Bernoulli **12**(5), 2006, 889–916

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

JEAN-PHILIPPE LEMOR[1], EMMANUEL GOBET[2] and XAVIER WARIN[3]

[1]*E-mail: jp.lemor@libertysurf.fr*
[2]*INP Grenoble, ENSIMAG, Laboratoire de Modélisation et Calcul (UMR 5523), BP 53, 38041 Grenoble Cedex 9, France. E-mail: emmanuel.gobet@imag.fr*
[3]*EDF R & D, Électricité de France, 1 avenue du Général de Gaulle, 92141 Clamart, France. E-mail: xavier.warin@edf.fr*

This study focuses on the numerical resolution of backward stochastic differential equations with data dependent on a jump-diffusion process. We propose and analyse a numerical scheme based on iterative regression functions which are approximated by projections on vector spaces of functions, with coefficients evaluated using Monte Carlo simulations. Regarding the error, we derive explicit bounds with respect to the time step, the number of paths simulated and the number of functions: this allows us to optimally adjust the parameters to achieve a given accuracy. We also present numerical tests related to option pricing with differential interest rates and locally risk-minimizing strategies (Föllmer–Schweizer decomposition).

Keywords: backward stochastic differential equations; empirical regressions

1. Introduction

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

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

where Y is a scalar cadlag adapted process, Z is a predictable $\mathbb{R}^q$-valued process (as a row vector) and L is a scalar cadlag martingale orthogonal to W (with $L_0 = 0$). Actually, in what follows, X could be any Markov process. Our results can be 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 a 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 comprehensive situation is treated in Barles *et al.* (1997), where in addition the driver is allowed todepend on the martingale L.

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

1.1. Applications

During the last decade, the importance of designing efficient numerical methods to solve (1) has increased significantly because of various applications. For references, see El Karoui *et al.* (1997). Solving (1) may give access to the Föllmer and Schweizer (1991) 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. In finance, this decomposition plays a crucial role in valuing and hedging claims (with payoff $\phi(X_T)$) in incomplete markets: this is the concept of locally risk-minimizing strategies. In this instance, Y stands for the price and Z is related to the hedging strategy. The current BSDE framework applies to this financial setting if the martingale part of the traded assets X is driven by W, meaning that the jumps are incorporated only in the volatility and the non-traded assets. Finally, for the connection between BSDEs and dynamic risk measures, see Peng (2004). On the relation with semi-linear partial differential equations (possibly with integral-differential operators), see Barles *et al.* (1997).

1.2. Where nonparametric regressions come in

The first approximation of (1) is a time discretization using a time step $h = T/N : (t_k = kh)_{0 \leqslant k \leqslant N}$ denotes the discretization times. We set $\Delta W_k = W_{t_{k+1}} - W_{t_k}$ ($\Delta W_{l,k}$ componentwise) and X^N a relative approximation of X at these discretization times, obtained, say, 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

$$Y^N_{t_k} = \mathbb{E}_{t_k}(Y^N_{t_{k+1}}) + h\mathbb{E}_{t_k} f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k}), \qquad hZ^N_{t_k} = \mathbb{E}_{t_k}(Y^N_{t_{k+1}} \Delta W^*_k), \tag{2}$$

where $\mathbb{E}_{t_k}$ is the conditional expectation with respect to $\mathcal{F}_{t_k}$ and * 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

$$y^N_k(x) = \mathbb{E}\Big(y^N_{k+1}(X^N_{t_{k+1}}) + hf(t_k, X^N_{t_k}, y^N_{k+1}(X^N_{t_{k+1}}), z^N_k(X^N_{t_k}))|X^N_{t_k} = x\Big),$$

$$hz^N_k(x) = \mathbb{E}\Big(y^N_{k+1}(X^N_{t_{k+1}})\Delta W^*_k|X^N_{t_k} = x\Big).$$

We are thus faced with the iterative computation of N unknown regression functions. There are several ways to approximate a regression function: for example, kernel methods (see Bosq and Lecoutre 1987) or projection methods on vector spaces of functions (see Györfi *et al.* 2002). However, in comparison with the classic nonparametric regression problem, there is a further difficulty in our case because the N regression functions are nested: the regression function computed at time t_{k+1} is used to compute a new regression function at time t_k. Thus we have to find a way of approximating the unknown regression functions that fulfils two constraints: it must lead to a *nice* propagation of the error during the backward iteration and its complexity must be reasonable regarding the accuracy (keep in mind that as N goes to infinity, more and more regression functions have to be estimated).

