ECOLE POLYTECHNIQUE

CENTRE DE MATHÉMATIQUES APPLIQUÉES

UMR CNRS 7641

91128 PALAISEAU CEDEX (FRANCE). Tél: 01 69 33 41 50. Fax: 01 69 33 30 11
http://www.cmap.polytechnique.fr/

Jean-Philippe LEMOR,
Emmanuel GOBET,
Xavier WARIN

Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations

R.I. 574 *May 2005*

Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations

Jean-Philippe Lemor, Emmanuel Gobet[*], Xavier Warin

May 26, 2005

Abstract. We are concerned with the numerical resolution of backward stochastic differential equations, whose data depend on a jump-diffusion process. We propose and analyze a numerical scheme based on iterative regressions on function bases, which coefficients are evaluated using Monte Carlo simulations. Regarding the error, we derive explicit bounds with respect to the time step, the number of simulated paths and the number of basis functions, which allows us to optimally adjust the parameters to achieve a given accuracy. We also present numerical experiments related to option pricing with differential interest rates and to locally risk-minimizing strategies (Föllmer-Schweizer decomposition).

Key words: backward stochastic differential equations, empirical regressions, Föllmer-Schweizer decomposition.

Mathematics Subject Classification (1991): 60H10, 62G08, 65C

Introduction

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ be a given filtered probability space on which are defined a standard Brownian motion W in $\mathbb{R}^q$ and a jump-diffusion process X in $\mathbb{R}^d$. We aim at numerically approximating a generalized backward stochastic differential equation (GBSDE) with a fixed terminal time T

$$-\mathrm{d}Y_t = f(t, X_t, Y_t, Z_t)\mathrm{d}t - Z_t \mathrm{d}W_t - \mathrm{d}L_t, \qquad Y_T = \phi(X_T), \tag{1}$$

[*]Corresponding author.

where Y is a scalar càdlàg adapted process, Z is a predictable $\mathbb{R}^q$-valued process (as a row vector) and L is a scalar càdlàg martingale orthogonal to W (with $L_0 = 0$). Actually in what follows, in principle X could also be any Markov process and the terminal condition could be a functional of the path $(X_t)_{0\leq t\leq T}$ (see Remark 1). Our results can be readily extended to higher dimensional Y, Z and L as well. Under suitable Lipschitz assumptions on the driver f and L_2-integrability conditions, there is an unique solution (Y, Z, L) in appropriate spaces of processes (with L_2 norms): for details, we refer to Pardoux and Peng (1990) for the Brownian filtration ($L \equiv 0$), and to El Karoui *et al.* (1997) for general filtrations. A more complete situation is handled in Barles *et al.* (1997), where in addition the driver is allowed to depend somehow on the martingale L.

The main focus of this work is to provide and analyze a simple algorithm, based on empirical regression methods using simulated paths of X, which approximates (Y, Z) solution of (1) (L could be obtained as a difference). For the convergence analysis, techniques from BSDEs and non-parametric regressions are mixed, which illustrates another interesting interface between probability and statistics. To encourage future collaborations between people of these different fields, we now give an overview of the related applications and issues.

Applications. During the last decade, the interest in designing efficient numerical methods to solve (1) has become very high, because of the various applications. We mention some of them.

1. Solving (1) may give an access to the Föllmer-Schweizer (FS in short) decomposition of $\phi(X_T) = Y_0 + \int_0^T \xi_t \mathrm{d}X_t + \tilde{L}_T$, with a martingale $\tilde{L}$ strongly orthogonal to the martingale part of X; in that case, the driver f is linear. Regarding the applications in finance, the FS decomposition plays a crucial role in valuing and hedging claims (of type $\phi(X_T)$) in incomplete markets, see Monat and Stricker (1995). Incompleteness may arise from non tradeable sources of risk; this is indeed the case with stochastic volatility models and jump models, both models being included in our jump-diffusion framework for X. A direct use of the FS decomposition leads to the locally risk-minimizing strategies and to the *minimal martingale measure* valuation. For a mean-variance hedging criterion, one also needs FS decompositions. For these issues, we refer to Pham *et al.* (1998) and for a survey, to Schweizer (1999). In these problems, we are interested in the prices (related to Y) and particularly in the hedging strategies (related to Z).

2. Non-linear drivers f allow to deal with market imperfections in finance: higher interest rate for borrowing (Bergman 1995), short sales constraints

(Jouini and Kallal 1995)... There is also a strong connection with non-linear pricing rules and dynamic risk measures (Peng 2003). For numerous references, see El Karoui *et al.* (1997).

3. In view of the connection with semi-linear PDEs, possibly with integral-differential operators (see Barles *et al.* 1997), our probabilistic approach may be an interesting alternative to deterministic PDE schemes.

To our knowledge this is the first time that the case of a non Brownian filtration (leading to $L \neq 0$) is considered for the numerical resolution of (1). This more general framework has been partly motivated by recent financial modeling: Lévy processes are now often incorporated to model jumps and spikes in energy markets (see for instance Benth *et al.* 2003); modeling credit risk makes an intensive use of Poisson processes (see Bielecki and Rutkowski 2002); point processes can be used to model the impact of rating and credit events on the dynamics of risky assets (see Becherer and Schweizer 2005).

Where non-parametric regressions come in. The first approximation of (1) is a time discretization using a time step $h = \frac{T}{N}$: the discretization times are denoted $(t_k = kh)_{0\leq k\leq N}$. We denote $\Delta W_k = W_{t_{k+1}} - W_{t_k}$ ($\Delta W_{l,k}$ componentwise) and X^N a relative approximation of X at these discretization times: say it is obtained through an Euler scheme on the jump-diffusion equation satisfied by X (see Jacod 2004 among others). Quite naturally, the solution (Y, Z) of (1) is approximated by (Y^N, Z^N) defined in a backward manner by $Y^N_{t_N} = \phi(X^N_{t_N})$ and

$$\begin{aligned} Y^N_{t_k} &= \mathbb{E}_{t_k}(Y^N_{t_{k+1}}) + h\mathbb{E}_{t_k} f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k}), &(2)\\ h\ Z^N_{t_k} &= \mathbb{E}_{t_k}(Y^N_{t_{k+1}} \Delta W^*_k), &(3)\end{aligned}$$

where $\mathbb{E}_{t_k}$ stands for the conditional expectation with respect to $\mathcal{F}_{t_k}$ and * for the transpose. In Theorem 1 below, we state the convergence of (Y^N, Z^N) towards (Y, Z) in the standard BSDE L_2-norm as N goes to infinity. As the terminal condition is a deterministic function of $X^N_{t_N}$ and as X^N is a Markov chain, it is easy to see that $Y^N_{t_k} = y^N_k(X^N_{t_k})$ and $Z^N_{t_k} = z^N_k(X^N_{t_k})$ where y^N_k and z^N_k are unknown regression functions defined in a backward manner by $y^N_N(\cdot) = \phi(\cdot)$ and:

$$\begin{aligned} y^N_k(x) =&\mathbb{E}\big(y^N_{k+1}(X^N_{t_{k+1}}) + hf(t_k, X^N_{t_k}, y^N_{k+1}(X^N_{t_{k+1}}), z^N_k(X^N_{t_k}))|X^N_{t_k} = x\big),\\ h\ z^N_k(x) =&\mathbb{E}\big(y^N_{k+1}(X^N_{t_{k+1}})\Delta W^*_k|X^N_{t_k} = x\big).\end{aligned}$$

We are thus faced to the iterative computation of N unknown regression functions. There are several ways to approximate a regression function : for example, we can think of either kernel methods (see for example Bosq and Lecoutre 1987), or projection methods on functions basis (see for instance Györfi *et al.*

2002). But compared to the classic non-parametric regression problem, in our case there is an extra difficulty because the N regression functions are intricate: the regression function estimated at time t_{k+1} is used to estimate a new regression function at time t_k. Thus we have to find a way of approximating the unknown regression functions which matches two constraints : it must lead to a *nice* propagation of the error during the backward in time iteration and its complexity must be reasonable regarding the accuracy (keeping in mind that N going to infinity, more and more regression functions have to be estimated).
Some approximation schemes have already been considered to solve this problem, in the case of Brownian filtration and diffusion processes for X. The first method consists in replacing X^N by a Markov chain with finite state space and known transition probabilities, leading to a regression function that can be exactly computed. This is either achieved by replacing the Brownian motion by a random walk (see Briand *et al.* 2001, or Ma *et al.* 2002) or by using quantization techniques (see Bally and Pagès 2002). The second method consists in directly computing a non-parametric approximation of the regression function. Bouchard and Touzi (2004) use a technique based on Malliavin calculus integration by part formulae (under an ellipticity assumption on the diffusion process X) whereas Egloff (2005) uses a least-squares method, both methods using Monte-Carlo simulations of X^N.

Our contributions. In this paper, we also approximate the unknown regression functions using projections on functions basis. Using M Monte-Carlo simulations of X^N, we solve at each discretization time t_k a least-squares problem to determine our approximation in the vector space spanned by a finite number of functions. The parameters of this numerical scheme are the number of time steps N, the number of Monte-Carlo simulations M and the number and kind of basis functions. In Gobet *et al.* (2005), for an analogous procedure we have already studied the influence of the parameters but unfortunately, the estimates as $M \to \infty$ (see Gobet *et al.* 2005, Theorem 3) involve the fourth moments of the L_2-orthonormalized basis functions. It turns out that these moments are difficult to estimate and in fact, they presumably converge to infinity as the size of the basis increases. Hence the practical use of these results is still questionable, in particular if one has to achieve a given accuracy by allowing a joint convergence of N, M and the number of basis functions to infinity. Here, our goal is to derive tractable error estimates that depend only on N, M and the number of basis functions.
Before going into details, we mention that these results enable us to derive an explicit rate of convergence for an algorithm which is very efficient. Beyond the fact that we are not restricted to Brownian randomness, we mention other advantages of our approach compared to existing schemes. Compared to Bally

and Pagès (2002), we do not need quantization grids and we allow more flexible choices of functions basis than Voronoï cells only. Compared to random walk approximations (Briand *et al.* 2001, Ma *et al.* 2002), we establish a rate of convergence. The algorithm is easier to implement than the one in Bouchard and Touzi (2004) and leads to a better accuracy. In addition, regarding to X our approach is distribution-free, which is a major advantage; in particular one does not require any non-degeneracy condition on X, as it is necessary in a Malliavin calculus approach. Finally, it improves results given in Egloff (2005), obtained for optimal stopping problems with a fixed number of dates N (actually, his error estimates increase geometrically with N and thus, are not relevant in the current framework).

Organization of the paper. The paper is organized as follows. In Section 1, we define rigorously the model, introduce notations used throughout the paper, explain the algorithm and state error bounds with respect to the time-step, the number of Monte-Carlo simulations and the number of basis functions. These bounds are proved in Section 2, combining BSDE techniques and non-parametric regression arguments. In Section 3 we propose a better (but less natural) alternative to the algorithm proposed in Section 1. Finally in Section 4 we make numerical experiments which illustrate the bounds derived in Section 1 for explicit choices of function bases.

Remark 1 *By considering extra state variables, the results obtained here can be extended to the case of a terminal condition $\phi(X)$ which is a Lipschitz functional of the path of X. For instance, if $\phi(X) = \phi(X_T, \min_{0\le t\le T} X_t)$ for a scalar process X, the regression functions at time t_k should depend on $X^N_{t_k}$ but also on $\min_{t_i\le t_k} X^N_{t_i}$. For details, see Gobet* et al. *(2005).*

1 The algorithm

1.1 Model

We follow the presentation of Barles *et al.* (1997). Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ be a stochastic basis, where the filtration satisfies the usual conditions of right-continuity and completeness. We suppose that the filtration is generated by the two mutually independent processes: a $\mathbb{R}^q$-valued Brownian motion W and a Poisson random measure μ on $\mathbb{R}_+ \times E$, where $E = \mathbb{R}^l\backslash\{0\}$ is equipped with its Borel field $\mathcal{E}$, with compensator $\nu(\mathrm{d}t, \mathrm{d}e) = \mathrm{d}t\lambda(\mathrm{d}e)$, such that $\{\tilde\mu([0,t]\times A) = (\mu-\nu)([0,t]\times A)\}_{t\ge 0}$ is a martingale for all $A \in \mathcal{E}$ with $\lambda(A) < +\infty$. λ is assumed to be a σ-finite measure on $(E, \mathcal{E})$ satisfying $\int_E (1 \wedge |e|^2)\lambda(\mathrm{d}e) < +\infty$. We consider the $\mathbb{R}^d$-valued jump-diffusion

$$X_t = x + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s + \int_0^t \int_E \beta(s, X_{s^-}, e)\tilde{\mu}(\mathrm{d}s, \mathrm{d}e), \quad (4)$$

which is uniquely defined under the following assumption.

(H1) The functions $b(t,x)$ and $\sigma(t,x)$ are uniformly Lipschitz continuous with respect to $(t,x) \in [0,T] \times \mathbb{R}^d$.

For some constant c, the function β satisfies $|\beta(t,x,e)| \leq c(1 \wedge |e|)$ and $|\beta(t,x,e) - \beta(t',x',e)| \leq c(|x-x'| + |t-t'|)(1 \wedge |e|)$ for any $(t,x), (t,x') \in [0,T] \times \mathbb{R}^d$ and $e \in E$.

We consider a time-discretization of X, which we denote X^N (we may think of Euler schemes, see Jacod 2004 and references therein). The latter is assumed to converge to X in L_2-norm, which is stated as

(H2) As N goes to infinity, one has $\sup_{0\leq k\leq N} \mathbb{E}|X_{t_k} - X^N_{t_k}|^2 \to 0$.

The GBSDE (1) is well-defined under the assumption

(H3) The driver f satisfies the following continuity estimate:

$$|f(t_2,x_2,y_2,z_2) - f(t_1,x_1,y_1,z_1)| \leq C(|t_2-t_1|^{1/2} + |x_2-x_1| + |y_2-y_1| + |z_2-z_1|)$$

for any $(t_1,x_1,y_1,z_1), (t_2,x_2,y_2,z_2) \in [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q$.

The terminal condition ϕ is Lipschitz continuous.

Actually, only the L_2-integrability of $\phi(X_T)$ is usually required, but here the smoothness of ϕ is imposed to derive explicit error estimates.
Finally, to ensure that the discrete GBSDE satisfy a Lipschitz continuity property with respect to the state variable X^N, one requires $(X^N_{t_k})_k$ to be a Markov chain and $X^{N,k_0,x}_{t_k}$ to satisfy[1]

(H4) For some constant $C > 0$, one has

a) $\mathbb{E}|X^{N,k_0,x}_{t_N} - X^{N,k_0,x'}_{t_N}|^2 + \mathbb{E}|X^{N,k_0,x}_{t_{k_0+1}} - X^{N,k_0,x'}_{t_{k_0+1}}|^2 \leq C|x-x'|^2$ for any x and x', uniformly in k_0 and N.

b) $\mathbb{E}|X^{N,k_0,x}_{t_{k_0+1}} - x|^2 \leq Ch(1+|x|^2)$ for any x, uniformly in k_0 and N.

This kind of assumption has been introduced in Gobet *et al.* (2005). The above property is quite natural since it is fulfilled by X itself under **(H1)**. Now, we state a convergence result regarding the time discretization.

