SIAM J. NUMER. ANAL.
Vol. 48, No. 1, pp. 257–277

SOLVING BSDE WITH ADAPTIVE CONTROL VARIATE*

EMMANUEL GOBET† AND CÉLINE LABART‡

Abstract. We present and analyze an algorithm to solve numerically BSDEs based on Picard's iterations and on a sequential control variate technique. Its convergence is geometric. Moreover, the solution provided by our algorithm is regular both w.r.t. time and space.

Key words. backward stochastic differential equations, adaptive control variate, semilinear parabolic PDE

AMS subject classifications. 60H10, 60H35, 65C05, 65G99

DOI. 10.1137/090755060

1. Introduction.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space on which is defined a q-dimensional standard Brownian motion W, whose natural filtration, augmented with $\mathbb{P}$-null sets, is denoted $(\mathcal{F}_t)_{0\leq t\leq T}$ (T is a fixed terminal time). We aim at numerically approximating the solution (Y, Z) of the following forward backward stochastic differential equation (FBSDE) with fixed terminal time T

$$-dY_t = f(t, X_t, Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = \Phi(X_T), \tag{1.1}$$

where $f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$, $\Phi : \mathbb{R}^d \to \mathbb{R}$ and X is the $\mathbb{R}^d$-valued process solution of

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \tag{1.2}$$

$b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, and $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times q}$. The main focus of this work is to provide and analyze an algorithm—based on Picard's iterations and an adaptive Monte Carlo method—to approximate the solution (Y, Z) of (1.1).

Several algorithms to solve BSDEs can be found in the literature. Ma, Protter, and Yong [22] present an algorithm to solve quasilinear PDEs (associated to forward BSDEs) using a finite difference approximation. More recently, Fahim, Touzi, and Warin [12] combine Monte Carlo and finite difference schemes to solve fully nonlinear PDEs. Concerning algorithms based on the dynamic programming equation, we refer to Bouchard and Touzi [7], Gobet, Lemor, and Warin [16], Bally and Pagès [3] and Delarue and Menozzi [9]. In [7], the authors compute the conditional expectations appearing in the dynamic programming equation by using Malliavin calculus techniques, whereas [16] proposes a scheme based on iterative regression functions which are approximated by projections on a reduced set of functions, the coefficients of the projection being evaluated using Monte Carlo simulations. [3] and [9] use quantization techniques for solving reflected BSDEs and forward BSDEs, respectively. Bender and Denk [4] propose a forward scheme which avoids the nesting of conditional expectations backwards through the time steps. Instead, it mimics Picard's type of iterations

*Received by the editors April 6, 2009; accepted for publication (in revised form) January 19, 2010; published electronically April 2, 2010.
http://www.siam.org/journals/sinum/48-1/75506.html
†Corresponding author. Laboratoire Jean Kuntzmann, Université de Grenoble and CNRS, BP 53, 38041 Grenoble cedex 9, France (emmanuel.gobet@imag.fr).
‡CMAP, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau, France and INRIA Paris-Rocquencourt, MathFi project (celine.labart@upmc.fr).

for BSDEs and, consequently, has nested conditional expectations along the iterations. This work has some connections to our approach, but it does not handle the error analysis for the conditional expectations. We also refer to Bender and Zhang [5], where an algorithm based on Picard iterations for coupled FBSDEs is derived.

Our algorithm works as follows. First, we use Picard's iterations to approximate the solution (Y, Z) of (1.1) by the solutions of a sequence of linear BSDEs converging geometrically fast to (Y, Z) (see El Karoui, Peng, and Quenez [11] for more details). Then, since we can link linear BSDEs and linear PDEs, we use the adaptive control variate method proposed by Gobet and Maire [17]. This method approximates the solutions of linear PDEs, which can be written as expectations of functionals of Markov processes via Feynman–Kac's formula. The authors use a control variate, changing at each step of the algorithm, to reduce the variance of the simulations. The convergence of this technique is also geometric w.r.t. the iterations (see [17] for more details). As a consequence, we can guess that combining these two methods in order to solve BSDEs will lead to a geometrically converging algorithm. As a difference with previous works, we provide an approximated solution to the semilinear PDE that has the same smoothness as the exact solution, which is quite satisfactory and potentially useful. As another difference with other Monte Carlo approaches, the final accuracy does not depend much on the number of simulations, but rather on an operator $\mathcal{P}$ used for approximating functions. It means that our algorithm benefits from the ability of Monte Carlo methods to solve high dimensional problems, without suffering for their relatively low rate of convergence.

The paper is organized as follows. In section 1, we give some definitions and notations and recall the link between BSDEs and semilinear PDEs. In section 2, we describe the two main ingredients (Picard's iterations and adaptive control variate) of our algorithm. In section 3, we define the norm used to measure the convergence of the algorithm, and in section 4, we present the operator $\mathcal{P}$ used in the algorithm to approximate functions, emphasizing the important properties that $\mathcal{P}$ should satisfy in order to make our algorithm converge. We give the main convergence result in section 5 and its proof in section 6. We present in section 7 an example of an operator $\mathcal{P}$ based on kernel estimators. Finally, in section 8, we expose some numerical results in the field of financial mathematics.

1.1. Definitions and notations.

- Let $C_b^{k,l}$ be the set of continuously differentiable functions $\phi : (t, x) \in [0, T] \times \mathbb{R}^d$ with continuous and uniformly bounded derivatives w.r.t. t (resp., w.r.t. x) up to order k (resp., up to order l). The function ϕ is also bounded.
- C_p^k denotes the set of C^{k-1} functions with piecewise continuous kth derivative.
- $C^{k+\alpha}$, $\alpha \in]0, 1]$ is the set of C^k functions whose kth derivative is Hölder continuous of order α.
- f_v function. Let $f_v : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ denote the following function:

$$f_v(t, x) = f(t, x, v(t, x), (\partial_x v\sigma)(t, x)),$$

  where f denotes the driver of BSDE (1.1), σ denotes the diffusion coefficient of the SDE satisfied by X, and $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ is C^1 in space.
- Euler scheme. When it exists, we approximate the solution of (1.2) by its N-time-steps Euler scheme, denoted X^N:

$$\text{(1.3)} \qquad \forall s \in [0, T], \;\; dX_s^N = b(\varphi^N(s), X^N_{\varphi^N(s)})ds + \sigma(\varphi^N(s), X^N_{\varphi^N(s)})dW_s,$$

$\varphi^N(s) := \sup\{t_j : t_j \le s\}$, and $\{0 = t_0 < t_1 < \cdots < t_N = T\}$ is a regular subdivision of the interval $[0, T]$.

- Transition density function $p(t, x; s, y)$. If σ is uniformly elliptic, the Markov process X admits a transition probability density $p(t, x; s, y)$. Concerning X^N (which is not Markovian except at times $(t_k)_k$), we know that for any $s > 0$ X^N_s has a probability density $p^N(0, x; s, y)$ w.r.t. the Lebesgue measure.
- $\Psi(s, y, g_1, g_2, W)$ and $\Psi^N(s, y, g_1, g_2, W)$. We define the two following functions:

$$\Psi(s, y, g_1, g_2, W) = \int_s^T g_1(r, X_r^{s,y}(W))dr + g_2(X_T^{s,y}(W)),$$
$$\Psi^N(s, y, g_1, g_2, W) = \int_s^T g_1(r, X_r^{N,s,y}(W))dr + g_2(X_T^{N,s,y}(W)),$$

where $X^{s,y}$ (resp., $X^{N,s,y}$) denotes the diffusion process solving (1.2) and starting from y at time s (resp., its approximation using an Euler scheme with N-time-steps), and W denotes the standard Brownian motion appearing in (1.2) and used to simulate X^N, as given in (1.3).

- Constants $c_{i,j}(\cdot)$. For any function ϕ in $C_b^{i,j}$, $c_{i,j}(\phi)$ denotes $\sum_{k,l=0}^{i,j} |\partial_x^k \partial_t^l \phi|_\infty$. For $i = j = 0$, we set $c_0(\phi) := c_{0,0}(\phi)$.
- Functions $K(T)$. $K(\cdot)$ denotes a generic function nondecreasing in T which may depend on d, μ, β, on the coefficients b and σ (through σ_0, σ_1, $c_{1,3}(\sigma)$, $c_{0,1}(\partial_t \sigma)$, $c_{1,3}(b)$), and on other constants appearing in the Appendix A (i.e., $K_{A.1}(T)$, α_1, α_2, $c_{A.2}$, $K_{A.2}(T)$, $c_{A.3}$, $K_{A.3}(T)$, $c_{A.4}$, $K_{A.4}(T)$, $c_{A.5}$, $K_{A.5}(T)$, and $K_{B.2}(T)$). The parameter β is defined in section 2.1, μ is defined in section 3.2, σ_0 and σ_1 are defined in Hypothesis 1.
- Functions $K_0(T)$. $K_0(T)$ are analogous to $K(T)$ except that they may also depend on the operator $\mathcal{P}$ (through $c_1(K_t)$ and $c_2(K_x)$, defined in section 7).

Hypothesis 1.

- The driver f is a bounded Lipschitz continuous function, i.e., for all (t_1, x_1, y_1, z_1), $(t_2, x_2, y_2, z_2) \in [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^{d\times q}$,

$$|f(t_1, x_1, y_1, z_1) - f(t_2, x_2, y_2, z_2)| \\ \le L_f(|t_1 - t_2| + |x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2|).$$

- σ is uniformly elliptic on $[0, T] \times \mathbb{R}^d$: there exist two positive constants σ_0, σ_1 such that (s.t.) for any vector ξ and any $(t, x) \in [0, T] \times \mathbb{R}^d$

$$\sigma_0 |\xi|^2 \le \sum_{i,j=1}^d [\sigma\sigma^*]_{i,j}(t, x)\xi_i \xi_j \le \sigma_1 |\xi|^2.$$

- Φ is bounded in $C^{2+\alpha}$, $\alpha \in]0, 1]$.
- b and σ are in $C_b^{1,3}$ and $\partial_t \sigma$ is in $C_b^{0,1}$.

1.2. Link with semilinear PDE. According to [23, Theorem 3.1] (see also [11, Proposition 4.3] and [9, Theorem 2.1]), we can link (Y, Z)—the solution of the BSDE (1.1)—to u, the solution of the following PDE:

$$\begin{cases} \partial_t u(t, x) + \mathcal{L}u(t, x) + f(t, x, u(t, x), (\partial_x u\sigma)(t, x)) = 0, \\ u(T, x) = \Phi(x), \end{cases} \tag{1.4}$$

where $\mathcal{L}$ is defined by

$$\mathcal{L}u(t,x) = \frac{1}{2}\sum_{i,j}[\sigma\sigma^*]_{ij}(t,x)\partial^2_{x_ix_j}u(t,x) + \sum_i b_i(t,x)\partial_{x_i}u(t,x).$$

THEOREM 1.1 (see [9], Theorem 2.1). *Under Hypothesis* 1, *the solution* u *of PDE* (1.4) *belongs to* $C_b^{1,2}$. *Moreover,* $(Y_t, Z_t)_{0\le t\le T}$*—solution of* (1.1)*—satisfies*

$$\forall t \in [0,T], \;\; (Y_t, Z_t) = (u(t,X_t), \partial_x u(t,X_t)\sigma(t,X_t)). \tag{1.5}$$

2. Description of the algorithm. As said in the introduction, the current algorithm is based on two ingredients: Picard's iterations and adaptive control variates. We present these ingredients in the following section before describing the main algorithm in section 2.3.