Some approximation schemes have already been considered for solving this problem, in the case of Brownian filtration and diffusion processes for X. The first method involves 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 achieved either by replacing the Brownian motion by a random walk (see Briand *et al.* 2001; Ma *et al.* 2002) or by using quantization techniques (see Bally and Pagès 2003). The second method involves directly computing a nonparametric approximation of the regression function. Bouchard and Touzi (2004) use a technique based on Malliavin calculus integration-by-parts 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. In our paper, for the first time (to our knowledge) the case of a non-Brownian filtration (leading to $L \neq 0$) is considered for the numerical resolution of (1).

1.3. Our contributions

We also approximate the unknown regression functions using projections on vector spaces of functions. Using M Monte Carlo simulations of X^N, we solve at each discretization time t_k a least-squares problem to determine the 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 type of functions. In Gobet *et al.* (2005), we have already studied the influence of the parameters for a similar procedure, 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 evaluate, presumably converging to infinity as the dimension of the vector space increases. Therefore the practical use of these results remains questionable, in particular if one has to achieve a given accuracy by a joint convergence of N, M and the number of functions to infinity. In this work, we derive tractable error estimates that depend only on N, M and the number of functions.

Thus we obtain 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 (2003), we do not need quantization grids and can use more flexible choices of vector spaces of functions than Voronoï cells alone. 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 that in Bouchard and Touzi (2004) and generally results in greater accuracy. Finally, a significant advantage is that our approach is distribution-free with respect to X. This means that in our error estimates, we make very little use of the specific form of the model for X, which could be any Markov process. No non-degeneracy condition on X is required, which is a significant difference from a Malliavin calculus approach. We mention similar results recently obtained by Egloff (2005), for optimal stopping problems with a fixed number of dates N. His approximations of regression functions are more general than ours. However, his error bounds increase geometrically with N, which does not fit our current framework (in addition, we have to overcome difficulties related to Z).

1.4. Organization of this paper

In Section 2, we rigorously define the model, introduce notation used throughout the paper, explain the algorithm, state our main results and discuss the trade-off between accuracy and complexity. The proofs are postponed to Section 4. In Section 3 we present numerical tests which illustrate the error bounds derived in Section 2 for explicit choices of vector spaces of functions.

2. The algorithm

2.1. 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 two mutually independent processes: an $\mathbb{R}^q$-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 $\{\tilde{\mu}([0, t] \times A) = (\mu - \nu)([0, t] \times A)\}_{t\geqslant 0}$ 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

$$X_t = x + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s + \int_0^t \int_E \beta(s, X_{s^-}, e)\tilde{\mu}(\mathrm{d}s, \mathrm{d}e), \tag{3}$$

which is uniquely defined under the following assumption.

Assumption 1. *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 $|\beta(t, x, e)| \leqslant c(1 \wedge |e|)$ and $|\beta(t, x, e) - \beta(t', x', e)| \leqslant c(|x - x'| + |t - t'|)(1 \wedge |e|)$ for any (t, x), $(t, x') \in [0, T] \times \mathbb{R}^d$ and $e \in E$.*

We denote by X^N a time-discretization of X (we may think of Euler schemes, see Jacod 2004 and references therein) and we suppose that X^N is a Markov chain, which satisfies

Assumption 2. *As N goes to infinity,* $\sup_{0\leqslant k\leqslant N}\mathbb{E}|X_{t_k} - X^N_{t_k}|^2 \to 0$.

