A PROBABILISTIC NUMERICAL METHOD FOR FULLY NONLINEAR PARABOLIC PDES[1]

BY ARASH FAHIM, NIZAR TOUZI AND XAVIER WARIN

CMAP, École Polytechnique Paris, CMAP, École Polytechnique Paris and EDF R&D & FiME (Finance des Marchés de l'Energie) Laboratory, Paris

We consider the probabilistic numerical scheme for fully nonlinear partial differential equations suggested in [*Comm. Pure Appl. Math.* **60** (2007) 1081–1110] and show that it can be introduced naturally as a combination of Monte Carlo and finite difference schemes 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 the approximation of 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.

1. Introduction. We consider the probabilistic numerical scheme for the approximation of the solution of a fully nonlinear parabolic Cauchy problem suggested in [12]. In the latter paper a representation of the solution of the partial differential equation (PDE) is derived in terms of the newly introduced notion of second order backward stochastic differential equations, assuming that the fully nonlinear parabolic Cauchy problem has a smooth solution. Then, similarly to the case of backward stochastic differential equations which are connected to semilinear PDEs, this representation suggests a backward probabilistic numerical scheme.

The representation result of [12] can be viewed as an extension of the Feynman–Kac representation result for the linear case, which is widely used 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 [13], El Karoui, Peng and Quenez [16], Bally and Pagès [2], Bouchard and Touzi [9] and Zhang [30]. 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 the time step. The discretization involves the computation of conditional expectations, which need to be further approximated in order to produce 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 [12] can be introduced naturally without appealing to the notion of backward stochastic differential equation. This is shown is Section 2, where the scheme is decomposed into three steps:

(i) the Monte Carlo step consists of 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 difference 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 that such conditional expectations be replaced 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 toward 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. We then consider the implementable scheme involving the Monte Carlo error, and we similarly prove a convergence result for general nonlinear PDEs and 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 relaxed in 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 showing that this condition is necessary, our numerical experiments show that the violation of this condition leads to a serious misperformance of the method; see Figure 5 below.

Our proofs rely on the monotonic scheme method developed by Barles and Souganidis [6] in the theory of viscosity solutions and the recent method of shaking coefficients of Krylov [21, 23] and [22] and Barles and Jakobsen [4, 5] and [3]. The use of the latter types of methods in the context of a stochastic scheme seems to be new. Note, 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.

The following are two related numerical methods based on finite differences in the context of Hamilton–Jacobi–Bellman nonlinear PDEs:

- Bonnans and Zidani [7] introduced a finite difference scheme which satisfies the crucial monotonicity condition of Barles and Souganidis [6] 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 which is encountered in finite difference 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.
- An alternative finite difference scheme is the semi-Lagrangian method which solves the monotonicity requirement by absorbing the dynamics of the underlying state in the finite difference approximation (see, e.g., Debrabant and Jakobsen [15], Camilli and Jacobsen [11], Camilli and Falcone [10] and Munos and Zidani [27]). Loosely speaking, this method is close in spirit to ours and corresponds to freezing the Brownian motion W_h, over each time step h, to its average order $\sqrt{h}$. However, it does not involve any simulation technique and requires the interpolation of the value function at each time step. Thus, it is also subject to the curse of dimensionality.

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 in question 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 [12], where such a parabolic envelope appears in the PDE, and where the Brownian motion plays the role of 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, together with 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 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.

NOTATION. For scalars $a, b \in \mathbb{R}$, we write $a \wedge b := \min\{a, b\}$, $a \vee b := \max\{a, b\}$, $a^- := \max\{-a, 0\}$ 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 by $\mathbb{S}_d$, 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 define $A \cdot B := \mathrm{Tr}[A^{\mathrm{T}} B]$. 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., *α-exponential growth*) if

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

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

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

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

2. Discretization. Let μ and σ be two maps from $\mathbb{R}_+ \times \mathbb{R}^d$ to $\mathbb{R}^d$ and $\mathbb{M}(d, d)$, respectively, and let $a := \sigma\sigma^{\mathrm{T}}$. We 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:

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

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

Under some conditions, a stochastic representation of the solution of this problem was provided in [12] 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 [12] 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 $h := T/n$, $t_i = ih$, $i = 0, \ldots, n$, and consider the one-step-ahead Euler discretization

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

of the diffusion X corresponding to the linear operator $\mathcal{L}^X$. Our analysis does not require any existence or uniqueness results for the underlying diffusion X. However, the subsequent formal discussion assumes existence and uniqueness in order to provide 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 have ignored the difficulties related to local martingale part, and where $\mathbb{E}_{t_i,x} := \mathbb{E}[\cdot | X_{t_i} = x]$ denotes the expectation operator conditional on $\{X_{t_i} = x\}$. Since v solves the PDE (2.1), this gives

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

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

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

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

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

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

$$(2.6)$$

$$\mathcal{D}_h \psi := (\mathcal{D}_h^0 \psi, \mathcal{D}_h^1 \psi, \mathcal{D}_h^2 \psi)^{\mathrm{T}},$$

and D^k is the kth order partial differential operator with respect to the space variable x. The differentiation in the above scheme is 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 has 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 means of an easy integration by parts argument.

LEMMA 2.1. *For every function $\varphi : Q_T \to \mathbb{R}$ 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 $H_h = (H_0^h, H_1^h, H_2^h)^{\mathrm{T}}$ and

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

PROOF. The main ingredient of the proof is the following, simple, observation. Let G be a one-dimensional Gaussian random variable with unit variance. For any function $f : \mathbb{R} \longrightarrow \mathbb{R}$ with exponential growth, we then have

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

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

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

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

and therefore

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

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

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

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

and for $j = i$,

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

This gives

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

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

REMARK 2.2.   Clearly, different choices are involved in the integration by parts in Lemma 2.1. One such possibility leads to the representation of $\mathcal{D}_2^h\varphi$ as

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

This representation shows that the backward scheme (2.4) is very similar to the probabilistic numerical algorithm suggested in [12].

Observe that the choice of the drift and diffusion coefficients μ and σ in the nonlinear PDE (2.1) is arbitrary. Thus far, they have been used only in order to define the underlying diffusion X, but our convergence result will place some restrictions on the choice of the diffusion coefficient; see Remark 3.4.

Once the linear operator $\mathcal{L}^X$ is chosen in the nonlinear PDE, the above algorithm handles the remaining nonlinearity by means of classical finite difference 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 $\frac{1}{2}(\delta_{\sqrt{h}} + \delta_{-\sqrt{h}})$. This then induces the following approximation:

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

    which is the centered finite difference 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 $\frac{1}{6}(\delta_{\{\sqrt{3h}\}} + 4\delta_{\{0\}} + \delta_{\{-\sqrt{3h}\}})$ so that $\mathbb{E}[w_j^n] = \mathbb{E}[W_h^n]$ for all integers $n \le 4$. This then induces the following approximation:

$$\mathcal{D}_h^2 \psi(t, x) := \mathbb{E}[\psi(t + h, X_h^{t,x})H_2^h]$$

$$\approx \frac{\psi(t, x + \sqrt{3h}) - 2\psi(t, x) + \psi(t, x - \sqrt{3h})}{3h},$$

which is the centered finite difference 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 difference algorithm. The Monte Carlo component of the scheme consists of the choice of an underlying diffusion process X. The finite difference component of the scheme consists of the approximation of the remaining nonlinearity by means of the integration by parts formula of Lemma 2.1.

3. Asymptotics of the discrete-time approximation.

3.1. *Main results*. Our first main convergence results follow the general methodology of Barles and Souganidis [6] and require that the nonlinear PDE (2.1) satisfies a comparison result in the sense of viscosity solutions.

We recall that an upper semicontinuous (resp., lower semicontinuous) function $\underline{v}$ (resp., $\overline{v}$) on $[0, T] \times \mathbb{R}^d$ is called a *viscosity subsolution* (resp., *supersolution*) of (2.1) if, for any $(t, x) \in [0, T) \times \mathbb{R}^d$ and any smooth function φ satisfying

$$0 = (\underline{v} - \varphi)(t, x) = \max_{[0,T] \times \mathbb{R}^d} (\underline{v} - \varphi)$$

$$\left(\text{resp., } 0 = (\overline{v} - \varphi)(t, x) = \min_{[0,T] \times \mathbb{R}^d} (\overline{v} - \psi) \right),$$

we have

$$-\mathcal{L}^X \varphi - F(t, x, \mathcal{D}\varphi(t, x)) \le (\text{resp., } \ge) \, 0.$$

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

$$\underline{v}(T, \cdot) \le \overline{v}(T, \cdot),$$

we have $\underline{v} \le \overline{v}$.

REMARK 3.2. Barles and Souganidis [6] use a stronger notion of comparison by accounting for the final condition, thus allowing for a possible boundary layer. In their context, a supersolution $\overline{v}$ and a subsolution $\underline{v}$ satisfy

$$(3.1) \qquad \min\{-\mathcal{L}^X \underline{v}(T, x) - F(T, x, \mathcal{D}\underline{v}(T, x)), \underline{v}(T, x) - g(x)\} \le 0,$$

$$(3.2) \qquad \max\{-\mathcal{L}^X \overline{v}(T, x) - F(T, x, \mathcal{D}\overline{v}(T, x)), \overline{v}(T, x) - g(x)\} \ge 0.$$

We observe that, by the nature of our equation, (3.1) and (3.2) imply that the sub-solution $\underline{v} \leq g$ and the supersolution $\overline{v} \geq g$, that is, the final condition holds in the usual sense, and no boundary layer can occur. To see this we suppose without loss of generality that $F(t, x, r, p, \gamma)$ is decreasing with respect to r (see Remark 3.13). Let φ be a function satisfying

$$0 = (\underline{v} - \varphi)(T, x) = \max_{[0,T] \times \mathbb{R}^d} (\underline{v} - \varphi)$$

and define $\varphi_K(t, \cdot) = \varphi(t, \cdot) + K(T - t)$ for $K > 0$. Then, $\underline{v} - \varphi_K$ also has a maximum at (T, x), and the subsolution property (3.1) implies that

$$\min\{-\mathcal{L}^X \varphi(T, x) - F(T, x, \mathcal{D}\varphi(T, x)) + K, \underline{v}(T, x) - g(x)\} \leq 0.$$

For a sufficiently large K, this provides the required inequality $\underline{v}(T, x) - g(x) \leq 0$. A similar argument shows that (3.1) implies that $\overline{v} - g \geq 0$.

In the sequel, we denote by F_r, F_p and F_γ the partial gradients of F with respect to r, p and γ, respectively. We also denote by F_γ^- the pseudoinverse of the nonnegative symmetric matrix F_γ. We recall that any Lipschitz function is differentiable a.e.

ASSUMPTION F.   (i) The nonlinearity F is Lipschitz continuous with respect to (x, r, p, γ) uniformly in t, and $|F(\cdot, \cdot, 0, 0, 0)|_\infty < \infty$;

(ii) F is elliptic and dominated by the diffusion of the linear operator $\mathcal{L}^X$, that is,

$$(3.3) \qquad\qquad \nabla_\gamma F \leq a \qquad \text{on } \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d;$$

(iii) $F_p \in \text{Image}(F_\gamma)$ and $|F_p^{\mathrm{T}} F_\gamma^- F_p|_\infty < +\infty$.