2.1. First ingredient: Picard's iterations. From [11, Corollary 2.1], we know that under standard assumptions on Φ and f, the sequence $(\hat{Y}^k, \hat{Z}^k)_k$ recursively defined by $(\hat{Y}_0 = 0, \hat{Z}_0 = 0)$ and

$$-d\hat{Y}^{k+1}_t = f(t, X_t, \hat{Y}^k_t, \hat{Z}^k_t)dt - \hat{Z}^{k+1}_t dW_t, \;\; \hat{Y}^{k+1}_T = \Phi(X_T)$$

converges to (Y,Z), $d\mathbb{P}\otimes dt$ a.s. (and in $\mathbb{H}^2_{T,\beta}(\mathbb{R}) \times \mathbb{H}^2_{T,\beta}(\mathbb{R}^q)$) as k goes to $+\infty$, where $\mathbb{H}^2_{T,\beta}(\mathbb{R}^q) := \{\phi \in \mathbb{H}^2_T(\mathbb{R}^q)$ such that $\mathbb{E}\int_0^T e^{\beta t}|\phi_t|^2 dt < \infty\}$ and β is such that $2(1+T)L_f < \beta$. This sequence of linear BSDEs can be linked to a sequence of linear PDEs; by writing $\hat{Y}^k_t = \hat{u}_k(t, X_t)$ and $\hat{Z}^k_t = \partial_x \hat{u}_k(t,X_t)\sigma(t,X_t)$, one has

$$\partial_t \hat{u}_{k+1} + \mathcal{L}\hat{u}_{k+1} + f(\cdot,\cdot,\hat{u}_k, \partial_x\hat{u}_k\sigma) = 0 \quad \text{and} \quad \hat{u}_{k+1}(T,\cdot) = \Phi(\cdot).$$

It means that the sequence of solutions of linear PDEs $(\hat{u}_k, \partial_x\hat{u}_k)_k$ converges (in an L_2 norm) to $(u, \partial_x u)$, solution of the semilinear PDE (1.4).

2.2. Second ingredient: Adaptive control variate. In their work [17], Gobet and Maire present an adaptive algorithm to solve linear PDEs of type

$$\partial_t v + \mathcal{L}v + g = 0 \quad \text{and} \quad v(T,\cdot) = \Phi(\cdot).$$

Thanks to the Feynman–Kac formula, we know that the probabilistic solution of this PDE is $v(t,x) = \mathbb{E}_{t,x}[\Phi(X_T) + \int_t^T g(s,X_s)ds] = \mathbb{E}(\Psi(t,x,g,\Phi,W))$.

Their idea is to compute a sequence of solutions $(v_k)_k$ by writing

$$v_{k+1} = v_k + \text{ Monte Carlo evaluations of the error } (v - v_k).$$

The probabilistic representation of the correction term $c_k := v - v_k$ is

$$c_k(t,x) = v(t,x) - v_k(t,x) = \mathbb{E}(\Psi(t,x,g+\partial_t v_k + \mathcal{L}v_k, \Phi - v_k(T,.), W)).$$

Their algorithm computes iterative approximations $(v_k)_k$ of the global solution v. These approximations rely on the computations of $\mathbb{E}(\Psi^N(t,x,\tilde{g},\tilde{\Phi},W))$ (for data $\tilde{g}$ and $\tilde{\Phi}$ possibly different from $g+\partial_t v_k+\mathcal{L}v_k$ and $\Phi - v_k(T,.)$) at some points $(t_i,x_i)_{1\le i\le n} \subset [0,T]\times\mathbb{R}^d$. We briefly recall below their algorithm and the associated convergence result.

ALGORITHM 1. *Begin with* $v_0 \equiv 0$ *and assume that an approximated solution* v_k *of class* $C_b^{1,2}$ *is built at step* k.

- *Evaluate $c_k(t_i, x_i)$ using M independent simulations,*

$$c_k^M(t_i, x_i) = \frac{1}{M} \sum_{m=1}^{M} \Psi^N(t_i, x_i, g + \partial_t v_k + \mathcal{L}^N v_k, \Phi - v_k, W^m),$$

  where
  $\mathcal{L}^N u(s,x) = \frac{1}{2} \sum_{i,j} [\sigma\sigma^*]_{ij}(\varphi^N(s), x)\partial^2_{x_i x_j} u(s,x) + \sum_i b_i(\varphi^N(s), x)\partial_{x_i} u(s,x).$
- *Build the global solution $c_k^M(\cdot)$ based on the values $[c_k^M(t_i, x_i)]_{1 \le i \le n}$ by using a linear approximation operator: $\mathcal{P} c_k^M(\cdot) = \sum_{i=1}^n c_k^M(t_i, x_i)\omega_i(\cdot)$ for some weight functions ω_i. Then, get $v_{k+1} = \mathcal{P}(v_k + c_k^M)$. (Examples of operator $\mathcal{P}$: interpolation, projection, Kernel-based estimator...)*

The main result of their paper is the following theorem.

THEOREM 2.1 (see Theorems 3.1 and 3.2 of [17]). *Define $\|v - v_k\|_2^2 := \sup_{1\le i \le n} |\mathbb{E}((v - v_k)(t_i, x_i))|^2 + \sup_{1 \le i \le n} Var(v_k(t_i, x_i))$. Then for any $k \ge 0$, one has*

$$\|v - v_{k+1}\|_2^2 \le \rho \|v - v_k\|_2^2 + C\|v - \mathcal{P}v\|_2^2 \left(\frac{1}{N} + \frac{1}{M}\right),$$

where $\rho < 1$ (depending on M, N, and $\mathcal{P}$) and $\|v - \mathcal{P}v\|_2$ is a suitable norm related to the approximation error $v - \mathcal{P}v$ (that we do not detail).

As for Picard's iterations, the algorithm converges at a geometric rate. Moreover, there is no need to take N and M large (in practice, in their experiments, they take $M = 10$ to get accurate approximations). The final error is strongly related to the ability of the operator $\mathcal{P}$ to approximate well the true solution v. In the following section, we present an algorithm combining these two ingredients and leading to the same features.

2.3. Algorithm. We recall that we aim at numerically solving BSDE (1.1), which is equivalent to solving the semilinear PDE (1.4). The current algorithm provides an approximation of the solution of this PDE. Then, by simulating the diffusion X through an Euler scheme, we deduce from (1.5) an approximation of the solution of BSDE (1.1). More precisely, let u_k (resp., (Y^k, Z^k)) denote the approximation of u (resp., (Y, Z)) at step k, and let X^N denote the approximation of X obtained with an N-time-steps Euler scheme. We write

$$(Y_t^k, Z_t^k) = (u_k(t, X_t^N), \partial_x u_k(t, X_t^N)\sigma(t, X_t^N)), \text{ for all } t \in [0, T], \tag{2.1}$$

where X^N is described in section 1.1. It remains to build u_{k+1}.

Adaptive control variate. As in Algorithm 1, we write

$$u_{k+1} = u_k + \text{ Monte Carlo evaluations of the error}(u - u_k).$$

Combining Itô's formula applied to $u(s, X_s)$ and to $u_k(s, X_s^N)$ between t and T and the semilinear PDE (1.4) satisfied by u, we get that the correction term $c_k := u - u_k$ is

$$c_k(t,x) = \mathbb{E}\left[\Psi\left(t, x, f_u, \Phi, W\right) - \Psi^N\left(t, x, -(\partial_t + \mathcal{L}^N)u_k, u_k(T, .), W\right) | \mathcal{G}_k\right].$$

Remark 2.2. As we will see later (see Remark 2.3), u_k depends on several random variables. $\mathcal{G}_k$ is the σ-algebra generated by the set of all random variables used to build u_k. In the above equation, we compute the expectation w.r.t. the law of X and X^N, and not w.r.t. the law of u_k, which is $\mathcal{G}_k$ measurable.

Picard's iteration. The correction term c_k cannot be used directly: we have to replace u and $\partial_x u$ (unknown terms) appearing in f with u_k and $\partial_x u_k$, as suggested by Picard's contraction principle:

$$\hat{c}_k(t,x) = \mathbb{E}\left[\Psi\left(t,x,f_{u_k},\Phi,W\right) - \Psi^N\left(t,x,-(\partial_t+\mathcal{L}^N)u_k,u_k,W\right)|\mathcal{G}_k\right]. \tag{2.2}$$

We still have to replace the expectation with a Monte Carlo summation, and Ψ with Ψ^N. Then, the algorithm computes iterative approximations of $(u_k)_k$ of the global solution u at some points $(t_i^k,x_i^k)_{1\le i\le n} \in [0,T]\times\mathbb{R}^d$, which may change over the iterations.

ALGORITHM 2. *We begin with $u_0 \equiv 0$. Assume that an approximated solution u_k of class $C^{1,2}$ is built at step k.*

- *Evaluate $\hat{c}_k(t_i^k,x_i^k)$ using M independent simulations:*

$$\hat{c}_k^M(t_i^k,x_i^k) = \frac{1}{M}\sum_{m=1}^{M}\left[\Psi^N\left(t_i,x_i,f_{u_k}+(\partial_t+\mathcal{L}^N)u_k,\Phi-u_k,W^{m,k,i}\right)\right].$$

- *Build the global solution $\hat{c}_k^M(\cdot)$ based on the values $[\hat{c}_k^M(t_i^k,x_i^k)]_{1\le i\le n}$ by using a linear approximation operator*

$$\mathcal{P}^k c(\cdot) = \sum_{i=1}^{n} c(t_i^k,x_i^k)\omega_i(\cdot), \tag{2.3}$$

where $(\omega_i^k)_i$ are some weight functions. Deduce the approximation of u at step $k+1$:

$$u_{k+1}(t,x) = \mathcal{P}^k(u_k+\hat{c}_k^M)(t,x), \tag{2.4}$$

where $(\mathcal{P}^k)_k$ satisfies Hypothesis 2 *(defined later in section* 4*).*

Remark 2.3. Since u_{k+1} is computed by using (2.4), u_{k+1} is a random function depending on the random variables needed to compute u_k and $W^{m,k,i}$, $1\le m\le M, 1\le i\le n$, appearing in the computation of $\hat{c}_k^M$. Moreover, $\mathcal{P}^k$ may be random. In such a case, u_{k+1} also depends on the random variables used to build $\mathcal{P}^k$.

DEFINITION 2.4 (definition of the σ-algebra $\mathcal{G}_k$). *Let $\mathcal{G}_{k+1}$ define the σ-algebra generated by the set of all random variables used to build u_{k+1}. Using* (2.4) *yields*

$$\mathcal{G}_{k+1} = \mathcal{G}_k \vee \sigma(\mathcal{A}_k,\mathcal{S}_k),$$

where $\mathcal{A}_k$ is the set of random points used at step k to build the estimator $\mathcal{P}^k$, $\mathcal{S}_k := \{W^{m,k,i}, 1\le m\le M, 1\le i\le n\}$, the set of independent Brownian motions used to simulate the paths $X^{m,k,N}(x_i^k)$, and $\mathcal{G}_k$ is the σ-algebra generated by the set of all random variables used to build u_k.

3. Choice of the norm to measure the convergence. The choice of the norm to measure the convergence is not harmless. To prove the convergence of the algorithm we combine results on BSDEs stated in a norm leading to the integration w.r.t. $e^{\beta s}ds$ (see [11]) and results on the bounds for solutions of linear PDEs in weighted Sobolev spaces (leading to the integration w.r.t. $e^{-\mu|x|}dx$), coming from [6], and recalled in Theorem B.2. Although rather technical, this section is crucial in order to analyze the convergence of the algorithm and it is interesting for itself.

3.1. Norm of the convergence.

DEFINITION 3.1 (definition of $\mathbb{H}^\mu_\beta(\mathbb{R}^q)$). *For any $\beta > 0$, we define $(\mathbb{H}^\mu_\beta(\mathbb{R}^q), \|\cdot\|_{\mu,\beta})$ the set of processes $V : \Omega \times [0,T] \times \mathbb{R}^d \to \mathbb{R}^q$ that are $\mathcal{P}r \otimes \mathcal{B}(\mathbb{R}^d)$-measurable (where $\mathcal{P}r$ is the σ-field of predictable subsets of $\Omega \times [0,T]$) and such that*

$$\|V\|^2_{\mu,\beta} := \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s}|V_s(x)|^2 e^{-\mu|x|}dxds\right] < \infty.$$

We use this norm to measure the error $(Y - Y^k, Z - Z^k)$ corresponding to the error we make at step k of the algorithm. Using (1.5) and (2.1), we get

$$\|Y - Y^k\|^2_{\mu,\beta} = \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} e^{\beta s}|u(s, X^x_s) - u_k(s, X^{N,x}_s)|^2 e^{-\mu|x|}dxds\right] \tag{3.1}$$

and $\|Z - Z^k\|^2_{\mu,\beta} = \mathbb{E}[\int_0^T \int_{\mathbb{R}^d} e^{\beta s}|(\partial_x u\sigma)(s, X^x_s) - (\partial_x u_k\sigma)(s, X^{N,x}_s)|^2 e^{-\mu|x|}dxds]$. Since u, u_k, $\partial_x u$, and $\partial_x u_k$ are bounded (see Theorem 1.1 and Hypothesis 2 defined later), $\|Y - Y^k\|^2_{\mu,\beta}$ and $\|Z - Z^k\|^2_{\mu,\beta}$ are finite.