We require a convergence of X^N towards X in the L_2-norm because we aim to approximate the solution of (1) in the usual BSDE norm. We refer to Jacod (2004) for other convergence results.

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

Assumption 3. *The driver f satisfies the continuity estimate*

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

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 satisfies 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] the following assumption:

Assumption 4. *For some constant $C > 0$,*

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

This kind of assumption has been introduced in Gobet *et al.* (2005) and is quite natural since it is fulfilled by X itself under Assumption 1. We now state a convergence result regarding the time discretization.

Theorem 1. *Under Assumptions 1–3, define the error*

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

where Y^N and Z^N are given by (2). *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, $e(N) = O(N^{-1})$.*

The proof, which is quite standard, is set out in full in Gobet and Lemor (2006).

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

2.2. Notation

We now introduce convenient notation for describing the algorithm.

2.2.1. Localization

We define localized versions of the Brownian increments and of the functions f, ϕ,

$$[\Delta W_{l,k}]_w = (-R_0\sqrt{h}) \vee \Delta W_{l,k} \wedge (R_0\sqrt{h}),$$

$$f^R(t, x, y, z) = f(t, -R_1 \vee x_1 \wedge R_1, \dots, -R_d \vee x_d \wedge R_d, y, z),$$

$$\phi^R(x) = \phi(-R_1 \vee x_1 \wedge R_1, \dots, -R_d \vee x_d \wedge R_d),$$

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

$$Y_{t_k}^{N,R} = \mathbb{E}_{t_k}(Y_{t_{k+1}}^{N,R}) + h\mathbb{E}_{t_k} f^R(t_k, X_{t_k}^N, Y_{t_{k+1}}^{N,R}, Z_{t_k}^{N,R}), \tag{4}$$

$$hZ_{t_k}^{N,R} = \mathbb{E}_{t_k}(Y_{t_{k+1}}^{N,R}[\Delta W_k]_w^*), \tag{5}$$

and $Y_{t_N}^{N,R} = \phi^R(X_{t_N}^N)$. With the same arguments as for (Y^N, Z^N), we can easily see that $Y_{t_k}^{N,R} = y_k^{N,R}(X_{t_k}^N)$ and $Z_{t_k}^{N,R} = z_k^{N,R}(X_{t_k}^N)$ for deterministic functions $y_k^{N,R}(\cdot)$ and $z_k^{N,R}(\cdot)$. Moreover, the functions $y_k^{N,R}$ and $\sqrt{h}z_k^{N,R}$ are Lipschitz continuous uniformly in R and N (see Proposition 1 below). But the main motivation for this localization is to provide *bounded* (unknown) regression functions $y_k^{N,R}$ and $z_k^{N,R}$: one has $\|y_k^{N,R}\|_\infty \leqslant C_y(R)$ and $\|z_{l,k}^{N,R}\|_\infty \leqslant C_z(R)$ (for details on these upper bounds, see again Proposition 1). This boundedness property plays an important role in the derivation of error bounds.