[1] as usual, $X^{N,k_0,x}_{t_k}$ denotes $X^N_{t_k}$ starting at x at time t_{k_0}

Theorem 1 *Under* **(H1-H2-H3)**, *define the error*

$$e(N) = \max_{0\le k\le N} \mathbb{E}|Y^N_{t_k} - Y_{t_k}|^2 + \mathbb{E}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} |Z^N_{t_k} - Z_t|^2 \mathrm{d}t,$$

where Y^N and Z^N are given by (2) *and* (3). *Then, $e(N)$ converges to* 0 *as $N \to \infty$. Furthermore, in the case of Brownian filtration ($\beta \equiv 0$ and $L \equiv 0$) and when X^N is the Euler scheme of X, one has $e(N) = O(N^{-1})$.*

The proof is somewhat standard and is not the core of our work. Hence we postpone it to the Appendix.

1.2 Notations

We now introduce convenient notations to describe the algorithm.
Localization. We define localized versions of the Brownian increments and of the functions f, ϕ:

$$\begin{aligned}
[\Delta W_{l,k}]_w &= \big(-R_0\sqrt{h}\ \big) \vee \Delta W_{l,k} \wedge \big(R_0\sqrt{h}\ \big),\\
f^R(t,x,y,z) &= f(t, -R_1 \vee x_1 \wedge R_1, \cdots, -R_d \vee x_d \wedge R_d, y, z),\\
\phi^R(x) &= \phi(-R_1 \vee x_1 \wedge R_1, \cdots, -R_d \vee x_d \wedge R_d),
\end{aligned}$$

where $R = (R_0, R_1, \dots, R_d) \in (\mathbb{R}^+)^{d+1}$ which influence is analyzed in the following section. These localizations enable us to slightly modify (2-3) and define $(Y^{N,R}, Z^{N,R})$ by:

$$Y^{N,R}_{t_k} = \mathbb{E}_{t_k}(Y^{N,R}_{t_{k+1}}) + h\mathbb{E}_{t_k} f^R(t_k, X^N_{t_k}, Y^{N,R}_{t_{k+1}}, Z^{N,R}_{t_k}), \tag{5}$$

$$hZ^{N,R}_{t_k} = \mathbb{E}_{t_k}(Y^{N,R}_{t_{k+1}}[\Delta W_k]^*_w), \tag{6}$$

and $Y^{N,R}_{t_N} = \phi^R(X^N_{t_N})$. With the same arguments as for (Y^N, Z^N), we easily see that $Y^{N,R}_{t_k} = y^{N,R}_k(X^N_{t_k})$ and $Z^{N,R}_{t_k} = z^{N,R}_k(X^N_{t_k})$ for deterministic functions $y^{N,R}_k(\cdot)$ and $z^{N,R}_k(\cdot)$. Moreover, the functions $y^{N,R}_k$ and $\sqrt{h}z^{N,R}_k$ are Lipschitz continuous, uniformly in R and N (see Proposition 1 later). But the main interest of this localization is to provide <u>bounded</u> (unknown) regression functions $y^{N,R}_k$ and $z^{N,R}_k$: one has $\|y^{N,R}_k\|_\infty \le C_y(R)$ and $\|z^{N,R}_{l,k}\|_\infty \le C_z(R)$ (for details on these upper bounds, see again Proposition 1). This boundedness property plays a important role in the derivation of error bounds.

Function basis. At each discretization time t_k, $0 \le k \le N-1$, we choose $q+1$ deterministic function bases $\big(p_{l,k}(\cdot)\big)_{0\le l\le q}$ and we look for an approximation of $y^{N,R}_k(\cdot)$ (resp. $z^{N,R}_{l,k}(\cdot)$) in the vector space spanned by the basis $p_{0,k}$ (resp. $p_{l,k}$). Each basis $p_{l,k}$ is considered as a vector of functions, which size equals $K_{l,k}$. The vector space of functions spanned by $p_{l,k}$ is denoted $\mathcal{P}_{l,k}$, i.e. $\mathcal{P}_{l,k} = \{\alpha \cdot p_{l,k}(\cdot), \alpha \in \mathbb{R}^{K_{l,k}}\}$.

Monte-Carlo simulations. The evaluation of the different projection coefficients α will be obtained using M independent Monte-Carlo simulations of $(X^N_{t_k})_{0\le k\le N}$ and $(\Delta W_k)_{0\le k\le N-1}$. We denote by $(X^{N,m}_{t_k})_{1\le m\le M,0\le k\le N}$ and $(\Delta W^m_k)_{1\le m\le M,0\le k\le N-1}$ these Monte-Carlo simulations.
To keep notations short, we write $p_{l,k}(X^{N,m}_{t_k}) = p^m_{l,k}$. We define by $B^M_{l,k}$ the matrix of size $M\times K_{l,k}$ which rows are $(p^m_{l,k})^*$. We denote by $K^M_{l,k}$ the rank of $B^M_{l,k}$ ($K^M_{l,k}$ is random and lower than $K_{l,k}$).

Truncations. We have mentioned that $y^{N,R}_k$ and $z^{N,R}_{l,k}$ are respectively bounded by $C_y(R)$ and $C_z(R)$, and it is useful to force our approximations to be bounded in the same way. This is the role of the following truncations. For a function ψ, we define two new functions $[\psi]_y$ and $[\psi]_z$ by

$$[\psi]_y(x) = -C_y(R)\vee\psi(x)\wedge C_y(R), \quad [\psi]_z(x) = -C_z(R)\vee\psi(x)\wedge C_z(R),$$

which are bounded respectively by $C_y(R)$ and $C_z(R)$. Our approximation of $y^{N,R}_k$ (resp. $z^{N,R}_{l,k}$) will belong to the space $[\mathcal{P}_{0,k}]_y = \{[\alpha\cdot p_{0,k}]_y(\cdot), \alpha\in\mathbb{R}^{K_{0,k}}\}$ (resp. $[\mathcal{P}_{l,k}]_z = \{[\alpha\cdot p_{l,k}]_z(\cdot), \alpha\in\mathbb{R}^{K_{l,k}}\}$).

Constants. In the following, we denote by C any finite constant which value may change from line to line but which is independent on N, M, the functions bases and the vector R. It depends only on b, σ, β, λ, f, ϕ, T and x.

1.3 Description of the algorithm

The functions $(y^{N,R}_k)_{0\le k\le N-1}$ and $(z^{N,R}_{l,k})_{1\le l\le q,0\le k\le N-1}$ are approximated by $(y^{N,R,M}_k)_{0\le k\le N-1}$ and $(z^{N,R,M}_{l,k})_{1\le l\le q,0\le k\le N-1}$, which are built in a backward manner.

- → Initialization : for $k=N$ take $y^{N,R,M}_N(\cdot) = \phi^R(\cdot)$.
- → Iteration : for $k = N-1,\cdots,0$, solve the q least-squares problems :

$$\alpha^M_{l,k} = \arg\inf_\alpha \frac{1}{M}\sum_{m=1}^M |y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}})\frac{[\Delta W^m_{l,k}]_w}{h} - \alpha\cdot p^m_{l,k}|^2. \tag{7}$$

Then compute $\alpha^M_{0,k}$ as the minimizer of

$$\frac{1}{M}\sum_{m=1}^M |y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}}) + hf^R(t_k, X^{N,m}_{t_k}, y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}}), [\alpha^M_{l,k}\cdot p^m_{l,k}]_z) - \alpha\cdot p^m_{0,k}|^2. \tag{8}$$

Here, we use the shorter notation $f^R(t_k,x,y,z_l) = f^R(t_k,x,y,(z_l)_{1\le l\le q})$. Then we define $y^{N,R,M}_k(\cdot)$ and $z^{N,R,M}_{l,k}(\cdot)$ by

$$y^{N,R,M}_k(\cdot) = [\alpha^M_{0,k}\cdot p_{0,k}]_y(\cdot), \quad z^{N,R,M}_{l,k}(\cdot) = [\alpha^M_{l,k}\cdot p_{l,k}]_z(\cdot).$$

In the least-squares problems (7-8), whenever convenient we can suppose (as done for example in the proof of Theorem 11.1 in Györfi *et al.* 2002) that for $0 \le l \le q$, $p_{l,k}$ is a complete orthonormal system in $\mathcal{P}_{l,k}$, with respect to the empirical scalar product $< \cdot, \cdot >_{k,M}$ defined by

$$< \psi_1, \psi_2 >_{k,M} = \frac{1}{M} \sum_{m=1}^{M} \psi_1(X_{t_k}^{N,m}) \psi_2(X_{t_k}^{N,m}).$$

Of course these orthonormal systems depend on the simulations $(X_{t_k}^{N,m})_{1 \le m \le M}$ and their ranks $(K_{l,k}^M)_{0 \le l \le q}$ satisfy $K_{l,k}^M \le K_{l,k}$. These orthonormal systems can be easily computed using a Singular Value Decomposition (see for instance Golub and Van Loan 1996)). With this choice, $\frac{(B_{l,k}^M)^* B_{l,k}^M}{M} = \mathrm{Id}$ $(0 \le l \le q)$ and the solutions of (7-8) are given by:

$$\alpha_{l,k}^M = \frac{1}{M} \sum_{m=1}^{M} p_{l,k}^m \; y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}) \frac{[\Delta W_{l,k}^m]_w}{h},$$

$$\alpha_{0,k}^M = \frac{1}{M} \sum_{m=1}^{M} p_{0,k}^m \{ y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}) + h f^R(t_k, X_{t_k}^{N,m}, y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}), [\alpha_{l,k}^M \cdot p_{l,k}^m]_z) \}.$$

1.4 Results

Firstly, the following (easy) Proposition states that the couple $(Y^{N,R}, Z^{N,R})$ is bounded and satisfies a Lipschitz property.

Proposition 1 *Under* **(H1-H2-H3)**, *there exists a constant C such that $\forall k, 0 \le k \le N$:*

$$|Y_{t_k}^{N,R}| \le C_y(R) = C\{\|\phi^R\|_\infty + \|f^R\|_\infty\}, \qquad |Z_{l,t_k}^{N,R}| \le C_z(R) = \frac{C_y(R)}{\sqrt{h}},$$

where $\|\phi^R\|_\infty = \sup_x |\phi^R(x)| \le C(1 + |R|)$ and $\|f^R\|_\infty = \sup_{(t,x)} |f^R(t, x, 0, 0)| \le C(1 + |R|)$.
In addition under **(H4)**, *for h small enough, the functions $y_k^{N,R}$ and $z_k^{N,R}$ defined by $y_k^{N,R}(X_{t_k}^N) = Y_{t_k}^{N,R}$ and $z_k^{N,R}(X_{t_k}^N) = Z_{t_k}^{N,R}$ satisfy $|y_k^{N,R}(x) - y_k^{N,R}(x')| + \sqrt{h}|z_k^{N,R}(x) - z_k^{N,R}(x')| \le C|x - x'|$ uniformly in k_0, N and R.*

Secondly, we state an error bound regarding the localization.

Proposition 2 *Under* **(H1-H2-H3-H4)**, *there exists a constant C such that for h small enough, one has*

$$\max_{0 \le k \le N} \mathbb{E}|Y_{t_k}^{N,R} - Y_{t_k}^N|^2 + h\mathbb{E} \sum_{k=0}^{N-1} |Z_{t_k}^{N,R} - Z_{t_k}^N|^2$$
$$\le C\mathbb{E}|\phi(X_{t_N}^N) - \phi^R(X_{t_N}^N)|^2 + C \frac{C_y(R)^2}{h} \sum_{k=0}^{N-1} \mathbb{E}\big(|\Delta W_k|^2 \mathbf{1}_{|\Delta W_k| \ge R_0\sqrt{h}}\big)$$

$$+ Ch\mathbb{E}\sum_{k=0}^{N-1}|f(t_k,X^N_{t_k},Y^N_{t_{k+1}},Z^N_{t_k})-f^R(t_k,X^N_{t_k},Y^N_{t_{k+1}},Z^N_{t_k})|^2.$$

As a consequence and since $|Y^N_{t_k}|+\sqrt{h}|Z^N_{t_k}|\leq C(1+|X^N_{t_k}|)$ (using the arguments of Proposition 1, see also Gobet *et al.* 2005), we easily get that

$$\begin{aligned}&\max_{0\leq k\leq N}\mathbb{E}|Y^{N,R}_{t_k}-Y^N_{t_k}|^2+h\mathbb{E}\sum_{k=0}^{N-1}|Z^{N,R}_{t_k}-Z^N_{t_k}|^2\\ \leq&\frac{C}{h}\max_{0\leq k\leq N}\mathbb{E}\left[(1+|X^N_{t_k}|^2)\mathbf{1}_{|X^N_{t_k}|>|R|}\right]+C\frac{1+|R|^2}{h}\exp(-R_0^2/4).\end{aligned}$$

Hence for appropriate thresholds R going to infinity, the localization error converges to 0. Rates are available if in addition $\sup_{0\leq k\leq N}\mathbb{E}|X^N_{t_k}|^p\leq C_p(1+|x|^p)$ for some $p>2$ (stronger moment conditions on the Lévy measure λ would lead to larger exponents p): indeed the upper bound becomes $\frac{C_p}{h(1+|R|)^{p-2}}+C\frac{1+|R|^2}{h}\exp(-R_0^2/4)$ and to get a contribution of order h (as in Theorem 1), it is enough to asymptotically set $R_i=h^{-2/(p-2)}$ $(i=1,\cdots,d)$ and $R_0=c\sqrt{\log(1/h)}$ (for c large enough). Hence, the convergence with respect to R is rather fast, especially if p can be taken large. In other words, setting R to a fixed large value gives a very good approximation, as it will be later observed in the numerical experiments.

The error on the unknown regression functions $(y^{N,R}_k)_{0\leq k\leq N-1}$ and $(z^{N,R}_{l,k})_{1\leq l\leq q,0\leq k\leq N-1}$ is now estimated in the following Theorem, which is our main result.