Remark 3.2. We point out that the expectation appearing in the above definition of $\|Y^k - Y\|^2_{\mu,\beta}$ and $\|Z^k - Z\|^2_{\mu,\beta}$ is computed w.r.t. the law of X, X^N and all the possible random variables used to compute u_k.

3.2. Some other useful norms. The following definitions introduce two norms strongly related to $\|\cdot\|_{\mu,\beta}$. They will be useful in the proof of the main result.

DEFINITION 3.3 (space $H^{m,\mu}_\beta$). *For any $m \le 2$, $\beta > 0$ and $\mu > 0$, let $H^{m,\mu}_\beta$ define the space of functions $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that*

$$\|v\|^2_{H^{m,\mu}_\beta} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \sum_{k \le m} |\partial^k_x v(s, x)|^2 dxds < \infty.$$

For $m = 0$, we set $H^{0,\mu}_\beta = H^\mu_\beta$.

DEFINITION 3.4 (function ν^t_μ). *For any $s, t \in [0,T]$ and any $x, y \in \mathbb{R}^d$ such that $t < s$ we define $\nu^t_\mu(s, y) := \int_{\mathbb{R}^d} e^{-\mu|x|}p(t, x; s, y)dx$, where μ is a positive constant.*

DEFINITION 3.5 (space $H^{m,\mu}_{\beta,\tilde{X}}$). *For any $m \le 2, \beta > 0, \mu > 0$ and any diffusion process $\tilde{X}_s$, $0 \le s \le T$ starting from x at time 0, let $H^{m,\mu}_{\beta,\tilde{X}}$ define the space of functions $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that*

$$\|v\|^2_{H^{m,\mu}_{\beta,\tilde{X}}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} e^{-\mu|x|} \sum_{k \le m} \mathbb{E}|\partial^k_x v(s, \tilde{X}^x_s)|^2 dxds < \infty.$$

For $m = 0$, we set $H^{0,\mu}_{\beta,\tilde{X}} = H^\mu_{\beta,\tilde{X}}$. By using the definition of ν, we also get $\|v\|^2_{H^\mu_{\beta,X}} = \int_0^T e^{\beta s} \int_{\mathbb{R}^d} \nu^0_\mu(s, y)|v(s, y)|^2 dyds$.

Remark 3.6. We can interpret $\|v\|^2_{H^\mu_{\beta,X}}$ as $\int_0^T e^{\beta s}\mathbb{E}|v(s, X^\mu_s)|^2 ds$, where $(X^\mu_s)_s$ stands for $(X_s)_s$ starting at time 0 with a random initial law having a density proportional to $e^{-\mu|x|}$.

Remark 3.7. In the above definition of $\|v\|^2_{H^{m,\mu}_{\beta,\tilde{X}}}$, we compute the expectation w.r.t. the law of $\tilde{X}$. If v is a random function, $\|v\|^2_{H^{m,\mu}_{\beta,\tilde{X}}}$ is random, too.

3.3. Norm equivalence results. The following proposition gives a norm equivalence result between $\|\cdot\|_{H^\mu_{\beta,X}}$ and $\|\cdot\|_{H^\mu_\beta}$. It can be compared to Bally and Matoussi [2, Proposition 5.1].

PROPOSITION 3.8 (norm equivalence). *Assume that the coefficients σ, b are bounded measurable functions on $[0,T]\times\mathbb{R}^d$, Lipschitz w.r.t. x, and that σ satisfies the ellipticity condition. There exist two constants $\underline{c}>0$ and $\overline{c}>0$ (depending on $T, d, \mu, K_{A.1}(T), \alpha_1, \alpha_2$) s.t. for every $v \in L_2([0,T]\times\mathbb{R}^d, e^{\beta t}dt \otimes e^{-\mu|x|}dx)$*

$$\underline{c}\|v\|^2_{H^\mu_\beta} \le \|v\|^2_{H^\mu_{\beta,X}} \le \overline{c}\|v\|^2_{H^\mu_\beta}.$$

Proof. Proving Proposition 3.8 boils down to showing that there exist two constants $c_i, i=1,2$ depending on μ, d, α_i s.t. $\forall y \in \mathbb{R}^d$

$$\frac{1}{2^d K(T)} e^{-\mu|y|} e^{c_1(s-t)} \le \nu^t_\mu(s,y) \le 2^d K(T) e^{c_2(s-t)} e^{-\mu|y|}. \tag{3.2}$$

The right-hand side (r.h.s.) (resp., the left-hand side (l.h.s.)) of the above inequality ensues from Proposition A.1 and Lemma C.1 with $c=\frac{1}{2\alpha_2}$ (resp., $c=\frac{1}{2\alpha_1}$). Moreover, we have $c_i = \frac{d\mu^2\alpha_i}{2}$, $i=1,2$. ∎

PROPOSITION 3.9. *Assume σ is uniformly elliptic and b and σ are in $C^{0,2}_b$. For every $v \in L_2([0,T]\times\mathbb{R}^d, e^{\beta t}dt\otimes e^{-\mu|x|}dx)$ it holds $\|v\|^2_{H^\mu_{\beta,X^N}} \le K(T)\|v\|^2_{H^\mu_{\beta,X}}$.*

Proof. Since $\|v\|^2_{H^\mu_{\beta,X^N}} = \int_0^T e^{\beta s}\int_{\mathbb{R}^d} e^{-\mu|x|}\int_{\mathbb{R}^d} v^2(s,y)p^N(0,x;s,y)dydxds$, we apply successively Proposition A.5, Lemma C.1, and the l.h.s. of (3.2). ∎

The following proposition bounds the solution of a linear PDE in norm $\|\cdot\|_{H^\mu_{\beta,X}}$.

PROPOSITION 3.10. *Assume σ is uniformly elliptic, $\sigma \in C^{1,1}_b$, b is $C^{1,1}$ and bounded, and $g \in L_2([0,T]\times\mathbb{R}^d, e^{\beta t}\nu^0_\mu(t,y)dt\otimes dy)$. We assume also that v satisfies the linear PDE $(\partial_t+\mathcal{L})v(t,x)+g(t,x)=0$ with a terminal condition $v(T,\cdot)=0$. Then,*

$$\|v\|^2_{H^{2,\mu}_{\beta,X}} + \|\partial_t v\|^2_{H^\mu_{\beta,X}} \le K_{3.10}(T)\|g\|^2_{H^\mu_{\beta,X}},$$

where $K_{3.10}(T)$ depends on $K_{B.2}(T)$, β, $\underline{c}$, $\overline{c}$, and T.

Proof. Let $\overline{v}(t,x) := e^{\beta t/2}v(t,x)$. Using the PDE satisfied by v, we get

$$-\partial_t\overline{v}(t,x) - \mathcal{L}\overline{v}(t,x) + \frac{\beta}{2}\overline{v}(t,x) = e^{\frac{\beta t}{2}}g(t,x), \quad \overline{v}(T,x)=0.$$

Then, we apply Theorem B.2 to $\overline{v}$ to obtain $\|v\|^2_{H^{2,\mu}_\beta} + \|\partial_t v\|^2_{H^\mu_\beta} \le K(T)\|g\|^2_{H^\mu_\beta}$, where $K(T)$ denotes a constant depending on $K_{B.2}(T)$ and on β. Proposition 3.8 (applied to v and its derivatives) ends the proof. ∎

4. Approximation operator $\mathcal{P}$. $\mathcal{P}$ denotes the sequence of approximation operators $(\mathcal{P}^k)_k$ satisfying the following hypothesis. In section 7, we give an example of such an operator $\mathcal{P}$ based on kernel estimators.

Hypothesis 2 (hypotheses on $(\mathcal{P}^k)_k$). Let $\epsilon_i(\mathcal{P}), i=1\cdots 4$ denote some constants depending on $\mathcal{P}$ and tending to 0 when the parameters of $\mathcal{P}$ tend to infinity. For any k, the random operator $\mathcal{P}^k$ satisfies the following properties.

1. Measurability and linearity. $\mathcal{P}^k$ is linear, and it writes (see (2.3))

$$\mathcal{P}^k c(\cdot) = \sum_{i=1}^n c(t^k_i, x^k_i)\omega_i(\cdot).$$

As said in Definition 2.4, the points $(t_i^k, x_i^k)_i$ are $\sigma(\mathcal{A}_k)$-measurable and $\mathcal{A}_k$ is independent of $\mathcal{A}_j$ for $k \neq j$. In addition, the weights $w_i^k(.) : \Omega \times [0,T] \times \mathbb{R}^d \mapsto \mathbb{R}$ are measurable w.r.t. $\sigma(\mathcal{A}_k) \times \mathcal{B}([0,T] \times \mathbb{R}^d)$.

2. Regularity. For any function $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $\mathcal{P}^k v \in C^{1,2}$ (equivalently, the weights $(w_i^k)_i$ are $C^{1,2}$ w.r.t. (t,x)).
3. Boundedness. For any bounded function $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $\mathcal{P}^k v$, $\partial_t(\mathcal{P}^k v)$, $\partial_x(\mathcal{P}^k v)$ and $\partial_x^2(\mathcal{P}^k v)$ are bounded by a constant depending on $c_0(v)$ and on the parameters of the operator $\mathcal{P}^k$ (the explicit form of the bounds is not needed).
4. Approximation. $\mathcal{P}^k$ approximates well a function and its spatial derivatives. Let v be a $C^{1,2}$ random function, independent of $\mathcal{A}_k$, from $[0,T] \times \mathbb{R}^d$ in $\mathbb{R}$. We also assume that v and $\partial_x v$ are bounded (by $c_{0,1}(v)$) and $\forall t, t' \in [0,T]$, $\forall x \in \mathbb{R}^d$, $|\partial_x v(t,x) - \partial_x v(t',x)| \leq c_{1/2}(v)\sqrt{|t'-t|}$. Then,
$$\mathbb{E}\|v - \mathcal{P}^k v\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\|\partial_x v - \partial_x(\mathcal{P}^k v)\|^2_{H^\mu_{\beta,X}} \leq \epsilon_1(\mathcal{P}) \left(\mathbb{E}\|v\|^2_{H^{2,\mu}_{\beta,X}} + \mathbb{E}\|\partial_t v\|^2_{H^\mu_{\beta,X}}\right) + \epsilon_2(\mathcal{P})(c^2_{1/2}(v) + c^2_{0,1}(v)).$$
5. Let v be a $C_b^{1,2}$ function from $[0,T] \times \mathbb{R}^d$ in $\mathbb{R}$. Then $\mathbb{E}\|v - \mathcal{P}^k v\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\|\partial_x v - \partial_x(\mathcal{P}^k v)\|^2_{H^\mu_{\beta,X}} \leq \epsilon_3(\mathcal{P})(c^2_{1/2}(v) + c^2_{1,2}(v))$.
6. Stability and centering property for random functions. For any random function v from $[0,T] \times \mathbb{R}^d$ in $\mathbb{R}$, one has $\mathbb{E}\|\mathcal{P}^k v\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\|\partial_x(\mathcal{P}^k v)\|^2_{H^\mu_{\beta,X}} \leq c_4(\mathcal{P})\mathbb{E}\|v\|^2_{H^\mu_{\beta,X}}$, where $c_4(\mathcal{P})$ is a constant depending on $\mathcal{P}$. If $\mathbb{E}(v(t,x)) = 0$, one has $\mathbb{E}\|\mathcal{P}^k v\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\|\partial_x(\mathcal{P}^k v)\|^2_{H^\mu_{\beta,X}} \leq \epsilon_4(\mathcal{P})\mathbb{E}\|v\|^2_{H^\mu_{\beta,X}}$.

5. Main result. The following theorem states the convergence of $(Y^k, Z^k)_k$ defined in (2.1) towards (Y,Z) in the $\|\cdot\|^2_{\mu,\beta}$ norm. See section 6 for a proof of it.