2.2.2. Vector spaces of functions

At each discretization time t_k, $0 \leqslant k \leqslant N-1$, we choose $q+1$ deterministic functions bases $(p_{l,k}(\cdot))_{0\leqslant l\leqslant q}$ and we look for an approximation of $y_k^{N,R}(\cdot)$ (or $z_{l,k}^{N,R}(\cdot)$) in the vector space spanned by the basis $p_{0,k}$ (or $p_{l,k}$). Each basis $p_{l,k}$ is considered as a vector of functions, of dimension $K_{l,k}$. The vector space of functions spanned by $p_{l,k}$ is denoted by $\mathcal{P}_{l,k}$, that is, $\mathcal{P}_{l,k} = \{\alpha \cdot p_{l,k}(\cdot), \alpha \in \mathbb{R}^{K_{l,k}}\}$. A common example is the hypercube basis HC used in Gobet *et al.* (2005). In this case, $p_{l,k}$ does not depend on l or k and its dimension is denoted by K. Choose a domain $D \subset \mathbb{R}^d$ centred on $X_0 = x$, that is, $D = \prod_{i=1}^d]x_i - a, x_i + a]$, and partition it into small hypercubes of edge δ. Thus, $D = \cup_{i_1,\dots,i_d} D_{i_1,\dots,i_d}$ where $D_{i_1,\dots,i_d} =]x_1 - a + i_1\delta, x_1 - a + (i_1+1)\delta] \times \cdots \times]x_d - a + i_d\delta, x_d - a + (i_d+1)\delta]$. Then we define $p_{l,k}$ as the indicator functions associated with this set of hypercubes: $p_{l,k}(\cdot) = (\mathbf{1}_{D_{i_1,\dots,i_d}}(\cdot))_{i_1,\dots,i_d}$.

2.2.3. Monte Carlo simulations

The evaluation of the different projection coefficients α will be obtained using M independent Monte Carlo simulations of $(X^N_{t_k})_{0\leqslant k\leqslant N}$ and $(\Delta W_k)_{0\leqslant k\leqslant N-1}$. We denote these Monte Carlo simulations by $(X^{N,m}_{t_k})_{1\leqslant m\leqslant M,0\leqslant k\leqslant N}$ and $(\Delta W^m_k)_{1\leqslant m\leqslant M,0\leqslant k\leqslant N-1}$.

For the sake of brevity of notation, we write $p_{l,k}(X^{N,m}_{t_k}) = p^m_{l,k}$. We define by $B^M_{l,k}$ the $M \times K_{l,k}$ matrix with rows $(p^m_{l,k})^*$. We denote by $K^M_{l,k}$ the rank of $B^M_{l,k}$ ($K^M_{l,k}$ is random and lower than $K_{l,k}$).

2.2.4. 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

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

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

2.2.5. Constants

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

2.3. Description of the algorithm

The functions $(y^{N,R}_k)_{0\leqslant k\leqslant N-1}$ and $(z^{N,R}_{l,k})_{1\leqslant l\leqslant q,0\leqslant k\leqslant N-1}$ are approximated by $(y^{N,R,M}_k)_{0\leqslant k\leqslant N-1}$ and $(z^{N,R,M}_{l,k})_{1\leqslant l\leqslant q,0\leqslant k\leqslant N-1}$, which are constructed in a backward manner.
Initialization. For $k = N$ take $y^{N,R,M}_N(\cdot) = \phi^R(\cdot)$.
Iteration. For $k = N-1, \dots, 0$, solve the q least-squares problems

$$\alpha^M_{l,k} = \arg\inf_\alpha \frac{1}{M}\sum_{m=1}^{M} \left| y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}})\frac{[\Delta W^m_{l,k}]_w}{h} - \alpha \cdot p^m_{l,k}\right|^2. \tag{6}$$

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

$$\frac{1}{M}\sum_{m=1}^{M} |y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}}) + hf^R(t_k, X^{N,m}_{t_k}, y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}})\text{h}, [\alpha^M_{l,k} \cdot p^m_{l,k}]_z) - \alpha \cdot p^m_{0,k}|^2. \tag{7}$$

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

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

In the least-squares problems (6)–(7), whenever convenient, we can suppose (as done, for example, in the proof of Theorem 11.1 in Györfi *et al.* 2002) that, for $0 \leqslant l \leqslant q$, $p_{l,k}$ is a complete orthonormal system in $\mathcal{P}_{l,k}$, with respect to the empirical scalar product $\langle \cdot, \cdot \rangle_{k,M}$ defined by $\langle \psi_1, \psi_2 \rangle_{k,M} = M^{-1} \sum_{m=1}^{M} \psi_1(X_{t_k}^{N,m}) \psi_2(X_{t_k}^{N,m})$. Of course these orthonormal systems depend on the simulations $(X_{t_k}^{N,m})_{1 \leqslant m \leqslant M}$ and their ranks $(K_{l,k}^M)_{0 \leqslant l \leqslant q}$ satisfy $K_{l,k}^M \leqslant K_{l,k}$. These orthonormal systems can easily be computed using a singular value decomposition (see Golub and Van Loan 1996). With this choice, $B_{l,k}^M$ is now of dimension $M \times K_{l,k}^M$ and $(B_{l,k}^M)^* B_{l,k}^M / M = \mathrm{Id}$ $(0 \leqslant l \leqslant q)$, and the solutions of (6)–(7) are given by:

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

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

2.4. Main convergence results

The error on the unknown regression functions $(y_k^{N,R})_{0 \leqslant k \leqslant N-1}$ and $(z_{l,k}^{N,R})_{1 \leqslant l \leqslant q, 0 \leqslant k \leqslant N-1}$ is now estimated in the following theorem, which is our main result.

Theorem 2. *Suppose that Assumptions 1–4 hold, and let* $\beta \in]0, 1]$. *Then there exists a constant C (independent of* β*) such that:*

$$\max_{0 \leqslant k \leqslant N} \mathbb{E} \frac{1}{M} \sum_{m=1}^{M} |y_k^{N,R}(X_{t_k}^{N,m}) - y_k^{N,R,M}(X_{t_k}^{N,m})|^2 + h \mathbb{E} \sum_{k=0}^{N-1} \frac{1}{M} \sum_{m=1}^{M} |z_k^{N,R}(X_{t_k}^{N,m}) - z_k^{N,R,M}(X_{t_k}^{N,m})|^2$$

$$\leqslant C \frac{C_y(R)^2}{M} \sum_{k=0}^{N-1} \sum_{l=0}^{q} \mathbb{E}(K_{l,k}^M) + C h^\beta$$

$$+ C \sum_{k=0}^{N-1} \left\{ \inf_\alpha \mathbb{E} |y_k^{N,R}(X_{t_k}^N) - \alpha \cdot p_{0,k}(X_{t_k}^N)|^2 + \sum_{l=1}^{q} \inf_\alpha \mathbb{E} |\sqrt{h} z_{l,k}^{N,R}(X_{t_k}^N) - \alpha \cdot p_{l,k}(X_{t_k}^N)|^2 \right\}$$

$$+ C \frac{C_y(R)^2}{h} \sum_{k=0}^{N-1} \left\{ \mathbb{E} \left(K_{0,k}^M \exp\left(- \frac{M h^{\beta+2}}{72 C_y(R)^2 K_{0,k}^M} \right) \exp\left(C K_{0,k+1} \log \frac{C C_y(R) (K_{0,k}^M)^{1/2}}{h^{(\beta+2)/2}} \right) \right) \right.$$

$$+ h \mathbb{E} \left(K_{l,k}^M \exp\left(- \frac{M h^{\beta+1}}{72 C_y(R)^2 R_0^2 K_{l,k}^M} \right) \exp\left(C K_{0,k+1} \log \frac{C C_y(R) R_0 (K_{l,k}^M)^{1/2}}{h^{(\beta+1)/2}} \right) \right)$$

$$\left. + \exp\left(C K_{0,k} \log \frac{C C_y(R)}{h^{(\beta+2)/2}} \right) \exp\left(- \frac{M h^{\beta+2}}{72 C_y(R)^2} \right) \right\},$$

where we adopt the convention that $K_{0,N} = 0$.

Remark 1. 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 $(X^{N,m})_{1\leqslant m\leqslant M}$, that is, we can bound $\max_{0\leqslant k\leqslant N}\mathbb{E}|y_k^{N,R}(X_{t_k}^N) - y_k^{N,R,M}(X_{t_k}^N)|^2 + h\mathbb{E}\sum_{k=0}^{N-1}|z_k^{N,R}(X_{t_k}^N) - z_k^{N,R,M}(X_{t_k}^N)|^2$. This extension is valid if we add to the upper bound a term $CC_y(R)^2\log(M)M^{-1} \times \sum_{k=0}^{N-1}\sum_{l=0}^{q}K_{l,k}$, which is essentially of the same order as the others (up to the log factor).

