$

In particular, because  $a_n^k = 0$ ,  $(A^{-1}B)^{n-1}a^k = 0$ . Iterating (3.29), this provides:

$$a^0 = ch(A^{-1}B)a^1 + c(A^{-1}B)b^0 = \dots = \sum_{k=0}^{n-2} c^{k+1} h^k (A^{-1}B)^{k+1} b^k,$$

and therefore:

$$u(0, x) = a_1^0 = c \sum_{k=0}^{n-2} (ch)^k (A^{-1}B)_{1,j}^{k+1} b^k. \quad (3.27)$$

Because of

$$(A^{-1}B)_{1,j}^k = \mathbf{1}_{\{j \geq 1+k\}} \binom{j-2}{k-1} \quad \text{for all } k \geq 1 \text{ and } j = 1, \dots, n.$$

we can write (3.30):

$$u(0, x) = c \sum_{k=0}^{n-2} (ch)^k \sum_{j=k+2}^n \binom{j-2}{k} b_j^{k-1}.$$

By changing the order of the summations in the above we conclude that:

$$u(0, x) = c \sum_{j=2}^n \sum_{k=0}^{j-2} (ch)^k \binom{j-2}{k} b_j^{k-1}. \quad (3.28)$$

Step 2 From our assumption that  $D^{2k+1}v$  is  $\mathbb{L}^\infty$ -bounded for every  $k \geq 0$ , it follows that

$$|b_j^k| \leq \mathbb{E} \left[ \int_{t_{i-1}}^{t_i} \left| \Delta^k v(s, X_s) - \Delta^k v(t_j, X_{t_j}) \right| ds \right] \leq Ch^{3/2}$$

for some constant  $C$ . We then deduce from (3.31) that:

$$|u(0, x)| \leq cCh^{3/2} \sum_{j=2}^n \sum_{k=0}^{j-2} (ch)^k \binom{j-2}{k}$$

So,

$$|u(0, x)| \leq cCh^{3/2} \sum_{j=2}^n (1+ch)^{j-2} = cCh^{3/2} \frac{(1+ch)^{n-1} - 1}{ch} \leq \text{Const} \sqrt{h}$$

## 4 Probabilistic Numerical Scheme

In order to implement the backward scheme (2.4), we still need to discuss the numerical computation of the conditional expectations involved in the definition of the operators  $\mathbf{T}_h$  in (2.5). In view of the Markov feature of the process  $X$ , these conditional expectations reduce to simple regressions. Motivated by the problem of American options in financial mathematics, various methods have been introduced in the literature for the numerical approximation of these regressions. We refer to [9] and [17] for a detailed discussion.

The chief object of this section is to investigate the asymptotic properties of our suggested numerical method when the expectation operator  $\mathbb{E}$  in (2.4) is replaced by some estimator  $\hat{\mathbb{E}}^N$  corresponding to a sample size  $N$ :

$$\tilde{\mathbf{T}}_h^N[\psi](t, x) := \hat{\mathbb{E}}^N \left[ \psi(t+h, \hat{X}_h^x) \right] + hF \left( \cdot, \hat{\mathcal{D}}_h \psi \right) (t, x), \quad (4.1)$$

$$\hat{\mathbf{T}}_h^N[\psi](t, x) := -K_h[\psi] \vee \tilde{\mathbf{T}}_h^N[\psi](t, x) \wedge K_h[\psi] \quad (4.2)$$

where

$$\hat{\mathcal{D}}_h \psi(t, x) := \hat{\mathbb{E}}^N \left[ \psi(t+h, \hat{X}_h^{t,x}) H_h(t, x) \right], \quad K_h[\psi] := \|\psi\|_\infty (1 + C_1 h) + C_2 h,$$

and  $C_1, C_2$  are derived from the proof of Lemma 3.3:

$$C_1 = \frac{1}{4} \|F_p^T F_\gamma^{-1} F_p + F_r\|_\infty \quad \text{and} \quad C_2 = \|F(t, x, 0, 0, 0)\|_\infty.$$

The above bounds are needed for technical reasons which were already observed in [9].

With these notations, the implementable numerical scheme is:

$$\hat{v}_N^h(t, x, \omega) = \hat{\mathbf{T}}_h^N[\hat{v}_N^h](t, x, \omega), \quad (4.3)$$

where  $\hat{\mathbf{T}}_h^N$  is defined in (4.1)-(4.2), and the presence of  $\omega$  throughout this section emphasizes the dependence of our estimator on the underlying sample.

Let  $\mathcal{R}$  be the family of random variables  $R$  of the form  $\psi(W_h)H_i(W_h)$  where  $\psi$  is a bounded function and  $H_i$ 's are the Hermite polynomials:

$$H_0(x) = 1, \quad H_1(x) = x \quad \text{and} \quad H_2(x) = x^T x - h \quad \forall x \in R^d.$$

**Assumption E** *There exist constants  $C, \lambda, \nu > 0$  such that  $\left\| \hat{\mathbb{E}}^N[R] - \mathbb{E}[R] \right\|_p \leq Ch^{-\lambda} N^{-\nu}$  for every  $R \in \mathcal{R}$ , for some  $p \geq 1$ .*

**Example 4.1** Consider the regression approximation based on the Malliavin integration by parts as introduced in Lions and Reigner [24], Bouchard, Ekeland and Touzi [3], and analyzed in the context of the simulation of backward stochastic differential equations by [9] and [12]. Then Assumption **E** is satisfied for every  $p > 1$  with the constants  $\lambda = \frac{d}{4p}$  and  $\nu = \frac{1}{2p}$ , see [9].

Our next main result establishes conditions on the sample size  $N$  and the time step  $h$  which guarantee the convergence of  $\hat{v}_N^h$  towards  $v$ .

**Theorem 4.1** *Let Assumptions **E** and **F** hold true, and assume that the fully nonlinear PDE (2.1) has strong comparison with growth  $q$ . Suppose in addition that*

$$\lim_{h \rightarrow 0} h^{\lambda+2} N_h^\nu = \infty. \quad (4.4)$$

*Assume that the final condition  $g$  is bounded Lipschitz, and the coefficients  $\mu$  and  $\sigma$  are bounded. Then, for almost every  $\omega$ :*

$$\hat{v}_{N_h}^h(\cdot, \omega) \longrightarrow v \quad \text{locally uniformly,}$$

where  $v$  is the unique viscosity solution of (2.1).

**Proof.** We adapt the argument of [7] to the present stochastic context. By Remark 3.8 and Lemma 3.4, we may assume without loss of generality that the strict monotonicity (3.5) holds.

By (4.2), we see that  $\hat{v}^h$  is uniformly bounded. So, we can define:

$$\hat{v}_*(t, x) := \liminf_{\substack{(t', x') \rightarrow (t, x) \\ h \rightarrow 0}} \hat{v}^h(t', x') \quad \text{and} \quad \hat{v}^*(t, x) := \limsup_{\substack{(t', x') \rightarrow (t, x) \\ h \rightarrow 0}} \hat{v}^h(t', x'). \quad (4.5)$$

Our objective is to prove that  $\hat{v}_*$  and  $\hat{v}^*$  are respectively viscosity supersolution and subsolution of (2.1). By the comparison assumption, we shall then conclude that they are both equal to the unique viscosity solution of the problem whose existence is given by Theorem 3.1. In particular, they are both deterministic functions.