THEOREM 5.1. *Assume that Hypotheses* 1 *and* 2 *hold. We assume also that f is a bounded Lipschitz function and $\Phi \in C_b^{2+\alpha}$. Then, there exists a constant $K(T)$ such that*
$$\|Y - Y^k\|^2_{\mu,\beta} + \|Z - Z^k\|^2_{\mu,\beta} \leq S_k + K(T)\frac{c^2_{0,2}(u)}{N}, \ \textit{where } S_k \leq \eta S_{k-1} + \epsilon \textit{ and}$$
$$\eta = K(T)\left(\frac{4(1+T)L_f^2}{\beta} + \epsilon_1(\mathcal{P})K_{3.10}(T)L_f^2 + \frac{\epsilon_4(\mathcal{P})}{M}\right),$$
$$\epsilon = K(T)\Bigg(\epsilon_3(\mathcal{P})c^2_{1,2}(u) + (\epsilon_2(\mathcal{P}) + \epsilon_3(\mathcal{P}))c_0^2(f) + \frac{\epsilon_4(\mathcal{P})c^2_{0,2}(u)}{MN} + \frac{c_4(\mathcal{P})}{N^2}(c_0^2(f) + c_0^2(\Phi))\Bigg).$$

COROLLARY 5.2. *Under the assumptions of Theorem* 5.1, *for β and $\mathcal{P}$-parameters large enough so that $\eta < 1$, we have*
$$\limsup_{k\to\infty} \|Y - Y^k\|^2_{\mu,\beta} + \|Z - Z^k\|^2_{\mu,\beta} \leq \frac{\epsilon}{1-\eta} + K(T)\frac{c^2_{0,2}(u)}{N}.$$

Remark 5.3. Using a sequential Monte Carlo method leads to an error of $\frac{1}{M}$ appearing in the contraction term η. If we had implemented a nonadaptive method (i.e., using only Picard's iterations), M^{-1} would have appeared in ϵ, and this would have led us to choose a much larger M, while practically $M = 100$ does the trick. A quick analysis of the form of ϵ shows that it is sufficient to take M and N of the same magnitude. (We observe in practice that taking $N = M = 100$ gives small related errors.) Then, the final algorithm accuracy relies heavily on the quality of the operator $\mathcal{P}$, through the convergence rates $\epsilon_2(\mathcal{P})$ and $\epsilon_3(\mathcal{P})$. These rates are related to those obtained in the approximation of a function v and its first partial derivative $\partial_x v$, when roughly speaking, the function v is $C_b^{1,2}$, and $\partial_x v$ is $\frac{1}{2}$-Hölder in time.

6. Proof of Theorem 5.1.

In this section, we assume that Hypotheses 1 and 2 hold.

We want to measure the error $\|Y - Y^k\|^2_{\mu,\beta} + \|Z - Z^k\|^2_{\mu,\beta}$. To do so, we split the different sources of errors: Euler scheme approximation, Picard's approximation, approximation operator $\mathcal{P}$, and Monte Carlo simulations. We prove Theorem 5.1 by combining Proposition 6.1, equation (6.2), and Propositions 6.3, 6.4, and 6.8.

6.1. First source of error: Euler scheme approximation.

PROPOSITION 6.1. *It holds*

$$\|Y - Y^k\|^2_{\mu,\beta} + \|Z - Z^k\|^2_{\mu,\beta} \le c^2_{0,2}(u)\frac{K(T)}{N} + K(T)(\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta}),$$

where $(\tilde{Y}^k_s, \tilde{Z}^k_s) := (u_k(s, X_s), (\partial_x u_k \sigma)(s, X_s))$.

Proof. We start from (3.1), and we introduce $\pm u(s, X_s^{N,x})$ in $|u(s, X^x_s) - u_k(s, X_s^{N,x})|$. We get $|u(s, X^x_s) - u_k(s, X_s^{N,x})|^2 \le 2|u(s, X^x_s) - u(s, X_s^{N,x})|^2 + 2|u(s, X_s^{N,x}) - u_k(s, X_s^{N,x})|^2$. Since u is a $C_b^{1,2}$ function (see Theorem 1.1), we bound the first term by $2c^2_{0,1}(u)|X^x_s - X^{N,x}_s|^2$, and we use the well known strong error bound for the Euler scheme to get $\|Y - Y^k\|^2_{\mu,\beta} \le c^2_{0,1}(u)\frac{K(T)}{N} + 2\mathbb{E}\,\|u - u_k\|^2_{H^\mu_{\beta,X^N}}$. The same technique enables us to bound $\|Z - Z^k\|^2_{\mu,\beta}$. Finally, we apply Proposition 3.9 and use the equality $\mathbb{E}\,\|u - u_k\|^2_{H^\mu_{\beta,X}} = \|Y - \tilde{Y}^k\|^2_{\mu,\beta}$ (see Remark 6.2) to end the proof. ∎

Remark 6.2. We recall that the expectation appearing in the definition of $\|Y - \tilde{Y}^k\|^2_{\mu,\beta}$ is computed w.r.t. the law of X and all the variables used to compute u_k, whereas the one appearing in the definition of $\|\cdot\|^2_{H^\mu_{\beta,X}}$ is computed w.r.t. the law of X; that's why we have that $\|Y - \tilde{Y}^k\|^2_{\mu,\beta} = \mathbb{E}\,\|u - u_k\|^2_{H^\mu_{\beta,X}}$. $\|u - u_k\|^2_{H^\mu_{\beta,X}}$ is $\mathcal{G}_k$-measurable.

Let us now work on the remaining term $\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta}$.

6.2. Second source of error: Picard's approximation.

We aim at bounding $\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta}$. To do so, we consider the sequence $(\overline{Y}^k, \overline{Z}^k)_{k\ge1}$, where $(\overline{Y}^k, \overline{Z}^k)$ is the solution of the linear BSDE

$$\overline{Y}^k_t = \Phi(X_T) + \int_t^T f(s, X_s, \tilde{Y}^{k-1}_s, \tilde{Z}^{k-1}_s)ds - \int_t^T \overline{Z}^k_s dW_s. \tag{6.1}$$

We split $\|Y - \tilde{Y}^k\|^2_{\mu,\beta}$ and $\|Z - \tilde{Z}^k\|^2_{\mu,\beta}$ by introducing $\overline{Y}^k$ and $\overline{Z}^k$:

$$\begin{aligned}\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta} \le\, & 2(\|Y - \overline{Y}^k\|^2_{\mu,\beta} + \|Z - \overline{Z}^k\|^2_{\mu,\beta}) \\ & + 2(\|\overline{Y}^k - \tilde{Y}^k\|^2_{\mu,\beta} + \|\overline{Z}^k - \tilde{Z}^k\|^2_{\mu,\beta}).\end{aligned} \tag{6.2}$$

Let us study $\|Y-\overline{Y}^k\|^2_{\mu,\beta}+\|Z-\overline{Z}^k\|^2_{\mu,\beta}$. The study of $\|\overline{Y}^k-\tilde{Y}^k\|^2_{\mu,\beta}$ and $\|\overline{Z}^k-\tilde{Z}^k\|^2_{\mu,\beta}$ is postponed to the next step.

PROPOSITION 6.3. *For any $\beta>0$ and any $\mu>0$, it holds*

$$\|Y-\overline{Y}^k\|^2_{\mu,\beta}+\|Z-\overline{Z}^k\|^2_{\mu,\beta}\le\frac{2(1+T)L_f^2}{\beta}\left(\|Y-\tilde{Y}^{k-1}\|^2_{\mu,\beta}+\|Z-\tilde{Z}^{k-1}\|^2_{\mu,\beta}\right),$$

where $(\overline{Y}^k,\overline{Z}^k)$ is the solution of (6.1).

Proof. Let us do the proof for the error on Y. Using a priori estimates (see [11, Proposition 2.1]) and the standard BSDE (1.1) leads to $\int_0^T e^{\beta s}\mathbb{E}[|Y_s-\overline{Y}^k_s|^2|\mathcal{G}_{k-1}]ds\le\frac{T}{\beta}\int_0^T e^{\beta s}\mathbb{E}[|f(s,X_s,Y_s,Z_s)-f(s,X_s,\tilde{Y}^{k-1}_s,\tilde{Z}^{k-1}_s)$ $|^2|\mathcal{G}_{k-1}]ds$. Since f is Lipschitz, $\int_0^T e^{\beta s}\mathbb{E}[|Y_s-\overline{Y}^k_s|^2|\mathcal{G}_{k-1}]ds\le\frac{2TL_f^2}{\beta}\int_0^T e^{\beta s}\mathbb{E}[|Y_s-\tilde{Y}^{k-1}_s|^2+|Z_s-\tilde{Z}^{k-1}_s|^2|\mathcal{G}_{k-1}]ds$. By integrating w.r.t. $e^{-\mu|x|}dx$ and taking the expectation, the result follows. ∎

It remains to measure the error $\|\overline{Y}^k-\tilde{Y}^k\|^2_{\mu,\beta}$ and $\|\overline{Z}^k-\tilde{Z}^k\|^2_{\mu,\beta}$.

6.3. Third source of error: Approximation operator $\boldsymbol{\mathcal{P}}$. In this part, we split the remaining errors $\|\overline{Y}^k-\tilde{Y}^k\|^2_{\mu,\beta}$ and $\|\overline{Z}^k-\tilde{Z}^k\|^2_{\mu,\beta}$ in three error terms (see (6.4) for the decomposition of Y). The first two are approximation errors connected to $\mathcal{P}$, and the last one is studied in the next section.

PROPOSITION 6.4. *It holds*

$$\begin{aligned}\|\overline{Y}^k-\tilde{Y}^k\|^2_{\mu,\beta}+\|\overline{Z}^k-\tilde{Z}^k\|^2_{\mu,\beta}\le{}&\eta(\|Y-\tilde{Y}^{k-1}\|^2_{\mu,\beta}+\|Z-\tilde{Z}^{k-1}\|^2_{\mu,\beta})\\&+3\mathbb{E}\,\|\mathcal{P}^{k-1}(\hat{c}_{k-1}-\hat{c}^M_{k-1})\|^2_{H^\mu_{\beta,X}}\\&+3c_0^2(\sigma)\mathbb{E}\,\|\partial_x\mathcal{P}^{k-1}(\hat{c}_{k-1}-\hat{c}^M_{k-1})\|^2_{H^\mu_{\beta,X}}+\epsilon,\end{aligned}$$

where $\eta=3\epsilon_1(\mathcal{P})K_{3.10}(T)L_f^2$ and $\epsilon=K(T)(\epsilon_3(\mathcal{P})c^2_{1,2}(u)+(\epsilon_2(\mathcal{P})+\epsilon_3(\mathcal{P}))c_0^2(f))$.

Before proving Proposition 6.4, we introduce two sequences $(\overline{u}_k)_k$ and $(H_k)_k$.

DEFINITION 6.5. *$\forall k\ge1$, $\overline{u}_k$ denotes the solution of the linear PDE*

$$(\partial_t+\mathcal{L})\overline{u}_k(t,x)+f_{u_{k-1}}(t,x)=0,\quad \overline{u}_k(T,x)=\Phi(x).$$

Then, $\forall t\in[0,T]$, $(\overline{Y}^k_t,\overline{Z}^k_t)=(\overline{u}_k(t,X_t),(\partial_x\overline{u}_k\sigma)(t,X_t))$, where $(\overline{Y}^k_t,\overline{Z}^k_t)_k$ is the solution of BSDE (6.1).

DEFINITION 6.6. *$\forall k\ge0$, H_k denotes the solution of the linear PDE*

$$(\partial_t+\mathcal{L})H_k(t,x)+f_{u_k}(t,x)-f_u(t,x)=0,\quad H_k(T,x)=0. \tag{6.3}$$

Clearly, one has $\overline{u}_k=u+H_{k-1}$ for any $k\ge1$.