Theorem 2 *Assume* **(H1-H2-H3-H4)** *and let* $\beta\in]0,1]$*. Then, there exists a constant* C *(independent on* β*) such that:*

$$\begin{aligned}&\max_{0\leq k\leq N}\mathbb{E}\frac{1}{M}\sum_{m=1}^{M}|y^{N,R}_k(X^{N,m}_{t_k})-y^{N,R,M}_k(X^{N,m}_{t_k})|^2\\ &+h\mathbb{E}\sum_{k=0}^{N-1}\frac{1}{M}\sum_{m=1}^{M}|z^{N,R}_k(X^{N,m}_{t_k})-z^{N,R,M}_k(X^{N,m}_{t_k})|^2\\ \leq&C\frac{C_y(R)^2}{M}\sum_{k=0}^{N-1}\sum_{l=0}^{q}\mathbb{E}(K^M_{l,k})+Ch^\beta\\ &+C\sum_{k=0}^{N-1}\{\inf_\alpha\mathbb{E}|y^{N,R}_k(X^N_{t_k})-\alpha\cdot p_{0,k}(X^N_{t_k})|^2+\sum_{l=1}^{q}\inf_\alpha\mathbb{E}|\sqrt{h}z^{N,R}_{l,k}(X^N_{t_k})-\alpha\cdot p_{l,k}(X^N_{t_k})|^2\}\\ &+C\frac{C_y(R)^2}{h}\sum_{k=0}^{N-1}\Big\{\mathbb{E}\big(K^M_{0,k}\exp(-\frac{Mh^{\beta+2}}{72C_y(R)^2K^M_{0,k}})\exp(CK_{0,k+1}\log\frac{C\ C_y(R)(K^M_{0,k})^{\frac{1}{2}}}{h^{\frac{\beta+2}{2}}}))\end{aligned}$$

$$+ h\mathbb{E}\big(K_{l,k}^M \exp(-\frac{Mh^{\beta+1}}{72C_y(R)^2R_0^2K_{l,k}^M}) \exp(CK_{0,k+1}\log\frac{C\ C_y(R)R_0(K_{l,k}^M)^{\frac{1}{2}}}{h^{\frac{\beta+1}{2}}}))$$
$$+ \exp(CK_{0,k}\log\frac{C\ C_y(R)}{h^{\frac{\beta+2}{2}}}) \exp(-\frac{Mh^{\beta+2}}{72C_y(R)^2})\Big\}.$$

Remark 2 *Using standard techniques of covering of functions classes (see for example the proof of Theorem 11.3 in Györfi* et al. *2002), we can state error estimates related to the law of X^N instead of the empirical law of $(X^{N,m})_{1\leq m\leq M}$, i.e. we can bound* $\max_{0\leq k\leq N}\mathbb{E}|y_k^{N,R}(X_{t_k}^N) - y_k^{N,R,M}(X_{t_k}^N)|^2 + h\mathbb{E}\sum_{k=0}^{N-1}|z_k^{N,R}(X_{t_k}^N) - z_k^{N,R,M}(X_{t_k}^N)|^2$*. This extension is valid if we add to the upper bound a term $CC_y(R)^2\frac{\log(M)}{M}\sum_{k=0}^{N-1}\sum_{l=0}^{q}K_{l,k}$, which is essentially of same order that the other ones (up to the log factor).*

Remark 3 *Of course, the inequality $K_{l,k}^M \leq K_{l,k}$ leads to simpler but rougher estimates; however, we think that in many cases it is possible to take advantage of better estimates on the law of $K_{l,k}^M$. This will be investigated in future works.*

The terms $\frac{C_y(R)^2}{M}\sum_{l=0}^{q}\mathbb{E}(K_{l,k}^M)$ and $\inf_\alpha\mathbb{E}\big(|y_k^{N,R}(X_{t_k}^N) - \alpha\cdot p_{0,k}(X_{t_k}^N)|^2\big) + \sum_{l=1}^{q}\inf_\alpha\mathbb{E}\big(|\sqrt{h}z_{l,k}^{N,R}(X_{t_k}^N) - \alpha\cdot p_{l,k}(X_{t_k}^N)|^2\big)\}$ are classic error terms which arise when one approximates a regression function from i.i.d. observations using projections on a finite functions basis (Györfi *et al.* 2002). They are summed up from $k=0$ until $k=N-1$ because we make N estimations, one at each time t_k. The other terms come from the lack of independence between the different estimation problems at each discretization time. From the contribution h^β, we understand why $\beta>0$ is necessary to ensure that the error tends to 0 and why $\beta>1$ is unnecessary because it gives a negligible term compared to h arising in Theorem 1.
Presumably, the optimal parameter β is equal to 1 (see in the next paragraph the discussion on the trade-off between complexity and accuracy). But regarding the exponential contributions, we see that allowing β to be close to 0 is less stringent than $\beta=1$ on the values of M for which the convergence holds.
Theorem 2 improves Theorem 3 in Gobet *et al.* (2005) because the error is not estimated in terms of the high moments of the orthonormalized basis functions, but directly in terms of the number of functions that are used in the algorithm. This result can therefore be easily used in practice.

1.5 Accuracy and complexity of the algorithm

Now, we can derive from Theorem 2 how to make N, M and the number of basis functions vary together. In this discussion, we neglect the influence of the localization parameter R, which is supposed to be large enough from the beginning (see the comments after Proposition 2). As already observed in Gobet *et al.* (2005), local basis functions enable us to take advantage of the Lipschitz property of functions $y_k^{N,R}$ and $\sqrt{h}z_{l,k}^{N,R}$. Let us consider the simplest example of a local function basis, i.e. the hypercubes basis, already used in Gobet *et al.* (2005) and still denoted **HC** here. To simplify, $p_{l,k}$ does not depend on l or k and its size is denoted by K. Choose a domain $D \subset \mathbb{R}^d$ centered on x, that is $D = \prod_{i=1}^d]x_i - a, x_i + a]$, and partition it into small hypercubes of edge δ. Thus, $D = \cup_{i_1,\cdots,i_d} D_{i_1,\cdots,i_d}$ where $D_{i_1,\cdots,i_d} =]x_1 - a + i_1\delta, x_1 - a + (i_1 + 1)\delta] \times \cdots \times]x_d - a + i_{d'}\delta, x_d - a + (i_d + 1)\delta]$. Then we define $p_{l,k}$ as the indicator functions associated to this set of hypercubes: $p_{l,k}(\cdot) = \big(\mathbf{1}_{D_{i_1,\cdots,i_d}}(\cdot)\big)_{i_1,\cdots,i_d}$. With this particular choice of function bases, we can make the projection error of Theorem 2 explicit and refer to Gobet *et al.* (2005) for details:

$$\inf_\alpha \mathbb{E}\big(|y_k^{N,R}(X_{t_k}^N) - \alpha \cdot p_{0,k}(X_{t_k}^N)|^2\big) \leq \ C\{\delta^2 + C_y(R)^2\mathbb{P}(X_{t_k}^N \in D^c)\}.$$

As for the impact of the localization parameter R, $\mathbb{P}(X_{t_k}^N \in D^c)$ becomes negligible with respect to the other errors if we choose D big enough (a feature which is confirmed by the next numerical experiments). Thus and as far the projection errors are concerned, to get a global (squared) error of order h^β we have to choose $\delta \approx h^{\frac{\beta+1}{2}}$, or equivalently a number of basis functions $K \approx h^{-\frac{d(\beta+1)}{2}}$ (considering a fixed domain D). Regarding now the number of simulations M, to avoid an explosive upper bound in Theorem 2, one should take $M \approx Ch^{-d(\beta+1)-(\beta+2)}\log(h^{-\frac{d(\beta+1)}{4}-\frac{\beta+1}{2}})$ for a constant C large enough (here, the ranks $K_{l,k}^M$ have simply been upper bounded by K).
The dominant term of the algorithm's complexity $\mathcal{C}$ associated to this choice of function basis is $\mathcal{C} = NMd\log(K)$, which corresponds to determine in which cells the simulation fall (this is the cost of a nearest neighbour algorithm in a tensored grid, i.e. $O(d\log(K))$ for one path at a given time). Hence, up to logarithmic factors, the complexity equals $\mathcal{C} = O(h^{-1-d(\beta+1)-(\beta+2)})$, while the squared error is of order $h^\beta = O(\mathcal{C}^{-\beta/(2+(\beta+1)(d+1))})$. The optimal value of $\beta \in]0,1]$ is achieved for $\beta = 1$, for which the squared error is of order $h = O(\mathcal{C}^{-1/(2d+4)})$.
It is now interesting to compare with the complexity of the algorithm presented in Bouchard and Touzi (2004). We compute the complexity of their algorithm in the most favourable case where X is a geometric Brownian motion; otherwise the algorithm is more difficult to implement and its complexity more difficult to evaluate because of the necessary calculation of Skorohod's integrals.
In this algorithm, one needs N independent sets $(\mathcal{M}_k)_{1\leq k\leq N}$ of M simulated paths of X^N (one set for each discretization time). At each discretization time t_k and for each path in the set $\mathcal{M}_k$, a calculation involving the M paths of the set $\mathcal{M}_{k+1}$ is performed. This leads to a complexity $\mathcal{C} = O(NM^2)$. The squared error associated to this complexity is given by Theorem 6.2 in Bouchard and Touzi (2004) and is of order $\frac{1}{N} + \frac{N^{2+\frac{d}{4}}}{\sqrt{M}}$. Expressing the squared error as a function of the complexity, we find $\mathcal{C}^{-\frac{1}{13+d}}$.

Thus, in the case of the geometric Brownian motion model, our algorithm is more efficient for $d \leq 9$ and less efficient otherwise. But in the case of a general model, the complexity of the algorithm presented in Bouchard and Touzi (2004) is really difficult to evaluate whereas in our case the complexity is independent of the model.
Finally, we present in Section 3 a slight modification of our algorithm, which gives a far better trade-off complexity/accuracy for all models and all dimension d.

2 Proofs

The section is devoted to the proofs of the results announced in Section 1.

2.1 Influence of the localization: proof of Proposition 2

We first estimate the localization error. Subtracting (2) from (5) and applying Young's inequality, we get for $0 \leq k \leq N-1$ and for all real $\gamma > 0$ (the value of γ will be chosen later)

$$\begin{aligned}|Y^{N,R}_{t_k} - Y^N_{t_k}|^2 \leq (1+\gamma h)|\mathbb{E}_{t_k}(Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}})|^2 \\ +C(h^2+\frac{h}{\gamma})\mathbb{E}_{t_k}|f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k}) - f^R(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k})|^2 \\ +C(h^2+\frac{h}{\gamma})\mathbb{E}_{t_k}|Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}|^2 + C(h^2+\frac{h}{\gamma})|Z^{N,R}_{t_k} - Z^N_{t_k}|^2. \qquad (9)\end{aligned}$$

Subtracting (3) from (6), it comes for $1 \leq l \leq q$:

$$Z^{N,R}_{l,t_k} - Z^N_{l,t_k} = \frac{1}{h}\mathbb{E}_{t_k}\big(\{Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}\}\Delta W_{l,k}\big) + \frac{1}{h}\mathbb{E}_{t_k}\big(Y^{N,R}_{t_{k+1}}\{[\Delta W_{l,k}]_w - \Delta W_{l,k}\}\big).$$

Applying the inequality $(a+b)^2 \leq 2(a^2+b^2)$, the Cauchy-Schwarz inequality to the first term of the right hand side and $|Y^{N,R}_{t_{k+1}}| \leq C_y(R)$ (see Proposition 1) we get:

$$\begin{aligned}|Z^{N,R}_{l,t_k} - Z^N_{l,t_k}|^2 \leq \frac{2}{h}\{\mathbb{E}_{t_k}\big(|Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}|^2\big) - \big(\mathbb{E}_{t_k}\{Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}\}\big)^2\} \\ + 2\frac{C_y(R)^2}{h^2}\mathbb{E}_{t_k}(|\Delta W_{l,k}|^2 \mathbf{1}_{|\Delta W_k| \geq R_0\sqrt{h}}). \qquad (10)\end{aligned}$$

Plugging this into (9), we finally obtain

$$\begin{aligned}\mathbb{E}|Y^{N,R}_{t_k} - Y^N_{t_k}|^2 \leq (1+\gamma h)\mathbb{E}|\mathbb{E}_{t_k}(Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}})|^2 + C(h^2+\frac{h}{\gamma})\mathbb{E}|Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}|^2 \\ + C(h^2+\frac{h}{\gamma})\mathbb{E}|f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k}) - f^R(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k})|^2\end{aligned}$$

$$+ C(h+\frac{1}{\gamma})\{\mathbb{E}\big(|Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}|^2\big) - \mathbb{E}\big(\mathbb{E}_{t_k}\{Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}\}\big)^2\}$$
$$+ \frac{CC_y(R)^2}{h}(h+\frac{1}{\gamma})\mathbb{E}(|\Delta W_k|^2 \mathbf{1}_{|\Delta W_k|\geq R_0\sqrt{h}}).$$

Now we take $\gamma = C$ and we obtain the following simplification:

$$\begin{aligned}
&\mathbb{E}|Y^{N,R}_{t_k} - Y^N_{t_k}|^2 \\
&\leq (1+Ch)\mathbb{E}|Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}|^2 + \frac{CC_y(R)^2}{h}\mathbb{E}(|\Delta W_k|^2 \mathbf{1}_{|\Delta W_k|\geq R_0\sqrt{h}}) \\
&\quad + Ch\mathbb{E}|f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k}) - f^R(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k})|^2.
\end{aligned}$$

An application of the discrete Gronwall lemma leads to the expected result for the difference between $Y^{N,R}$ and Y^N. For the difference between $Z^{N,R}$ and Z^N, we use (10) to get

$$\begin{aligned}
&h\mathbb{E}\sum_{k=0}^{N-1}|Z^{N,R}_{t_k} - Z^N_{t_k}|^2 \\
&\leq C\sum_{k=0}^{N-1}\{\mathbb{E}\big(|Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}|^2\big) - \mathbb{E}\big(\mathbb{E}_{t_k}\{Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}\}\big)^2\} \\
&\quad + \frac{CC_y(R)^2}{h}\sum_{k=0}^{N-1}\mathbb{E}(|\Delta W_k|^2 \mathbf{1}_{|\Delta W_k|\geq R_0\sqrt{h}}) \\
&\leq C\mathbb{E}|\phi^R(X^N_{t_N}) - \phi(X^N_{t_N})|^2 + \frac{CC_y(R)^2}{h}\sum_{k=0}^{N-1}\mathbb{E}(|\Delta W_k|^2 \mathbf{1}_{|\Delta W_k|\geq R_0\sqrt{h}}) \\
&\quad + C\sum_{k=0}^{N-1}\{\mathbb{E}\big(|Y^{N,R}_{t_k} - Y^N_{t_k}|^2\big) - \mathbb{E}\big(\mathbb{E}_{t_k}\{Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}\}\big)^2\}. \qquad (11)
\end{aligned}$$

Taking the expectation in (9) leads to ($\forall\gamma > 0$)

$$\begin{aligned}
&\mathbb{E}|Y^{N,R}_{t_k} - Y^N_{t_k}|^2 - \mathbb{E}|\mathbb{E}_{t_k}(Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}})|^2 \\
&\leq \big(\gamma h + C(h^2+\frac{h}{\gamma})\big)\mathbb{E}|Y^{N,R}_{t_{k+1}} - Y^N_{t_{k+1}}|^2 + C(h^2+\frac{h}{\gamma})\mathbb{E}|Z^{N,R}_{t_k} - Z^N_{t_k}|^2 \\
&\quad + C(h^2+\frac{h}{\gamma})\mathbb{E}|f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k}) - f^R(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k})|^2.
\end{aligned}$$

Plug this inequality into (11) and take $\gamma = 3C^2$ to get the result for $h\mathbb{E}\sum_{k=0}^{N-1}|Z^{N,R}_{t_k} - Z^N_{t_k}|^2$, for h small enough. □

2.2 Influence of the localization: proof of Proposition 1

The aim is to prove that $(Y^{N,R}, Z^{N,R})$ is bounded. Applying Young's inequality to (5), we get using the Lipschitz property of f^R $(\forall \gamma > 0)$

$$\begin{aligned}|Y^{N,R}_{t_k}|^2 \le (1+\gamma h)|\mathbb{E}_{t_k}(Y^{N,R}_{t_{k+1}})|^2 + Ch^2(1+\frac{1}{\gamma h})\mathbb{E}_{t_k}|Y^{N,R}_{t_{k+1}}|^2 \\ + Ch^2(1+\frac{1}{\gamma h})|Z^{N,R}_{t_k}|^2 + Ch^2(1+\frac{1}{\gamma h})|f^R(t_k, X^N_{t_k}, 0, 0)|^2.\end{aligned}$$

Using

$$h|Z^{N,R}_{l,t_k}|^2 \le \{\mathbb{E}_{t_k}|Y^{N,R}_{t_{k+1}}|^2 - |\mathbb{E}_{t_k}(Y^{N,R}_{t_{k+1}})|^2\} \tag{12}$$

and adjusting γ in the same way than for Proposition 2, we get $|Y^{N,R}_{t_k}|^2 \le (1+Ch)\mathbb{E}_{t_k}|Y^{N,R}_{t_{k+1}}|^2 + Ch|f^R(t_k, X^N_{t_k}, 0, 0)|^2$. We get the result for $Y^{N,R}$ by applying the discrete Gronwall lemma and deduce the result for $Z^{N,R}$ thanks to (12).
The Lipschitz property can be established using analogous techniques (see also the proof of Proposition 3 in Gobet *et al.* 2005): for this, the assumption **(H4)**a) is crucial. □

2.3 Proof of Theorem 2

The proof is technical and we divide it into several steps. Firstly, we introduce additional notations, which are closely related to non-parametric regression arguments. Secondly, we state a result concerning the propagation of the error from time t_{k+1} to time t_k (see Proposition 3). Finally, the different contributions in the propagation error are evaluated in Proposition 4.