We shall only report the proof of the supersolution property, the subsolution property follows from the same type of argument.

In order to prove that  $\hat{v}_*$  is a supersolution of (2.1), we consider  $(t_0, x_0) \in [0, T) \times \mathbb{R}^n$  together with a test function  $\varphi \in C^2([0, T) \times \mathbb{R}^n)$ , so that

$$0 = \min\{\hat{v}_* - \varphi\} = (\hat{v}_* - \varphi)(t_0, x_0).$$

By classical manipulations, we can find a sequence  $(t_n, x_n, h_n) \rightarrow (t_0, x_0, 0)$  so that  $\hat{v}^{h_n}(t_n, x_n) \rightarrow \hat{v}_*(t_0, x_0)$  and

$$(\hat{v}^{h_n} - \varphi)(t_n, x_n) = \min\{\hat{v}^{h_n} - \varphi\} =: C_n \rightarrow 0.$$

Then,  $\hat{v}^{h_n} \geq \varphi + C_n$ , and it follows from the monotonicity of the operator  $\mathbf{T}_h$  that:

$$\mathbf{T}_{h_n}[\hat{v}^{h_n}] \geq \mathbf{T}_{h_n}[\varphi + C_n].$$

By the definition of  $\hat{v}^{h_n}$  in (4.3), this provides:

$$\hat{v}^{h_n}(t, x) \geq \mathbf{T}_{h_n}[\varphi + C_n](t, x) - (\mathbf{T}_{h_n} - \hat{\mathbf{T}}_{h_n})[\hat{v}_n^h](t, x),$$

where, for ease of notations, the dependence on  $N_h$  has been dropped. Because  $\hat{v}^{h_n}(t_n, x_n) = \varphi(t_n, x_n) + C_n$ , the last inequality gives:

$$\varphi(t_n, x_n) + C_n - \mathbf{T}_{h_n}[\varphi + C_n](t_n, x_n) + h_n R_n \geq 0, \quad R_n := h_n^{-1}(\mathbf{T}_{h_n} - \hat{\mathbf{T}}_{h_n})[\hat{v}^{h_n}](t_n, x_n).$$

We claim that

$$R_n \longrightarrow 0 \quad \mathbb{P} - \text{a.s. along some subsequence.} \quad (4.6)$$

Then, after passing to the subsequence, dividing both sides by  $h_n$ , and sending  $n \rightarrow \infty$ , it follows from Lemma 3.1 that:

$$-\mathcal{L}^X \varphi - F(\cdot, \varphi, D\varphi, D^2\varphi) \geq 0,$$

which is the required supersolution property.

It remains to show (4.6). We start by bounding  $R_n$  with respect to the error of estimation of conditional expectation. By Lemma 3.3,  $|\mathbf{T}_{h_n}[\hat{v}^{h_n}]|_\infty \leq K_{h_n}$  and so by (4.2), we can write:

$$\left| (\mathbf{T}_{h_n} - \hat{\mathbf{T}}_{h_n})[\hat{v}^{h_n}](t_n, x_n) \right| \leq \left| (\mathbf{T}_{h_n} - \tilde{\mathbf{T}}_{h_n})[\hat{v}^{h_n}](t_n, x_n) \right|. \quad (4.7)$$

By the Lipschitz-continuity of  $F$ , we have:

$$\left| (\mathbf{T}_{h_n} - \hat{\mathbf{T}}_{h_n})[\hat{v}^{h_n}](t_n, x_n) \right| \leq C(\mathcal{E}_0 + h_n \mathcal{E}_1 + h_n \mathcal{E}_2).$$

where:

$$\mathcal{E}_i = |(\mathbb{E} - \hat{\mathbb{E}})[\hat{v}^{h_n}(t_n + h_n, X_{h_n}^{x_n}) H_i^{h_n}(t_n, x_n)]|$$

$$\left| (\mathbf{T}_{h_n} - \hat{\mathbf{T}}_{h_n})[\hat{v}^{h_n}](t_n, x_n) \right| \leq C \left( \left| (\mathbb{E} - \hat{\mathbb{E}})[R_n^0] \right| + \left| (\mathbb{E} - \hat{\mathbb{E}})[R_n^1] \right| + h_n^{-1} \left| (\mathbb{E} - \hat{\mathbb{E}})[R_n^2] \right| \right).$$

where  $R_n^i = \hat{v}^{h_n}(t_n + h_n, x_n + \sigma(x)W_h) H_i(W_h)$ ,  $i = 1, 2, 3$  and  $H_i$  is Hermite polynomial of degree  $i$ . This leads the following estimate for the error  $R_n$ :

$$|R_n| \leq \frac{C}{h_n} \left( \left| (\mathbb{E} - \hat{\mathbb{E}})[R_n^0] \right| + \left| (\mathbb{E} - \hat{\mathbb{E}})[R_n^1] \right| + h_n^{-1} \left| (\mathbb{E} - \hat{\mathbb{E}})[R_n^2] \right| \right). \quad (4.8)$$

Because  $R_n^i \in \mathcal{R}$  by Assumption **E** we have,:

$$\|R_n\|_p \leq C h_n^{-\lambda-2} N_{h_n}^{-\nu},$$

so by (4.4) we have  $\|R_n\|_p \longrightarrow 0$  which implies (4.6).  $\square$

We finally discuss the choice of the sample size so as to keep the same rate for the error bound.

**Theorem 4.2** *Let the nonlinearity  $F$  be as in Assumption **HJB**, and consider a regression operator satisfying Assumption **E**. Let the sample size  $N_h$  be such that*

$$\lim_{h \rightarrow 0} h^{\lambda + \frac{21}{10}} N_h^\nu > 0. \quad (4.9)$$

*Then, for any bounded Lipschitz final condition  $g$ , we have the following  $\mathbb{L}^p$ -bounds on the rate of convergence:*

$$\|v - \hat{v}^h\|_p \leq Ch^{1/10}.$$

**Proof.** By Remark 3.8 and Lemma 3.4, we may assume without loss of generality that the strict monotonicity (3.5) holds true.

We proceed as in the proof of Theorem 3.2 to see that

$$v - \hat{v}^h \leq v - v^h + v_h - \hat{v}^h = \varepsilon + R(h, \varepsilon) + v_h - \hat{v}^h.$$

Since  $\hat{v}^h$  satisfies (4.3),

$$h^{-1} \left( \hat{v}^h - \mathbf{T}_h[\hat{v}^h] \right) \geq -R_h[\hat{v}^h] \quad \text{where} \quad R_h[\varphi] := \frac{1}{h} \left| \left( \mathbf{T}_h - \hat{\mathbf{T}}_h \right) [\varphi] \right|,$$

where, in the present context,  $R_h[\hat{v}^h]$  is a non-zero stochastic term. By Proposition 3.1, it follows from the last inequality that:

$$v - \hat{v}^h \leq C \left( \varepsilon + R(h, \varepsilon) + R_h[\hat{v}^h] \right),$$

where the constant  $C > 0$  depends only on the Lipschitz coefficient of  $F$ ,  $\beta$  in Lemma 3.5 and the constant in Lemma 3.6.