Proof of Proposition 6.4. Let us work on $\|\overline{Y}^k-\tilde{Y}^k\|^2_{\mu,\beta}=\mathbb{E}\,\|\overline{u}_k-u_k\|^2_{H^\mu_{\beta,X}}$. Using (2.4) gives $\overline{u}_k-u_k=\overline{u}_k-\mathcal{P}^{k-1}(u_{k-1}+\hat{c}^M_{k-1})=\overline{u}_k-\mathcal{P}^{k-1}(u_{k-1}+\hat{c}_{k-1})+\mathcal{P}^{k-1}(\hat{c}_{k-1}-\hat{c}^M_{k-1})$. Since $u_{k-1}+\hat{c}_{k-1}=\overline{u}_k=u+H_{k-1}$ (see (2.2)), we get

$$\overline{u}_k-u_k=(u-\mathcal{P}^{k-1}u)+(H_{k-1}-\mathcal{P}^{k-1}H_{k-1})+\mathcal{P}^{k-1}(\hat{c}_{k-1}-\hat{c}^M_{k-1}). \tag{6.4}$$

Then, since u and its derivatives are bounded and since H_{k-1} satisfies PDE (6.3), we combine Lemma 6.7 and the features of $\mathcal{P}^{k-1}$ (see Hypothesis 2) to get

$$
\begin{aligned}
\mathbb{E}\,\|\overline{u}_k - u_k\|^2_{H^\mu_{\beta,X}} \le\ & 3\epsilon_3(\mathcal{P})(c^2_{1/2}(u) + c^2_{1,2}(u)) \\
& + 3\epsilon_1(\mathcal{P})\left(\mathbb{E}\,\|H_{k-1}\|^2_{H^{2,\mu}_{\beta,X}} + \mathbb{E}\,\|\partial_t H_{k-1}\|^2_{H^\mu_{\beta,X}}\right) \\
& + 3\epsilon_2(\mathcal{P})(c^2_{1/2}(H_{k-1}) + c^2_{0,1}(H_{k-1})) \\
& + 3\mathbb{E}\,\|\mathcal{P}^{k-1}(\hat{c}_{k-1} - \hat{c}^M_{k-1})\|^2_{H^\mu_{\beta,X}}.
\end{aligned}
\tag{6.5}
$$

We bound the second term on the r.h.s. by using Proposition 3.10: $\mathbb{E}\,\|H_{k-1}\|^2_{H^{2,\mu}_{\beta,X}} + \mathbb{E}\,\|\partial_t H_{k-1}\|^2_{H^\mu_{\beta,X}} \le K_{3.10}(T)\mathbb{E}\,\|f_{u_{k-1}} - f_u\|^2_{H^\mu_{\frac{\beta}{2},X}}$. Since f is Lipschitz, the result follows. The proof is similar for $\|\overline{Z}^k - \tilde{Z}^k\|_{\mu,\beta}$, except that (6.5) contains $c^2_0(\sigma)\mathbb{E}\,\|\partial_x\mathcal{P}^{k-1}(\hat{c}_{k-1} - \hat{c}^M_{k-1})\|^2_{H^\mu_{\beta,X}}$ (and not $\mathbb{E}\,\|\mathcal{P}^{k-1}(\hat{c}_{k-1} - \hat{c}^M_{k-1})\|^2_{H^\mu_{\beta,X}}$). Lemma 6.7 (resp., Proposition B.3) enables us to replace constants $c^2_{1/2}(H_{k-1}) + c^2_{0,1}(H_{k-1})$ (resp., $c^2_{1/2}(u)$) by $K(T)c^2_0(f)$. (resp., $K(T)(c^2_0(f) + c^2_2(\Phi))$). Since $\Phi(\cdot) = u(T,\cdot)$, we get $c^2_2(\Phi) \le c^2_{0,2}(u)$, and the result follows. ∎

LEMMA 6.7. *For all k, H_k and $\partial_x H_k$ are bounded by a constant of the form $K(T)c_0(f)$. Moreover, for all $t, t' \in [0,T]$, $|\partial_x H_k(t',x) - \partial_x H_k(t,x)| \le c_{1/2}(H_k)\sqrt{|t'-t|}$ with $c_{1/2}(H_k) = K(T)c_0(f)$*

Proof of the lemma. First, we prove that H_k and $\partial_x H_k$ are bounded. Using Feynman–Kac's formula yields $H_k(t,x) = \mathbb{E}\left[\int_t^T (f_{u_k}(s, X^{t,x}_s) - f_u(s, X^{t,x}_s))ds|\mathcal{G}_k\right]$. Since f is bounded, we get that H is bounded by $2Tc_0(f)$. To prove that $\partial_x H_k$ is bounded, we write $H_k(t,x) = \int_t^T \int_{\mathbb{R}^d} \tilde{f}_k(s,y)p(t,x;s,y)dyds$, where p denotes the transition density of X and $\tilde{f}_k := f_{u_k} - f_u$. We differentiate H_k w.r.t. x, and we use Proposition A.2 to get $|\partial_x H_k(t,x)| \le 2K(T)c_0(f)\int_t^T \frac{ds}{\sqrt{s-t}}\int_{\mathbb{R}^d} e^{-\frac{c|x-y|^2}{s-t}}\frac{dy}{(s-t)^{\frac{d}{2}}}$. Our statement on $\partial_x H_k$ readily follows. The second assertion ensues from Proposition B.3. ∎

6.4. Fourth source of error: Monte Carlo simulations. In this section, we study $\mathbb{E}\,\|\mathcal{P}^k(\hat{c}_k - \hat{c}^M_k)\|^2_{H^\mu_{\beta,X}}$ and $\mathbb{E}\,\|\partial_x\mathcal{P}^k(\hat{c}_k - \hat{c}^M_k)\|^2_{H^\mu_{\beta,X}}$.

PROPOSITION 6.8. *Let Δc_k denote $\hat{c}_k - \hat{c}^M_k$. It holds*

$$
\begin{aligned}
\mathbb{E}\,\|\mathcal{P}^k(\Delta c_k)\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\,\|\partial_x\mathcal{P}^k(\Delta c_k)\|^2_{H^\mu_{\beta,X}} \le\ & \epsilon_4(\mathcal{P})\frac{K(T)}{M}(\|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta}) \\
& + \epsilon_4(\mathcal{P})\frac{K(T)}{MN}c^2_{0,2}(u) \\
& + c_4(\mathcal{P})\frac{K(T)}{N^2}(c^2_0(f) + c^2_0(\Phi)).
\end{aligned}
$$

Proof. We split Δc_k into two terms: the bias and the noise

$$
\mathcal{P}^k(\Delta c_k) = \mathcal{P}^k(\mathbb{E}(\Delta c_k|\mathcal{G}_k)) + \mathcal{P}^k(\Delta c_k - \mathbb{E}(\Delta c_k|\mathcal{G}_k)) := \mathcal{P}^k(\mathbb{E}(\Delta c_k|\mathcal{G}_k)) + \mathcal{P}^k(\varepsilon_k),
$$

where $\varepsilon_k := \Delta c_k - \mathbb{E}(\Delta c_k|\mathcal{G}_k) = \hat{c}^M_k - \mathbb{E}(\hat{c}^M_k|\mathcal{G}_k)$. The same decomposition holds for $\partial_x\mathcal{P}^k(\Delta c_k)$. It remains to apply Propositions 6.9 and 6.10 to end the proof. ∎

6.4.1. Bias terms $\mathcal{P}^k(\mathbb{E}(\Delta c_k|\mathcal{G}_k))$ and $\partial_x\mathcal{P}^k(\mathbb{E}(\Delta c_k|\mathcal{G}_k))$.

PROPOSITION 6.9. *It holds* $\mathbb{E}\|\mathcal{P}^k(\mathbb{E}(\Delta c_k|\mathcal{G}_k))\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\|\partial_x\mathcal{P}^k(\mathbb{E}(\Delta c_k|\mathcal{G}_k))\|^2_{H^\mu_{\beta,X}} \le c_4(\mathcal{P})\frac{K(T)}{N^2}$.

Proof. First, we use the properties of $\mathcal{P}$ (see point 6, Hypothesis 2) to get $\mathbb{E}\|\mathcal{P}^k(\mathbb{E}(\Delta c_k|\mathcal{G}_k))\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\|\partial_x\mathcal{P}^k(\mathbb{E}(\Delta c_k|\mathcal{G}_k))\|^2_{H^\mu_{\beta,X}} \le c_4(\mathcal{P})\|\mathbb{E}(\Delta c_k|\mathcal{G}_k)\|^2_{H^\mu_{\beta,X}}$. Then, we define $\Delta(s,y) := \Phi(X_T^{s,y}) - \Phi(X_T^{N,s,y}) + \int_s^T (f_{u_k}(r, X_r^{s,y}) - f_{u_k}(r, X_r^{N,s,y}))dr$. We have

$$\mathbb{E}(\Delta c_k|\mathcal{G}_k) = \hat{c}_k - \mathbb{E}(\hat{c}_k^M|\mathcal{G}_k) = \mathbb{E}(\Delta|\mathcal{G}_k).$$

Let us bound $\mathbb{E}(\Delta|\mathcal{G}_k)$. First, we work on the first term: $\mathbb{E}[\Phi(X_T^{s,y}) - \Phi(X_T^{N,s,y})] = \int_{\mathbb{R}^d}\Phi(y')(p(s,y;T,y') - p^N(s,y;T,y'))dy'$. By using Proposition A.4, we get $|\mathbb{E}[\Phi(X_T^{s,y}) - \Phi(X_T^{N,s,y})]| \le \frac{K(T)}{N}\sqrt{T-s}\int_{\mathbb{R}^d}|\Phi(y')|\frac{1}{(T-s)^{\frac{d}{2}}}\exp(-\frac{c_{A.4}|y'-y|^2}{T-s})dy'$. Then, $|\mathbb{E}[\Phi(X_T^{s,y}) - \Phi(X_T^{N,s,y})]| \le c_0(\Phi)\frac{K(T)}{N}\sqrt{T-s}$. The same proof holds for $\int_s^T(f_{u_k}(r,X_r^{s,y}) - f_{u_k}(r,X_r^{N,s,y}))dr$: $\mathbb{E}|f_{u_k}(r,X_r^{s,y}) - f_{u_k}(r,X_r^{N,s,y})| \le \frac{K(T)}{N}\frac{T-s}{\sqrt{r-s}}c_0(f)$. Integrating both sides w.r.t. r between s and T leads to $|\mathbb{E}[\Delta(s,y)|\mathcal{G}_k]| \le \frac{K(T)}{N}\sqrt{T-s}(c_0(\Phi) + (T-s)c_0(f))$, and the result follows. ∎

6.4.2. Noise terms $\mathcal{P}^k(\varepsilon_k)$ and $\partial_x\mathcal{P}^k(\varepsilon_k)$.

PROPOSITION 6.10. *It holds* $\mathbb{E}\|\mathcal{P}^k(\varepsilon_k)\|^2_{H^\mu_{\beta,X}} + \mathbb{E}\|\partial_x\mathcal{P}^k(\varepsilon_k)\|^2_{H^\mu_{\beta,X}} \le \epsilon_4(\mathcal{P})\frac{K(T)}{M}(\frac{c^2_{0,2}(u)}{N} + \|Y - \tilde{Y}^k\|^2_{\mu,\beta} + \|Z - \tilde{Z}^k\|^2_{\mu,\beta})$.

Proof. Since $\mathbb{E}(\varepsilon_k|\mathcal{G}_k) = 0$, the properties of $\mathcal{P}$ (point 6, Hypothesis 2) give $\mathbb{E}(\|\mathcal{P}^k\varepsilon_k\|^2_{H^\mu_{\beta,X}}|\mathcal{G}_k) + +\mathbb{E}(\|\partial_x\mathcal{P}^k(\varepsilon_k)\|^2_{H^\mu_{\beta,X}}|\mathcal{G}_k) \le \epsilon_4(\mathcal{P})\mathbb{E}(\|\varepsilon_k\|^2_{H^\mu_{\beta,X}}|\mathcal{G}_k)$. Since $\mathbb{E}(\|\varepsilon_k\|^2_{H^\mu_{\beta,X}}|\mathcal{G}_k) = \int_0^T e^{\beta s}\int_{\mathbb{R}^d}\nu^0_\mu(s,y)\mathbb{E}(|\varepsilon_k(s,y)|^2|\mathcal{G}_k)dsdy = \int_0^T e^{\beta s}\int_{\mathbb{R}^d}\nu^0_\mu(s,y)$ $\mathrm{Var}(\varepsilon_k(s,y)|\mathcal{G}_k)dsdy$, it remains to bound

$$\mathrm{Var}(\varepsilon_k(s,y)|\mathcal{G}_k) = \frac{1}{M}\mathrm{Var}(\Psi^N(s,y,f_{u_k} + (\partial_t + \mathcal{L}^N)u_k, \Phi - u_k, W)|\mathcal{G}_k).$$

By applying Itô's formula to $u(r,X_r^N)$ between s and T and to $u_k(r,X_r^N)$ between s and T, we get $\Psi^N(s,y,f_{u_k} + (\partial_t + \mathcal{L}^N)u_k, \Phi - u_k, W) = (u - u_k)(s,y) + \int_s^T f_{u_k}(r,X_r^N) + (\partial_t + \mathcal{L}^N)u(r,X_r^N)dr + \int_s^T(\partial_x u(r,X_r^N) - \partial_x u_k(r,X_r^N))\sigma(\varphi^N(r), X^N_{\varphi^N(r)})dW_r$. Then,

$$\begin{aligned}\mathrm{Var}(\varepsilon_k(s,y)|\mathcal{G}_k) \le{}& \frac{2}{M}\mathbb{E}\left(\left(\int_s^T f_{u_k}(r,X_r^N) + (\partial_t + \mathcal{L}^N)u(r,X_r^N)dr\right)^2|\mathcal{G}_k\right)\\ &+ \frac{2}{M}\mathbb{E}\left(\left(\int_s^T(\partial_x u(r,X_r^N) - \partial_x u_k(r,X_r^N))\sigma(\varphi^N(r), X^N_{\varphi^N(r)})dW_r\right)^2|\mathcal{G}_k\right).\end{aligned}$$