Remark 2. Of course, the inequality $K_{l,k}^M \leqslant K_{l,k}$ 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 work.

The terms $C_y(R)^2M^{-1}\sum_{l=0}^{q}\mathbb{E}(K_{l,k}^M)$ and $\inf_\alpha \mathbb{E}(|y_k^{N,R}(X_{t_k}^N) - \alpha\cdot p_{0,k}(X_{t_k}^N)|^2) + \sum_{l=1}^{q}\inf_\alpha \mathbb{E}(|\sqrt{h}z_{l,k}^{N,R}(X_{t_k}^N) - \alpha\cdot p_{l,k}(X_{t_k}^N)|^2)\}$ are classic error terms which arise when one approximates a regression function from independent and identically distributed observations using projections on a finite set of functions (see, for example, Györfi *et al.* 2002: Theorem 11.1). They are summed 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.

Theorem 2 improves Theorem 3 in Gobet *et al.* (2005) because the error is not estimated in terms of the fourth 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 easily be used in practice. Indeed, it is easy to establish the following corollary and we refer to Gobet and Lemor (2006) for a complete proof.

Corollary 1. *Suppose that Assumptions 1–4 hold and that $p_{l,k}$ is the hypercube basis* HC *of edge δ for which $\inf_\alpha \mathbb{E}|y_k^{N,R}(X_{t_k}^N) - \alpha\cdot p_{0,k}(X_{t_k}^N)|^2 + \sum_{l=1}^{q}\inf_\alpha \mathbb{E}|\sqrt{h}z_{l,k}^{N,R}(X_{t_k}^N) - \alpha p_{l,k}(X_{t_k}^N)|^2 \leqslant C\delta^2$ (see Gobet* et al. *2005). Neglecting the impact of the localization, in order to obtain an error of order h^β, it is enough to choose δ and M as $\delta = C_1^{-1}h^{(\beta+1)/2}$ and $M = C_1h^{-(\beta+2+d(\beta+1))}|\log(h)|$ for a sufficiently large constant C_1.*

2.5. Accuracy and complexity of the algorithm

Here we wish to compare the accuracy–complexity ratio of this algorithm with that of other numerical schemes.

If we denote by $\mathcal{C}$ the complexity of the algorithm, or equivalently the computational effort, it is easy to see that the squared error obtained with the basis HC and $\beta=1$ is of order $\mathcal{C}^{-1/(4+2d)}$ (Table 1) we refer to Gobet and Lemor (2006) for the details. Therefore this algorithm has the same complexity as the less natural algorithm described and applied to reflected BSDEs in Gobet and Lemor (2006) where extra simulations of $(X_{t_{k+1}}^{N,m})_{1\leqslant m\leqslant M}$ are needed, but it has a worse trade-off between accuracy and complexity. Nevertheless,

Table 1. Squared error for different algorithms with respect to the complexity $\mathcal{C}$ with basis HC and $\beta = 1$ (see Gobet and Lemor 2006 for details)

Section 2	Gobet and Lemor (2006)	Bouchard and Touzi (2004)
$\mathcal{C}^{-1/(4+2d)}$	$\mathcal{C}^{-1/(4+d)}$	$\mathcal{C}^{-1/(13+d)}$

Note: The squared error for the Bouchard and Touzi algorithm is for the case where X is Brownian motion or geometric Brownian motion.

both algorithms behave the same in the numerical tests we performed, and we think that our estimations in Theorem 2 are not optimal.

Compared to Bouchard and Touzi (2004), in the case of the geometric Brownian motion model, our algorithm is theoretically more efficient for $d \leqslant 9$ 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.