REMARK 3.3.   Assumption F(iii) is equivalent to

$$(3.4) \qquad |m_F^-|_\infty < \infty, \qquad \text{where } m_F := \min_{w \in \mathbb{R}^d} \{F_p \cdot w + w^{\mathrm{T}} F_\gamma w\}.$$

This is immediately seen by recalling that, by the symmetric feature of F_γ, any $w \in \mathbb{R}^d$ has an orthogonal decomposition $w = w_1 + w_2 \in \text{Ker}(F_\gamma) \oplus \text{Image}(F_\gamma)$, and, by the nonnegativity of F_γ,

$$F_p \cdot w + w^{\mathrm{T}} F_\gamma w = F_p \cdot w_1 + F_p \cdot w_2 + w_2^{\mathrm{T}} F_\gamma w_2$$

$$= -\tfrac{1}{4} F_p^{\mathrm{T}} F_\gamma^- F_p + F_p \cdot w_1 + |\tfrac{1}{2}(F_\gamma^-)^{1/2} \cdot F_p - F_\gamma^{1/2} w_2|^2.$$

REMARK 3.4.   Condition (3.3) 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 σ. Since $\sigma\sigma^{\mathrm{T}} \in \mathcal{S}_d^+$, this implies, in particular, that our main results do not apply to general degenerate nonlinear parabolic PDEs. Second, the diffusion of the linear opera-tor σ is required to dominate the nonlinearity F which implicitly places a lower bound on the choice of the diffusion σ.

EXAMPLE 3.5. 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. If we then restrict the choice of the diffusion to be constant, it follows from Assumption F that $\frac{1}{3}a^2 \leq \sigma^2 \leq b^2$, which implies that $a^2 \leq 3b^2$. If the parameters a and b do not satisfy the latter condition, then the diffusion σ has to be chosen to be state- and time-dependent.

THEOREM 3.6 (Convergence). *Let Assumption F hold true, $|\mu|_1, |\sigma|_1 < \infty$ and σ be invertible. Also, assume that the fully nonlinear PDE* (2.1) *has comparison for bounded functions. For every bounded Lipschitz function g, there then exists a bounded function v such that*

$$v^h \longrightarrow v \qquad locally\ uniformly.$$

In addition, v is the unique bounded viscosity solution of problem (2.1)–(2.2).

REMARK 3.7. Under the boundedness condition on the coefficients μ and σ, the restriction to a bounded terminal data g in the above Theorem 3.6 can be relaxed by an immediate change of variable. Let g be a function with α-exponential growth for some $\alpha > 0$. Fix some $M > 0$, and let ρ be an arbitrary smooth positive function with

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

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

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

The nonlinear PDE problem (2.1)–(2.2) satisfied by v then becomes the following nonlinear PDE for u:

(3.5)
$$-\mathcal{L}^X u - \tilde{F}(\cdot, u, Du, D^2 u) = 0 \qquad \text{on } [0, T) \times \mathbb{R}^d,$$
$$v(T, \cdot) = \tilde{g} := \rho^{-1}g \qquad \text{on } \mathbb{R}^d,$$

where

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

Recall that the coefficients μ and σ are assumed to be bounded. It is then easy to see that $\tilde{F}$ satisfies the same conditions as F. Since $\tilde{g}$ is bounded, the convergence Theorem 3.6 applies to the nonlinear PDE (3.5).

REMARK 3.8. Theorem 3.6 states that inequality (3.3) (i.e., diffusion must dominate the nonlinearity in γ) is sufficient for the convergence of the Monte Carlo/finite difference scheme. We do not know whether this condition is neces-

sary:

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

We next provide bounds on the rate of convergence of the Monte Carlo/finite difference scheme in the context of nonlinear PDEs of the Hamilton–Jacobi–Bellman type in the same context as [5]. The following assumptions are stronger than Assumption F and imply that the nonlinear PDE (2.1) satisfies a comparison result for bounded functions.

ASSUMPTION HJB.   The nonlinearity F satisfies Assumption F(ii)–(iii) and is of the Hamilton–Jacobi–Bellman type:

$$\frac{1}{2}a \cdot \gamma + b \cdot p + F(t, x, r, p, \gamma) = \inf_{\alpha \in \mathcal{A}} \{\mathcal{L}^\alpha(t, x, r, p, \gamma)\},$$

$$\mathcal{L}^\alpha(t, x, r, p, \gamma) := \frac{1}{2}\operatorname{Tr}[\sigma^\alpha \sigma^{\alpha \mathrm{T}}(t, x)\gamma] + b^\alpha(t, x)p + c^\alpha(t, x)r + f^\alpha(t, x),$$

where the functions μ, σ, σ^α, b^α, c^α and f^α 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 $\alpha \in \mathcal{A}$,

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

REMARK 3.9.   The Assumption HJB+ is satisfied if $\mathcal{A}$ is a separable topological space and $\sigma^\alpha(\cdot)$, $b^\alpha(\cdot)$, $c^\alpha(\cdot)$ and $f^\alpha(\cdot)$ are continuous maps from $\mathcal{A}$ to $C_b^{1/2,1}$, the space of bounded maps which are Lipschitz in x and $\frac{1}{2}$-Hölder in t.

THEOREM 3.10 (Rate of convergence).   *Assume that the final condition g is bounded and Lipschitz continuous. There then exists a constant $C > 0$ such that*:

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

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

We also observe that in the PDE and finite difference 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 equations, notice that $|\Delta x|$ is of the order of $h^{1/2}$. Therefore, the above upper and lower bounds on the rate of convergence correspond to those for 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 of Theorem 3.6 by building on Theorem 2.1 and Remark 2.1 of Barles and Souganidis [6], which requires the scheme to be consistent, monotone and stable. Moreover, since we are assuming (weak) comparison for the equation, we also need to prove that our scheme produces a limit which satisfies the terminal condition in the usual sense; see Remark 3.2.

Throughout this section, all the conditions of Theorem 3.6 are in force.

LEMMA 3.11. *Let φ be a smooth function with bounded derivatives. Then, for all $(t, x) \in [0, T] \times \mathbb{R}^d$,*

$$\lim_{\substack{(t',x') \to (t,x) \\ (h,c) \to (0,0) \\ t'+h \leq T}} \frac{[c + \varphi](t', x') - \mathbf{T}_h[c + \varphi](t', x')}{h}$$

$$= -(\mathcal{L}^X \varphi + F(\cdot, \varphi, D\varphi, D^2\varphi))(t, x).$$

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

LEMMA 3.12. *Let $\varphi, \psi : [0, T] \times \mathbb{R}^d \longrightarrow \mathbb{R}$ be two Lipschitz functions. Then, for some $C > 0$,*

$$\varphi \leq \psi \quad \Longrightarrow \quad \mathbf{T}_h[\varphi](t, x) \leq \mathbf{T}_h[\psi](t, x) + Ch\mathbb{E}[(\psi - \varphi)(t + h, \hat{X}_h^{t,x})],$$

where C depends only on the constant K in (3.4).

PROOF. By Lemma 2.1, the operator $\mathbf{T}_h$ can be written as

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

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

$$\mathbf{T}_h[\psi](t, x) - \mathbf{T}_h[\varphi](t, x) = \mathbb{E}[f(\hat{X}_h^{t,x})] + hF_\tau(\theta) \cdot \mathcal{D}_h f(\hat{X}_h^{t,x})$$

$$= \mathbb{E}[f(\hat{X}_h^{t,x})(1 + hF_\tau(\theta) \cdot H_h(t, x))]$$

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

$$\mathbf{T}_h[\psi] - \mathbf{T}_h[\varphi] = \mathbb{E}[f(\hat{X}_h^{t,x})(1 + hF_r + F_p.(\sigma^{\mathrm{T}})^{-1}W_h$$

$$+ h^{-1}F_\gamma \cdot (\sigma^{\mathrm{T}})^{-1}(W_h W_h^{\mathrm{T}} - hI)\sigma^{-1})],$$

where the dependence on θ and x has been omitted for the sake of notational simplicity. Since $F_\gamma \leq a$ by (3.4) of Assumption F, we have $1 - a^{-1} \cdot F_\gamma \geq 0$, and

therefore

$$
\begin{aligned}
\mathbf{T}_h[\psi] &- \mathbf{T}_h[\varphi] \\
&\geq \mathbb{E}\big[f(\hat{X}_h^{t,x})(hF_r + F_p \cdot (\sigma^{\mathrm{T}})^{-1}W_h + h^{-1}F_\gamma \cdot (\sigma^{\mathrm{T}})^{-1}W_h W_h^{\mathrm{T}}\sigma^{-1})\big] \\
&= \mathbb{E}\bigg[f(\hat{X}_h^{t,x})\bigg(hF_r + hF_p \cdot (\sigma^{\mathrm{T}})^{-1}\frac{W_h}{h} + hF_\gamma \cdot (\sigma^{\mathrm{T}})^{-1}\frac{W_h W_h^{\mathrm{T}}}{h^2}\sigma^{-1}\bigg)\bigg].
\end{aligned}
$$

Let $m_F^- := \max\{-m_F, 0\}$, where the function m_F is defined in (3.4). Under Assumption F, we have $K := |m_F^-|_\infty < \infty$, so

$$
F_p.\sigma^{\mathrm{T}-1}\frac{W_h}{h} + hF_\gamma \cdot \sigma^{\mathrm{T}-1}\frac{W_h W_h^{\mathrm{T}}}{h^2}\sigma^{-1} \geq -K.
$$

We can write

$$
\mathbf{T}_h[\psi] - \mathbf{T}_h[\varphi] \geq \mathbb{E}[f(\hat{X}_h^{t,x})(hF_r - hK)] \geq -C'h\mathbb{E}[f(\hat{X}_h^{t,x})]
$$

for some constant $C > 0$, where the last inequality follows from (3.4). □

The following observation will be used in the proof of Theorem 3.10 below.

REMARK 3.13. The monotonicity result of the previous Lemma 3.12 is slightly different from that required in [6]. However, as observed in Remark 2.1 in [6], their convergence theorem holds under this approximate monotonicity. From the previous proof, we observe that if the function F satisfies the condition

$$
(3.6) \qquad F_r - \tfrac{1}{4}F_p^{\mathrm{T}}F_\gamma^- F_p \geq 0,
$$

then the standard monotonicity condition

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

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 (3.6). Indeed, direct calculation shows that the PDE inherited by u is

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

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

where $\overline{F}(t,x,r,p,\gamma) = e^{\theta(T-t)}F(t,x,e^{-\theta(T-t)}r, e^{-\theta(T-t)}p, e^{-\theta(T-t)}\gamma) + \theta r$. It is then easily seen that $\overline{F}$ satisfies the same conditions as F, together with (3.6), for sufficiently large θ.