To bound the first term, we introduce $\pm f_u(r,X_r^N)$. Since f is Lipschitz continuous with constant L_f, and u satisfies $(\partial_t + \mathcal{L})u + f = 0$, we get $|f_{u_k}(r,X_r^N) + (\partial_t + \mathcal{L}^N)u(r,X_r^N)| \le L_f(|u_k - u|(r,X_r^N) + |\partial_x u_k\sigma - \partial_x u\sigma|(r,X_r^N)) + \sum_{i=1}^d |b_i(r,X_r^N) - b_i(\varphi^N(r), X^N_{\varphi^N(r)})||\partial_{x_i}u(r,X_r^N)| + \frac{1}{2}\sum_{i,j=1}^d |[\sigma\sigma^*]_{ij}(r,X_r^N) - [\sigma\sigma^*]_{ij}(\varphi^N(r), X^N_{\varphi^N(r)})||\partial^2_{x_ix_j}u(r,X_r^N)|$. Since b, σ, and u are bounded with bounded derivatives, we get $|f_{u_k}(r,X_r^N) + (\partial_t + \mathcal{L}^N)u(r,X_r^N)| \le L_f(|u - u_k|(r,X_r^N) + |\partial_x u_k\sigma -$

$\partial_x u\sigma|(r, X_r^N)) + K(T)c_{0,2}(u)(\frac{T}{N} + |X_r^N - X_{\varphi^N(r)}^N|)$. Thus,

$$\mathrm{Var}(\varepsilon_k(s,y)|\mathcal{G}_k) \leq \frac{K(T)}{M}\left(\frac{c_{0,2}^2(u)}{N} + \int_s^T \mathbb{E}(\Theta(r, X_r^N)|\mathcal{G}_k)dr\right),$$

where $\Theta(r,z) = |u-u_k|^2(r,z) + |\partial_x u - \partial_x u_k|^2(r,z)$. Finally, it remains to compute $\int_0^T e^{\beta s}\int_{\mathbb{R}^d}\nu_\mu^0(s,y)\int_s^T \mathbb{E}(\Theta(r,X_r^N)|\mathcal{G}_k)drdyds$. To do so, we write $\mathbb{E}(\Theta(r,X_r^N)|\mathcal{G}_k) = \int_{\mathbb{R}^d} dz\Theta(r,z)p^N(s,y;r,z)$, and we use successively the r.h.s. of (3.2), Proposition A.5, Lemma C.1, and the l.h.s. of (3.2). We get $\mathbb{E}(\|\varepsilon_k\|^2_{H^\mu_{\beta,X}}|\mathcal{G}_k) \leq \frac{K(T)}{M}(\frac{c_{0,2}^2(u)}{N} + \mathbb{E}(\|u-u_k\|^2_{H^\mu_{\beta,X}}|\mathcal{G}_k) + \mathbb{E}(\|\partial_x u - \partial_x u_k\|^2_{H^\mu_{\beta,X}}|\mathcal{G}_k))$. It remains to use the ellipticity condition on σ to end the proof. ◻

7. An example of operator $\mathcal{P}$ based on kernel estimators. In this section we present an operator satisfying Hypothesis 2. It is based on a nonparametric regression technique called local averaging. We refer to Györfi et al. [18, Chapter 2] and Härdle [19] for more details on nonparametric regression.

DEFINITION 7.1. *We approximate a function $v(t,x)$ by*

$$\mathcal{P}^k v(t,x) = \frac{r_n^k(t,x)}{f_n^k(t,x)} g(2^{d+1}T\lambda(B)f_n^k(t,x)), \tag{7.1}$$

where

- $r_n^k(t,x) = \frac{1}{nh_th_x^d}\sum_{i=1}^n K_t(\frac{t-T_i^k}{h_t})K_x(\frac{x-X_i^k}{h_x})v(T_i^k, X_i^k)$;
- $f_n^k(t,x) = \frac{1}{nh_th_x^d}\sum_{i=1}^n K_t(\frac{t-T_i^k}{h_t})K_x(\frac{x-X_i^k}{h_x})$;
- *the points $(T_i^k, X_i^k)_{1\leq i\leq n}$ are uniformly distributed on $[0,T]\times B$, where $B := [-a,a]^d$;*
- $\lambda(B) = (2a)^d$;
- *g is such that*

$$g(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 & \text{if } y > 1, \\ -y^4 + 2y^2 & \text{if } y \in [0,1]; \end{cases} \tag{7.2}$$

- *the kernel function K_t is defined on the compact support $[-1,1]$, bounded, even, nonnegative, C_p^2, and $\int_{\mathbb{R}} K_t(u)du = 1$;*
- *the kernel function K_x is defined on the compact support $[-1,1]^d$, bounded, and such that $\forall\ y = (y_1,\ldots,y_d) \in \mathbb{R}^d$, $K_x(y) = \Pi_{j=1}^d K_x^j(y_j)$, where for $j = 1,\ldots,d$ $K_x^j : \mathbb{R}\to\mathbb{R}$ is an even nonnegative C_p^2 function and $\int_{\mathbb{R}} K_x^j(y)dy = 1$;*
- *δ_n denotes $\frac{1}{nh_th_x^d}$, and $T\lambda(B)\delta_n \ll 1$;*
- *$h_x \ll a$ and $h_t \ll \frac{T}{2}$. Since we study the convergence when h_t and h_x tend to 0, we assume in the following that $h_t \leq 1$ and $h_x \leq 1$.*

Remark 7.2. Since we want to solve the PDE on $[0,T]\times\mathbb{R}^d$, we have to choose the interval $[-a,a]$ large enough. Besides, $g(2^{d+1}T\lambda(B)f_n^k(t,x))$ is nothing more than a regularizing term. If we had defined $\mathcal{P}^k v(t,x)$ as $\frac{r_n^k(t,x)}{f_n^k(t,x)}$, $\mathcal{P}^k v(t,x)$ would not have satisfied the regularity property of Hypothesis 2.

Now, let us check that $\mathcal{P}^k$ satisfies the properties of Hypothesis 2. Since we pick new random points at each iteration, the required measurability property is easily satisfied (with $\mathcal{A}_k := (T_i^k, X_i^k)_{1\leq i\leq n}$). The linearity property ensues from the definition of r_n, and the regularity property comes from the definition of g (see Remark 7.2). Concerning the boundedness, the following proposition holds.

LEMMA 7.3. *For any bounded function v, $\mathcal{P}^k v$, $\partial_x \mathcal{P}^k v$, $\partial_x^2 \mathcal{P}^k v$, and $\partial_t \mathcal{P}^k v$ are bounded by some constants depending on h_t, h_x, $c_0(v)$, $c_1(K_t)$, $c_2(K_x)$, T and $\lambda(B)$.*

Proof. It is sufficient to note that $x \longmapsto \frac{g(x)}{x}$ is bounded with bounded derivatives up to order 2. To get more details on the bounds, we refer the reader to [20, Proposition 11.8]. ∎

The convergence rates are proved in the companion paper [14] and are stated in the next theorem.

THEOREM 7.4. *Assume that the coefficients σ, b are bounded measurable functions on $[0,T] \times \mathbb{R}^d$, Lipschitz w.r.t. x, and that σ satisfies the ellipticity condition. The sequence $(\mathcal{P}^k)_k$ satisfies the Hypothesis 2 with*

$$\begin{aligned}
\epsilon_1(\mathcal{P}) &= K_0(T)\left(h_t^2 + h_x^2 + \frac{T\lambda(B)\delta_n}{h_x^2}\right),\\
\epsilon_2(\mathcal{P}) &= K_0(T)\left(h_t + e^{-\mu a}\frac{a^{d-1}}{h_x} + e^{-\frac{\mu a}{\sqrt{d}}} + \frac{T\lambda(B)\delta_n}{h_x^2}\right),\\
\epsilon_3(\mathcal{P}) &= \epsilon_1(\mathcal{P}) + \epsilon_2(\mathcal{P}),\\
\epsilon_4(\mathcal{P}) &= K_0(T)\frac{T\lambda(B)\delta_n}{h_x^2}, \qquad c_4(\mathcal{P}) = \frac{K_0(T)}{h_x^2}.
\end{aligned}$$

8. Numerical results.

8.1. Choice of the parameters a, n, h_x, h_t, M, and N and complexity of the algorithm.

By using the operator $\mathcal{P}$ described in section 7, the constants η and ϵ defined in Theorem 5.1 become $\eta = (K(T)\frac{4(1+T)L_f^2}{\beta} + K_0(T)(1 + K_{3.10}(T)L_f^2)(h_t^2 + h_x^2 + \frac{T\lambda(B)\delta_n}{h_x^2}(1+M^{-1}))$ and $\epsilon = K_0(T)(c_{1,2}^2(u) + c_0^2(f) + c_0^2(\Phi))(h_t + h_x^2 + \frac{T\lambda(B)\delta_n}{h_x^2}(1 + (MN)^{-1}) + \frac{e^{-\mu a}a^{d-1}}{h_x} + e^{-\frac{\mu a}{\sqrt{d}}} + \frac{1}{N^2 h_x^2})$. From this, we deduce the assumptions to be imposed on the coefficients to get a converging algorithm: $h_t^2 \ll 1$, $h_x^2 \ll 1$, $T\lambda(B)\delta_n \ll h_x^2$, $N^{-2} \ll h_x^2$, a large enough and M as small as we want. In particular, it means $n \gg T\lambda(B)h_t^{-1}h_x^{-d-2}$.

The complexity of Algorithm 2 is of order $K_{\max} nM(n+N)$, where $K_{\max}$ denotes the number of iteration steps. Numerically, one can choose $N \ll n$, then the complexity is equivalent to $K_{\max}Mn^2$. We don't need to choose M large, however, getting an accurate approximating operator requires a quite large value of n. The limiting factor of the complexity undoubtedly is n.

8.2. Application to financial problems.

In this section we present two applications of Algorithm 2 to financial problems. As presented in [11], BSDEs appear in numerous financial problems, like the pricing and the hedging of European and American options.

Example 1. In order to check the convergence of the algorithm, we consider a problem for which the exact solution is known. Let us consider the BSDE (1.1)–(1.2) in dimension $d = 1$ and with the following parameters: $b(t,x) = \mu_0 - \frac{\sigma^2}{2}$, $\sigma(t,x) = \sigma$, $f(t,x,y,z) = -ry - \frac{\mu_0 - r}{\sigma}z$, and $\Phi(x) = (e^x - K)^+$. Although the driver and the terminal condition are not bounded and Φ is not $C^{2+\alpha}$, it seems that the convergence of our algorithm still holds.

We explicitly know the value of the solution (Y,Z) of the BSDE: $Y_t = F(t, e^{X_t})$ and $Z_t = \partial_x F(t, e^{X_t})\sigma e^{X_t}$, where F is the price function of a standard call option in the Black and Scholes model.

TABLE 8.1
Evolution of $e(Y^k - Y)$ *and* $e(Z^k - Z)$ *w.r.t.* k.

	$e(Y^k - Y)$	$e(Z^k - Z)$
$k = 1$	0.0743476	0.0265350
$k = 2$	0.0014802	0.0104687
$k = 3$	0.0010029	0.0082452
$k = 4$	0.0008865	0.0076881
$k = 5$	0.0008373	0.0075321

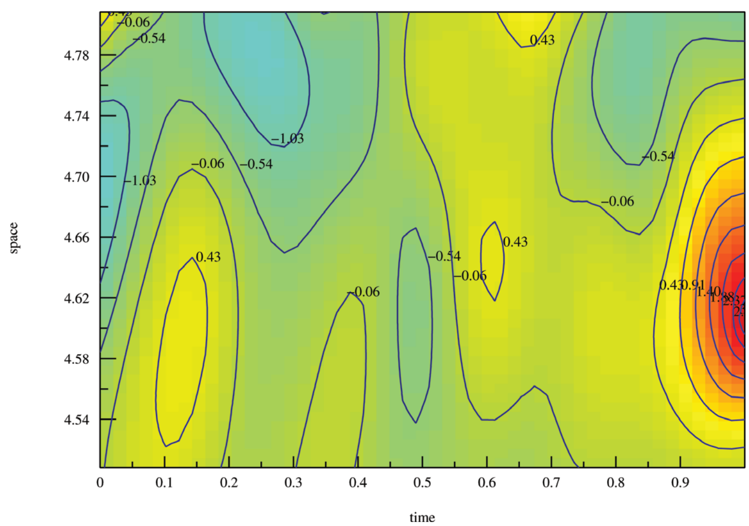

FIG. 8.1. *Evolution of the pointwise error on* Y.