Similarly, we follow the line of argument of the proof of Theorem 3.2 to show that a lower bound holds true, and therefore:

$$|v - \hat{v}^h| \leq C \left( \varepsilon^{1/3} + R(h, \varepsilon) + R_h[\hat{v}^h] \right),$$

We now use (4.9) and proceed as in the last part of the proof of Theorem 4.1 to deduce from (4.8) and Assumption **F** that

$$\|R_h[\hat{v}^h]\|_p \leq Ch^{1/10}.$$

With this choice of the sample size  $N$ , the above error estimate reduces to

$$\|\hat{v}^h - v\|_p \leq C \left( \varepsilon^{1/3} + R(h, \varepsilon) + h^{1/10} \right),$$

and the additional term  $h^{1/10}$  does not affect the minimization with respect to  $\varepsilon$ .  $\square$

**Example 4.2** Let us illustrate the convergence results of this section in the context of the Malliavin integration by parts repression method of [24] and [9] where  $\lambda = \frac{d}{4p}$  and  $\nu = \frac{1}{2p}$  for every  $p > 1$ . So, for the convergence result we need to choose  $N_h$  of the order of  $h^{-\alpha_0}$  with  $\alpha_0 > \frac{d}{2} + 4p$ . For the  $L^p$ -rate of convergence result, we need to choose  $N_h$  of the order of  $h^{-\alpha_1}$  with  $\alpha_1 \geq \frac{d}{2} + \frac{21p}{5}$ .

## 5 Numerical Results

In this section, we provide an application of the Monte Carlo-finite differences scheme suggested in this paper in the context of two different types of problems. We first consider the classical mean curvature flow equation as the simplest front propagation example. We test our backward probabilistic scheme on the example where the initial data is given by a sphere, for which an easy explicit solution is available. A more interesting geometric example in space dimensions 2 is also considered. We next consider the Hamilton-Jacobi-Bellman equation characterizing the classical optimal investment problem in financial mathematics. Here, we again test our scheme in dimension two where an explicit solution is available, and we consider more involved examples in space dimension 5, in addition to the time variable.

In all examples considered in this section the operator  $F(t, x, r, p, \gamma)$  does not depend on the  $r$ -variable. We shall then drop this variable from our notations, and we simply write the scheme as:

$$\begin{aligned} v^h(T, \cdot) &:= g \quad \text{and} \\ v^h(t_i, x) &:= \mathbb{E}[v^h(t_{i+1}, \hat{X}_h^x)] + hF(t_i, x, \mathcal{D}_h v^h(t_i, x)) \end{aligned} \tag{5.1}$$

where

$$\mathcal{D}_h \psi := (\mathcal{D}_h^1 \psi, \mathcal{D}_h^2 \psi),$$

and  $\mathcal{D}_h^1$  and  $\mathcal{D}_h^2$  are defined in Lemma 2.1. We recall from Remark 2.1 that:

$$\begin{aligned} \mathcal{D}_{2h}^2 \varphi(t_i, x) &= \mathbb{E} \left[ \varphi(t_i + 2h, \hat{X}_{2h}^{t_i, x}) (\sigma^T)^{-1} \frac{(W_{t_i+h} - W_{t_i})(W_{t_i+h} - W_{t_i})^T - h\mathbf{I}_d}{h^2} \sigma^{-1} \right] \\ &= \mathbb{E} \left[ \mathcal{D}_h^1 \varphi(t_i + h, \hat{X}_h^{t_i, x}) (\sigma^T)^{-1} \frac{W_{t_i+h} - W_{t_i}}{h} \right] \end{aligned}$$

The second representation is the one reported in [10] where the present backward probabilistic scheme was first introduced. These two representations induce two different numerical schemes because once the expectation operator  $\mathbb{E}$  is replaced by an approximation  $\hat{E}^N$ , equality does not hold anymore in the latter equation for finite  $N$ . In our numerical examples below, we provide results for both methods. The numerical schemes based on the first (resp. second) representation will be referred to as *scheme 1* (resp. *2*). An important outcome of our numerical experiments is that scheme 2 turns out to have a significantly better performance than scheme 1.

**Remark 5.1** The second scheme needs some final condition for  $\mathcal{D}_h^1 \varphi(T, X_h^{T-h, x})$ . Since  $g$  is smooth in all our examples, we set this final condition to  $\nabla g$ . Since the second scheme turns out to have a better performance, we may also use the final condition for  $Z$  suggested by the first scheme.

We finally discuss the choice of the regression estimator in our implemented examples. Two methods have been used:

- The first method is the basis projection *a la* Longstaff and Schwartz [25], as developed in [17]. We use regression functions with localized support : on each support the regression functions are chosen linear and the size of the support is adaptative according to the Monte Carlo distribution of the underlying process.
- The second method is based on the Malliavin integration by parts formula as suggested in [24] and further developed in [3]. In particular, the optimal exponential localization function  $\phi^k(y) = \exp(-\eta^k y)$  in each direction  $k$  is chosen as follows. The optimal parameter  $\eta_k$  is provided in [3] and should be chosen for each conditional expectation depending on  $k$ . Our numerical experiments however revealed that such optimal parameters do not provide sufficiently good performance, and more accurate results are obtained by choosing  $\eta_k = 5/\sqrt{\Delta t}$  for all values of  $k$ .

### 5.1 Mean curvature flow problem

The mean curvature flow equation describes the motion of a surface where each point moves along the inward normal direction with speed proportional to the mean curvature at that point. This geometric problem can be characterized as the zero-level set  $S(t) := \{x \in \mathbb{R}^d : v(t, x) = 0\}$  of a function  $v(t, x)$  depending on time and space satisfying the geometric partial differential equation:

$$v_t - \Delta v + \frac{Dv \cdot D^2 v Dv}{|Dv|^2} = 0 \quad \text{and} \quad v(0, x) = g(x) \quad (5.2)$$

and  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  is a bounded Lipschitz-continuous function. We refer to [27] for more details on the mean curvature problem and the corresponding stochastic representation.

To model the motion of a sphere in  $\mathbb{R}^d$  with radius  $2R > 0$ , we take  $g(x) := 2R - |x|^2$  so that  $g$  is positive inside the sphere and negative outside. We first solve the sphere problem in dimension 3. In this case, it is well-known that the surface  $S(t)$  is a sphere with a radius  $R(t)$  equal to  $2\sqrt{R^2 - t}$  for  $t \in (0, R^2)$ . Reversing time, we rewrite (5.2) for  $t \in (0, T)$  with  $T = R^2$ :

$$-v_t - \frac{1}{2}\sigma^2 \Delta v + F(x, Dv, D^2 v) = 0 \quad \text{and} \quad v(T, x) = g(x), \quad (5.3)$$

where

$$F(x, z, \gamma) := \gamma \left( \frac{1}{2}\sigma^2 - 1 \right) + \frac{z \cdot \gamma z}{|z|^2}.$$

We implement our Monte Carlo-finite differences scheme to provide an approximation  $\hat{v}^h$  of the function  $v$ . As mentioned before, we implement four methods: Malliavin integration by parts-based or basis projection-based regression, and scheme 1 or 2 for the representation of the Hessian.