LEMMA 3.14. Let $\varphi, \psi : [0,T] \times \mathbb{R}^d \longrightarrow \mathbb{R}$ be two $\mathbb{L}^\infty$-bounded functions. There then 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, then 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,

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

where

$$A_h = \frac{1}{2}(F_\gamma^-)^{1/2}F_p - F_\gamma^{1/2}\sigma^{\mathrm{T}-1}\frac{W_h}{h}.$$

Since $1 - \mathrm{Tr}[a^{-1}F_\gamma] \geq 0$, $|F_r|_\infty < \infty$ and $|F_p^{\mathrm{T}}F_\gamma^- F_p|_\infty < \infty$ by Assumption F, it follows that

$$|\mathbf{T}_h[\varphi] - \mathbf{T}_h[\psi]|_\infty \leq |f|_\infty(1 - a^{-1} \cdot F_\gamma + h\mathbb{E}[|A_h|^2] + Ch).$$

However, $\mathbb{E}[|A_h|^2] = \frac{h}{4}F_p^{\mathrm{T}}F_\gamma^- F_p + a^{-1} \cdot F_\gamma$. Therefore, by Assumption F,

$$|\mathbf{T}_h[\varphi] - \mathbf{T}_h[\psi]|_\infty \leq |f|_\infty\left(1 + \frac{h}{4}F_p^{\mathrm{T}}F_\gamma^- F_p + Ch \right) \leq |f|_\infty(1 + \bar{C}h).$$

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 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 3.12 with $\varphi \equiv v^h(t_{i+1}, \cdot)$ and $\psi \equiv 0$, we see that

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

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. □

REMARK 3.15. The approximate function v^h defined by (2.4) is only defined on $\{ih|i = 0, \ldots, N\} \times \mathbb{R}^d$. Our methodology requires that it be extended to any $t \in [0, T]$. This can be achieved by any interpolation, as long as the regularity property of v^h mentioned in Lemma 3.16 below is preserved. For instance, we may simply use linear interpolation.

LEMMA 3.16. *The function v^h is Lipschitz in x, uniformly in h.*

PROOF. We report the following calculation in the one-dimensional case $d = 1$ in order to simplify the presentation.

1. For fixed $t \in [0, T - h]$, we argue as in the proof of Lemma 3.12 to see that for $x, x' \in \mathbb{R}^d$ with $x > x'$,

(3.10) $$v^h(t, x) - v^h(t, x') = A + hB,$$

where, defining $\delta^{(k)} := D^k v^h(t+h, \hat{X}_h^{t,x}) - D^k v^h(t+h, \hat{X}_h^{t,x'})$ for $k = 0, 1, 2$,

$$A := \mathbb{E}[\delta^{(0)}] + h(F(t, x', \mathcal{D}v^h(t+h, \hat{X}_h^{t,x})) - F(t, x', \mathcal{D}v^h(t+h, \hat{X}_h^{t,x'})))$$
$$= \mathbb{E}[(1 + hF_r)\delta^{(0)} + hF_p\delta^{(1)} + hF_\gamma\delta^{(2)}],$$

$$|B| := |F(t, x, \mathcal{D}v^h(t+h, \hat{X}_h^{t,x})) - F(t, x', \mathcal{D}v^h(t+h, \hat{X}_h^{t,x}))| \le |F_x|_\infty|x - x'|,$$

by Assumption F(i). By Lemma 2.1, we write, for $k = 1, 2$,

$$\mathbb{E}[\delta^{(k)}] = \mathbb{E}[\delta^{(0)} H_k^h(t, x) + v^h(t+h, \hat{X}_h^{t,x'})(H_k^h(t, x) - H_k^h(t, x'))]$$
$$= \mathbb{E}\Big[\delta^{(0)} H_k^h(t, x)$$
$$+ Dv^h(t+h, \hat{X}_h^{t,x'})\Big(\frac{W_h}{h}\Big)^{k-1}(\sigma(t, x)^{-k} - \sigma(t, x')^{-k})\sigma(t, x')\Big].$$

Dividing both sides of (3.10) by $x - x'$ and taking limsup, it then follows from the above equalities that

$$\limsup_{|x-x'|\searrow 0} \frac{|v^h(t, x) - v^h(t, x')|}{(x - x')}$$
$$\le \mathbb{E}\Big[\Big|\limsup_{|x-x'|\searrow 0} \frac{v^h(t+h, \hat{X}_h^{t,x}) - v^h(t+h, \hat{X}_h^{t,x'})}{(x - x')}$$
$$\times \Big(1 + hF_r + F_p\frac{W_h}{\sigma(t, x)} + F_\gamma\frac{W_h^2 - h}{\sigma(t, x)^2 h}\Big)$$
$$+ Dv^h(t+h, \hat{X}_h^{t,x})$$
$$\times \Big(W_h F_\gamma\frac{-2\sigma_x(t, x)}{\sigma(t, x)^2} + hF_p\frac{\sigma_x(t, x)}{\sigma(t, x)}\Big)\Big|\Big] + Ch.$$

2. Assume $v^h(t+h, \cdot)$ is Lipschitz with constant L_{t+h}. Then,

$$\limsup_{|x-x'|\searrow 0} \frac{|v^h(t, x) - v^h(t, x')|}{(x - x')}$$
$$\le L_{t+h}\mathbb{E}\Big[\Big|(1 + \mu_x(t, x)h + \sigma_x(t, x)\sqrt{h}N)$$
$$\times \Big(1 + hF_r + \frac{F_p\sqrt{h}N}{\sigma(t, x)} + \frac{F_\gamma N^2}{\sigma(t, x)^2} - \frac{F_\gamma}{\sigma(t, x)^2}\Big)$$
$$+ \sqrt{h}N F_\gamma\frac{-2\sigma_x(t, x)}{\sigma(t, x)^2} + hF_p\frac{\sigma_x(t, x)}{\sigma(t, x)}\Big|\Big] + Ch.$$

Observe that

$$F_p \frac{\sigma_x}{\sigma} = \sigma_x \frac{F_p}{\sqrt{F_\gamma}} \frac{\sqrt{F_\gamma}}{\sigma} \mathbb{1}_{\{F_\gamma \neq 0\}}.$$

Since all terms on the right-hand side are bounded, under our assumptions, it follows that $|F_p \frac{\sigma_x}{\sigma}|_\infty < \infty$ (we emphasize that the geometric structure imposed in Assumption F(iii) provides this result in any dimension). Then,

$$\limsup_{|x-x'|\searrow 0} \frac{|v^h(t,x) - v^h(t,x')|}{(x-x')}$$

$$\leq L_{t+h} \Big(\mathbb{E}\Big[\Big| (1 + \mu_x(t,x)h + \sigma_x(t,x)\sqrt{h}N)$$

$$\times \Big(1 + \frac{F_p\sqrt{h}N}{\sigma(t,x)} + \frac{F_\gamma N^2}{\sigma(t,x)^2} - \frac{F_\gamma}{\sigma(t,x)^2} \Big)$$

$$+ \sqrt{h}N F_\gamma \frac{-2\sigma_x(t,x)}{\sigma(t,x)^2} \Big| \Big] + Ch \Big) + Ch.$$

3. Let $\tilde{\mathbb{P}}$ be the probability measure equivalent to $\mathbb{P}$ defined by the density

$$Z := 1 - \alpha + \alpha N^2, \qquad \text{where } \alpha = \frac{F_\gamma}{\sigma(t,x)^2}.$$

Then,

$$\limsup_{|x-x'|\searrow 0} \frac{|v^h(t,x) - v^h(t,x')|}{(x-x')}$$

$$\leq L_{t+h} \Big(\mathbb{E}^{\tilde{\mathbb{P}}}\Big[\Big| (1 + \mu_x(t,x)h + \sigma_x(t,x)\sqrt{h}N)$$

$$\times \Big(1 + Z^{-1}F_p \frac{\sqrt{h}N}{\sigma(t,x)} \Big)$$

$$+ Z^{-1}\sqrt{h}N F_\gamma \frac{-2\sigma_x(t,x)}{\sigma(t,x)^2} \Big| \Big] + Ch \Big) + Ch.$$

By the Cauchy–Schwarz inequality, we have

$$\limsup_{|x-x'|\searrow 0} \frac{|v^h(t,x) - v^h(t,x')|}{x-x'}$$

$$\leq L_{t+h} \Big(\mathbb{E}^{\tilde{\mathbb{P}}}\Big[\Big| (1 + \mu_x(t,x)h + \sigma_x(t,x)\sqrt{h}N)$$

$$\times \Big(1 + Z^{-1}F_p \frac{\sqrt{h}N}{\sigma(t,x)} \Big)$$

$$+ Z^{-1}\sqrt{h}N F_\gamma \frac{-2\sigma_x(t,x)}{\sigma(t,x)^2} \Big|^2 \Big]^{1/2} + Ch \Big) + Ch.$$

By reexpressing the expectation in terms of probability $\mathbb{P}$, we have

$$
\begin{aligned}
\limsup_{|x-x'|\searrow 0} & \frac{|v^h(t,x) - v^h(t,x')|}{x - x'} \\
&\leq L_{t+h}\bigg(\mathbb{E}\bigg[Z\bigg|(1 + \mu_x(t,x)h + \sigma_x(t,x)\sqrt{h}N) \\
&\qquad\qquad \times \bigg(1 + Z^{-1}F_p \frac{\sqrt{h}N}{\sigma(t,x)}\bigg) \\
&\qquad\qquad\quad + Z^{-1}\sqrt{h}N F_\gamma \frac{-2\sigma_x(t,x)}{\sigma(t,x)^2}\bigg|^2\bigg]^{1/2} + Ch\bigg) + Ch.
\end{aligned}
$$

By expanding the quadratic term inside the expectation, we observe that the expectation of all the terms having $\sqrt{h}$ is zero. Therefore,

$$
\begin{aligned}
\limsup_{|x-x'|\searrow 0} & \frac{|v^h(t,x) - v^h(t,x')|}{(x - x')} \\
&\leq L_{t+h}\bigg(\mathbb{E}^{\tilde{\mathbb{P}}}\bigg[\bigg|(1 + \mu_x(t,x)h + \sigma_x(t,x)\sqrt{h}N) \\
&\qquad\qquad \times \bigg(1 + Z^{-1}F_p \frac{\sqrt{h}N}{\sigma(t,x)}\bigg) \\
&\qquad\qquad\quad + Z^{-1}\sqrt{h}N F_\gamma \frac{-2\sigma_x(t,x)}{\sigma(t,x)^2}\bigg|^2\bigg]^{1/2} + Ch\bigg) + Ch \\
&\leq L_{t+h}\big((1 + C'h)^{1/2} + Ch\big) + Ch,
\end{aligned}
$$

which leads to

$$
\limsup_{|x-x'|\searrow 0} \frac{|v^h(t,x) - v^h(t,x')|}{(x - x')} \leq Ce^{C'T/2}
$$

for some constants $C, C' > 0$. $\qquad\square$

Finally, we prove that the terminal condition is preserved by our scheme as the time step shrinks to zero.