2.3.1 Extra notations for the proofs

Monte-Carlo simulations. We recall that the algorithm uses M Monte-Carlo simulations of the Brownian increments ΔW and of an approximation X^N of X, $(X^N_{t_k})_k$ being a Markov chain. In addition to $(X^{N,m}_{t_{k+1}}, \Delta W^m_k)$ and for the proofs, we will use at each time t_k extra random variables $(\tilde{X}^{N,m}_{t_{k+1}}, \Delta\tilde{W}^m_k)$ which are, conditionally to $X^{N,m}_{t_k}$, an independent copy of $(X^{N,m}_{t_{k+1}}, \Delta W^m_k)$ (and also independent of everything else). For instance, when X has no jump part $(\beta \equiv 0)$ and an Euler scheme is used for X^N, $X^{N,m}_{t_{k+1}}$ and $\tilde{X}^{N,m}_{t_{k+1}}$ are defined by

$$\begin{aligned}X^{N,m}_{t_{k+1}} &= X^{N,m}_{t_k} + b(t_k, X^{N,m}_{t_k})h + \sigma(t_k, X^{N,m}_{t_k})\Delta W^m_k,\\ \tilde{X}^{N,m}_{t_{k+1}} &= X^{N,m}_{t_k} + b(t_k, X^{N,m}_{t_k})h + \sigma(t_k, X^{N,m}_{t_k})\Delta\tilde{W}^m_k,\end{aligned}$$

where $(\Delta W^m_k)_{k,m}$ and $(\Delta\tilde{W}^m_k)_{k,m}$ are i.i.d.

Norms. For a function ψ, we define $(0 \leq k \leq N)$:

$$\|\psi\|_{k,M}^2 = \frac{1}{M}\sum_{m=1}^{M} |\psi(X_{t_k}^{N,m})|^2, \qquad \|\psi\|_{k,\tilde{M}}^2 = \frac{1}{M}\sum_{m=1}^{M} |\psi(\tilde{X}_{t_k}^{N,m})|^2.$$

Projection coefficients. Remind that coefficients $(\alpha_{l,k}^M)_{0\leq l\leq q}$ are defined by:

$$\alpha_{l,k}^M = \arg\inf_{\alpha} \frac{1}{M}\sum_{m=1}^{M} |y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m})\frac{[\Delta W_{l,k}^m]_w}{h} - \alpha\cdot p_{l,k}^m|^2$$

and $\alpha_{0,k}^M$ as the minimizer of

$$\frac{1}{M}\sum_{m=1}^{M} |y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}) + hf^R(t_k, X_{t_k}^{N,m}, y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}), [\alpha_{l,k}^M\cdot p_{l,k}^m]_z) - \alpha\cdot p_{0,k}^m|^2.$$

We will need other coefficients in the proofs below. Thus, we define the projection coefficients $(\tilde{\alpha}_{l,k}^M)_{1\leq l\leq q}$

$$\tilde{\alpha}_{l,k}^M = \arg\inf_{\alpha} \frac{1}{M}\sum_{m=1}^{M} |y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m})\frac{[\Delta \tilde{W}_{l,k}^m]_w}{h} - \alpha\cdot p_{l,k}^m|^2, \tag{13}$$

$\tilde{\alpha}_{0,k}^M$ as the minimizer of

$$\frac{1}{M}\sum_{m=1}^{M} |y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}) + hf^R(t_k, X_{t_k}^{N,m}, y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}), [\tilde{\alpha}_{l,k}^M\cdot p_{l,k}^m]_z) - \alpha\cdot p_{0,k}^m|^2, \tag{14}$$

the projection coefficients $(\tilde{\beta}_{l,k}^M)_{1\leq l\leq q}$

$$\tilde{\beta}_{l,k}^M = \arg\inf_{\alpha} \frac{1}{M}\sum_{m=1}^{M} |y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})\frac{[\Delta \tilde{W}_{l,k}^m]_w}{h} - \alpha\cdot p_{l,k}^m|^2, \tag{15}$$

and $\tilde{\beta}_{0,k}^M$ as the minimizer of

$$\frac{1}{M}\sum_{m=1}^{M} |y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m}) + hf^R(t_k, X_{t_k}^{N,m}, y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m}), z_{l,k}^{N,R}(X_{t_k}^{N,m})) - \alpha\cdot p_{0,k}^m|^2. \tag{16}$$

We emphasize the differences between these projection coefficients: for $(\alpha_{l,k}^M)_{0\leq l\leq q}$ and $(\tilde{\alpha}_{l,k}^M)_{0\leq l\leq q}$, the function $y_{k+1}^{N,R,M}$ is fixed and we estimate $\alpha_{l,k}^M$ from $(X_{t_k}^{N,m}, X_{t_{k+1}}^{N,m}, \Delta W_k^m)_{1\leq m\leq M}$ whereas we estimate $\tilde{\alpha}_{l,k}^M$ from $(X_{t_k}^{N,m}, \tilde{X}_{t_{k+1}}^{N,m}, \Delta \tilde{W}_k^m)_{1\leq m\leq M}$. As for $(\tilde{\beta}_{l,k}^M)_{0\leq l\leq q}$, we also estimate from

$(X_{t_k}^{N,m}, \tilde{X}_{t_{k+1}}^{N,m}, \Delta\tilde{W}_k^m)_{1\le m\le M}$ but knowing the true functions $y_{k+1}^{N,R}(\cdot)$ and $z_k^{N,R}(\cdot)$. We also note that $\alpha_{0,k}^M = \alpha_{0,k}^{1,M} + \alpha_{0,k}^{2,M}$ where

$$\alpha_{0,k}^{1,M} = \arg\inf_\alpha \frac{1}{M}\sum_{m=1}^M |y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}) - \alpha\cdot p_{0,k}^m|^2,$$

$$\alpha_{0,k}^{2,M} = \arg\inf_\alpha \frac{1}{M}\sum_{m=1}^M |hf^R(t_k, X_{t_k}^{N,m}, y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}), [\alpha_{l,k}^M\cdot p_{l,k}^m]_z) - \alpha\cdot p_{0,k}^m|^2.$$

We define $(\tilde{\beta}_{0,k}^{i,M})_{1\le i\le 2}$ and $(\tilde{\alpha}_{0,k}^{i,M})_{1\le i\le 2}$ in the same way.

Conditional expectations. We write $\mathcal{F}^M$ for the σ-algebra generated by $((X_{t_k}^{N,m})_{0\le k\le N}, (\Delta W_k^m)_{0\le k\le N-1})_{1\le m\le M}$ and $\mathbb{E}^M$ for the conditional expectation with respect to $\mathcal{F}^M$. We denote by $\mathbb{E}_k^M$ (resp. $\mathbb{P}_k^M$) the conditional expectation (resp. conditional probability) with respect to the σ-field generated by $((X_{t_i}^{N,m})_{0\le i\le k}, (\Delta W_i^m)_{0\le i\le k-1})_{1\le m\le M}$.

Error terms. According to β and the projection coefficients, we define the following events which probabilities are evaluated later in Proposition 4:

$$\begin{aligned}
A_{0,k}^M =& \{\frac{1}{M}\sum_{m=1}^M |p_{0,k}^m.(\alpha_{0,k}^{1,M} - \tilde{\alpha}_{0,k}^{1,M})|^2 < h^{\beta+2}\},\\
A_{l,k}^M =& \{\frac{1}{M}\sum_{m=1}^M |p_{l,k}^m.(\alpha_{l,k}^M - \tilde{\alpha}_{l,k}^M)|^2 < h^\beta\},\\
A_k^M =& \{\forall\psi\in[\mathcal{P}_{0,k}]_y - y_k^{N,R} : \|\psi\|_{k,\tilde{M}} - \|\psi\|_{k,M} < h^{\frac{\beta+2}{2}}\}.
\end{aligned}$$

We also deal with the following quantities, which are bounded in Proposition 4 as well:

$$\begin{aligned}
T_{1,k}^M &= \mathbb{E}\|\mathbb{E}^M(\tilde{\beta}_{0,k}^M)\cdot p_{0,k} - y_k^{N,R}\|_{k,M}^2,\\
T_{2,k}^M &= \mathbb{E}\|\{\tilde{\alpha}_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\alpha}_{0,k}^{1,M})\}\cdot p_{0,k}\|_{k,M}^2,\\
T_{3,l,k}^M &= \mathbb{E}\|\mathbb{E}^M(\tilde{\beta}_{l,k}^M)\cdot p_{l,k} - z_{l,k}^{N,R}\|_{k,M}^2,\\
T_{4,l,k}^M &= \mathbb{E}\|\{\tilde{\alpha}_{l,k}^M - \mathbb{E}^M(\tilde{\alpha}_{l,k}^M)\}\cdot p_{l,k}\|_{k,M}^2.
\end{aligned}$$

Covering numbers. In the proofs below we use random covering numbers. We refer to Györfi *et al.* (2002) for a complete description. However for sake of completeness, we briefly recall here that if $\mathcal{G}$ is a class of functions and $x_1^M = (x_1,\ldots,x_M)$ are M points of $\mathbb{R}^d$, $\mathcal{N}_2(\epsilon,\mathcal{G},x_1^M)$ ($\epsilon > 0$) is the minimal $p\in\mathbb{N}$ such that there exist functions $g_1,\cdots,g_p$, such that for all $g\in\mathcal{G}$ we can find a $j\in\{1,\cdots,p\}$ with

$$\Big(\frac{1}{M}\sum_{m=1}^M |g(x_m) - g_j(x_m)|^2\Big)^{\frac{1}{2}} < \epsilon.$$

To simplify, we adopt the notation

$$\mathcal{N}_2(\epsilon, k) = \mathcal{N}_2(\epsilon, [\mathcal{P}_{0,k}]_y, (X_{t_k}^{N,m}, \tilde{X}_{t_k}^{N,m})_{1\le m\le M}).$$

2.3.2 Propagation of the error

Our main tool is the following result.

Proposition 3 *Under the assumptions of Theorem 2, for* $0 \le k \le N-1$ *one has*

$$\begin{aligned}
&\mathbb{E}\|y_k^{N,R} - y_k^{N,R,M}\|_{k,M}^2 \\
&\le (1+Ch)\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}^2 + C\{T_{1,k}^M + T_{2,k}^M\} + Ch\sum_{l=1}^{q}\{T_{3,l,k}^M + T_{4,l,k}^M\} \\
&\quad + C\frac{C_y(R)^2}{h}\{\mathbb{P}([A_{0,k}^M]^c) + h\sum_{l=1}^{q}\mathbb{P}([A_{l,k}^M]^c) + \mathbb{P}([A_{k+1}^M]^c)\} + Ch^{\beta+1}.
\end{aligned}$$

If we admit for a while this result and Proposition 4 which estimates each contribution, it is easy to complete the error's estimation on Y in Theorem 2. Then, the same calculations we made in Proposition 2 enable us to deduce an error's estimate on Z from that on Y.

Proof of Proposition 3. First by using $[y_k^{N,R}]_y = y_k^{N,R}$ and that $[\cdot]_y$ is 1-Lipschitz, we get

$$\mathbb{E}\|y_k^{N,R,M} - y_k^{N,R}\|_{k,M}^2 \le \mathbb{E}\|\alpha_{0,k}^M \cdot p_{0,k} - y_k^{N,R}\|_{k,M}^2.$$

Now we introduce $\tilde{\beta}_{0,k}^M$ (see (16)) and noting that $\mathbb{E}^M(\tilde{\beta}_{0,k}^M)$ is the minimizer of $\frac{1}{M}\sum_{m=1}^M |y_k^{N,R}(X_{t_k}^{N,m}) - \alpha \cdot p_{0,k}^m|^2$ we apply the Pythagoras Theorem to get

$$\mathbb{E}\|\alpha_{0,k}^M \cdot p_{0,k} - y_k^{N,R}\|_{k,M}^2 = \mathbb{E}\|\{\alpha_{0,k}^M - \mathbb{E}^M(\tilde{\beta}_{0,k}^M)\}\cdot p_{0,k}\|_{k,M}^2 + T_{1,k}^M.$$

Now as $\mathbb{E}^M(\alpha_{0,k}^M) = \alpha_{0,k}^M$ and $\alpha_{0,k}^M = \alpha_{0,k}^{1,M} + \alpha_{0,k}^{2,M}$ we first apply Young's inequality, then Jensen's inequality to get

$$\begin{aligned}
\mathbb{E}\|\{\alpha_{0,k}^M - \mathbb{E}^M(\tilde{\beta}_{0,k}^M)\}\cdot p_{0,k}\|_{k,M}^2 &\le (1+\gamma h)\mathbb{E}\|\{\alpha_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\beta}_{0,k}^{1,M})\}\cdot p_{0,k}\|_{k,M}^2 \\
&\quad + (1+\frac{1}{\gamma h})\mathbb{E}\|\{\alpha_{0,k}^{2,M} - \mathbb{E}^M(\tilde{\beta}_{0,k}^{2,M})\}\cdot p_{0,k}\|_{k,M}^2 \\
&\le (1+\gamma h)\mathbb{E}\|\{\alpha_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\beta}_{0,k}^{1,M})\}\cdot p_{0,k}\|_{k,M}^2 \\
&\quad +(1+\frac{1}{\gamma h})\mathbb{E}\|\{\alpha_{0,k}^{2,M} - \tilde{\beta}_{0,k}^{2,M}\}\cdot p_{0,k}\|_{k,M}^2. \qquad (17)
\end{aligned}$$