Given the approximation  $\hat{v}^h$ , we deduce an approximation of the surface  $\hat{S}^h(t) := \{x \in \mathbb{R}^3 : \hat{v}^h(t, x) = 0\}$  by using a dichotomic gradient descent method using the estimation of the gradient  $\mathcal{D}^1 v$  estimated along the resolution. The dichotomy is stopped when the solution is localized within 0.01 accuracy.

**Remark 5.2** Of course the use of the gradient is not necessary in the present context where we know that  $S(t)$  is a sphere at any time  $t \in [0, T)$ . The algorithm described above is designed to handle any type of geometry.

**Remark 5.3** In our numerical experiments, the nonlinearity  $F$  is truncated so that it is bounded by an arbitrary value taken equal to 200.

Our numerical results show that Malliavin and basis projection methods give similar results. However, for a given number of sample paths, the basis projection method of [17] are slightly more accurate. Therefore, all results reported for this example correspond to the basis projection method.

Figure 1 provides results obtained with one million particles and  $10 \times 10 \times 10$  mesh with a time step equal to 0.0125. The diffusion coefficient  $\sigma$  is taken to be either 1 or 1.8. We observe that results are better with  $\sigma = 1$ . We also observe that the error increases near time 0.25 corresponding to an acceleration of the dynamics of the phenomenon, and suggesting that a thinner time step should be used at the end of simulation.

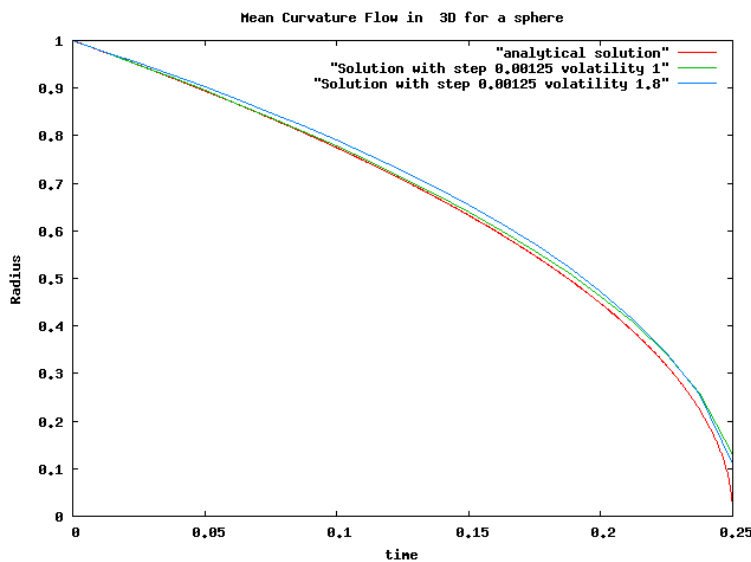


Figure 1: Solution of the mean curvature flow for the sphere problem

Figure 2 plots the difference between our calculation and the reference for scheme 1 and volatility 1 and 1.8 for varying time step. We notice that some points at time  $T = 0.25$  are missing due to a non convergence of the gradient method for a diffusion  $\sigma = 1.8$ . The corresponding results with scheme 2 are reported in figure 3. We observe that results for scheme 2 are slightly better than results for scheme 1.

We finally report in Figure 4 some numerical results for the mean curvature flow problem in dimension 2 with a more interesting geometry: the initial surface (i.e. the zero-level set for  $v$ ) consists of two disks with unit radius, with centers positioned at -1.5 and 1.5 and connected by a stripe of unit width. We give the resulting deformation with scheme 2 for a diffusion  $\sigma = 1$ , a time step  $h = 0.0125$ , and one million particles. Once again, the

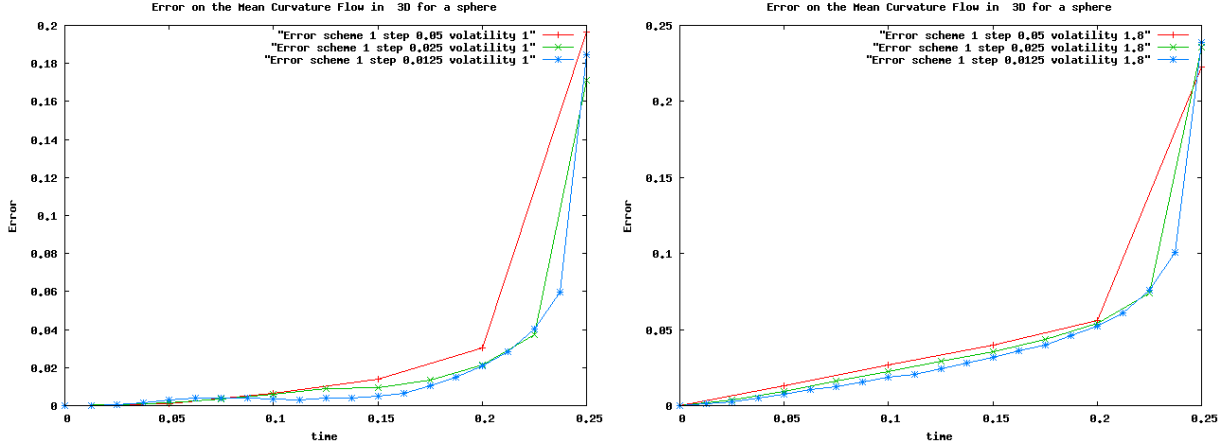


Figure 2: Mean curvature flow problem for different time step and diffusion: scheme 1

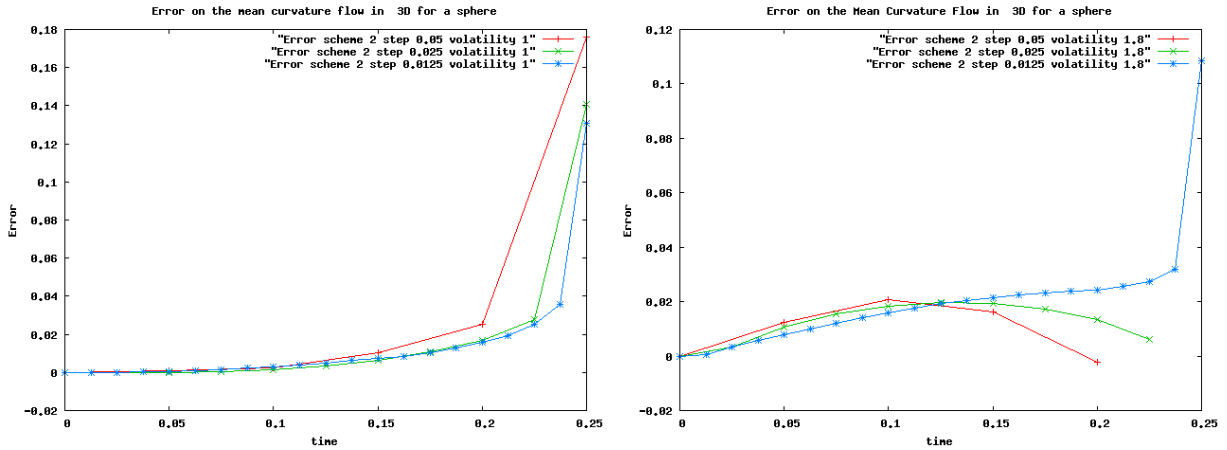


Figure 3: Mean curvature flow problem for different time step and diffusions: scheme 2

Malliavin integration by parts based regression method and the basis projection method with  $10 \times 10$  meshes produce similar results. We used 1024 points to describe the surface. One advantage of this method is the total parallelization that can be performed to solve the problem for different points on the surface : for the results given parallelization by Message Passing (MPI) was achieved.