LEMMA 3.17.   *For each $x \in \mathbb{R}^d$ and $t_k = kh$ with $k = 1, \ldots, n$, we have*
$$
|v^h(t_k, x) - g(x)| \leq C(T - t_k)^{1/2}.
$$

PROOF.   1. By the same argument as in the proof of Lemma 3.14, we have, for $j \geq i$,

$$
\begin{aligned}
v^h(t_j, \hat{X}_{t_j}^{t_i,x}) &= \mathbb{E}_{t_j}[v^h(t_{j+1}, \hat{X}_{t_{j+1}}^{t_i,x})(1 - \alpha_j + \alpha_j N_j^2)] \\
&\quad + h\big(F_0^j + F_r^j \mathbb{E}_{t_j}[v^h(t_{j+1}, \hat{X}_{t_{j+1}}^{t_i,x})] + F_p^j \cdot \mathbb{E}_{t_j}[Dv^h(t_{j+1}, \hat{X}_{t_{j+1}}^{t_i,x})]\big),
\end{aligned}
$$

where $F_0^j := F(t_j, \hat{X}_{t_j}^{t_i,x}, 0, 0, 0)$, α_j, F_r^j, F_p^j are $\mathcal{F}_{t_j}$-adapted random variables defined as in the proof of Lemma 3.14 at t_j, and $N_j = \frac{W_{t_{j+1}} - W_{t_j}}{\sqrt{h}}$ has a standard Gaussian distribution. Combining the above formula for j from i to $n-1$, we see that

$$v^h(t_i, x)\mathbb{E}[g(\hat{X}_T^{t_i,x})P_{i,n}]$$

$$= +h\mathbb{E}\sum_{j=i}^{n-1} F_0^j + F_r^j \mathbb{E}_{t_j}[v^h(t_{j+1}, \hat{X}_{t_{j+1}}^{t_i,x})] + F_p^j \cdot \mathbb{E}_{t_j}[Dv^h(t_{j+1}, \hat{X}_{t_{j+1}}^{t_i,x})],$$

where $P_{i,k} := \prod_{j=i}^{k-1}(1 - \alpha_j + \alpha_j N_j^2) > 0$ a.s. for all $1 \leq i < k \leq n$ and $P_{i,i} = 1$. Obviously, $\{P_{i,k}, i \leq k \leq n\}$ is a martingale for all $i \leq n$, a property which will be used later. Since $|F(\cdot, \cdot, 0, 0, 0)|_\infty < +\infty$, and using Assumption F and Lemmas 3.16 and 3.14, we have

$$(3.11) \qquad |v^h(t_i, x) - g(x)| \leq \left|\mathbb{E}[(g(\hat{X}_T^{t_i,x}) - g(x))P_{i,n}]\right| + C(T - t_i).$$

2. Let $\{g_\varepsilon\}_\varepsilon$ be the family of smooth functions obtained from g by convolution with a family of mollifiers $\{\rho_\varepsilon\}$, that is, $g_\varepsilon = g * \rho_\varepsilon$. Note that we have

$$(3.12) \quad |g_\varepsilon - g|_\infty \leq C\varepsilon, \qquad |Dg_\varepsilon|_\infty \leq |Dg|_\infty \quad \text{and} \quad |D^2 g_\varepsilon|_\infty \leq \varepsilon^{-1}|Dg|_\infty.$$

Then,

$$
\begin{aligned}
(3.13) \quad & \left|\mathbb{E}[(g(\hat{X}_T^{t_i,x}) - g(x))P_{i,n}]\right| \\
& \leq \mathbb{E}[|g(\hat{X}_T^{t_i,x}) - g_\varepsilon(\hat{X}_T^{t_i,x})P_{i,n}|] \\
& \quad + \left|\mathbb{E}[(g_\varepsilon(\hat{X}_T^{t_i,x}) - g_\varepsilon(x))P_{i,n}]\right| + |g_\varepsilon - g|_\infty \\
& \leq C\varepsilon + \left|\mathbb{E}[(g_\varepsilon(\hat{X}_T^{t_i,x}) - g_\varepsilon(x))P_{i,n}]\right| \\
& \leq C\varepsilon + \left|\mathbb{E}\left[P_{i,n}\int_{t_i}^T \left(Dg_\varepsilon \hat{b} + \frac{1}{2}\text{Tr}[D^2 g_\varepsilon \hat{a}]\right)(s, \hat{X}_s^{t_i,x})\,ds\right]\right| \\
& \quad + \left|\mathbb{E}\left[P_{i,n}\int_{t_i}^T Dg_\varepsilon(\hat{X}_s^{t_i,x})\hat{\sigma}(s)\,dW_s\right]\right|,
\end{aligned}
$$

where $\hat{b}(s) = b(t_j, \hat{X}_{t_j}^{t_i,x})$ and $\hat{\sigma}(s) = \sigma(t_j, \hat{X}_{t_j}^{t_i,x})$ for $t_j \leq s < t_{j+1}$, and $\hat{a} = \hat{\sigma}^T\hat{\sigma}$. We next estimate each term separately.

2a. First, since $\{P_{i,k}, i \leq k \leq n\}$ is a martingale,

$$
\begin{aligned}
& \left|\mathbb{E}\left[P_{i,n}\int_{t_i}^T Dg_\varepsilon(\hat{X}_s^{t_i,x})\hat{\sigma}(s)\,dW_s\right]\right| \\
& = \left|\sum_{j=i}^{n-1}\mathbb{E}\left[P_{i,n}\int_{t_j}^{t_{j+1}} Dg_\varepsilon(\hat{X}_s^{t_i,x})\hat{\sigma}(s)\,dW_s\right]\right|
\end{aligned}
$$

$$\leq \sum_{j=i}^{n-1} \left| \mathbb{E}\left[P_{i,j+1} \int_{t_j}^{t_{j+1}} Dg_\varepsilon(\hat{X}_s^{t_i,x})\hat{\sigma}(s)\,dW_s \right] \right|$$

$$= \sum_{j=i}^{n-1} \left| \mathbb{E}\left[P_{i,j}\hat{\sigma}(t_j)\mathbb{E}_{t_j}\left[P_{j,j+1} \int_{t_j}^{t_{j+1}} Dg_\varepsilon(\hat{X}_s^{t_i,x})\,dW_s \right] \right] \right|.$$

Notice that

$$\mathbb{E}_{t_j}\left[P_{j,j+1} \int_{t_j}^{t_{j+1}} Dg_\varepsilon(\hat{X}_s^{t_i,x})\,dW_s \right] = \mathbb{E}_{t_j}\left[(W_{t_{j+1}} - W_{t_j})^2 \int_{t_j}^{t_{j+1}} Dg_\varepsilon(\hat{X}_s^{t_i,x})\,dW_s \right]$$

$$= \mathbb{E}_{t_j}\left[\int_{t_j}^{t_{j+1}} 2W_s Dg_\varepsilon(\hat{X}_s^{t_i,x})\,ds \right].$$

Using Lemma 2.1 and (3.12), this gives

$$\left| \mathbb{E}\left[P_{i,n} \int_{t_i}^{T} Dg_\varepsilon(\hat{X}_s^{t_i,x})\hat{\sigma}(s)\,dW_s \right] \right|$$

$$\leq 2 \sum_{j=i}^{n-1} \left| \mathbb{E}\left[P_{i,j+1}\hat{\sigma}(t_j)^2 \frac{\alpha_j}{h} \mathbb{E}_{t_j}\left[\int_{t_j}^{t_{j+1}} s D^2 g_\varepsilon(\hat{X}_s^{t_i,x})\,ds \right] \right] \right|$$

$$(3.14) \qquad \leq C\varepsilon^{-1} \sum_{j=i}^{n-1} h \leq C'(T - t_i)\varepsilon^{-1}.$$

2b. By (3.12) and the boundedness of b and σ, we also estimate that

$$(3.15) \quad \left| Dg_\varepsilon(\hat{X}_s^{t_i,x})\hat{b}(s, \hat{X}_s^{t_i,x}) + \tfrac{1}{2}\operatorname{Tr}[D^2 g_\varepsilon(\hat{X}_s^{t_i,x})\hat{a}(s, \hat{X}_s^{t_i,x})] \right| \leq C + C\varepsilon^{-1}.$$

2c. Plugging (3.14) and (3.15) into (3.13), we obtain

$$\left| \mathbb{E}\left[(g_\varepsilon(\hat{X}_T^{t_i,x}) - g_\varepsilon(x))P_{i,n} \right] \right| \leq C(T - t_i) + C(T - t_i)\varepsilon^{-1},$$

which, by (3.11), provides

$$|v^h(t_i, x) - g(x)| \leq C\varepsilon + C(T - t_i)\varepsilon^{-1} + C(T - t_i).$$

The required result follows from the choice $\varepsilon = \sqrt{T - t_i}$. □

COROLLARY 3.18. *The function v^h is $1/2$-Hölder continuous on t uniformly on h.*

PROOF. The proof of $\frac{1}{2}$-Hölder continuity with respect to t could easily be provided by replacing g and $v^h(t_k, \cdot)$ in the assertion of Lemma 3.17 by $v^h(t, \cdot)$ and $v^h(t', \cdot)$, respectively, and considering the scheme from 0 to time t' with time step equal to h. Therefore, we can write

$$|v^h(t, x) - v^h(t', x)| \leq C(t' - t)^{1/2},$$

where C can be chosen independently of t' for $t' \leq T$. □

3.3. *Derivation of the rate of convergence.* The proof of Theorem 3.10 is based on Barles and Jakobsen [5], which uses switching systems approximation and the Krylov method of shaking coefficients [21].