To bring out the result of Theorem 5.1, we compute in Table 8.1 an approximation of $\|Y - Y^k\|^2_{\mu,\beta}$ (denoted $e(Y - Y^k)$) and $\|Z - Z^k\|^2_{\mu,\beta}$ (denoted $e(Z - Z^k)$) when k increases: $e(Y^k - Y) = \int_0^T \int_{x_-}^{x_+} |u(s,y) - u_k(s,y)|^2 e^{-|y|} dyds$, where $e^{x_-} = 90$, $e^{x_+} = 120$, $\beta = 0$, and $\mu = 1$. The option parameters are $\mu_0 = 0.1$, $\sigma = 0.2$, $r = 0.02$, $T = 1$, and $K = 100$. The algorithm parameters are $n = 2500$, $N = M = 100$, $h_x = h_t = 0.1$, and $2a = 1.2$. We use a truncated Gaussian kernel, i.e., $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathbf{1}_{|x| \le 6}$. We notice that $e(Y^k - Y)$ (resp., $e(Z^k - Z)$) is almost divided by 100 (resp., by 4) between the first and the fifth iterations. The huge difference between the two reduction coefficients is due to the fact Z is linked to the "derivative" of Y: it is well known that a function is always better approximated than its derivatives. Moreover, the errors drastically decrease between the first and the second iterations. From iteration 2, the algorithm does not so much improve the result anymore. Figure 8.1 represents the level-sets of the pointwise error on Y at iteration 10. Time goes from 0 to 1, and space varies between 4.5 and 4.81 (which means that the starting point for e^X belongs to $[90, 120]$). We notice that the error is quite small, except around the point $t = 1$ and $e^x = 100$. This corresponds to the fact that at maturity time, the solution Y_T is equal to $\Phi(X_T)$, which is continuous (but not C^1) in $(1, K)$.

Example 2. Let us consider the pricing of a basket call option in dimension 3 in the Black and Scholes model with the following parameters: $b(t,x) = rx$, $\sigma(t,x) = \sigma x$, f is kept the same as above, and $\Phi(x) = (\frac{x_1+x_2+x_3}{3} - K)_+$. We aim at computing

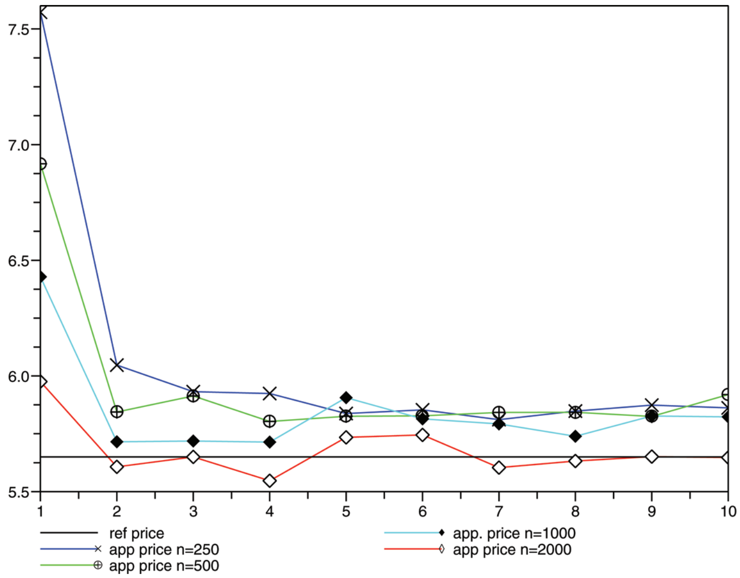

FIG. 8.2. *Evolution of Y_0 w.r.t. the iterations (in dimension 3).*

the price at time 0 and for $X_0 = (100, 100, 100)$. The reference price is given by the approximated formula given in [8]. The option parameters are $\mu_0 = r = 0.02$, $\sigma = 0.2$, $T = 1$, $\beta = 0$, and $\mu = 1$. The algorithm parameters are $N = M = 100$, $K_{\max} = 10$, and $2a = 127$ for each space component, i.e., we solve the BSDE on $[0, T] \times [55, 182]^3$. As above, we use a truncated Gaussian kernel. Figure 8.2 shows the evolution of Y_0 (i.e., the price of the option at time 0) w.r.t. the number of iterations k, when $n = 250$, $n = 500$, $n = 1000$, and $n = 2000$. The values of h_t and h_x depend on the value of n. They are computed with the following formula: $h_t = (\frac{T}{n})^{1/3}$ and $h_x = 2a(\frac{T}{n})^{1/3}$. We notice that the larger n is, the better is the approximation, and the faster is the convergence.

Example 3. We consider a basket Call option in dimension 5. We keep the same option parameters as in Example 2, except that $\phi(x) = (\frac{\sum_{i=1}^{5} x_i}{5} - K)_+$ and $N = M = 10$. The values of h_t and h_x are computed as above: $h_t = (\frac{T}{n})^{1/5}$ and $h_x = 2a(\frac{T}{n})^{1/5}$. We compute the price at time 0 and $X_0 = (100, 100, 100, 100, 100)$ and compare it with the reference price given by the approximated formula of [8]. Figure 8.3 shows the evolution of Y_0 w.r.t. the number of iterations k, when $n = 10^5$ and $n = 2.10^5$.

This algorithm also enables us to price and hedge contingent claims with constraints on the wealth or portfolio processes. For example, we can use it to hedge with higher interest rate for borrowing, which boils down to solving a nonlinear BSDE. We also can deal with the pricing and hedging of American options by using a penalization method. Since pricing American options is equivalent to solving a reflected BSDE, we approximate the solutions of the RBSDEs by a sequence of standard BSDEs with penalizations (see [10] for more details). We refer to [20, Chapter 15] for more precisions on these applications and for other financial applications.

As mentioned at the beginning of section 8, the limiting factor of the complexity of the algorithm is n, the number of points at which we compute the solution. To keep a good precision, n should increase with the dimension d. When $d = 5$, each

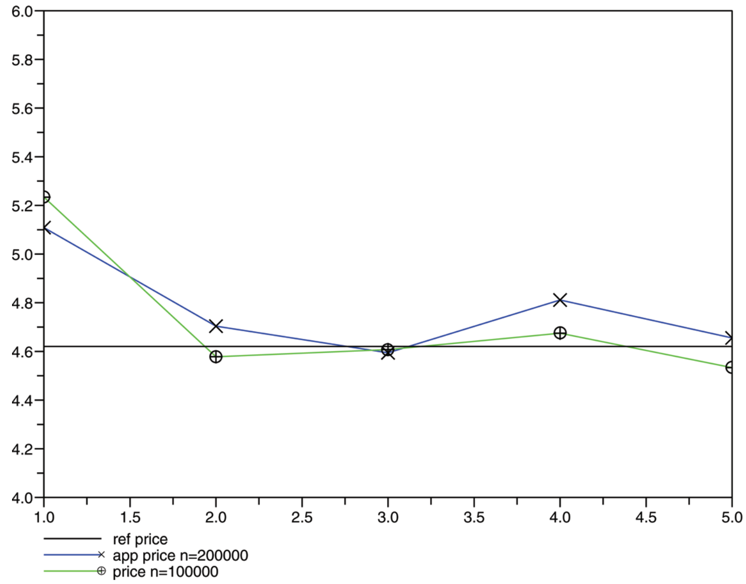

FIG. 8.3. *Evolution of Y_0 w.r.t. the iterations (in dimension 5).*

iteration requires 24 hours of computation on a standard personal laptop PC. To get over this time issue, one could parallelize the algorithm. This will be the purpose of future work.

Appendix A. Properties of the transition density of a diffusion process.

PROPOSITION A.1 (Aronson [1]). *Assume that the coefficients σ and b are bounded measurable functions and that σ is elliptic. There exist positive constants $K_{A.1}(T)$, α_1, α_2 s.t. for any $s,t \in [0,T](s > t)$ and any $x,y \in \mathbb{R}^d$*

$$\frac{K_{A.1}^{-1}(T)}{(2\pi\alpha_1(s-t))^{\frac{d}{2}}} e^{-\frac{|x-y|^2}{2\alpha_1(s-t)}} \leq p(t,x;s,y) \leq \frac{K_{A.1}(T)}{(2\pi\alpha_2(s-t))^{\frac{d}{2}}} e^{-\frac{|x-y|^2}{2\alpha_2(s-t)}}.$$

The constant $K_{A.1}(T)$ depends on d and on the suprema of the coefficients σ, b, and α_1, and α_2 depend on σ_0 and σ_1.

PROPOSITION A.2 (Ladyzenskaja et al. [21], pages 376–377). *Assume that σ is uniformly elliptic and that the coefficients σ, b are bounded, Hölder continuous of order α in x and $\frac{\alpha}{2}$ in t. There exist two positive constants $c_{A.2}$ (depending on σ_0, σ_1) and $K_{A.2}(T)$ (depending on d, α and on the suprema of σ, b), s.t. for any $s,t \in [0,T]$ $(s > t)$ and any $x,y \in \mathbb{R}^d$*

$$|\partial_t^r \partial_x^q p(t,x;s,y)| \leq K_{A.2}(T)(s-t)^{-\frac{(d+2r+q)}{2}} e^{-c_{A.2}\frac{|x-y|^2}{s-t}}, \text{ where } 2r+q \leq 2.$$

The following corollary bounds $\partial_{tx}^2 p(t,x;s,y)$. It ensues from a result bounding $\partial_x^{m+a}\partial_y^b p(t,x;s,y)$, $0 \leq |a|+|b| \leq 2, m = 0,1$, stated by Friedman (see [13, page 261]). We refer to [20, Proposition 6.9] for a detailed proof.

COROLLARY A.3. *Assume that σ and b are in $C_b^{1,2}$ and that σ is uniformly elliptic. It holds for any $s,t \in [0,T]$ $(s > t)$ and any $x,y \in \mathbb{R}^d$*

$$|\partial_{tx}^2 p(t,x;s,y)| \leq K_{A.3}(T)(s-t)^{-\frac{d+3}{2}} e^{-c_{A.3}\frac{|x-y|^2}{s-t}},$$

where $c_{A.3}$ depends on σ_0, σ_1, and $K_{A.3}(T)$ depends on d and on the suprema of σ, b .

PROPOSITION A.4 (Gobet and Labart [15]). *Assume σ is uniformly elliptic, b and σ are in $C_b^{1,3}$, and $\partial_t\sigma$ is in $C_b^{0,1}$. Then, $\forall(s,x,y) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^d$, there exist two positive constants $c_{A.4}$ and $K_{A.4}(T)$ s.t.*

$$|p(0,x;s,y) - p^N(0,x;s,y)| \leq \frac{K_{A.4}(T)T}{Ns^{\frac{d+1}{2}}}\exp\left(-\frac{c_{A.4}|x-y|^2}{s}\right),$$

where $c_{A.4}$ depends on σ_0,σ_1 and $K_{A.4}(T)$ depends on the dimension d, on the upper bounds of σ, b and their derivatives.

PROPOSITION A.5 (Gobet and Labart [15]). *Assume σ is uniformly elliptic and b and σ are in $C_b^{0,2}$. Then for any $x,y \in \mathbb{R}^d$, $s \in [0,T]$, there exist two positive constants $c_{A.5}$ and $K_{A.5}(T)$ s.t.*

$$p^N(0,x;s,y) \leq \frac{K_{A.5}(T)}{s^{d/2}}e^{-c_{A.5}\frac{|x-y|^2}{s}},$$

where $c_{A.5}$ depends on σ_0,σ_1 and $K_{A.5}(T)$ depends on d, on the upper bounds of b,σ and their derivatives.

Appendix B. Bounds for linear PDEs in weighted Sobolev spaces.

DEFINITION B.1 (Space $H^{m,\mu}$ and $L_2(0,T;F)$).