We deal separately with the two terms of the right-hand side of (17). For the first term, we introduce $\tilde{\alpha}_{0,k}^{1,M}$ (see (14)) and use the definition of $A_{0,k}^M$ and $T_{2,k}^M$ to get:

$$\begin{aligned}
&\mathbb{E}\|\{\alpha_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\beta}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^2 \\
&\leq (1+h^{-1})\mathbb{E}\|\{\alpha_{0,k}^{1,M} - \tilde{\alpha}_{0,k}^{1,M}\} \cdot p_{0,k}\|_{k,M}^2 + (1+h)\mathbb{E}\|\{\tilde{\alpha}_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\beta}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^2 \\
&\leq Ch^{\beta+1} + \frac{C}{h}C_y(R)^2\mathbb{P}\big([A_{0,k}^M]^c\big) + (1+h)\bigg(\mathbb{E}\|\{\tilde{\alpha}_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\alpha}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^2 \\
&\qquad + \mathbb{E}\|\{\mathbb{E}^M(\tilde{\alpha}_{0,k}^{1,M}) - \mathbb{E}^M(\tilde{\beta}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^2\bigg) \\
&\leq Ch^{\beta+1} + \frac{C}{h}C_y(R)^2\mathbb{P}\big([A_{0,k}^M]^c\big) + (1+h)T_{2,k}^M \\
&\qquad + (1+h)\mathbb{E}\frac{1}{M}\sum_{m=1}^M |\mathbb{E}^M\{y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}) - y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})\}|^2. \qquad (18)
\end{aligned}$$

At the second and third inequalities, we have used the contraction property of the projection on $\big((p_{0,k}^m)^*\big)_{1\leq m\leq M}$. Using once more this contraction property, we get for the second term of the right hand side of (17):

$$\begin{aligned}
&\mathbb{E}\|\{\alpha_{0,k}^{2,M} - \tilde{\beta}_{0,k}^{2,M}\} \cdot p_{0,k}\|_{k,M}^2 \qquad (19) \\
\leq\ & \frac{Ch^2}{M}\mathbb{E}\sum_{m=1}^M |y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}) - y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})|^2 + Ch^2\mathbb{E}\|z_k^{N,R,M} - z_k^{N,R}\|_{k,M}^2.
\end{aligned}$$

Let us deal with the last term of the right hand side of (19). For $1 \leq l \leq q$ we can first write that

$$\mathbb{E}\|z_{l,k}^{N,R,M} - z_{l,k}^{N,R}\|_{k,M}^2 \leq \mathbb{E}\|\alpha_{l,k}^M \cdot p_{l,k} - z_{l,k}^{N,R}\|_{k,M}^2$$

as $[z_{l,k}^{N,R}]_z = z_{l,k}^{N,R}$. Next, we introduce $\tilde{\beta}_{l,k}^M$ (see (15)) and as $\mathbb{E}^M(\tilde{\beta}_{l,k}^M)$ is the minimizer of $\frac{1}{M}\sum_{m=1}^M |z_{l,k}^{N,R}(X_{t_k}^{N,m}) - \alpha \cdot p_{l,k}^m|^2$, we get

$$\begin{aligned}
&\mathbb{E}\|\alpha_{l,k}^M \cdot p_{l,k} - z_{l,k}^{N,R}\|_{k,M}^2 \\
=\ & \mathbb{E}\|\mathbb{E}^M(\tilde{\beta}_{l,k}^M).p_{l,k} - z_{l,k}^{N,R}\|_{k,M}^2 + \mathbb{E}\|\{\mathbb{E}^M(\tilde{\beta}_{l,k}^M) - \alpha_{l,k}^M\} \cdot p_{l,k}\|_{k,M}^2 \\
=\ & T_{3,l,k}^M + 3\mathbb{E}\|\{\mathbb{E}^M(\tilde{\beta}_{l,k}^M) - \mathbb{E}^M(\tilde{\alpha}_{l,k}^M)\} \cdot p_{l,k}\|_{k,M}^2 + 3\mathbb{E}\|\{\mathbb{E}^M(\tilde{\alpha}_{l,k}^M) - \tilde{\alpha}_{l,k}^M\} \cdot p_{l,k}\|_{k,M}^2 \\
&\qquad + 3\mathbb{E}\|\{\tilde{\alpha}_{l,k}^M - \alpha_{l,k}^M\} \cdot p_{l,k}\|_{k,M}^2 \\
\leq\ & T_{3,l,k}^M + 3\mathbb{E}\|\{\mathbb{E}^M(\tilde{\beta}_{l,k}^M) - \mathbb{E}^M(\tilde{\alpha}_{l,k}^M)\} \cdot p_{l,k}\|_{k,M}^2 + 3T_{4,l,k}^M + Ch^\beta \\
&\qquad + C\frac{C_y(R)^2}{h}\mathbb{P}([A_{l,k}^M]^c). \qquad (20)
\end{aligned}$$

Now, an application of the contraction property associated to the projection on $\big((p_{l,k}^m)^*\big)$ and of the Cauchy-Schwarz inequality yields

$$
\begin{aligned}
&\frac{3}{M}\mathbb{E}\sum_{m=1}^{M}|\{\mathbb{E}^M(\tilde{\beta}_{l,k}^M)-\mathbb{E}^M(\tilde{\alpha}_{l,k}^M)\}\cdot p_{l,k}^m|^2\\
\leq&\frac{3}{M}\mathbb{E}\sum_{m=1}^{M}|\mathbb{E}^M\{(y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}))\frac{[\Delta\tilde{W}_{l,k}^m]_w}{h}\}|^2\\
\leq&\frac{3}{Mh}\mathbb{E}\sum_{m=1}^{M}\{|y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m})|^2\\
&\qquad\qquad-|\mathbb{E}^M(y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}))|^2\}.
\end{aligned}
\tag{21}
$$

Putting (18), (19), (20) and (21) in (17) we get:

$$
\begin{aligned}
&\mathbb{E}\|y_k^{N,R}-y_k^{N,R,M}\|_{k,M}^2\\
\leq\ &T_{1,k}^M+(1+\gamma h)\bigg(Ch^{\beta+1}+\frac{C}{h}C_y(R)^2\mathbb{P}([A_{0,k}^M]^c)+(1+h)T_{2,k}^M\\
&+(1+h)\frac{1}{M}\mathbb{E}\sum_{m=1}^{M}|\mathbb{E}^M(y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}))|^2\bigg)\\
&+Ch^2(1+\frac{1}{\gamma h})\frac{1}{M}\mathbb{E}\sum_{m=1}^{M}|y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m})|^2\\
&+Ch^2(1+\frac{1}{\gamma h})\mathbb{E}\sum_{l=1}^{q}\{T_{3,l,k}^M+T_{4,l,k}^M\}\\
&+Ch^2(1+\frac{1}{\gamma h})h^\beta+Ch^2(1+\frac{1}{\gamma h})\frac{C}{h}C_y(R)^2\sum_{l=1}^{q}\mathbb{P}([A_{l,k}^M]^c)\\
&+C(h+\frac{1}{\gamma})\frac{1}{M}\mathbb{E}\sum_{m=1}^{M}\{|y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m})|^2\\
&\qquad\qquad-|\mathbb{E}^M(y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}))|^2\}.
\end{aligned}
$$

Taking now $\gamma=C$ we get the following simplification

$$
\begin{aligned}
&\mathbb{E}\|y_k^{N,R}-y_k^{N,R,M}\|_{k,M}^2\\
\leq&(1+Ch)\mathbb{E}\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,\tilde{M}}^2+C\frac{C_y(R)^2}{h}\big(\mathbb{P}([A_{0,k}^M]^c)+h\sum_{l=1}^{q}\mathbb{P}([A_{l,k}^M]^c)\big)\\
&+T_{1,k}^M+CT_{2,k}^M+Ch\sum_{l=1}^{q}\{T_{3,l,k}^M+T_{4,l,k}^M\}+Ch^{\beta+1}\\
&+Ch\mathbb{E}\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,M}^2+C\frac{h}{M}\mathbb{E}\sum_{m=1}^{M}|y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})-y_{k+1}^{N,R}(X_{t_{k+1}}^{N,m})|^2.
\end{aligned}
\tag{22}
$$

Since $y_{k+1}^{N,R}$ is Lipschitz continuous, the last term of the right hand side above is bounded by Ch^2 (here, we use **(H4)**b)). To get the result of Proposition 3, the first term of the right hand side should be changed in $\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}^2$. Thus, we use the definition of A_{k+1}^M to write:

$$\begin{aligned}
&\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,\tilde{M}}^2 \\
&= \mathbb{E}\big(\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,\tilde{M}} - \|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M} + \|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}\big)^2 \\
&\leq (1+h^{-1})\mathbb{E}\big(\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,\tilde{M}} - \|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}\big)_+^2 \\
&\quad + (1+h)\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}^2 \\
&\leq Ch^{\beta+1} + C\frac{C_y(R)^2}{h}\mathbb{P}([A_{k+1}^M]^c) + (1+h)\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}^2. \qquad (23)
\end{aligned}$$

Plugging (23) into (22) gives the result. □

2.3.3 Other estimates

Proposition 4 *Under the assumptions of Theorem 2, for $0 \leq k \leq N-1$ one has*

$$\begin{aligned}
&\mathbb{P}([A_{0,k}^M]^c) \leq 2\mathbb{E}\bigg(K_{0,k}^M \exp(-\frac{Mh^{\beta+2}}{72C_y(R)^2K_{0,k}^M})\mathcal{N}_2(\frac{h^{\frac{\beta+2}{2}}}{3\sqrt{2K_{0,k}^M}}, k+1)\bigg), \\
&\mathbb{P}([A_{l,k}^M]^c) \leq 2\mathbb{E}\bigg(K_{l,k}^M \exp(-\frac{Mh^{\beta+1}}{72C_y(R)^2R_0^2K_{l,k}^M})\mathcal{N}_2(\frac{h^{\frac{\beta+1}{2}}}{3\sqrt{2K_{l,k}^M}R_0}, k+1)\bigg), \\
&\mathbb{P}([A_{k}^M]^c) \leq 2\mathbb{E}\bigg(\mathcal{N}_2(\frac{h^{\frac{\beta+2}{2}}}{3\sqrt{2}}, k)\exp(-\frac{Mh^{\beta+2}}{72C_y(R)^2})\bigg), \\
&T_{1,k}^M = \mathbb{E}\big(\inf_\alpha \|y_k^{N,R} - \alpha\cdot p_{0,k}\|_{k,M}^2\big) \leq \inf_\alpha \mathbb{E}|y_k^{N,R}(X_{t_k}^N) - \alpha\cdot p_{0,k}(X_{t_k}^N)|^2, \\
&T_{2,k}^M \leq \frac{C_y(R)^2}{M}\mathbb{E}(K_{0,k}^M), \\
&T_{3,l,k}^M = \frac{1}{h}\mathbb{E}\big(\inf_\alpha \|\sqrt{h}z_{l,k}^{N,R} - \alpha\cdot p_{l,k}\|_{k,M}^2\big) \leq \frac{1}{h}\inf_\alpha \mathbb{E}|\sqrt{h}z_{l,k}^{N,R}(X_{t_k}^N) - \alpha\cdot p_{l,k}(X_{t_k}^N)|^2, \\
&T_{4,l,k}^M \leq \frac{C_y(R)^2}{hM}\mathbb{E}(K_{l,k}^M), \\
&\mathcal{N}_2(\epsilon, k+1) \leq C\exp(CK_{0,k+1}\log\frac{C\ C_y(R)}{\epsilon}).
\end{aligned}$$

Proof of the bound for $\mathbb{P}([A_{0,k}^M]^c)$. As already mentioned, we suppose without loss of generality that $\frac{(B_{0,k}^M)^*B_{0,k}^M}{M} = \mathrm{Id}$ and that $B_{0,k}^M$ is a matrix of size $M\times K_{0,k}^M$. Under this assumption, we can rewrite (see our notations for $\mathbb{P}_k^M$)

$$\mathbb{P}([A_{0,k}^M]^c) = \mathbb{E}\bigg(\mathbb{P}_k^M\big(|\alpha_{0,k}^{1,M} - \tilde{\alpha}_{0,k}^{1,M}|_2^2 \geq h^{\beta+2}\big)\bigg). \qquad (24)$$

By making the norm $|\cdot|_2$ in (24) explicit, we get:

$$\begin{aligned}
&\mathbb{P}_k^M\big(|\alpha_{0,k}^{1,M}-\tilde\alpha_{0,k}^{1,M}|_2^2\ge h^{\beta+2}\big)\\
=\ &\mathbb{P}_k^M\Big(\sum_{i=1}^{K_{0,k}^M}\Big|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^m\{y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(\tilde X_{t_{k+1}}^{N,m})\}\Big|^2\ge h^{\beta+2}\Big)\\
\le\ &\sum_{i=1}^{K_{0,k}^M}\mathbb{P}_k^M\Big(\Big|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^m\{y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m})-y_{k+1}^{N,R,M}(\tilde X_{t_{k+1}}^{N,m})\}\Big|^2\ge\frac{h^{\beta+2}}{K_{0,k}^M}\Big)\\
\le\ &\sum_{i=1}^{K_{0,k}^M}\mathbb{P}_k^M\Big(\exists\psi\in[\mathcal{P}_{0,k+1}]_y:\Big|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^m\{\psi(\tilde X_{t_{k+1}}^{N,m})-\psi(X_{t_{k+1}}^{N,m})\}\Big|^2\ge\frac{h^{\beta+2}}{K_{0,k}^M}\Big)\\
=\ &\sum_{i=1}^{K_{0,k}^M}\mathbb{P}_k^M\Big(\exists\psi\in[\mathcal{P}_{0,k+1}]_y:\Big|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^mU_m\{\psi(\tilde X_{t_{k+1}}^{N,m})-\psi(X_{t_{k+1}}^{N,m})\}\Big|\ge\sqrt{\frac{h^{\beta+2}}{K_{0,k}^M}}\Big)
\end{aligned}$$

where (U_m) is a sequence of i.i.d. Bernoulli random variables, taking values 1 and -1 with probability $\frac{1}{2}$, which are independent of everything else. The last equality comes from the fact that $\tilde X_{t_{k+1}}^{N,m}$ and $X_{t_{k+1}}^{N,m}$ have the same law. Now, we introduce a covering $\mathcal{G}$ of $[\mathcal{P}_{0,k+1}]_y$ such that $\forall\psi\in[\mathcal{P}_{0,k+1}]_y$, $\exists\ g\in\mathcal{G}$ such that

$$\frac{1}{2M}\sum_{m=1}^{M}\{|\psi(X_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})|^2+|\psi(\tilde X_{t_{k+1}}^{N,m})-g(\tilde X_{t_{k+1}}^{N,m})|^2\}\le\frac{h^{\beta+2}}{18K_{0,k}^M}.$$