## 5.2 Continuous-time portfolio optimization

We next report an application to the continuous-time portfolio optimization problem in financial mathematics. Let  $\{S_t, t \in [0, T]\}$  be an Itô process modeling the price evolution of  $n$  financial securities. The investor chooses an adapted process  $\{\theta_t, t \in [0, T]\}$  with values in  $\mathbb{R}^n$ , where  $\theta_t^i$  is the amount invested in the  $i$ -th security held at time  $t$ . In addition, the investor has access to a non-risky security (bank account) where the remaining part of his wealth is invested. The non-risky asset  $S^0$  is defined by a an adapted interest rates process  $\{r_t, t \in [0, T]\}$ , i.e.  $dS_t^0 = S_t^0 r_t dt, t \in [0, 1]$ . Then, the dynamics of the wealth process is

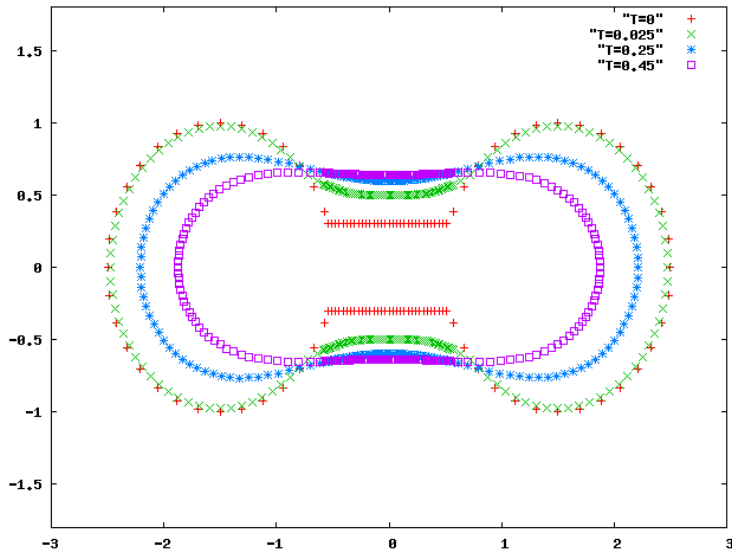


Figure 4: Mean curvature flow problem in 2D

described by:

$$dX_t^\theta = \theta_t \cdot \frac{dS_t}{S_t} + (X_t^\theta - \theta_t \cdot 1) \frac{dS_t^0}{S_t^0} = \theta_t \cdot \frac{dS_t}{S_t} + (X_t^\theta - \theta_t \cdot 1) r_t dt.$$

Let  $\mathcal{A}$  be the collection of all adapted processes  $\theta$  with values in  $\mathbb{R}^n$ , which are integrable with respect to  $S$  and such that the process  $X^\theta$  is uniformly bounded from below. Given an absolute risk aversion coefficient  $\eta > 0$ , the portfolio optimization problem is defined by:

$$v_0 := \sup_{\theta \in \mathcal{A}} \mathbb{E} \left[ -\exp \left( -\eta X_T^\theta \right) \right]. \quad (5.4)$$

Under fairly general conditions, this linear stochastic control problem can be characterized as the unique viscosity solution of the corresponding HJB equation. The main purpose of this subsection is to implement our Monte Carlo-finite differences scheme to derive an approximation of the solution of the fully nonlinear HJB equation in non-trivial situations where the state has a few dimensions. We shall first start by a two-dimensional example where an explicit solution of the problem is available. Then, we will present some results in a five dimensional situation.

### 5.2.1 A two dimensional problem

Let  $d = 1$ ,  $r_t = 0$  for all  $t \in [0, 1]$ , and assume that the security price process is defined by the Heston model [18]:

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{Y_t} S_t dW_t^{(1)} \\ dY_t &= k(m - Y_t) dt + c \sqrt{Y_t} \left( \rho dW_t^{(1)} + \sqrt{1 - \rho^2} dW_t^{(2)} \right), \end{aligned}$$

where  $W = (W^{(1)}, W^{(2)})$  is a Brownian motion in  $\mathbb{R}^2$ . In this context, it is easily seen that the portfolio optimization problem (5.4) does not depend on the state variable  $s$ . Given

an initial state at the time origin  $t$  given by  $(X_t, Y_t) = (x, y)$ , the value function  $v(t, x, y)$  solves the HJB equation:

$$\begin{aligned} v(T, x, y) &= -e^{-\eta x} \text{ and } 0 = -v_t - k(m - y)v_y - \frac{1}{2}c^2 y v_{yy} - \sup_{\theta \in \mathbb{R}} \left( \frac{1}{2}\theta^2 y v_{xx} + \theta(\mu v_x + \rho c y v_{xy}) \right) \\ &= -v_t - k(m - y)v_y - \frac{1}{2}c^2 y v_{yy} + \frac{(\mu v_x + \rho c y v_{xy})^2}{2y v_{xx}}. \end{aligned} \quad (5.5)$$

A quasi explicit solution of this problem was provided by Zariphopoulou [29]:

$$v(t, x, y) = -e^{-\eta x} \left\| \exp \left( -\frac{1}{2} \int_t^T \frac{\mu^2}{\tilde{Y}_s} ds \right) \right\|_{\mathbb{L}^{1-\rho^2}} \quad (5.6)$$

where the process  $\tilde{Y}$  is defined by

$$\tilde{Y}_t = y \quad \text{and} \quad d\tilde{Y}_t = (k(m - \tilde{Y}_t) - \mu c \rho) dt + c \sqrt{\tilde{Y}_t} dW_t.$$

In order to implement our Monte Carlo-finite differences scheme, we re-write (5.5) as:

$$-v_t - k(m - y)v_y - \frac{1}{2}c^2 y v_{yy} - \frac{1}{2}\sigma^2 v_{xx} + F(y, Dv, D^2v) = 0, \quad v(T, x, y) = -e^{-\eta x}, \quad (5.7)$$

where  $\sigma > 0$  and the nonlinearity  $F : \mathbb{R} \times \mathbb{R}^2 \times \mathbb{S}_2$  is given by:

$$F(y, z, \gamma) = \frac{1}{2}\sigma^2 \gamma_{11} + \frac{(\mu z_1 + \rho c y \gamma_{12})^2}{2y \gamma_{11}}.$$

Notice that the nonlinearity  $F$  does not satisfy Assumption **F**, we consider the truncated nonlinearity:

$$F_{\varepsilon, n} := \frac{1}{2}\sigma^2 \gamma_{11} - \sup_{\varepsilon \leq \theta \leq n} \left( \frac{1}{2}\theta^2 (y \vee \varepsilon) v_{xx} + \theta(\mu v_x + \rho c (y \vee \varepsilon) v_{xy}) \right),$$

for some  $\varepsilon, n > 0$  jointly chosen with  $\sigma$  so that Assumption **F** holds true. Under this form, the forward two-dimensional diffusion is defined by:

$$dX_t^{(1)} = \sigma dW_t^{(1)}, \quad \text{and} \quad dX_t^{(2)} = k(m - X_t^{(2)}) dt + c \sqrt{X_t^{(2)}} dW_t^{(2)}. \quad (5.8)$$

In order to guarantee the non-negativity of the discrete-time approximation of the process  $X^{(2)}$ , we use the implicit Milstein scheme [19]:

$$X_n^{(2)} = \frac{X_{n-1}^{(2)} + km\Delta t + c \sqrt{X_{n-1}^{(2)}} \xi_n \sqrt{\Delta t} + \frac{1}{4}c^2 \Delta (\xi_n^2 - 1)}{1 + k\Delta t} \quad (5.9)$$

where  $(\xi_n)_{n \geq 1}$  is a sequence of independent random variable with distribution  $\mathbf{N}(0, 1)$ .