3.3.1. *Comparison result for the scheme.* Because F does not satisfy the standard monotonicity condition (3.7) of Barles and Souganidis [6], we need to introduce the nonlinearity $\overline{F}$ of Remark 3.13 so that $\overline{F}$ satisfies (3.6). Let u^h be the family of functions defined by

$$(3.16) \qquad u^h(T, \cdot) = g \quad \text{and} \quad u^h(t_i, x) = \overline{\mathbf{T}}_h[u^h](t_i, x),$$

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

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

and set

$$(3.17) \qquad \overline{v}^h(t_i, x) := e^{-\theta(T - t_i)} u^h(t_i, x), \qquad i = 0, \ldots, n.$$

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

LEMMA 3.19. *Under Assumption F, we have*

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

PROOF. By the definition of $\overline{F}$, we directly calculate that

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

Since $1 + h\theta = e^{\theta h} + O(h^2)$, this shows that $\overline{v}^h(t, x) = \mathbf{T}_h[\overline{v}^h](t, x) + O(h^2)$. By Lemma 3.14, we conclude that

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

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

By Remark 3.13, the operator $\overline{\mathbf{T}}_h$ satisfies the standard monotonicity condition (3.7):

$$(3.18) \qquad \varphi \leq \psi \quad \Longrightarrow \quad \overline{\mathbf{T}}_h[\varphi] \leq \overline{\mathbf{T}}_h[\psi].$$

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

PROPOSITION 3.20.   *Let Assumption F hold true and set $\beta := |F_r|_\infty$. Consider two arbitrary bounded functions φ and ψ satisfying*

(3.19) $$h^{-1}(\varphi - \overline{\mathbf{T}}_h[\varphi]) \le g_1 \quad and \quad h^{-1}(\psi - \overline{\mathbf{T}}_h[\psi]) \ge g_2$$

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

(3.20)
$$
\begin{aligned}
(\varphi - \psi)(t_i, x) &\le e^{\beta(T - t_i)} |(\varphi - \psi)^+(T, \cdot)|_\infty \\
&\quad + (T - h)e^{\beta(T - t_i)} |(g_1 - g_2)^+|_\infty.
\end{aligned}
$$

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

LEMMA 3.21.   *Let Assumption F hold true and let $\beta := |F_r|_\infty$. Then, for every $a, b \in \mathbb{R}_+$ and all bounded functions $\varphi \le \psi$, the function $\delta(t) := e^{\beta(T-t)}(a + b(T - t))$ satisfies*

$$\overline{\mathbf{T}}_h[\varphi + \delta](t, x) \le \overline{\mathbf{T}}_h[\psi](t, x) + \delta(t) - hb, \qquad t \le T - h, \; x \in \mathbb{R}^d.$$

PROOF.   Because δ does not depend on x, we have $\mathcal{D}_h[\varphi + \delta] = \mathcal{D}_h\varphi + \delta(t + h)e_1$, where $e_1 := (1, 0, 0)$. It then follows from the regularity of $\overline{F}$ that there exists some ξ such that

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

and

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

Since $\overline{\mathbf{T}}_h$ satisfies the standard monotonicity condition (3.18), this gives

$$\overline{\mathbf{T}}_h[\varphi + \delta](t, x) \le \overline{\mathbf{T}}_h[\psi](t, x) + \delta(t) + \zeta(t),$$

$$\text{where } \zeta(t) := (1 + \beta h)\delta(t + h) - \delta(t).$$

It remains to prove that $\zeta(t) \le -hb$. From the smoothness of δ, we have $\delta(t + h) - \delta(t) = h\delta'(\bar{t})$ for some $\bar{t} \in [t, t + h)$. Then, since δ is decreasing in t, we see that

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

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

PROOF OF PROPOSITION 3.20. We may directly refer to the similar result of [5]. However, in our context, we give the following, simpler, proof. Observe that we may assume without loss of generality that

$$(3.21) \qquad \varphi(T, \cdot) \le \psi(T, \cdot) \quad \text{and} \quad g_1 \le g_2.$$

Indeed, we can otherwise consider the function

$$\bar{\psi} := \psi + e^{\beta(T-t)}(a + b(T - t)),$$

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

and β is the parameter defined in the previous Lemma 3.21, so that $\bar{\psi}(T, \cdot) \ge \varphi(T, \cdot)$ and, by Lemma 3.21, $\bar{\psi}(t, x) - \overline{\mathbf{T}}_h[\bar{\psi}](t, x) \ge h(g_1 \vee g_2)$. Hence, (3.21) holds true for φ and $\bar{\psi}$.

We now prove the required result by induction. First, $\varphi(T, \cdot) \le \psi(T, \cdot)$, by (3.21). We next assume that $\varphi(t + h, \cdot) \le \psi(t + h, \cdot)$ for some $t + h \le T$. Since $\overline{\mathbf{T}}_h$ satisfies the standard monotonicity condition (3.18), it follows from (3.21) that

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

On the other hand, under (3.21), the hypothesis of the lemma implies that

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

Then, $\varphi(t, \cdot) \le \psi(t, \cdot)$. $\quad\square$

3.3.2. *Proof of Theorem* 3.10(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 coefficients, which is Lipschitz in x, 1/2-Hölder continuous in t and approximates uniformly the solution v:

$$v - \varepsilon \le v^\varepsilon \le v.$$

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

$$(3.22) \qquad \underline{w}^\varepsilon(t, x) := v^\varepsilon * \rho^\varepsilon, \qquad \text{where } \rho^\varepsilon(t, x) := \frac{1}{\varepsilon^{d+2}} \rho\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right)$$

so that, from the convexity of the operator F,

$$(3.23) \qquad \underline{w}^\varepsilon \text{ is a subsolution of (2.1)}, \qquad |\underline{w}^\varepsilon - v| \le 2\varepsilon.$$

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

$$\underline{w}^\varepsilon \text{ is } C^\infty, \quad \text{and} \quad |\partial_t^{\beta_0} D^\beta \underline{w}^\varepsilon| \le C\varepsilon^{1-2\beta_0-|\beta|_1}$$

$$(3.24)$$

$$\text{for any } (\beta_0, \beta) \in \mathbb{N} \times \mathbb{N}^d \setminus \{0\},$$

where $|\beta|_1 := \sum_{i=1}^d \beta_i$, and $C > 0$ is some constant. As a consequence of the consistency result of Lemma 3.11 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 $h \to 0$. The next key ingredient is to estimate the rate of convergence of $\mathcal{R}_h[\underline{w}^\varepsilon]$ to zero.

LEMMA 3.22. *For a family* $\{\varphi_\varepsilon\}_{0<\varepsilon<1}$ *of smooth functions satisfying* (3.24), *we have*

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

The proof of this result is included at the end of this section. From the previous estimate, together with the subsolution property of w^ε, we see that $\underline{w}^\varepsilon \le \overline{\mathbf{T}}_h[\underline{w}^\varepsilon] + Ch^2\varepsilon^{-3}$. It then follows from Proposition 3.20 that

$$(3.25) \qquad \underline{w}^\varepsilon - \overline{v}^h \le C|(\underline{w}^\varepsilon - \overline{v}^h)(T, \cdot)|_\infty + Ch\varepsilon^{-3} \le C(\varepsilon + h\varepsilon^{-3}).$$

We now use (3.23) and (3.25) to conclude that

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

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

$$(3.26) \qquad\qquad\qquad v - \overline{v}^h \le Ch^{1/4}.$$

3.3.3. *Proof of Theorem* 3.10(ii). The results of the previous section, together with the reinforced Assumption HJB+, allow the switching system method of Barles and Jakobsen [5] to be applied, which provides the following lower bound on the error:

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

for some constants $C, C' > 0$. The required rate of convergence again follows from Lemma 3.19, which states that the difference $v^h - \overline{v}^h$ is dominated by the above rate of convergence.

PROOF OF LEMMA 3.22. *Step 1.* 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

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

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

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

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

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

Using the boundedness of the coefficients μ and σ, it follows from (3.24) that for $\varepsilon \in (0,1)$,

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

Step 2. This implies that

$$|\mathcal{R}_h[\varphi_\varepsilon](t,x)|$$

(3.27)
$$\leq \left|\frac{\mathbb{E}\varphi_\varepsilon(t+h,\hat{X}_h^{t,x}) - \varphi_\varepsilon(t,x)}{h} - \mathcal{L}^X\varphi_\varepsilon(t,x)\right|$$
$$+ |F(x,\varphi_\varepsilon(t,x),D\varphi_\varepsilon(t,x),D^2\varphi_\varepsilon(t,x)) - F(\cdot,\mathcal{D}_h[\varphi_\varepsilon](t,x))|$$
$$\leq R_0(h,\varepsilon) + C\sum_{k=0}^2 |\mathbb{E}D^k\varphi_\varepsilon(t+h,\hat{X}_h^{t,x}) - D^k\varphi_\varepsilon(t,x)|,$$

by the Lipschitz continuity of the nonlinearity F.

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

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

which, together with (3.27), provides the required result. $\square$

3.4. *The rate of convergence in the linear case.* In this subsection, we specialize the discussion to the linear one-dimensional case

$$(3.28) \qquad\qquad\qquad F(\gamma) = c\gamma$$

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

$$(3.29) \quad v(t,x) = \mathbb{E}\big[g\big(x + \sqrt{1+2c}\,W_{T-t}\big)\big] \qquad \text{for } (t,x) \in [0,T] \times \mathbb{R}^d.$$

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

$$(3.30) \qquad D^k v(t,x) = \mathbb{E}\big[g^{(k)}\big(x + \sqrt{1+2c}\,W_{T-t}\big)\big], \qquad t < T, \ x \in \mathbb{R}.$$

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

Of course, we can use the classical Monte Carlo estimate to produce an approximation of the function v of (3.29). The objective of this section is to analyze the error of the numerical scheme outlined in the previous sections:

$$\begin{aligned} &v^h(T, \cdot) = g, \\ &v^h(t_{i-1}, x) = \mathbb{E}[v^h(t_i, x + W_h)] + ch\mathbb{E}[v^h(t_i, x + W_h)H_2^h], \qquad i \leq n, \end{aligned}$$

(3.31)

where we set $\sigma = 1$ and $\mu = 0$ for simplicity.

PROPOSITION 3.23.   *Consider the linear operator F of* (3.28) *and assume that $D^{(2k+1)}v$ is bounded for every $k \geq 0$. Then,*

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

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 $i \leq n - 1$,

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

(3.32)

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

Step 1. Set

$$a_i^k := \mathbb{E}[\Delta^k u(t_i, X_{t_i})],$$

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

introduce the matrices

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

and observe that (3.32) implies that the vectors $a^k := (a_1^k, \ldots, a_n^k)^{\mathrm{T}}$ and $b^k := (b_1^k, \ldots, b_n^k)^{\mathrm{T}}$ satisfy $A a^k = ch B a^{k+1} + cB b^k$ for all $k \geq 0$, and therefore

$$a^k = ch A^{-1} B a^{k+1} + cA^{-1} B b^k,$$

(3.33)

$$\text{where } A^{-1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

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

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

In particular, because $a_n^k = 0$, $(A^{-1}B)^{n-1} a^k = 0$. Iterating (3.33), this gives

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

and therefore

(3.34)
$$u(0, x) = a_1^0 = c \sum_{k=0}^{n-2} (ch)^k (A^{-1}B)_{1,j}^{k+1} b^k.$$

Because of

$$(A^{-1}B)_{1,j}^k = \mathbb{1}_{\{j \geq 1+k\}} \binom{j-2}{k-1} \qquad \text{for all } k \geq 1 \text{ and } j = 1, \ldots, n,$$

we can write (3.34) as follows:

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

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

(3.35)
$$u(0, x) = c \sum_{j=2}^{n} \sum_{k=0}^{j-2} (ch)^k \binom{j-2}{k} b_j^{k-1}.$$

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

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

for some constant C. We then deduce from (3.35) that

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

So,

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

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 objective 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:

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

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

where

$$\hat{\mathcal{D}}_h\psi(t,x) := \hat{\mathbb{E}}^N[\psi(t+h, \hat{X}_h^{t,x})H_h(t,x)], \qquad K_h[\psi] := \|\psi\|_\infty(1+C_1h) + C_2h,$$

where

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

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

With this notation, the implementable numerical scheme is

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

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

Let $\mathcal{R}_b$ be the family of random variables R of the form $\psi(W_h)H_i(W_h)$, where ψ is a function with $|\psi|_\infty \leq b$ and the H_i's are Hermite polynomials:

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

ASSUMPTION E. There exist constants $C_b, \lambda, \nu > 0$ such that

$$\|\hat{\mathbb{E}}^N[R] - \mathbb{E}[R]\|_p \leq C_b h^{-\lambda} N^{-\nu}$$

for every $R \in \mathcal{R}_b$, 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 [8] and analyzed in the context of the simulation of backward stochastic differential equations by [9] and [14]. Assumption E is then 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 $\hat{v}_N^h$ toward v.