We can assume without loss of generality that elements of $\mathcal{G}$ are bounded by $C_y(R)$. Note that $\mathcal{G}$ depends on $(X_{t_{k+1}}^{N,m},\tilde X_{t_{k+1}}^{N,m})_{1\le m\le M}$ but not on $(U_m)_{1\le m\le M}$, and that the cardinal of $\mathcal{G}$ is equal to $\mathcal{N}_2(\sqrt{\frac{h^{\beta+2}}{18K_{0,k}^M}},k+1)$. Taking advantage of the Cauchy-Schwarz inequality $|\frac{1}{M}\sum_{m=1}^M p_{0,k,i}^m\lambda_m|^2\le\frac{1}{M}\sum_{m=1}^M\lambda_m^2$ (under the assumption $\frac{(B_{0,k}^M)^*B_{0,k}^M}{M}=\mathrm{Id}$), we easily get

$$\begin{aligned}
&\mathbb{P}_k^M\Big(\exists\psi\in[\mathcal{P}_{0,k+1}]_y:\Big|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^mU_m\{\psi(\tilde X_{t_{k+1}}^{N,m})-\psi(X_{t_{k+1}}^{N,m})\}\Big|\ge\sqrt{\frac{h^{\beta+2}}{K_{0,k}^M}}\Big)\\
\le\ &\mathbb{P}_k^M\Big(\exists g\in\mathcal{G}:\Big|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^mU_m\{g(\tilde X_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})\}\Big|\ge\frac{1}{3}\sqrt{\frac{h^{\beta+2}}{K_{0,k}^M}}\Big)\\
\le\ &\mathcal{N}_2\Big(\sqrt{\frac{h^{\beta+2}}{18K_{0,k}^M}},k+1\Big)\max_{g\in\mathcal{G}}\mathbb{P}_k^M\Big(\Big|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^mU_m\{g(\tilde X_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})\}\Big|\ge\frac{1}{3}\sqrt{\frac{h^{\beta+2}}{K_{0,k}^M}}\Big).
\end{aligned}$$

To bound this last probability, we additionally condition by $(X_{t_{k+1}}^{N,m},\tilde X_{t_{k+1}}^{N,m})_{1\le m\le M}$ and denote by $\tilde{\mathbb{P}}_{k,k+1}^M$ the resulting conditional probability. We note that, if $X_m=p_{0,k,i}^mU_m\{g(\tilde X_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})\}$, $\tilde{\mathbb{E}}_{k,k+1}^M(X_m)=0$

and $|X_m| \leq 2C_y(R)|p^m_{0,k,i}|$. A combination of Hoeffding's inequality and of $\frac{1}{M}\sum_{m=1}^M |p^m_{0,k,i}|^2 = 1$ gives:

$$\tilde{\mathbb{P}}^M_{k,k+1}\big(\big|\frac{1}{M}\sum_{m=1}^M p^m_{0,k,i} U_m\{g(\tilde{X}^{N,m}_{t_{k+1}}) - g(X^{N,m}_{t_{k+1}})\}\big| \geq \frac{1}{3}\sqrt{\frac{h^{\beta+2}}{K^M_{0,k}}}\big)$$
$$\leq 2\exp(-\frac{2Mh^{\beta+2}}{144C_y(R)^2 K^M_{0,k}\frac{1}{M}\sum_{m=1}^M |p^m_{0,k,i}|^2}) = 2\exp(-\frac{Mh^{\beta+2}}{72C_y(R)^2 K^M_{0,k}}).$$

The estimate on $\mathbb{P}([A^M_{0,k}]^c)$ is now proved. □

Proof of the bound for $\mathbb{P}([A^M_{l,k}]^c)$. The calculations are the same as for $\mathbb{P}([A^M_{0,k}]^c)$ except that we need here a covering $\mathcal{G}$ of $[\mathcal{P}_{0,k+1}]_y$ such that $\forall \psi \in [\mathcal{P}_{0,k+1}]_y$, $\exists g \in \mathcal{G}$ which satisfies

$$\frac{1}{2M}\sum_{m=1}^M \{|\psi(\tilde{X}^{N,m}_{t_{k+1}}) - g(\tilde{X}^{N,m}_{t_{k+1}})|^2 + |\psi(X^{N,m}_{t_{k+1}}) - g(X^{N,m}_{t_{k+1}})|^2\} \leq \frac{h^{\beta+1}}{18K^M_{l,k}R_0^2}.$$

□

Proof of the bound for $\mathbb{P}([A^M_k]^c)$. We partially follow the proof of Theorem 11.2 in Györfi *et al.* (2002) and define the vector $(Z_m)_{1\leq m\leq 2M}$ by $(Z_m, Z_{M+m}) = (\tilde{X}^{N,m}_{t_k}, X^{N,m}_{t_k})$ if $U_m = 1$ and $(Z_m, Z_{M+m}) = (X^{N,m}_{t_k}, \tilde{X}^{N,m}_{t_k})$ if $U_m = -1$ where $(U_m)_{1\leq m\leq M}$ is a sequence of i.i.d. Bernoulli variables, independent of everything else, taking values 1 and -1 with probability $\frac{1}{2}$. As for $\mathbb{P}([A^M_{0,k}]^c)$, we introduce a covering $\mathcal{G}$ (whose elements are bounded by $2C_y(R)$) of $[\mathcal{P}_{0,k}]_y - y^{N,R}_k$ such that for all $\psi \in [\mathcal{P}_{0,k}]_y - y^{N,R}_k$, there exists a $g \in \mathcal{G}$ such that

$$\frac{1}{2M}\sum_{m=1}^M \{|\psi(\tilde{X}^{N,m}_{t_k}) - g(\tilde{X}^{N,m}_{t_k})|^2 + |\psi(X^{N,m}_{t_k}) - g(X^{N,m}_{t_k})|^2\} \leq \frac{h^{\beta+2}}{18}.$$

Thus, we can write that $\mathbb{P}([A^M_k]^c)$ is equal to

$$\mathbb{P}\big(\exists \psi \in [\mathcal{P}_{0,k}]_y - y^{N,R}_k : \{\frac{1}{M}\sum_{m=1}^M |\psi(Z_m)|^2\}^{\frac{1}{2}} - \{\frac{1}{M}\sum_{m=1}^M |\psi(Z_{M+m})|^2\}^{\frac{1}{2}} \geq h^{\frac{\beta+2}{2}}\big)$$
$$\leq \mathbb{P}\big(\exists g \in \mathcal{G} : \{\frac{1}{M}\sum_{m=1}^M |g(Z_m)|^2\}^{\frac{1}{2}} - \{\frac{1}{M}\sum_{m=1}^M |g(Z_{M+m})|^2\}^{\frac{1}{2}} \geq \frac{h^{\frac{\beta+2}{2}}}{3}\big).$$

Introducing the conditional probability $\tilde{\mathbb{P}}^M_{k-1,k}$ and $\mathcal{N}_2(\frac{h^{\frac{\beta+2}{2}}}{3\sqrt{2}}, k)$ (the cardinal of $\mathcal{G}$), simple computations lead to (see Györfi *et al.* 2002, p.191)

$$\tilde{\mathbb{P}}^M_{k-1,k}\big(\exists g \in \mathcal{G} : \{\frac{1}{M}\sum_{m=1}^M |g(Z_m)|^2\}^{\frac{1}{2}} - \{\frac{1}{M}\sum_{m=1}^M |g(Z_{M+m})|^2\}^{\frac{1}{2}} \geq \frac{h^{\frac{\beta+2}{2}}}{3}\big)$$

$$\leq 2\ \mathcal{N}_2\Big(\frac{h^{\frac{\beta+2}{2}}}{3\sqrt{2}}, k\Big) \sup_{g\in\mathcal{G}} \exp\left(- \frac{M h^{\beta+2} \sum_{m=1}^M \{|g(\tilde{X}_{t_k}^{N,m})|^2 + |g(X_{t_k}^{N,m})|^2\}}{18 \sum_{m=1}^M \{|g(\tilde{X}_{t_k}^{N,m})|^2 - |g(X_{t_k}^{N,m})|^2\}^2}\right).$$

The above exponential is bounded by $\exp\big(-\frac{M h^{\beta+2}}{72 C_y(R)^2}\big)$, because

$$\sum_{m=1}^M \{|g(\tilde{X}_{t_k}^{N,m})|^2 - |g(X_{t_k}^{N,m})|^2\}^2 \leq \sum_{m=1}^M \big(|g(\tilde{X}_{t_k}^{N,m})|^4 + |g(X_{t_k}^{N,m})|^4\big)$$
$$\leq 4 C_y(R)^2 \sum_{m=1}^M \big(|g(\tilde{X}_{t_k}^{N,m})|^2 + |g(X_{t_k}^{N,m})|^2\big). \qquad (25)$$

Bringing together all the previous estimates gives the required upper bound for $\mathbb{P}([A_k^M]^c)$. □

Proof of the identities for $\mathbb{E}T_{1,k}^M$ and $\mathbb{E}T_{3,l,k}^M$. Observe that $\mathbb{E}^M(\tilde{\beta}_{0,k}^M)$ minimizes $\frac{1}{M}\sum_{m=1}^M |y_k^{N,R}(X_{t_k}^{N,m}) - \alpha \cdot p_{0,k}^m|^2 = \|y_k^{N,R} - \alpha \cdot p_{0,k}\|_{k,M}^2$, i.e. $T_{1,k}^M = \inf_\alpha \|y_k^{N,R} - \alpha \cdot p_{0,k}\|_{k,M}^2$. The same arguments apply to $\mathbb{E}T_{3,l,k}^M$. □

Proof of the bounds for $\mathbb{E}T_{2,k}^M$ and $\mathbb{E}T_{4,l,k}^M$. We prove only the estimate for $\mathbb{E}T_{2,k}^M$, techniques being the same for $\mathbb{E}T_{4,l,k}^M$. We adapt the proof of Theorem 11.1 in Györfi *et al.* (2002) and suppose without loss of generality that $\frac{(B_{0,k}^M)^* B_{0,k}^M}{M} = \mathrm{Id}$ as before. We can thus write for $1 \leq m \leq M$

$$\begin{aligned}
&\mathbb{E}^M\big(|p_{0,k}^m.\{\tilde{\alpha}_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\alpha}_{0,k}^{1,M})\}|^2\big) \qquad (26)\\
=&\mathbb{E}^M\Big((p_{0,k}^m)^* \frac{(B_{0,k}^M)^*}{M}\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^* \frac{B_{0,k}^M}{M} p_{0,k}^m\Big)\\
=&(p_{0,k}^m)^* \frac{(B_{0,k}^M)^*}{M} \mathbb{E}^M\big(\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^*\big) \frac{B_{0,k}^M}{M} p_{0,k}^m
\end{aligned}$$

where V is the vector of $\mathbb{R}^M$ with coordinates $y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m})$. We bound this last expression by considering $\|\mathbb{E}^M\big(\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^*\big)\|_2$. As $\alpha_{0,k+1}^M$ is $\mathcal{F}^M$-measurable, we get

$$\begin{aligned}
&\mathbb{E}^M\big(y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}) y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m'})\big)\\
=&\mathbb{E}^M\big([\alpha_{0,k+1}^M \cdot p_{0,k+1}(\tilde{X}_{t_{k+1}}^{N,m})]_y [\alpha_{0,k+1}^M \cdot p_{0,k+1}(\tilde{X}_{t_{k+1}}^{N,m'})]_y\big)\\
=&\mathbb{E}^M\big([\alpha_{0,k+1}^M \cdot p_{0,k+1}(\tilde{X}_{t_{k+1}}^{N,m})]_y\big) \mathbb{E}^M\big([\alpha_{0,k+1}^M \cdot p_{0,k+1}(\tilde{X}_{t_{k+1}}^{N,m'})]_y\big)
\end{aligned}$$

from which we deduce that the non-diagonal terms of the matrix $\mathbb{E}^M\big(\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^*\big)$ are equal to 0. The introduction of the projection coefficients $\tilde{\alpha}_{0,k}^M$ ensures this crucial property which is not true with projection coefficients $\alpha_{0,k}^M$. As for the diagonal terms, they are bounded by $C_y(R)^2$. Thus $\|\mathbb{E}^M\big(\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^*\big)\|_2 \leq C_y(R)^2$. Finally, in view of (26), we get

$$\begin{aligned}
&\mathbb{E}^M\left(\frac{1}{M}\sum_{m=1}^{M}|p^m_{0,k}.\{\tilde\alpha^{1,M}_{0,k}-\mathbb{E}^M(\tilde\alpha^{1,M}_{0,k})\}|^2\right)\\
&\leq \frac{1}{M}\sum_{m=1}^{M}|\frac{B^M_{0,k}}{M}p^m_{0,k}|_2^2 C_y(R)^2\\
&\leq \frac{C_y(R)^2}{M^2}\sum_{m=1}^{M}|p^m_{0,k}|^2=\frac{C_y(R)^2}{M^2}\mathrm{Tr}[(B^M_{0,k})^*B^M_{0,k}]=\frac{C_y(R)^2}{M}K^M_{0,k}.
\end{aligned}$$

The estimates for $\mathbb{E}T^M_{2,k}$ readily follows. □

Proof of the bound for $\mathcal{N}_2(\epsilon,k+1)$. One can directly apply Theorem 9.4 in Györfi *et al.* (2002) (and Theorem 9.5 in the same reference to bound the Vapnik-Chervonenkis dimension of a functions vector space) which writes:

$$\mathcal{N}_2(\epsilon,k+1)\leq 3\left(\frac{2e(2C_y(R))^2}{\epsilon^2}\log(\frac{3e(2C_y(R))^2}{\epsilon^2})\right)^{K_{0,k+1}},$$

whence our result. □

3 Another algorithm

We have used in the proof of Proposition 3 projections coefficients $(\tilde\alpha^M_{l,k})_{0\leq l\leq q}$, which have enabled us to overcome the lack of independence between least-squares problems at different discretization times t_k. Up to simulating few extra random variables, we could design an algorithm where, at a discretization time t_k, we would calculate the coefficients $(\tilde\alpha^M_{l,k})_{0\leq l\leq q}$ instead of the coefficients $(\alpha^M_{l,k})_{0\leq l\leq q}$. We now describe this algorithm:

→ Initialization : for $k=N$ take $y^{N,R,M}_N(\cdot)=\phi^R(\cdot)$.

→ Let $k<N-1$. For each path m, simulate $(\tilde X^{N,m}_{t_{k+1}},\Delta\tilde W^m_k)$ which are, conditionally to $X^{N,m}_{t_k}$, an independent copy of $(X^{N,m}_{t_{k+1}},\Delta W^m_k)$ (and independent of everything else; see paragraph 2.3.1). Solve the least-squares problems (13) to get $(\tilde\alpha^M_{l,k})_{1\leq l\leq q}$. Then define $z^{N,R,M}_{l,k}(\cdot)=[\tilde\alpha^M_{l,k}\cdot p_{l,k}]_z(\cdot)$. Next, solve the least-squares problem (14) to get $\tilde\alpha^M_{0,k}$ and define $y^{N,R,M}_k(\cdot)=[\tilde\alpha^M_{0,k}\cdot p_{0,k}]_y(\cdot)$.

→ Iterate until time t_0.