Our numerical results correspond to the following values of the parameter:  $\mu = 0.15$ ,  $c = 0.2$ ,  $k = 0.1$ ,  $m = 0.3$ ,  $Y_0 = m$ ,  $\rho = 0$ . The initial value of the portfolio is  $x_0 = 1$ , the maturity  $T$  is taken equal to one year. With this parameters, the value function is computed from the quasi-explicit formula (5.6) to be  $v_0 = -0.3534$ .

We also choose  $n = 40$  for the truncation of the nonlinearity. This choice turned out to be critical as an initial choice of  $n = 10$  produced an important bias in the results.

The two schemes have been tested with the Malliavin and basis projection methods. The latter was applied with  $40 \times 10$  basis functions. We provide numerical results corresponding to 2 millions particles. Our numerical results show that the Malliavin and the basis projection methods produce very similar results, and achieve a good accuracy: with 2 millions particles, we calculate the variance of our estimates by performing 100 independent calculations:

- the results of the Malliavin method exhibit a standard deviation smaller than 0.005 for scheme one (except for a step equal to 0.025 and a volatility equal to 1.2 where standard deviation jumped to 0.038), 0.002 for scheme two ,
- the results of the basis projection method exhibit a standard deviation smaller than 0.002 for scheme 1 and 0.0009 for scheme two.

Figure 5 provides the plots of the errors obtained by the integration by parts-based regression with Schemes one and two. All solutions have been calculated as the average of 100 calculations. We first observe that for a small diffusion coefficient  $\sigma = 0.2$ , the numerical performance of the algorithm is very poor: surprisingly, the error increases as the time step shrinks to zero and the method seems to be biased. This numerical result hints that the requirement that the diffusion should dominate the nonlinearity in Theorem 3.1, i.e. the right hand-side inequality in (3.2) of Assumption **F**, might be a sharp condition. We also observe that Scheme one has a persistent bias even for a very small time step, while

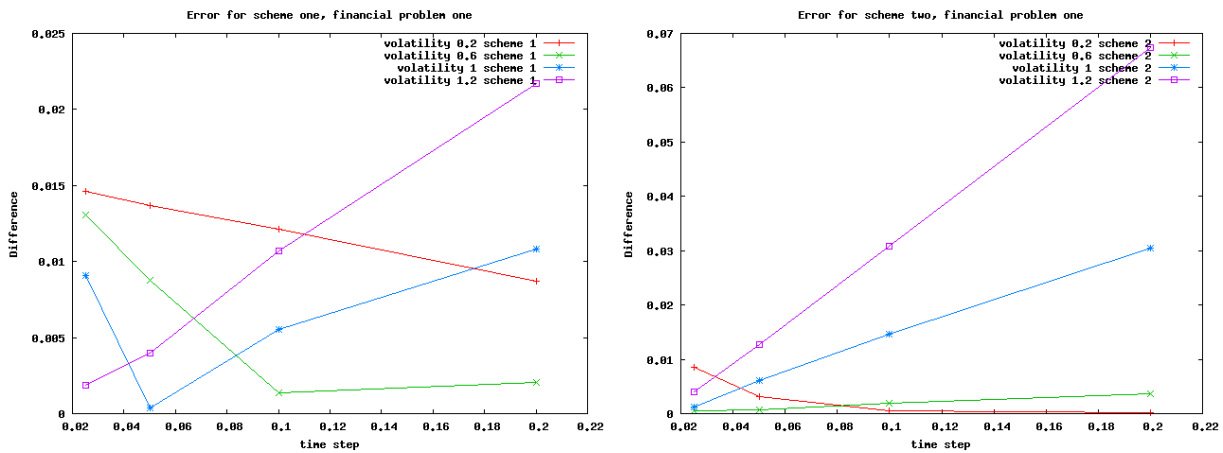


Figure 5: Difference between calculation and reference for scheme one and two

Scheme two exhibits a better convergence towards the solution.

### 5.2.2 A five dimensional example

We now let  $n = 2$ , and we assume that the interest rate process is defined by the Ornstein-Uhlenbeck process:

$$dr_t = \kappa(b - r_t)dt + \zeta dW_t^{(0)}.$$

While the price process of the second security is defined by a Heston model, the first security's price process is defined by a CEV-SV models, see e.g. [26] for a presentation of these models and there simulation:

$$\begin{aligned} dS_t^{(i)} &= \mu_i S_t^{(i)} dt + \sigma_i \sqrt{Y_t^{(i)}} S_t^{(i)\beta_i} dW_t^{(i,1)}, \quad \beta_2 = 1, \\ dY_t^{(i)} &= k_i (m_i - Y_t^{(i)}) dt + c_i \sqrt{Y_t^{(i)}} dW_t^{(i,2)} \end{aligned}$$

where  $(W^{(0)}, W^{(1,1)}, W^{(1,2)}, W^{(2,1)}, W^{(2,2)})$  is a Brownian motion in  $\mathbb{R}^5$ , and for simplicity we considered a zero-correlation between the security price process and its volatility process.

Since  $\beta_2 = 1$ , the value function of portfolio optimization problem (5.4) does not depend on the  $s^{(2)}$ -variable. Given an initial state  $(X_t, r_t, S_t^{(1)}, Y_t^{(1)}, Y_t^{(2)}) = (x, r, s_1, y_1, y_2)$  at the time origin  $t$ , the value function  $v(t, x, r, s_1, y_1, y_2)$  satisfies the HJB equation:

$$\begin{aligned} 0 &= -v_t - (\mathbf{L}^r + \mathbf{L}^Y + \mathbf{L}^{S^1})v - rxv_x \\ &\quad - \sup_{\theta=(\theta_1, \theta_2)} \left\{ \theta_1 \cdot (\mu - r\mathbf{1})v_x + \theta_1 \sigma_1^2 y_1 s_1^{2\beta_1-1} v_{xs_1} + \frac{1}{2} (\theta_1^2 \sigma_1^2 y_1 s_1^{2\beta_1-2} + \theta_2^2 \sigma_2^2 y_2) v_{xx} \right\} \\ &= -v_t - (\mathbf{L}^r + \mathbf{L}^Y + \mathbf{L}^{S^1})v - rxv_x \\ &\quad + \frac{((\mu_1 - r)v_x + \sigma_1^2 y_1 s_1^{2\beta_1-1} v_{xs_1})^2}{2\sigma_1^2 y_1 s_1^{2\beta_1-2} v_{xx}} + \frac{((\mu_2 - r)v_x)^2}{2\sigma_2^2 y_2 v_{xx}} \end{aligned} \quad (5.10)$$

where

$$\mathbf{L}^r v = \kappa(b - r)v_r + \frac{1}{2}\zeta^2 v_{rr}, \quad \mathbf{L}^Y v = \sum_{i=1}^2 k_i (m_i - y_i) v_{y_i} + \frac{1}{2} c_i^2 y_i v_{y_i y_i},$$

$$\text{and } \mathbf{L}^{S^1} v = \mu_1 s_1 v_{s_1} - \frac{1}{2} \sigma_1^2 s_1 y_1 v_{s_1 s_1}.$$