THEOREM 4.2. *Let Assumptions* E *and* F *hold true and assume that the fully nonlinear PDE* (2.1) *has comparison with growth* q. *Suppose, in addition, that*

$$\text{(4.4)} \qquad \lim_{h \to 0} h^{\lambda+2} N_h^\nu = \infty.$$

Assume that the final condition g *is bounded Lipschitz, and the coefficients* μ *and* σ *are bounded. Then, for almost every* ω,

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

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

PROOF. We adapt the argument of [6] to the present stochastic context. By Remark 3.13 and Lemma 3.19, we may assume without loss of generality that the strict monotonicity (3.6) holds.

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

$$\text{(4.5)} \qquad \begin{aligned} \hat{v}_*(t, x) &:= \liminf_{\substack{(t', x') \to (t, x) \\ h \to 0}} \hat{v}^h(t', x') \quad \text{and} \\ \hat{v}^*(t, x) &:= \limsup_{\substack{(t', x') \to (t, x) \\ h \to 0}} \hat{v}^h(t', x'). \end{aligned}$$

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 3.6. In particular, they are both deterministic functions.

We shall only include the proof of the supersolution property as the subsolution property follows by 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 $\varphi \in C^2([0, T) \times \mathbb{R}^n)$, such that

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

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

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

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 gives

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

where, for ease of notation, 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

$$\varphi(t_n, x_n) + C_n - \mathbf{T}_{h_n}[\varphi + C_n](t_n, x_n) + h_n R_n \geq 0,$$

where $R_n := h_n^{-1}(\mathbf{T}_{h_n} - \hat{\mathbf{T}}_{h_n})[\hat{v}^{h_n}](t_n, x_n)$.

We claim that

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

Then, after passing to the subsequence, dividing both sides by h_n and sending $n \to \infty$, it follows from Lemma 3.11 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 3.14, $|\mathbf{T}_{h_n}[\hat{v}^{h_n}]|_\infty \leq K_{h_n}$ and so by (4.2), we can write

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

By the Lipschitz continuity of F, we have

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

where

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

Therefore,

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

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 the Hermite polynomial of degree i. This leads to the following estimate for the error R_n:

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

Because $R_n^i \in \mathcal{R}_b$ with bound obtained in Lemma 3.14, 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). $\quad\square$

Finally, we discuss the choice of sample size so as to keep the same rate for the error bound.

THEOREM 4.3. *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*

$$(4.9) \qquad \lim_{h\to 0} h^{\lambda+21/10} N_h^\nu > 0.$$

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

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

PROOF. By Remark 3.13 and Lemma 3.19, we may assume without loss of generality that the strict monotonicity (3.6) holds true.

We proceed as in the proof of Theorem 3.10 to see that

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

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

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

where, in the present context, $R_h[\hat{v}^h]$ is a nonzero stochastic term. By Proposition 3.20, it follows from the last inequality that

$$v - \hat{v}^h \le C\big(\varepsilon + R(h,\varepsilon) + R_h[\hat{v}^h]\big),$$

where the constant $C > 0$ depends only on the Lipschitz coefficient of F, β in Lemma 3.21 and the constant in Lemma 3.22.

Similarly, we follow the line of argument of the proof of Theorem 3.10 to show that a lower bound holds true and therefore

$$|v - \hat{v}^h| \le C\big(\varepsilon^{1/3} + R(h,\varepsilon) + R_h[\hat{v}^h]\big).$$

We now use (4.9) and proceed as in the last part of the proof of Theorem 4.2 to deduce from (4.8) and Assumption F that

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

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

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

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

$\hfill\square$

EXAMPLE 4.4. Let us illustrate the convergence results of this section in the context of the Malliavin integration by parts regression 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 difference 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 two-dimensional space 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 two dimensions, where an explicit solution is available, and we consider more involved examples in five spatial dimensions, 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 notation and simply write the scheme as

(5.1)
$$v^h(T, .) := g \quad \text{and}$$
$$v^h(t_i, x) := \mathbb{E}[v^h(t_{i+1}, \hat{X}_h^x)] + h F(t_i, x, \mathcal{D}_h v^h(t_i, x)),$$

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 2.2 that

(5.2)
$$\mathcal{D}_{2h}^2 \varphi(t_i, x) = \mathbb{E}\left[\varphi(t_i + 2h, \hat{X}_{2h}^{t_i, x})(\sigma^{\mathrm{T}})^{-1} \right.$$
$$\left. \times \frac{(W_{t_i+h} - W_{t_i})(W_{t_i+h} - W_{t_i})^{\mathrm{T}} - h\mathbf{I}_d}{h^2} \sigma^{-1} \right]$$
$$= \mathbb{E}\left[\mathcal{D}_h^1 \varphi(t_i + h, \hat{X}_h^{t_i, x})(\sigma^{\mathrm{T}})^{-1} \frac{W_{t_i+h} - W_{t_i}}{h} \right].$$

The second representation is the one reported in [12], 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 no longer holds 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 ∇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 in the style of 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 adaptive 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 [8]. 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 η_k is provided in [8] and should be chosen for each conditional expectation depending on k. However, our numerical experiments 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

$$(5.3) \qquad v_t - \Delta v + \frac{Dv \cdot D^2 v \, Dv}{|Dv|^2} = 0 \quad \text{and} \quad v(0, x) = g(x),$$

where $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ is a bounded Lipschitz continuous function. We refer to [28] 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 $g(x) := 4R^2 - |x|^2$ so that g is positive inside the sphere and negative outside. We first solve the sphere problem in three dimensions. In this case, it is well known that the

surface $S(t)$ is a sphere with radius $R(t) = 2\sqrt{R^2 - t}$ for $t \in (0, R^2)$. Reversing time, we rewrite (5.3) for $t \in (0, T)$ with $T = R^2$:

$$(5.4) \qquad -v_t - \tfrac{1}{2}\sigma^2 \Delta v + F(x, Dv, D^2 v) = 0 \quad \text{and} \quad v(T, x) = g(x),$$

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 difference 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 to equal 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] is 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 a $10 \times 10 \times 10$ mesh with a time step equal to 0.0125. The diffusion coefficient σ 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 2 plots the difference between our calculation and the reference for scheme 1 and volatility 1 and 1.8 for a varying time step. The corresponding results with scheme 2 are reported in Figure 3. We observe that some points at time $T = 0.25$ are missing due to a nonconvergence of the gradient method for a diffusion $\sigma = 1.8$. We observe that results for scheme 2 are slightly better than results for scheme 1. With $\sigma = 1$, it takes 150 seconds on a Intel Nehalem processor (2.9 GHz) to obtain the result at time $t = 0.15$ with the regression method, while it takes 1500 seconds with the Malliavin method (notice that the dichotomy used with the gradient method is a very inefficient method).
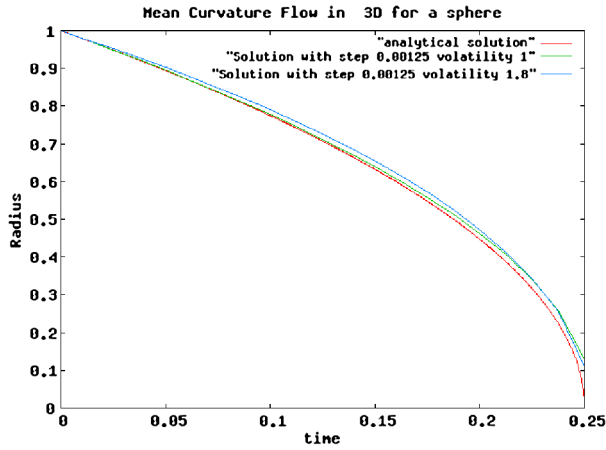
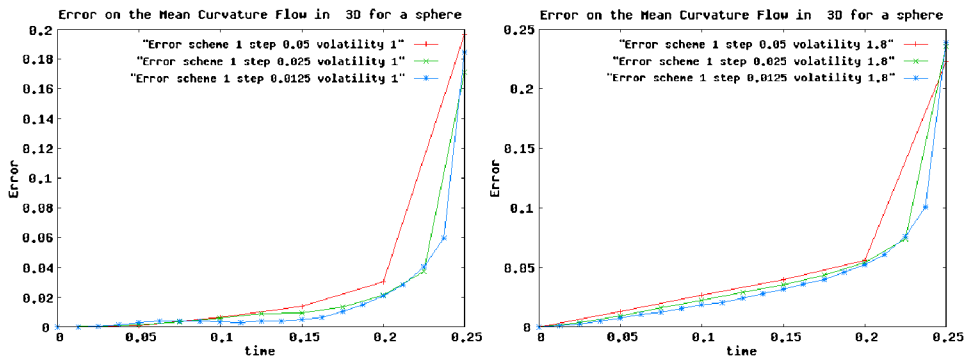FIG. 1. *Solution of the mean curvature flow for the sphere problem.*

FIG. 2. *Mean curvature flow problem for different time step and diffusion*: *scheme* 1.
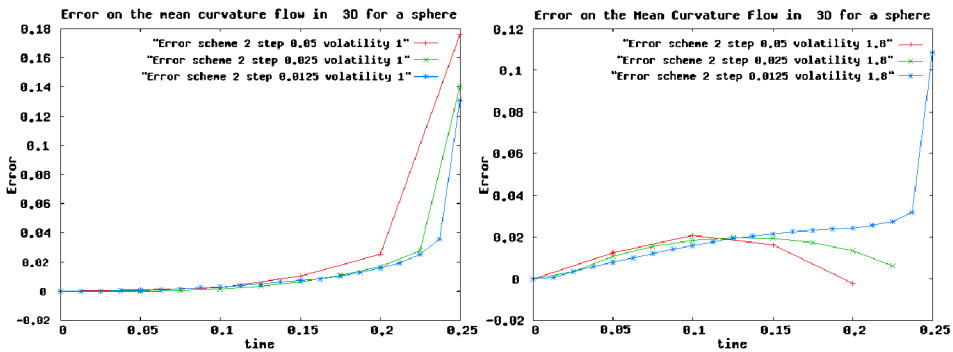