- *For any $m \leq 2$ and $\mu > 0$, $H^{m,\mu}$ defines the space of functions $v : \mathbb{R}^d \to \mathbb{R}$ s.t.*

$$\|v\|_{H^{m,\mu}} = \left(\sum_{k\leq m}\int_{\mathbb{R}^d} e^{-\mu|x|}|\partial^k v(x)|^2dx\right)^{\frac{1}{2}} < \infty.$$

  For $m = 0$, $H^\mu := H^{0,\mu}$.
- *Space $L_2(0,T;F)$. Let $(F,\|\cdot\|_F)$ be a Banach space. $L_2(0,T;F)$ defines the space of functions ϕ from $[0,T]$ into F s.t. $\|\phi\|^2_{L_2(0,T;F)} = \int_0^T\|\phi(t)\|_F^2dt < \infty$.*

The following result is a direct application of Bensoussan and Lions [6, Theorem 6.12 page 130].

THEOREM B.2 (Bensoussan and Lions [6]). *Assume σ is uniformly elliptic, σ is in $C_b^{1,1}$, b is in $C^{1,1}$, b is bounded, and k is bounded from below. We also assume that $\overline{g} \in L_2(0,T;H^\mu)$. Then, the solution $\overline{v}$ of $(\partial_t+\mathcal{L})\overline{v}(t,x) + k(t,x)\overline{v}(t,x) + \overline{g}(t,x) = 0$ with terminal condition $\overline{v}(T,\cdot) = 0$, is in $L_2(0,T;H^{2,\mu})$ and $\partial_t\overline{v} \in L_2(0,T;H^\mu)$. Furthermore, we have*

$$\|\partial_t\overline{v}\|_{L_2(0,T;H^\mu)} + \|\overline{v}\|_{L_2(0,T;H^{2,\mu})} \leq K_{B.2}(T)\|\overline{g}\|_{L_2(0,T;H^\mu)}.$$

PROPOSITION B.3. *Let v be the solution of $(\partial_t+\mathcal{L})v(t,x) + g(t,x) = 0$ with terminal condition $v(T,\cdot) = \Pi(\cdot)$. Assume $g \in C_b^{0,0}$, $\Pi \in C_b^2$ and σ and b are in $C_b^{1,2}$. Then, for all $t,t' \in [0,T]$, it holds $|\partial_x v(t',x) - \partial_x v(t,x)| \leq c_{1/2}(v)\sqrt{|t'-t|}$, where $c_{1/2}(v)$ is of the form $K(T)(c_0(g)+c_2(\Pi))$.*

Proof. First, we do the proof in the special case $\Pi \equiv 0$. Assume without loss of generality that $t < t'$. Feynman–Kac's formula gives $v(t,x) = \mathbb{E}[\int_t^T g(s,X_s)ds]$. Using the transition density function of X_s leads to $\partial_{x_i}v(t,x) = \int_{\mathbb{R}^d}dy\int_t^T g(s,y)\partial_{x_i}p(t,x;s,y)ds$. Then,

$$\begin{aligned}\partial_{x_i}v(t',x) - \partial_{x_i}v(t,x) = \int_{\mathbb{R}^d}dy\int_{t'}^T & g(s,y)[\partial_{x_i}p(t',x;s,y) - \partial_{x_i}p(t,x;s,y)]ds \\ & -\int_t^{t'} g(s,y)\partial_{x_i}p(t,x;s,y)ds := A_1 - A_2.\end{aligned}$$

Let us establish upper bounds for A_1 and A_2.

Upper bound for A_2. Applying Proposition A.2 yields $|A_2| \leq K(T)c_0(g)\int_{\mathbb{R}^d} dy \int_t^{t'} \frac{ds}{(s-t)^{\frac{(d+1)}{2}}} e^{-c\frac{|x-y|^2}{s-t}} \leq K(T)c_0(g)\sqrt{t'-t}$.

Upper bound for A_1. Applying Taylor's formula gives $\partial_{x_i}p(t',x;s,y) - \partial_{x_i}p(t,x;s,y) = (t'-t)\int_0^1 \partial^2_{tx}p(\lambda t' + (1-\lambda)t, x; s, y)d\lambda$. Then, we use Corollary A.3: $|A_1| \leq K(T)c_0(g)(t'-t)\int_{\mathbb{R}^d} dy \int_{t'}^T ds \int_0^1 (s-(\lambda t'+(1-\lambda)t))^{-\frac{d+3}{2}} e^{-c\frac{|x-y|^2}{(s-(\lambda t'+(1-\lambda)t))}} d\lambda \leq K(T)c_0(g)(t'-t)\int_{t'}^T ds \int_0^1 (s-(\lambda t' + (1-\lambda)t))^{-\frac{3}{2}} d\lambda$. Since $\lambda t' + (1-\lambda)t \in [t,t']$, we can write $(s-(\lambda t' + (1-\lambda)t))^{-\frac{3}{2}} \leq ((s-(\lambda t' + (1-\lambda)t))\sqrt{s-t'})^{-1}$. It remains to prove that $\int_{t'}^T ds \int_0^1 ((s-(\lambda t' + (1-\lambda)t))\sqrt{s-t'})^{-1} d\lambda \leq K(T)\sqrt{t'-t}^{-1}$. $\int_{t'}^T \frac{ds}{(s-(\lambda t'+(1-\lambda)t))\sqrt{s-t'}} = 2\int_0^{\sqrt{T-t'}} \frac{du}{u^2+(1-\lambda)(t'-t)} = \frac{2}{\sqrt{(1-\lambda)(t'-t)}} \int_0^{\sqrt{\frac{T-t'}{(1-\lambda)(t'-t)}}} \frac{dv}{1+v^2}$. Thus, $\int_{t'}^T ds \int_0^1 \frac{1}{(s-(\lambda t'+(1-\lambda)t))\sqrt{s-t'}} d\lambda \leq \frac{\pi}{\sqrt{t'-t}} \int_0^1 \frac{d\lambda}{\sqrt{1-\lambda}}$, and the result follows. In the general case, we introduce $v_0(t,x) := v(t,x) - \Pi(x)$. v_0 satisfies the PDE $(\partial_t + \mathcal{L})v_0(t,x) + \mathcal{L}\Pi(x) + g(t,x) = 0$ with terminal condition $v_0(T,\cdot) = 0$. Then, the first part of the proposition yields that for all $t,t' \in [0,T]$, it holds $|\partial_x v_0(t',x) - \partial_x v_0(t,x)| \leq c_{1/2}(v_0)\sqrt{|t'-t|}$, where $c_{1/2}(v_0)$ is of the form $K(T)(c_0(g) + c_2(\Pi))$. Since $\partial_x v_0(t',x) - \partial_x v_0(t,x) = \partial_x v(t',x) - \partial_x v(t,x)$, we get the result. ∎

Appendix C. Technical results.

LEMMA C.1. *Let $I = \int_{\mathbb{R}^d} dx e^{-\mu|x|} \frac{1}{(s-t)^{\frac{d}{2}}} \exp(-c\frac{|x-y|^2}{s-t})$, where $c > 0$. For any $s,t \in [0,T]$ and any $x,y \in \mathbb{R}^d$ such that $t < s$, the following assertion holds*

$$\frac{1}{2^d}\left(\frac{\pi}{c}\right)^{d/2} e^{-\frac{d\mu^2}{4c}(s-t)} e^{-\mu|y|} \leq I \leq 2^d \left(\frac{\pi}{c}\right)^{d/2} e^{\frac{d\mu^2}{4c}(s-t)} e^{-\mu|y|}.$$

Proof. Using a change of variables in I yields $I = \left(\frac{\pi}{c}\right)^{d/2}\mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}+y|}]$. Furthermore, $e^{-\mu|y|}\mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}|}] \leq \mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}+y|}] \leq e^{-\mu|y|}\mathbb{E}[e^{\mu\frac{1}{\sqrt{2c}}|W_{s-t}|}]$. The components $(W^i_{s-t})_{1\leq i\leq d}$ of W_{s-t} are independently and identically distributed, then $\mathbb{E}[e^{\mu\frac{1}{\sqrt{2c}}|W_{s-t}|}] \leq \left(\mathbb{E}[e^{\mu\frac{1}{\sqrt{2c}}|W^1_{s-t}|}]\right)^d \leq 2^d(\mathbb{E}[\mathrm{ch}(\mu\frac{1}{\sqrt{2c}}W^1_{s-t})])^d \leq 2^d e^{\frac{d\mu^2(s-t)}{4c}}$. The last term being bounded, we get the upper bound for I. Since $\mathbb{E}[e^{-\mu|\frac{1}{\sqrt{2c}}W_{s-t}|}] \geq \frac{1}{\mathbb{E}[e^{\mu|\frac{1}{\sqrt{2c}}W_{s-t}|}]}$, the same proof gives the lower bound. ∎

REFERENCES

[1] D. ARONSON, *Bounds for the fundamental solution of a parabolic equation*, Bull. Amer. Math. Soc., 73 (1967), pp. 890–903.

[2] V. BALLY AND A. MATOUSSI, *Weak solutions for SPDEs and backward doubly stochastic differential equations*, J. Theoret. Probab., 14 (2001), pp. 125–164.

[3] V. BALLY AND G. PAGÈS, *Error analysis of the optimal quantization algorithm for obstacle problems*, Stochastic Process. Appl., 106 (2003), pp. 1–40.

[4] C. BENDER AND R. DENK, *A forward scheme for backward SDEs*, Stochastic Process. Appl., 117 (2007), pp. 1793–1812.

[5] C. BENDER AND J. ZHANG, *Time discretization and Markovian iteration for coupled FBSDEs*, Ann. Appl. Probab., 18 (2008), pp. 143–177.

[6] A. BENSOUSSAN AND J. LIONS, *Applications des Inéquations Variationnelles en Contrôle Stochastique*, Dunod, Paris, 1978.

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

[8] R. Carmona and V. Durrleman, *Generalizing the Black-Scholes formula to multivariate contingent claims*, J. Comput. Finance, 9 (2005), pp. 43–67.

[9] F. Delarue and S. Menozzi, *A forward-backward stochastic algorithm for quasi-linear PDEs*, Ann. Appl. Probab., 16 (2006), pp. 140–184.

[10] N. El Karoui, C. Kapoudjian, E. Pardoux, S. Peng, and M. Quenez, *Reflected solutions of backward SDE's, and related obstacle problems for PDE's*, Ann. Probab., 25 (1997), pp. 702–737.

[11] N. El Karoui, S. Peng, and M. Quenez, *Backward stochastic differential equations in finance*, Math. Finance, 7 (1997), pp. 1–71.

[12] A. Fahim, N. Touzi, and X. Warin, *A Probabilistic Numerical Method for Fully Nonlinear Parabolic PDEs*, http://hal.archives-ouvertes.fr/hal-00367103/fr/.

[13] A. Friedman, *Stochastic Differential Equations and Applications*, Academic Press, New York, 1975.

[14] E. Gobet and C. Labart, *A note on the operator $\mathcal{P}^k v$*, http://hal.archives-ouvertes.fr/hal-00373349/fr/.

[15] E. Gobet and C. Labart, *Sharp estimates for the convergence of the density of the Euler scheme in small time*, Electron. Comm. Probab., 13 (2008), pp. 352–363.

[16] E. Gobet, J. Lemor, and X. Warin, *A regression-based Monte Carlo method to solve backward stochastic differential equations*, Ann. Appl. Probab., 15 (2005), pp. 2172–2202.

[17] E. Gobet and S. Maire, *Sequential control variates for functionals of Markov processes*, SIAM J. Numer. Anal., 43 (2005), pp. 1256–1275.

[18] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer, New York, 2002.

[19] W. Härdle, *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, UK, 1992.

[20] C. Labart, *BSDE: Analysis of the discretisation error and simulation using adaptive Monte Carlo methods; Domain perturbations for American options*, Ph.D. thesis, CMAP, Ecole Polytechnique, Palaiseau, 2007.

[21] O. Ladyzenskaja, V. Solonnikov, and N. Ural'ceva, *Linear and quasi-linear equations of parabolic type*, Translations of Mathematical Monographs 23, American Mathematical Society, Providence, RI, 1968.

[22] J. Ma, P. Protter, and J. Yong, *Solving forward backward stochastic differential equations explicitly-a four step scheme*, Probab. Theory Related Fields, 98 (1994), pp. 339–359.

[23] E. Pardoux and S. Peng, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, Lecture Notes in Control and Inform. Sci., 176 (1992), pp. 200–217.