In order to implement our Monte Carlo-finite differences scheme, we re-write (5.10) as:

$$\begin{aligned} -v_t - (\mathbf{L}^r + \mathbf{L}^Y + \mathbf{L}^{S^1})v - \frac{1}{2}\sigma^2 v_{xx} + F((x, r, s_1, y_1, y_2), Dv, D^2v) &= 0, \\ v(T, x, r, s_1, y_1, y_2) &= -e^{-\eta x}, \end{aligned} \quad (5.11)$$

where  $\sigma > 0$ , and the nonlinearity  $F : \mathbb{R} \times \mathbb{R}^2 \times \mathbb{S}_2$  is given by:

$$F(x, z, \gamma) = \frac{1}{2}\sigma^2 \gamma_{11} - x_1 x_2 z_1 + \frac{((\mu_1 - x_2)z_1 + \sigma_1^2 x_4 x_3^{2\beta_1-1} \gamma_{1,3})^2}{2\sigma_1^2 x_4 x_3^{2\beta_1-2} \gamma_{11}} + \frac{((\mu_2 - x_2)z_1)^2}{2\sigma_2^2 x_5 \gamma_{11}}.$$

We next consider the truncated nonlinearity:

$$\begin{aligned} F_{\varepsilon, n} &:= \frac{1}{2}\sigma^2 \gamma_{11} - x_1 x_2 z_1 + \sup_{\varepsilon \leq |\theta| \leq n} \left\{ (\theta \cdot (\mu - r\mathbf{1})v_x + \theta_1 \sigma_1^2 (y_1 \vee \varepsilon)(s_1 \vee \varepsilon)^{2\beta_1-1} v_{xs_1} \right. \\ &\quad \left. + \frac{1}{2} (\theta_1^2 \sigma_1^2 (y_1 \vee \varepsilon)(s_1 \vee \varepsilon)^{2\beta_1-2} + \theta_2^2 \sigma_2^2 (y_2 \vee \varepsilon)) v_{xx} \right\}, \end{aligned}$$

where  $\varepsilon, n > 0$  are jointly chosen with  $\sigma$  so that Assumption **F** holds true. Under this form, the forward two-dimensional diffusion is defined by:

$$\begin{aligned} dX_t^{(1)} &= \sigma dW_t^{(0)} & , & & dX_t^{(2)} &= \kappa(b - X_t^{(2)})dt + \zeta dW_t^{(1)}, \\ dX_t^{(3)} &= \mu_1 X_t^{(3)}dt + \sigma_1 \sqrt{X_t^{(4)}} X_t^{(3)\beta_1} dW_t^{(1,1)} & , & & dX_t^{(4)} &= k(m_1 - X_t^{(4)})dt + c_1 \sqrt{X_t^{(4)}} dW_t^{(1,2)}, \\ dX_t^{(5)} &= k(m_2 - X_t^{(5)})dt + c_2 \sqrt{X_t^{(5)}} dW_t^{(2,2)}. \end{aligned} \tag{5.12}$$

The component  $X_t^{(2)}$  is simulated according to the exact discretization:

$$X_{t_n}^{(2)} = b + e^{-k\Delta t} \left( X_{t_{n-1}}^{(2)} - b \right) + \zeta \sqrt{\frac{1 - \exp(-2\kappa\Delta t)}{2\kappa}} \xi_n,$$

where  $(\xi_n)_{n \geq 1}$  is a sequence of independent random variable with distribution  $\mathbf{N}(0, 1)$ . The following scheme for the price of the asset guarantees non-negativity (see [1]) :

$$\ln X_n^{(3)} = \ln X_{n-1}^{(3)} + \left( \mu_1 - \frac{1}{2} \sigma_1^2 \left( X_{n-1}^{(3)} \right)^{2(\beta_1-1)} X_{n-1}^{(4)} \right) \Delta t + \sigma_1 \left( X_{n-1}^{(3)} \right)^{\beta_1-1} \sqrt{X_{n-1}^{(4)}} \Delta W_n^{(1,2)}$$

where  $\Delta W_n^{(1,2)} := W_n^{(1,2)} - W_{n-1}^{(1,2)}$ . We take the following parameters  $\mu_1 = 0.10$ ,  $\sigma_1 = 0.3$ ,  $\beta_1 = 0.5$  for the first asset,  $k_1 = 0.1$ ,  $m_1 = 1.$ ,  $c_1 = 0.1$  for the diffusion process of the first asset. The second asset is defined by the same parameters as in the two dimensional example:  $\mu_2 = 0.15$ ,  $c_2 = 0.2$ ,  $m = 0.3$  and  $Y_0^{(2)} = m$ . As for the interest rate model we take  $b = 0.07$ ,  $X_0^{(2)} = b$ ,  $\zeta = 0.3$ .

The initial values of the portfolio the assets prices are all set to 1. For this test case we first use the basis projection regression method with  $4 \times 4 \times 4 \times 4 \times 10$  meshes and three millions particles. Figure 6 contains the plot of the solution obtained by Scheme 2, with different time steps. We only provide results for the implementation of Scheme 1 with a coarse time step, because the method was diverging with a thinner time step. We observe that there is still a difference for very thin time step with the three considered values of the diffusion. This seems to indicate that more particles and more meshes are needed. While doing many calculation we observed that for the thinner time step mesh, the solution sometimes diverges. We therefore report the results corresponding to thirty millions particles with  $4 \times 4 \times 4 \times 4 \times 40$  meshes. First we notice that with this discretization all results are converging as time step goes to zero: the exact solution seems to be very closed to  $-0.258$ . During our experiments with thirty millions particles, the scheme was always converging with a very low variance on the results.

**Remark 5.4** With thirty millions particles, the memory needed forced us to use 64-bit processors with more than four gigabytes of memory.