FIG. 3. *Mean curvature flow problem for different time step and diffusions*: *scheme* 2.
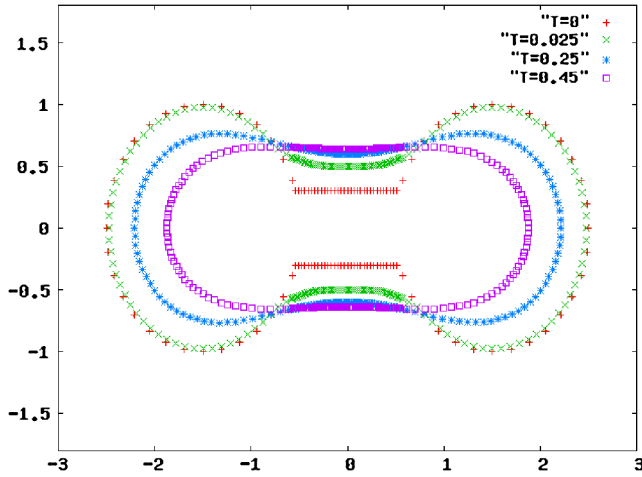
FIG. 4. *Mean curvature flow problem in two dimensions.*

In Figure 4, we finally report some numerical results for the mean curvature flow problem in two dimensions with a more interesting geometry: the initial surface (i.e., the zero-level set for v) consists of two discs with unit radius, with centers positioned at -1.5 and 1.5 and connected by a strip 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 Malliavin-integration-by-parts-based regression method and the basis projection method with 10×10 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 interface (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 θ_t^i is the amount invested in the ith security held at time t. In addition, the investor has access to a nonrisky security (bank account), where the remaining part of his wealth is invested. The nonrisky asset S^0 is defined by an adapted interest rates process $\{r_t, t \in [0, T]\}$, that is, $dS_t^0 = S_t^0 r_t \, dt, t \in [0, 1]$. Then, the dynamics of the wealth process is described by

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

where $\mathbb{1} = (1, \ldots, 1) \in \mathbb{R}^d$. Let $\mathcal{A}$ be the collection of all adapted processes θ with values in $\mathbb{R}^d$ which are integrable with respect to S and such that the process X^θ is

uniformly bounded from below. Given an absolute risk aversion coefficient $\eta > 0$, the portfolio optimization problem is defined by

$$(5.5) \qquad\qquad v_0 := \sup_{\theta \in \mathcal{A}} \mathbb{E}[-\exp(-\eta X_T^\theta)].$$

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 difference scheme to derive an approximation of the solution of the fully nonlinear HJB equation in nontrivial situations where the state has a few dimensions. We shall first start with a two-dimensional example where an explicit solution of the problem is available. We will then 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]:

$$dS_t = \mu S_t \, dt + \sqrt{Y_t} S_t \, dW_t^{(1)},$$

$$dY_t = k(m - Y_t) \, dt + c\sqrt{Y_t}\left(\rho \, dW_t^{(1)} + \sqrt{1 - \rho^2} \, dW_t^{(2)}\right),$$

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 (5.5) 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:

$$v(T, x, y) = -e^{-\eta x} \quad \text{and}$$

$$(5.6) \qquad
\begin{aligned}
0 = {}& -v_t - k(m - y)v_y - \frac{1}{2}c^2 y v_{yy} \\
& - \sup_{\theta \in \mathbb{R}}\left(\frac{1}{2}\theta^2 y v_{xx} + \theta(\mu v_x + \rho c y v_{xy})\right) \\
= {}& -v_t - k(m - y)v_y - \frac{1}{2}c^2 y v_{yy} + \frac{(\mu v_x + \rho c y v_{xy})^2}{2 y v_{xx}}.
\end{aligned}$$

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

$$(5.7) \qquad v(t, x, y) = -e^{-\eta x}\left\| \exp\left(-\frac{1}{2}\int_t^T \frac{\mu^2}{\tilde{Y}_s} \, ds\right)\right\|_{\mathbb{L}^{1-\rho^2}},$$

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

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

In order to implement our Monte Carlo/finite difference scheme, we rewrite (5.6) as

$$-v_t - k(m-y)v_y - \tfrac{1}{2}c^2 y v_{yy} - \tfrac{1}{2}\sigma^2 v_{xx} + F(y, Dv, D^2 v) = 0,$$

(5.8)

$$v(T, x, y) = -e^{-\eta x},$$

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

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

Notice that the nonlinearity F does not to satisfy Assumption F; we consider the truncated nonlinearity

$$F_{\varepsilon,M}(y, z, \gamma) := \frac{1}{2}\sigma^2 \gamma_{11} - \sup_{\varepsilon \leq \theta \leq M} \left(\frac{1}{2}\theta^2 (y \vee \varepsilon)\gamma_{11} + \theta(\mu z_1 + \rho c(y \vee \varepsilon)\gamma_{12}) \right)$$

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

$$(5.9)\quad dX_t^{(1)} = \sigma \, dW_t^{(1)} \quad \text{and} \quad dX_t^{(2)} = k(m - X_t^{(2)}) \, dt + c\sqrt{X_t^{(2)}} \, dW_t^{(2)}.$$

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

$$(5.10)\qquad X_n^{(2)} = \frac{X_{n-1}^{(2)} + km\Delta t + c\sqrt{X_{n-1}^{(2)}}\xi_n\sqrt{\Delta t} + \tfrac{1}{4}c^2 \Delta(\xi_n^2 - 1)}{1 + k\Delta t},$$

where $(\xi_n)_{n \geq 1}$ is a sequence of independent random variables 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 $x_0 = 1$, and the maturity T is taken equal to one year. With these parameters, the value function is computed from the quasi-explicit formula (5.7) to be $v_0 = -0.3534$.

We also choose $M = 40$ for the truncation of the nonlinearity. This choice turns out to be critical, as an initial choice of $M = 10$ 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 40×10 basis functions. We provide numerical results corresponding to two million particles. Our numerical results show that the Malliavin and basis projection methods produce very similar results and achieve good accuracy. With two million 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 1 (except for a step equal to 0.025 and a volatility equal to 1.2, where the standard deviation jumped to 0.038) and smaller than 0.002 for scheme 2, with a computing time of 378 seconds for 40 time steps;
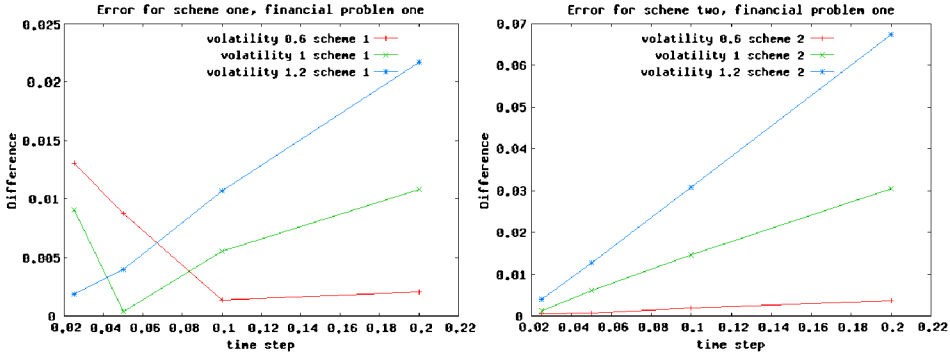
FIG. 5. *Difference between calculation and reference for schemes 1 and 2.*

- the results of the basis projection method exhibit a standard deviation smaller than 0.002 for scheme 1 and smaller than 0.0009 for scheme 2, with a computing time of 114 seconds for 40 time steps.

Figure 5 provides the plots of the errors obtained by the integration-by-parts-based regression with schemes 1 and 2. All solutions have been calculated as averages 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 3.6 might be a sharp condition. We also observe that scheme 1 has a persistent bias, even for a very small time step, while scheme 2 exhibits a better convergence toward the solution.

5.2.2. *A five-dimensional example.* We now let $n = 2$ and 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 model (see, e.g., [26] for a presentation of these models and their simulation):

$$dS_t^{(i)} = \mu_i S_t^{(i)}\,dt + \sigma_i \sqrt{Y_t^{(i)}}\,S_t^{(i)^{\beta_i}}\,dW_t^{(i,1)}, \qquad \beta_2 = 1,$$

$$dY_t^{(i)} = k_i\big(m_i - Y_t^{(i)}\big)\,dt + c_i\sqrt{Y_t^{(i)}}\,dW_t^{(i,2)},$$

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

Since $\beta_2 = 1$, the value function of the portfolio optimization problem 5.5 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

$$
0 = -v_t - (\mathbf{L}^r + \mathbf{L}^Y + \mathbf{L}^{S^1})v - rxv_x
$$

$$
- \sup_{\theta_1,\theta_2} \left\{ \theta_1 \cdot (\mu - r\mathbb{1})v_x + \theta_1 \sigma_1^2 y_1 s_1^{2\beta_1-1} v_{xs_1} \right.
$$

<div align="right">(5.11)</div>

$$
\left. + \frac{1}{2}(\theta_1^2 \sigma_1^2 y_1 s_1^{2\beta_1-2} + \theta_2^2 \sigma_2^2 y_2)v_{xx} \right\}
$$

$$
= -v_t - (\mathbf{L}^r + \mathbf{L}^Y + \mathbf{L}^{S^1})v - rxv_x
$$

$$
+ \frac{((\mu_1 - r)v_x + \sigma_1^2 y_1 s_1^{2\beta_1-1} v_{xs_1})^2}{2\sigma_1^2 y_1 s_1^{2\beta_1-2} v_{xx}} + \frac{((\mu_2 - r)v_x)^2}{2\sigma_2^2 y_2 v_{xx}},
$$

where

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

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

In order to implement our Monte Carlo/finite difference scheme, we rewrite (5.11) as