Compared to the algorithm of Section 1, we draw twice more simulations. The global complexity is therefore multiplied at most by a factor 2, which is not costly at all. Nevertheless the accuracy is improved, at least theoretically (see comments below). It is easy to adapt the proof of Proposition 3 to analyze this new algorithm (for full details, see Lemor 2005). Theorem 2 simplifies to give:

Theorem 3 *Assume* **(H1-H2-H3-H4)** *and let* $\beta \in]0,1]$. *Then, there exists a constant* C *(independent on* β*) such that:*

$$\begin{aligned}
&\max_{0\le k\le N} \mathbb{E}\frac{1}{M}\sum_{m=1}^{M} |y_k^{N,R}(X_{t_k}^{N,m}) - y_k^{N,R,M}(X_{t_k}^{N,m})|^2 \\
&+ h\mathbb{E}\sum_{k=0}^{N-1}\frac{1}{M}\sum_{m=1}^{M} |z_k^{N,R}(X_{t_k}^{N,m}) - z_k^{N,R,M}(X_{t_k}^{N,m})|^2 \\
\le & C\frac{C_y(R)^2}{M}\sum_{k=0}^{N-1}\sum_{l=0}^{q}\mathbb{E}(K_{l,k}^M) + Ch^\beta \\
&+ C\sum_{k=0}^{N-1}\{\inf_\alpha \mathbb{E}|y_k^{N,R}(X_{t_k}^N) - \alpha\cdot p_{0,k}(X_{t_k}^N)|^2 \\
&\qquad + \sum_{l=1}^{q}\inf_\alpha \mathbb{E}|\sqrt{h}z_{l,k}^{N,R}(X_{t_k}^N) - \alpha\cdot p_{l,k}(X_{t_k}^N)|^2\} \\
&+ C\frac{C_y(R)^2}{h}\sum_{k=0}^{N-1}\exp(CK_{0,k}\log\frac{C\ C_y(R)}{h^{\frac{\beta+2}{2}}})\exp(-\frac{Mh^{\beta+2}}{72C_y(R)^2}).
\end{aligned}$$

Firstly, as for Theorem 2, analogous estimates are valid for $\max_{0\le k\le N}\mathbb{E}|y_k^{N,R}(X_{t_k}^N) - y_k^{N,R,M}(X_{t_k}^N)|^2 + h\mathbb{E}\sum_{k=0}^{N-1}|z_k^{N,R}(X_{t_k}^N) - z_k^{N,R,M}(X_{t_k}^N)|^2$ (see Remark 2).

Alg. of Section 1	Alg. of Section 3	Alg. Bouchard and Touzi (2004)
$\mathcal{C}^{-\frac{1}{4+2d}}$	$\mathcal{C}^{-\frac{1}{4+d}}$	$\mathcal{C}^{-\frac{1}{13+d}}$ (if X=Brownian motion or geometric BM)

Table 1: Squared error for different algorithms with respect to the complexity $\mathcal{C}$.

Secondly, it is not surprising to see that the terms coming from the events $(A_{l,k}^M)_{0\le l\le q}$ (which express the closeness between $(\alpha_{l,k}^M)_{0\le l\le q}$ and $(\tilde{\alpha}_{l,k}^M)_{0\le l\le q}$) have disappeared. Mimicking the calculations done after Theorem 2, this leads to a squared error of order $\mathcal{C}^{-\frac{1}{4+d}}$. The table 1 sums up the complexity's evaluations for the different algorithms and it appears that this new algorithm gives a far better complexity to achieve a given accuracy.

4 Numerical experiments

To test asymptotic results by letting N, M and the number of basis functions go to infinity, one needs to use a functions basis for which the regression error arising in Theorem 2 is explicit. We use the hypercubes basis **HC** (see paragraph 1.5).

We consider the case of pricing an option with a differential of interest rates (Bergman 1995). We suppose that X follows the Black-Scholes model in dimension $d = 1$, $\frac{\mathrm{d}X_t}{X_t} = \mu \mathrm{d}t + \sigma \mathrm{d}W_t$, with parameters $\mu = 0.05$, $\sigma = 0.2$ and $X_0 = 100$. For the terminal condition, we take that of a Call Spread option, that is $(X_T - K_1)^+ - 2(X_T - K_2)^+$ with $K_1 = 95$ and $K_2 = 105$. The non-linear driver f is defined by $f(t, x, y, z) = -\theta z - ry + (y - \frac{z}{\sigma})^-(R - r)$, where the two interest rates are $r = 0.01$, $R = 0.06$ and $\theta = (\mu - r)/\sigma$. The maturity of the option is $T = 0.25$. According to Gobet *et al.* (2005), the relative solution Y_0 is equal to **2.95**.

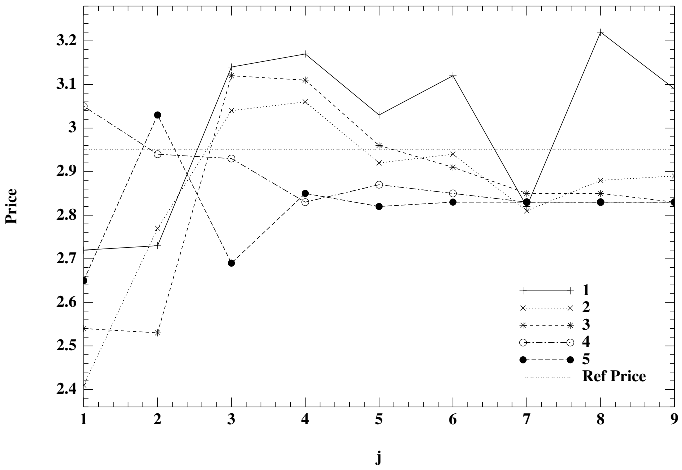

Figure 1: basis **HC**, $\beta = 0.2$, $\alpha_M^* = 3.4$, $\alpha_M = 1, 2, 3, 4, 5$.

Here the numerical issue is to determine if our algorithm asymptotically recovers this value when one modifies all the parameters N, M and δ (the edge of the hypercubes). Regarding N, one starts from $N_0 = 2$ and $N = N_0(\sqrt{2})^{(j-1)}$ where $j = 1 \cdots$ is the number of different values of N to be tested. As mentioned before, we neglect the influences of the Brownian increments threshold R_0 and of the domain width R_1 on which the basis **HC** is defined. This domain is fixed once for all to $[40, 180]$. As far the regression errors are concerned, the choice $\delta = \frac{50}{(\sqrt{2})^{\frac{\beta+1}{2}(j-1)}} \sim C\ h^{\frac{\beta+1}{2}}$ makes the algorithm converge at rate h^β (for the squared error). Now it remains to adjust M as a function

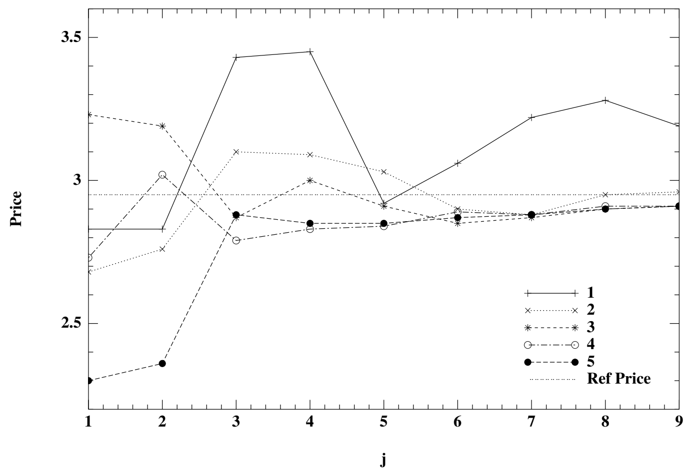

Figure 2: basis **HC**, $\beta = 1$, $\alpha_M^* = 5$, $\alpha_M = 1, 2, 3, 4, 5$.

of N and δ, or equivalently h and β. In Theorem 2, the convergence to 0 of the term $C\frac{C_y(R)^2}{M}\sum_{k=0}^{N-1}\sum_{l=0}^{q}\mathbb{E}(K_{l,k}^M)$ leads to $Mh^{1+\frac{\beta+1}{2}} \to \infty$. Furthermore, the inspection of the other terms involving the covering numbers imposes the stronger condition $Mh^{\beta+2+(\beta+1)} = Mh^{3+2\beta} \to \infty$. The following experiments are aimed at testing the empirical validity of this threshold rule. For this, we set $M = 2(\sqrt{2})^{\alpha_M(j-1)}$ for different values of α_M and check the algorithm convergence according to $\alpha_M < \alpha_M^*$ or $\alpha_M > \alpha_M^*$ where $\alpha_M^* = 3 + 2\beta$ is the (theoretical) critical convergence threshold. In practice, we perform tests for $\beta = 0.2$, $\beta = 1$ and report the average value given by the algorithm on 50 runs. From Figure 1, the algorithm's price seems to diverge for $\alpha_M = 1$ whereas the prices $\alpha_M > 1$ seem to converge towards the reference price but very slowly, in accord with the choice of $\beta = 0.2$. From Figure 2 ($\beta = 1$) we note that this time the algorithm's price for $\alpha_M \geq 3$ clearly converges towards the reference price but we observe on Figure 3 that too big values of α_M are undesirable : this does not speed up the convergence with respect to j because some error terms (actually the bias) only depend on N and δ but not on M whereas the calculation time becomes very large. Moreover, one cannot anymore use the confidence interval for the price given by the empirical standard deviation because when M tends too quickly to infinity, this empirical standard deviation tends very fast to 0 and does not reflect anymore the bias terms depending on N and δ. As usual, it is important to well balance the bias and variance terms. Finally, we observe on this last example that the empirical levels of convergence of α_M are better than the ones expected from the condition $Mh^{3+2\beta} \to \infty$: this indicates that the bounds of Theorem 2 may be not optimal. One possible

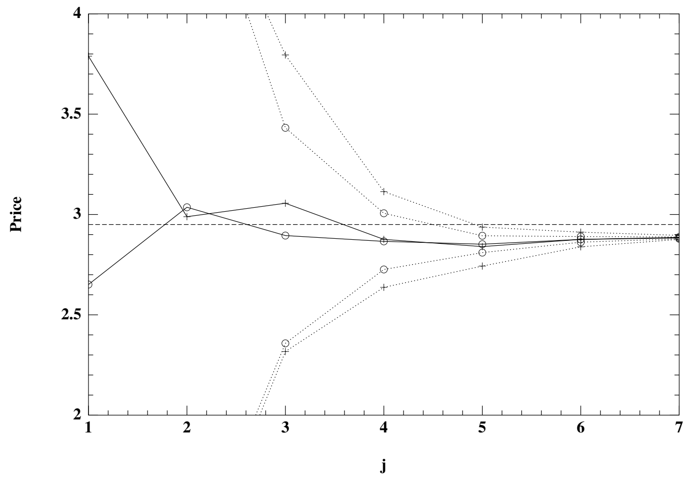

Figure 3: basis **HC**, $\beta = 1$, $\alpha_M^* = 5$, $\alpha_M = 6$: curves with cross markers, $\alpha_M = 7$: curves with circle markers. In plain line, prices and in dotted lines upper and lower 0.95 confidence intervals.

reason could be the sub-optimality of the upper bound (25): indeed, as a matter of fact we do not use the closeness between $X_{t_k}^{N,m}$ and $\tilde{X}_{t_k}^{N,m}$. Another reason may be that essentially only the empirical basis dimension $K_{l,k}^M$ is involved in Theorem 2 and this dimension can be much smaller than the theoretical one $K_{l,k}$ (see Remark 3).

A numerical comparison between the algorithms of Section 1 and 3 on several examples shows that both algorithms behave similarly (they have the same convergence levels), even if the bound of Theorem 3 is better than the one in Theorem 2.

On Figure 4, we test again $\beta = 0.2$ but with the basis **HC(1,0)** which is analogous to **HC**: it consists in using the local polynomial basis $1, x$ on each hypercube to approximate $y^{N,R}$ instead of just 1 in the case of basis **HC**, while for $z^{N,R}$ there is no modification. We refer to Lemor (2005) for more details on these function bases. On this example, the basis **HC(1,0)** speeds up the overall convergence. This phenomenon is fundamental to our opinion, compared to the quantization method that can be viewed as using only indicator function bases (basis **HC**).

Finally, we look at a last example taken from Heath *et al.* (2001). In this article, the authors approximate via PDE's methods the local risk-minimization price (see Föllmer and Schweizer 1991) of a put option in a stochastic volatility model (Heston type model). The dynamics of the asset price X and of the square of

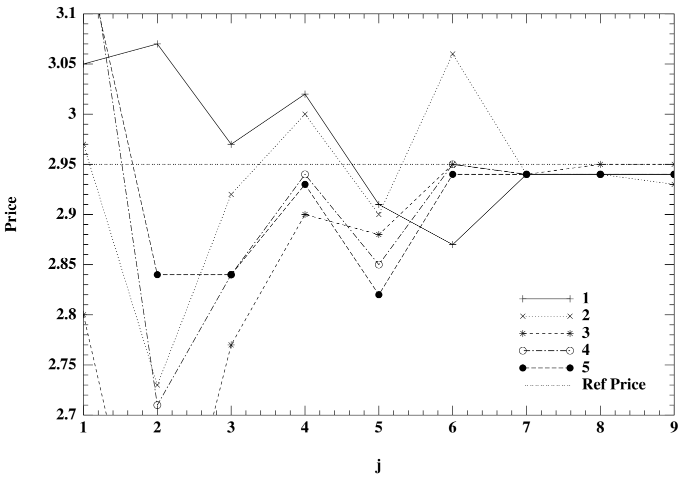

Figure 4: basis $\mathbf{HC(1,0)}$, $\beta = 0.2$, $\alpha_M^* = 3.4$, $\alpha_M = 1, 2, 3, 4, 5$.

the volatility F are:

$$\begin{aligned}\frac{\mathrm{d}X_t}{X_t} =& \gamma F_t \mathrm{d}t + \sqrt{F_t}\mathrm{d}W_t,\\ \mathrm{d}F_t =& \kappa(\theta - F_t)\mathrm{d}t + \Sigma\sqrt{F_t}\mathrm{d}W_t'\end{aligned}$$

with W, W' two independent Brownian motions. It is easy to see that the local risk-minimization price Y must satisfy the following GBSDE:

$$\begin{cases} -\mathrm{d}Y_t = & -(r_t Y_t + \frac{Z_t}{\sqrt{F_t}}(\gamma F_t - r_t))\mathrm{d}t - Z_t \mathrm{d}W_t - \mathrm{d}L_t, \\ Y_T = & (K - X_T)_+. \end{cases}$$

Taking r to be 0 this leads to the driver $f(t, x, F, y, z) = -\gamma z\sqrt{F}$ which is not Lipschitz (for related results, see El Karoui and Huang 1997). We nevertheless apply our algorithm for $\alpha_M = 3$ and present the results on Figure 5. As in Heath *et al.* (2001), we take $\kappa = 5$, $\theta = 0.04$, $\Sigma = 0.6$, $\gamma = 2.5$, $X_0 = K = 100$, $F_0 = 0.04$. The functions basis is **HC**, in dimension 2 (one dimension for the asset price and one for the stochastic volatility). The reference price (for $r = 0$) is taken from Heath *et al.* (2001) and is **7.69**. We observe that, in this non-Lipschitz case, the price still converges towards the reference price.