### 5.2.3 Conclusion on numerical results

We achieved calculation with both schemes with regression and Malliavin methods. We observe that the second scheme give better results than the first whatever the method is used to calculate the conditional expectation. On our test, regression methods achieved

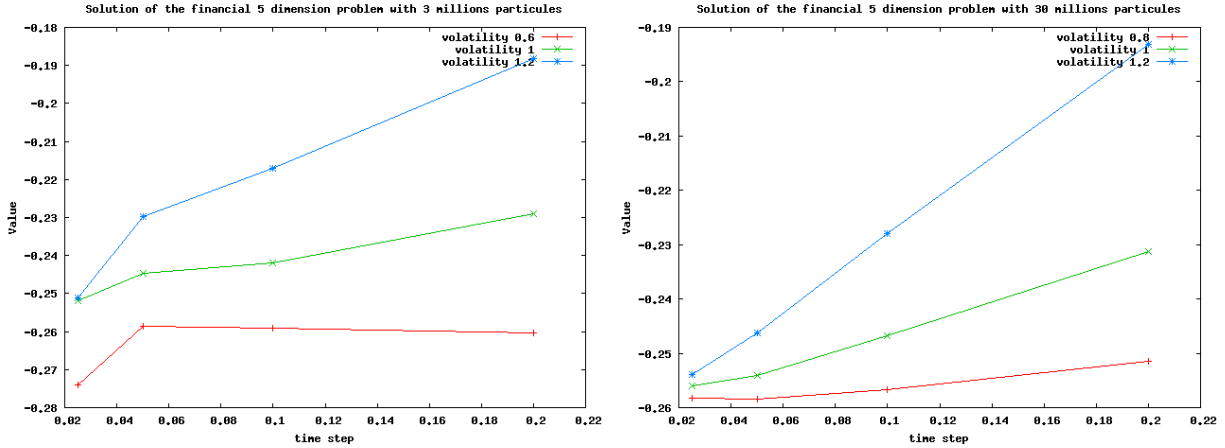


Figure 6: Five dimensional financial problem and its results for different volatilities with 3 millions and 30 millions particles

better performance both in term of results and time of calculation for a given number of particles. In the proposed scheme, we gave results with different volatilities that the user can choose. If we know that the volatility chosen has to be greater than a given value to be able to get things work in demonstration ( equation 3.2 in assumption F) we notice that a too high volatility force us to take to many meshes and too many particles giving too high a calculation time. At last we could think to use some time varying volatility depending on the evolution of the phenomenon we want to follow and the use of different volatilities could be useful to estimate some errors on the solution given and lead to adaptative methods.

## References

- [1] L.B.G. Andersen, R. Brotherton-Ratcliffe, “Extended LIBOR market models with stochastic volatility” *Journal of Computational Finance*, Vol.9,1, 1-40, 2005
- [2] V. Bally, “An Approximation Schemes For BSDEs and Applications to Control and Nonlinear PDEs”. Prepublication 95-15 du Laboratoire de Statistique et Processus de l’Universite du Maine, (1995).
- [3] B. Bouchard, I. Ekeland, N. Touzi, “On the Malliavin approach to Monte Carlo approximation of conditional expectations”. *Finance and Stochastics*, 8, 45-71 (2004).
- [4] G. Barles, E. R. Jakobsen, “On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations”. *Mathematical Modelling and Numerical Analysis*, ESAIM, M2AM, Vol. 36 (2002), No. 1, 33-54.
- [5] G. Barles, E. R. Jakobsen, “Error Bounds For Monotone Approximation Schemes For Hamilton-Jacobi-Bellman Equations”. *SIAM J. Numer. Anal.*, Vol. 43 (2005), No. 2, 540-558.

- [6] G. Barles, E. R. Jakobsen, “Error Bounds For Monotone Approximation Schemes For Parabolic Hamilton-Jacobi-Bellman Equations”. *Math. Comp.*, 76(240):1861-1893, 2007.
- [7] G. Barles, P. E. Souganidis, “Convergence of Approximation Schemes for Fully Non-linear Second Order Equation”. *Asymptotic Anal.*, 4, pp. 271–283, 1991.
- [8] F. Bonnans, H. Zidani, “Consistency of Generalized Finite Difference Schemes for the Stochastic HJB Equation”. *SIAM J. Numerical Analysis* 41-3(2003), 1008-1021
- [9] B. Bouchard, N. Touzi, “Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations”. *Stochastic Processes and their Applications* 111, 175–206 (2004).
- [10] P. Cheridito, H. M. Soner, N. Touzi, N. Victoir, “Second Order Backward Stochastic Differential Equations and Fully Non-Linear Parabolic PDEs”. *Communications on Pure and Applied Mathematics*, Volume 60, Issue 7, Date: July 2007, Pages: 1081-1110.
- [11] D. Chevance, “Ph.D. thesis”. Univ. Provence, (1996).
- [12] D. Crisan, K. Manolarakis, N. Touzi, “On the Monte Carlo simulation Of BSDE’s: An improvement on the Malliavin weights”. to appear.
- [13] F. Delarue and S. Menozzi, “A Forward-Backward Stochastic Algorithm for Quasi-Linear PDEs”. *Annals of Applied Probability*, Vol. 16, No. 1, 140-184, 2006.
- [14] H. Dong, N. V. Krylov, “On The Rate Of Convergence Of Finite-Difference Approximations For Bellman’s Equations With Constant Coefficients”. *St. Petersburg Math. J.*, Vol. 17 (2005), No. 2, 108-132.
- [15] N. El Karoui, S. Peng, M. C. Quenez, “Backward Stochastic Differential Equations in Finance”. *Mathematical Finance*, 7, 1-71 (1997).
- [16] 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).
- [17] E. Gobet, J. P. Lemor, X. Warin “A regression-based Monte-Carlo method to solve backward stochastic differential equations”. *Annals of Applied Probability*, Vol.15(3), pp.2172-2002 - 2005.
- [18] S. L. Heston, “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options”. *The Review of Financial Studies*, Vol.6(2), 327-343, 1993.
- [19] C. Kahl, P. Jackel “Fast strong approximation Monte Carlo schemes for stochastic volatility models” *Quantitative Finance*, Vol.6(6), 513-536, 2006.
- [20] R. V. Kohn, S. Serfaty, “Second-order PDEs and Deterministic Games”. Preprint, to appear in *Proceedings ICIAM07*.

- [21] N. V. Krylov, “On The Rate Of Convergence Of Finite-Difference Approximations For Bellman’s Equations”. St. Petersburg Math. J., Vol. 9 (1997), No. 3, 245-256.
- [22] N. V. Krylov, “On The Rate Of Convergence Of Finite-Difference Approximations For Bellman’s Equations With Lipschitz Coefficients”. Applied Math. and Optimization, Vol. 52 (2005), No. 3, 365-399.
- [23] N. V. Krylov, “On The Rate Of Convergence Of Finite-Difference Approximations For Bellman’s Equations With Variable Coefficients”. Probab. Theory Ralat. fields, 117:1-16, 2000.
- [24] Lions, P.L., Regnier, H. (2001). Calcul du prix et des sensibilités d’une option américaine par une méthode de Monte Carlo. Preprint.
- [25] Longstaff, F.A. and Schwartz, R.S. (2001). Valuing American options by simulation: a simple least-square approach, *Review of Financial Studies* 14, 113–147.
- [26] R. Lord, R. Koekoek, D Van Dijk, A comparison of biased simulation schemes for stochastic volatility models forthcoming in *Quantitative Finance*, 2008
- [27] H. M. Soner and N. Touzi, “A stochastic representation for mean curvature type geometric flows”. The Annals of Probability, Vol 31(3), 1145-1165 (2003).
- [28] J. Zhang, “A numerical scheme for backward stochastic differential equations”. Annals of Applied Probability, 14 (1), 459-488, (2004).
- [29] T. Zariphopoulou, “A solution approach to valuation with unhedgeable risks”. Finance and Stochastics, 5 (2001) 61-82.