<div align="right">(5.12)</div>

$$
-v_t - (\mathbf{L}^r + \mathbf{L}^Y + \mathbf{L}^{S^1})v - \tfrac{1}{2}\sigma^2 v_{xx}
$$

$$
+ F((x, r, s_1, y_1, y_2), Dv, D^2 v) = 0,
$$

$$
v(T, x, r, s_1, y_1, y_2) = -e^{-\eta x},
$$

where $\sigma > 0$, and the nonlinearity $F : \mathbb{R}^5 \times \mathbb{R}^5 \times \mathbb{S}_2$ is given by

$$
F(u, z, \gamma) = \frac{1}{2}\sigma^2 \gamma_{11} - x_1 x_2 z_1
$$

$$
+ \frac{((\mu_1 - x_2)z_1 + \sigma_1^2 x_4 x_3^{2\beta_1-1} \gamma_{1,3})^2}{2\sigma_1^2 x_4 x_3^{2\beta_1-2} \gamma_{11}} + \frac{((\mu_2 - x_2)z_1)^2}{2\sigma_2^2 x_5 \gamma_{11}},
$$

where $u = (x_1, \ldots, x_5)$. We next consider the truncated nonlinearity

$$
F_{\varepsilon,M}(u, z, \gamma) := \frac{1}{2}\sigma^2 \gamma_{11} - x_1 x_2 z_1
$$

$$
+ \sup_{\varepsilon \le |\theta| \le M} \left\{ (\theta \cdot (\mu - r\mathbb{1})z_1 + \theta_2^2 \sigma_2^2 (x_5 \vee \varepsilon))\gamma_{11} \right.
$$

$$
+ \theta_1 \sigma_1^2 (x_4 \vee \varepsilon)(x_3 \vee \varepsilon)^{2\beta_1-1} \gamma_{13}
$$

$$
\left. + \frac{1}{2}\theta_1^2 \sigma_1^2 (x_3 \vee \varepsilon)(x_4 \vee \varepsilon)^{2\beta_1-2} \right\},
$$

where ε, $M > 0$ are jointly chosen with σ so that Assumption F holds true. In this setting, the forward two-dimensional diffusion is defined by

$$
\begin{aligned}
dX_t^{(1)} &= \sigma \, dW_t^{(0)}, \\
dX_t^{(2)} &= \kappa \big(b - X_t^{(2)}\big) \, dt + \zeta \, dW_t^{(1)}, \\
(5.13) \qquad dX_t^{(3)} &= \mu_1 X_t^{(3)} \, dt + \sigma_1 \sqrt{X_t^{(4)}} X_t^{(3)\beta_1} \, dW_t^{(1,1)}, \\
dX_t^{(4)} &= k_1 \big(m_1 - X_t^{(4)}\big) \, dt + c_1 \sqrt{X_t^{(4)}} \, dW_t^{(1,2)}, \\
dX_t^{(5)} &= k_2 \big(m_2 - X_t^{(5)}\big) \, dt + c_2 \sqrt{X_t^{(5)}} \, dW_t^{(2,2)}.
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
\ln X_n^{(3)} = \ln X_{n-1}^{(3)} &+ \big(\mu_1 - \tfrac{1}{2}\sigma_1^2 \big(X_{n-1}^{(3)}\big)^{2(\beta_1 - 1)} X_{n-1}^{(4)}\big)\Delta t \\
&+ \sigma_1 \big(X_{n-1}^{(3)}\big)^{\beta_i - 1} \sqrt{X_{n-1}^{(4)}} \Delta W_n^{(1,2)},
\end{aligned}
$$

where $\Delta W_n^{(1,2)} := W_n^{(1,2)} - W_{n-1}^{(1,2)}$. We take the parameters $\mu_1 = 0.10$, $\sigma_1 = 0.3$ and $\beta_1 = 0.5$ for the first asset, $k_1 = 0.1$, $m_1 = 1$ and $c_1 = 0.1$ 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 asset 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 which, for example, takes 520 seconds for 20 time steps. 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 found to diverge with a thinner time step. We observe that there is still a difference for a 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 carrying out many calculations, we observed that for the thinner time step mesh, the solution sometimes diverges. We therefore report the results corresponding to thirty million particles with $4 \times 4 \times 4 \times 4 \times 40$ meshes. First, we note that with this discretization, all results are converging as the time step goes to zero: the exact solution seems to be very close to -0.258. During our experiments with thirty million particles, the scheme was always converging,
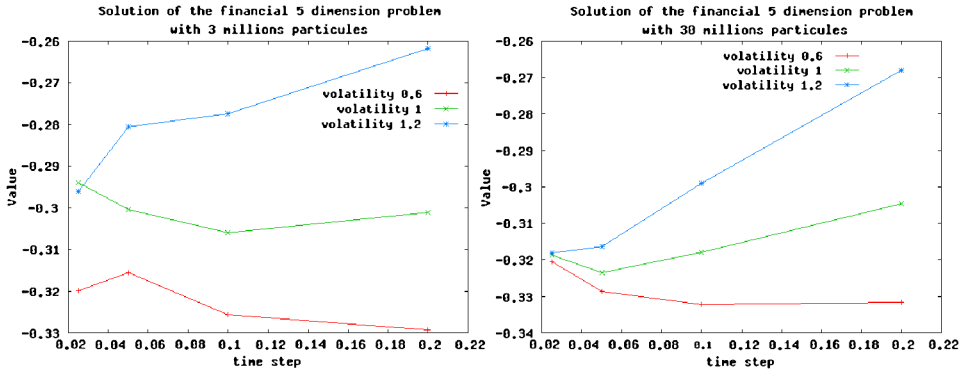
FIG. 6.   *The five-dimensional financial problem and its results for different volatilities with* 3 *million particles* (*left*) *and* 30 *million particles* (*right*).

with a very low variance for the results. A single calculation takes 5100 seconds with 20 time steps.

REMARK 5.4.   With thirty million particles, the required memory forced us to use 64-bit processors with more than 4 GB of memory.

5.2.3. *Conclusion on numerical results.*   The Monte Carlo/finite difference algorithm has been implemented with both schemes suggested by (5.2), using the basis projection and Malliavin regression methods. Our numerical experiments reveal that the second scheme performs better, both in term of results and time of calculation for a given number of particles, independently of the regression method used.

We also provided numerical results for different choices of the diffusion parameter in the Monte Carlo step. We observed that small diffusion coefficients lead to poor results, which hints that the condition that the diffusion must dominate the nonlinearity in Assumption F(iii) may be sharp. On the other hand, we also observed that large diffusions require high-level refinements of the meshes and large numbers of particles, leading to lengthy computational times.

Finally, we note that a reasonable choice for the diffusion could be time and state dependent, as in the classical importance sampling method. We have not attempted any experiments in this direction, and we eventually hope to have some theoretical results on how to choose optimal drift and diffusion coefficients for the Monte Carlo step.

REFERENCES

[1] ANDERSEN, L. B. G. and BROTHERTON-RATCLIFFE, R. (2005). Extended LIBOR market models with stochastic volatility. *Journal of Computational Finance* **9** 1–40.

[2] BALLY, V. (1997). Approximation scheme for solutions of BSDE. In *Backward Stochastic Differential Equations* (*Paris*, 1995–1996). *Pitman Res. Notes Math. Ser.* **364** 177–191. Longman, Harlow. MR1752682

[3] BARLES, G. and JAKOBSEN, E. R. (2002). On the convergence rate of approximation schemes for Hamilton–Jacobi–Bellman equations. *M2AN Math. Model. Numer. Anal.* **36** 33–54. MR1916291

[4] BARLES, G. and JAKOBSEN, E. R. (2005). Error bounds for monotone approximation schemes for Hamilton–Jacobi–Bellman equations. *SIAM J. Numer. Anal.* **43** 540–558 (electronic). MR2177879

[5] BARLES, G. and JAKOBSEN, E. R. (2007). Error bounds for monotone approximation schemes for parabolic Hamilton–Jacobi–Bellman equations. *Math. Comp.* **76** 1861–1893 (electronic). MR2336272

[6] BARLES, G. and SOUGANIDIS, P. E. (1991). Convergence of approximation schemes for fully nonlinear second order equations. *Asymptot. Anal.* **4** 271–283. MR1115933

[7] BONNANS, J. F. and ZIDANI, H. (2003). Consistency of generalized finite difference schemes for the stochastic HJB equation. *SIAM J. Numer. Anal.* **41** 1008–1021 (electronic). MR2005192

[8] BOUCHARD, B., EKELAND, I. and TOUZI, N. (2004). On the Malliavin approach to Monte Carlo approximation of conditional expectations. *Finance Stoch.* **8** 45–71. MR2022978

[9] BOUCHARD, B. and TOUZI, N. (2004). Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Process. Appl.* **111** 175–206. MR2056536

[10] CAMILLI, F. and FALCONE, M. (1995). An approximation scheme for the optimal control of diffusion processes. *RAIRO Modél. Math. Anal. Numér.* **29** 97–122. MR1326802

[11] CAMILLI, F. and JAKOBSEN, E. R. (2009). A finite element like scheme for integro-partial differential Hamilton–Jacobi–Bellman equations. *SIAM J. Numer. Anal.* **47** 2407–2431. MR2525605

[12] CHERIDITO, P., SONER, H. M., TOUZI, N. and VICTOIR, N. (2007). Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs. *Comm. Pure Appl. Math.* **60** 1081–1110. MR2319056

[13] CHEVANCE, D. (1996). Résolution numériques des équations différentielles stochastiques rétrogrades. Ph.D. thesis, Univ. Provence.

[14] CRISAN, D., MANOLARAKIS, K. and TOUZI, N. (2010). On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights. *Stochastic Process. Appl.* **120** 1133–1158. MR2639741

[15] DEBRABANT, K. and JAKOBSEN, E. R. (2009). Semi-Lagrangian schemes for linear and fully non-linear diffusion equations. Preprint.

[16] EL KAROUI, N., PENG, S. and QUENEZ, M. C. (1997). Backward stochastic differential equations in finance. *Math. Finance* **7** 1–71. MR1434407

[17] GOBET, E., LEMOR, J.-P. and WARIN, X. (2005). A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Probab.* **15** 2172–2202. MR2152657

[18] HESTON, S. L. (1993). A closed–form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* **6** 327–343.

[19] KAHL, C. and JÄCKEL, P. (2006). Fast strong approximation Monte Carlo schemes for stochastic volatility models. *Quant. Finance* **6** 513–536. MR2278638

[20] KOHN, R. V. and SERFATY, S. (2010). A deterministic-control-based approach to fully nonlinear parabolic and elliptic equations. *Comm. Pure Appl. Math.* **63** 1298–1350. MR2681474

[21] KRYLOV, N. V. (1997). On the rate of convergence of finite-difference approximations for Bellman's equations. *Algebra i Analiz* **9** 245–256. MR1466804

[22] KRYLOV, N. V. (2000). On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients. *Probab. Theory Related Fields* **117** 1–16. MR1759507

[23] KRYLOV, N. V. (2005). The rate of convergence of finite-difference approximations for Bellman equations with Lipschitz coefficients. *Appl. Math. Optim.* **52** 365–399. MR2174020

[24] LIONS, P. L. and 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. *The Review of Financial Studies* **14** 113–147.

[26] LORD, R., KOEKKOEK, R. and VAN DIJK, D. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quant. Finance* **10** 177–194. MR2642962

[27] MUNOS, R. and ZIDANI, H. (2005). Consistency of a simple multidimensional scheme for Hamilton–Jacobi–Bellman equations. *C. R. Math. Acad. Sci. Paris* **340** 499–502. MR2135230

[28] SONER, H. M. and TOUZI, N. (2003). A stochastic representation for mean curvature type geometric flows. *Ann. Probab.* **31** 1145–1165. MR1988466

[29] ZARIPHOPOULOU, T. (2001). A solution approach to valuation with unhedgeable risks. *Finance Stoch.* **5** 61–82. MR1807876

[30] ZHANG, J. (2004). A numerical scheme for BSDEs. *Ann. Appl. Probab.* **14** 459–488. MR2023027

A. FAHIM
N. TOUZI
CMAP, ÉCOLE POLYTECHNIQUE
ROUTE DE SACLAY
91128 PALAISEAU
FRANCE
E-MAIL: arash.fahim@polytechnique.edu
        nizar.touzi@polytechnique.edu

X. WARIN
ÉLECTRICITÉ DE FRANCE—EDF R&D
L'AVENUE DU GÉNÉRAL DE GAULLE
92141 CLAMART
FRANCE
E-MAIL: xavier.warin@edf.fr
URL: www.fime-lab.org