5 Conclusion

We have proposed a simple algorithm to solve GBSDEs. The dynamic programming equation resulting from the time discretization of the equation (1) is solved

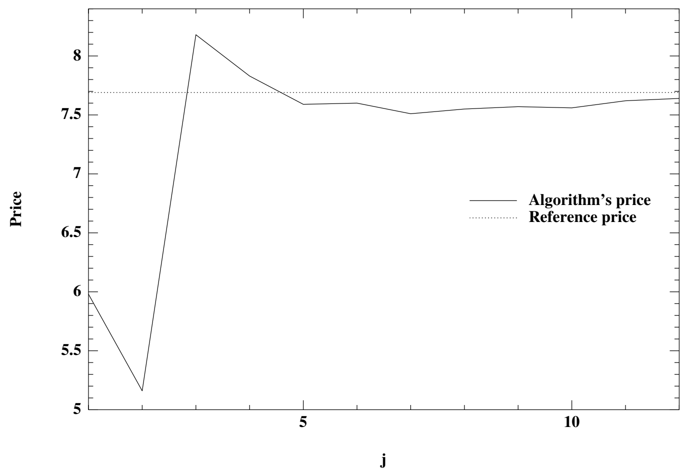

Figure 5: basis **HC**, $\beta = 1$, $\alpha_M = 3$.

using a sequence of empirical regression problems based on simulations of the underlying Markov process. The extension to path-dependent terminal conditions is straightforward (see Remark 1). We have derived explicit error bounds which allow to optimally choose the parameters of the method to achieve a given accuracy. This is a major improvement compared to previous works. However our numerical experiments reveal that the convergence can be faster than what our theoretical estimates predict. The explanation of this phenomenon concerns future researches. Additional works are also necessary to consider in (1) another martingale than W and to let the driver depend on L.

6 Appendix: proof of Theorem 1

We only prove the result for $\max_{0\leq k\leq N}\mathbb{E}|Y_{t_k} - Y^N_{t_k}|^2$, that for $\mathbb{E}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}|Z^N_{t_k} - Z_t|^2\mathrm{d}t$ following from the same kind of calculations than in the proof of Proposition 2.
Firstly, we know from El Karoui *et al.* (1997) that the solution (Y, Z, L) satisfies

$$\mathbb{E}\big(\max_{t\in[0,T]} Y_t^2 + \int_0^T |Z_t|^2\mathrm{d}t + [L]_T\big) < +\infty. \tag{27}$$

Then, from (1-2) we get $Y_{t_k} - Y^N_{t_k} = \mathbb{E}_{t_k}(Y_{t_{k+1}} - Y^N_{t_{k+1}}) + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}$ $\{f(s, X_s, Y_s, Z_s) - f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k})\}\mathrm{d}s$. A combination of Young's inequality (with a parameter $\gamma > 0$ to be chosen later) and of the Lipschitz property of f gives

$$\mathbb{E}|Y_{t_k}-Y^N_{t_k}|^2 \leq (1+\gamma h)\mathbb{E}|\mathbb{E}_{t_k}(Y_{t_{k+1}}-Y^N_{t_{k+1}})|^2 + C(h+\frac{1}{\gamma})\mathbb{E}\int_{t_k}^{t_{k+1}} |Z_s - Z^N_{t_k}|^2 \mathrm{d}s$$
$$+C(h+\frac{1}{\gamma})(h^2+\int_{t_k}^{t_{k+1}}\mathbb{E}|X_s-X^N_{t_k}|^2\mathrm{d}s+\int_{t_k}^{t_{k+1}}\mathbb{E}|Y_s-Y^N_{t_{k+1}}|^2\mathrm{d}s). \quad (28)$$

Now define $\overline{Z}_{t_k}$ by

$$h\overline{Z}_{t_k} := \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}} Z_s\mathrm{d}s = \mathbb{E}_{t_k}\left(\{Y_{t_{k+1}}+\int_{t_k}^{t_{k+1}} f(s,X_s,Y_s,Z_s)\mathrm{d}s\}\Delta W^*_k\right).$$

Clearly

$$\mathbb{E}\int_{t_k}^{t_{k+1}}|Z_s-Z^N_{t_k}|^2\mathrm{d}s = \mathbb{E}\int_{t_k}^{t_{k+1}}|Z_s-\overline{Z}_{t_k}|^2\mathrm{d}s + h\mathbb{E}|\overline{Z}_{t_k}-Z^N_{t_k}|^2. \quad (29)$$

The Cauchy-Schwarz inequality yields

$$|\mathbb{E}_{t_k}(\{Y_{t_{k+1}}-Y^N_{t_{k+1}}\}\Delta W_{l,k})|^2 \leq h\{\mathbb{E}_{t_k}(|Y_{t_{k+1}}-Y^N_{t_{k+1}}|^2) - |\mathbb{E}_{t_k}(Y_{t_{k+1}}-Y^N_{t_{k+1}})|^2\}$$

and consequently

$$h\mathbb{E}|\overline{Z}_{t_k}-Z^N_{t_k}|^2 \leq C\mathbb{E}\{\mathbb{E}_{t_k}(|Y_{t_{k+1}}-Y^N_{t_{k+1}}|^2) - \mathbb{E}|\mathbb{E}_{t_k}(Y_{t_{k+1}}-Y^N_{t_{k+1}})|^2\}$$
$$+ Ch\mathbb{E}\int_{t_k}^{t_{k+1}} f(s,X_s,Y_s,Z_s)^2\mathrm{d}s. \quad (30)$$

Plugging (29-30) into (28), we get:

$$\begin{aligned}
\mathbb{E}|Y_{t_k}-Y^N_{t_k}|^2 \leq{}& (1+\gamma h)\mathbb{E}|\mathbb{E}_{t_k}(Y_{t_{k+1}}-Y^N_{t_{k+1}})|^2 + C(h+\frac{1}{\gamma})\mathbb{E}\int_{t_k}^{t_{k+1}}|Z_s-\overline{Z}_{t_k}|^2\mathrm{d}s\\
&+ C(h+\frac{1}{\gamma})(h^2+\int_{t_k}^{t_{k+1}}\mathbb{E}|X_s-X^N_{t_k}|^2\mathrm{d}s+\int_{t_k}^{t_{k+1}}\mathbb{E}|Y_s-Y^N_{t_{k+1}}|^2\mathrm{d}s)\\
&+ C(h+\frac{1}{\gamma})\mathbb{E}\{\mathbb{E}_{t_k}(|Y_{t_{k+1}}-Y^N_{t_{k+1}}|^2) - |\mathbb{E}_{t_k}(Y_{t_{k+1}}-Y^N_{t_{k+1}})|^2\}\\
&+ Ch(h+\frac{1}{\gamma})\mathbb{E}\int_{t_k}^{t_{k+1}} f(s,X_s,Y_s,Z_s)^2\mathrm{d}s.
\end{aligned}$$

Now write $\mathbb{E}|Y_s-Y^N_{t_{k+1}}|^2 \leq 2\mathbb{E}|Y_s-Y_{t_{k+1}}|^2+2\mathbb{E}|Y_{t_{k+1}}-Y^N_{t_{k+1}}|^2$ and analogously for $X_s - X^N_{t_k}$, take $\gamma = C$: for h small enough, it gives

$$\begin{aligned}
\mathbb{E}|Y_{t_k}-Y^N_{t_k}|^2 \leq{}& (1+Ch)\mathbb{E}|Y_{t_{k+1}}-Y^N_{t_{k+1}}|^2 + Ch^2 + Ch\max_{0\leq k\leq N}\mathbb{E}|X_{t_k}-X^N_{t_k}|^2\\
&+ C\mathbb{E}\int_{t_k}^{t_{k+1}}|Z_s-\overline{Z}_{t_k}|^2\mathrm{d}s + C\int_{t_k}^{t_{k+1}}\mathbb{E}|X_s-X_{t_k}|^2\mathrm{d}s\\
&+ C\int_{t_k}^{t_{k+1}}\mathbb{E}|Y_s-Y_{t_{k+1}}|^2\mathrm{d}s + Ch\mathbb{E}\int_{t_k}^{t_{k+1}} f(s,X_s,Y_s,Z_s)^2\mathrm{d}s
\end{aligned}$$

and by Gronwall's lemma

$$\max_{0\le k\le N}\mathbb{E}|Y_{t_k}-Y^N_{t_k}|^2 \le Ch + C\max_{0\le k\le N}\mathbb{E}|X_{t_k}-X^N_{t_k}|^2$$
$$+C\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}\{|Z_s-\overline{Z}_{t_k}|^2+|X_s-X_{t_k}|^2+|Y_s-Y_{t_{k+1}}|^2\}\mathrm{d}s.$$

The contribution $\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}\mathbb{E}|Y_s - Y_{t_{k+1}}|^2\mathrm{d}s$ is a $O(h)$: indeed it is upper bounded by $3\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}\mathrm{d}s(t_{k+1}-s)\int_s^{t_{k+1}}\mathbb{E}f(u,X_u,Y_u,Z_u)^2\mathrm{d}u + 3\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}\mathrm{d}s\int_s^{t_{k+1}}\mathbb{E}|Z_u|^2\mathrm{d}u + 3\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}\mathrm{d}s\mathbb{E}([L]_{t_{k+1}}-[L]_s)$, which equals a $O(h)$ owing to the a priori estimates on (Z,L) (see (27)). In the same way, the contribution related to $X_s - X_{t_k}$ is of order $O(h)$. Finally, it gives

$$\max_{0\le k\le N}\mathbb{E}|Y_{t_k}-Y^N_{t_k}|^2 \le Ch + C\max_{0\le k\le N}\mathbb{E}|X_{t_k}-X^N_{t_k}|^2$$
$$+C\sum_{k=0}^{N-1}\mathbb{E}\int_{t_k}^{t_{k+1}}|Z_s-\overline{Z}_{t_k}|^2\mathrm{d}s.$$

Without extra assumptions, the above approximation's error related to the predictable process Z converges to 0: combining this with **(H2)**, we conclude $\max_{0\le k\le N}\mathbb{E}|Y_{t_k}-Y^N_{t_k}|^2\to 0$.
In the Brownian filtration case ($\beta\equiv 0$) and when X^N is the Euler scheme of X, clearly $\mathbb{E}|X_{t_k}-X^N_{t_k}|^2 = O(h)$ uniformly in k. Furthermore, Zhang (2004) establishes that the error on Z equals $O(h)$. Hence, $\max_{0\le k\le N}\mathbb{E}|Y_{t_k}-Y^N_{t_k}|^2 = O(h)$. □

Acknowledgements. The authors thank *Association Nationale de la Recherche Technique*, *École Polytechnique* and *Électricité de France* for their financial supports. This work is part of the first author's PhD thesis, under the supervision of the two other authors.

References

[1] V. Bally and G. Pagès (2002). Error analysis of the optimal quantization algorithm for obstacle problems. *Stochastic Process. Appl.*, 106(1):1–40, 2003.

[2] D. Becherer and M. Schweizer (2005). Classical solutions to reaction diffusion systems for hedging problems with interacting Ito and point processes. *Annals of Applied Probability*, 15(2):1111-1144, 2005.

[3] F.E. Benth, L. Ekeland, R. Hauge and B.F. Nielsen (2003). A note on arbitrage-free pricing of forward contracts in energy markets. *Applied Mathematical Finance*, 10(4):325–336, 2003.

[4] Y.Z. Bergman (1995). Option pricing with differential interest rates. *Rev. of Financial Studies*, 8(2):475–500, 1995.

[5] T.R. Bielecki and M. Rutkowski (2002). Credit risk: modelling, valuation and hedging. Springer Finance, Springer-Verlag, Berlin, 2002.

[6] G. Barles, R. Buckdahn, and E. Pardoux (1997). Backward stochastic differential equations and integral-partial differential equations. *Stochastics and Stochastics Reports*, 60:57–83, 1997.

[7] D. Bosq and J.P. Lecoutre (1987). Théorie de l'estimation fonctionnelle. *Economica*, Paris, 1987.

[8] B. Bouchard and N. Touzi (2004). Discrete time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications*, 111:175–206, 2004.

[9] P. Briand, B. Delyon, and J. Mémin (2001). Donsker-type theorem for BSDEs. *Electron. Comm. Probab.*, 6:1–14 (electronic), 2001.

[10] D. Egloff (2005). Monte Carlo algorithms for optimal stopping and statistical learning. To appear in *Annals of Applied Probability*, 2005.

[11] N. El Karoui and S.J. Huang (1997). A general result of existence and uniqueness of backward stochastic differential equations. In *Backward stochastic differential equations (Paris, 1995–1996)*, Pitman Res. Notes Math. Ser. 364, 27–36, 1997.

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

[13] H. Föllmer and M. Schweizer (1991). Hedging of Contingent Claims under Incomplete Information. In *Applied Stochastic Analysis*, eds. M.H.A. Davis and R.J. Elliott, 389–414, Gordon and Breach, London (1991).

[14] E. Gobet, J.P. Lemor and X. Warin (2005). A regression-based Monte-Carlo method for backward stochastic differential equations. To appear in *Annals of Applied Probability*, 2005.

[15] G. Golub and C.F. Van Loan (1996). *Matrix computations.3rd ed.* Baltimore, MD: The Johns Hopkins Univ. Press. xxvii, 694 p. , 1996.

[16] L. Györfi, M. Kohler, A. Krzyzak and H. Walk (2002). A distribution-free theory of nonparametric regression. Springer Series in Statistics, 2002.

[17] D. Heath, E. Platen and M. Schweizer (2001). Comparison of Two Quadratic Approaches to Hedging in Incomplete Markets. *Mathematical Finance*, 11:385–413, 2001.

[18] J. Jacod (2004). The Euler scheme for Lévy driven stochastic differential equations: limit theorems. *The Annals of Probability*, 32(3):1830–1872, 2004.

[19] E. Jouini and H. Kallal (1995). Arbitrage in securities markets with short-sales constraints. *Math. Finance*, 5(3):197–232, 1995.

[20] J.P. Lemor (2005) Approximation par projections et simulations Monte-Carlo des équations différentielles stochastiques rétrogrades. PhD Thesis, Ecole Polytechnique, 2005.

[21] J. Ma, P. Protter, J. San Martín, and S. Soledad (2002). Numerical method for backward stochastic differential equations. *Ann. Appl. Probab.*, 12(1):302–316, 2002.

[22] P. Monat and C. Stricker (1995). Föllmer-Schweizer decomposition and mean-variance hedging of general claims. *The Annals of Probability*, 23:605–628, 1995.

[23] E. Pardoux and S.G. Peng (1990). Adapted solution of a backward stochastic differential equation. *Systems Control Lett.*, 14(1):55–61, 1990.

[24] S. Peng (2004). Nonlinear expectations, nonlinear evaluations and risk measures. In *Stochastic methods in finance*, Lecture Notes in Math. 1856, 165–253, Springer Berlin, 2004.

[25] H. Pham, T. Rheinländer and M. Schweizer (1998). Mean-variance hedging for continuous processes: new proofs and examples. *Finance Stochast.*, 2:173–198, 1998.

[26] M. Schweizer (2001). A guided tour through quadratic hedging approaches. In *Option Pricing, Interest Rates and Risk Management* (E. Jouini, J. Cvitani and M. Musiela, eds.), 538–574, Cambridge Univ. Press, 2001.

[27] J. Zhang (2004). A numerical scheme for BSDEs. *Ann. Appl. Probab.*, 14(1):459–488, 2004.

J.P. LEMOR - Électricité de France - EDF R&D, site de Clamart - 1 avenue du Général De Gaulle, 92141 Clamart - FRANCE (e-mail: jp.lemor@libertysurf.fr).

E. GOBET - École Polytechnique - Centre de Mathématiques Appliquées - 91128 Palaiseau Cedex - FRANCE (e-mail: emmanuel.gobet@polytechnique.fr). Corresponding author

X. WARIN - Électricité de France - EDF R&D, site de Clamart - 1 avenue du Général De Gaulle, 92141 Clamart - FRANCE (e-mail: xavier.warin@edf.fr).