3. Numerical tests

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

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 $d = 1$, $\mathrm{d}X_t/X_t = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, 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 nonlinear driver f is defined by $f(t, x, y, z) = -\theta z - ry + (y - z/\sigma)^-(R - r)$, 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.

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 $j = 1, \ldots$ 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 and for all to [40, 180]. As observed in Corollary 1, the choice $\delta = 50/(\sqrt{2})^{(j-1)(\beta+1)/2} \sim Ch^{(\beta+1)/2}$ (50 being arbitrary and chosen to start with only three functions) makes the algorithm converge at rate h^β (for the squared error). Now it remains to adjust M as a function of N and δ, or equivalently h and β. According to Corollary 1, M must be such that $Mh^{\beta+2+(\beta+1)} = Mh^{3+2\beta} \to \infty$ up to a logarithmic factor. The following tests 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 $\alpha_M \geqslant 3$ clearly converges towards the reference price but we observe in Figure 3 that excessively large 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. As usual, it is important to properly balance the bias and variance terms.

Finally, we observe in this last example that the empirical levels of convergence of α_M are better than those expected from the condition $Mh^{3+2\beta} \to \infty$: this indicates that the bound of Theorem 2 is not optimal.

In Figure 4 we again test $\beta = 0.2$ but with the basis HC(1,0) which is analogous to HC: it involves 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 functions. In this example, the basis HC(1,0) speeds up the overall convergence. This gives, in our opinion, a fundamental improvement, compared to the quantization method that can be viewed as using only indicator function bases.

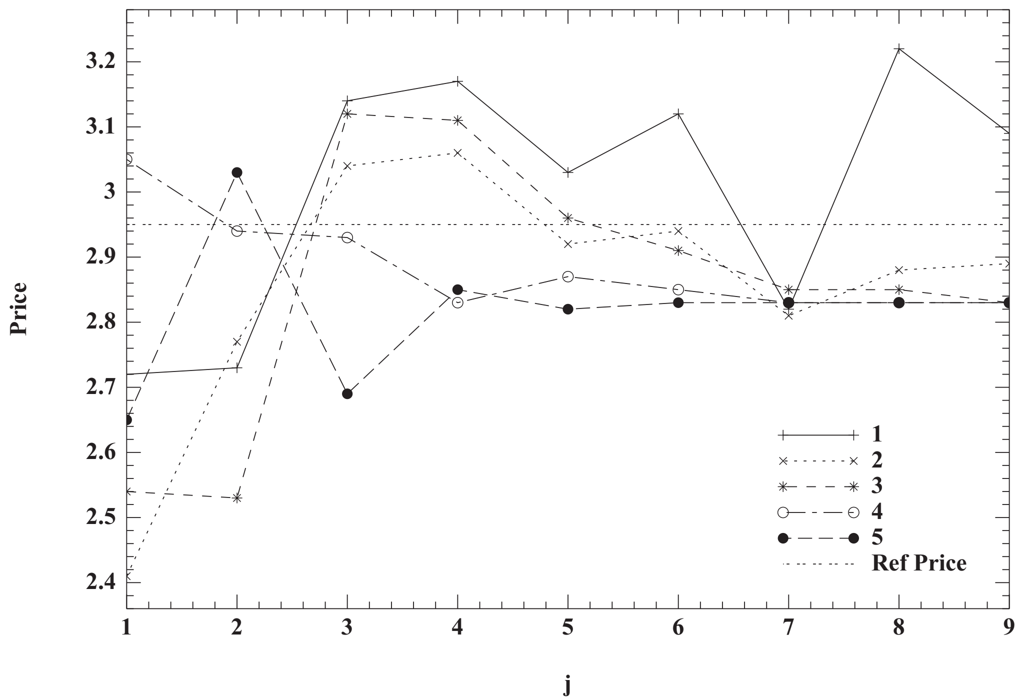

Figure 1. Basis HC, $\beta = 0.2$, $\alpha_M^* = 3.4$, $\alpha_M = 1, 2, 3, 4, 5$.

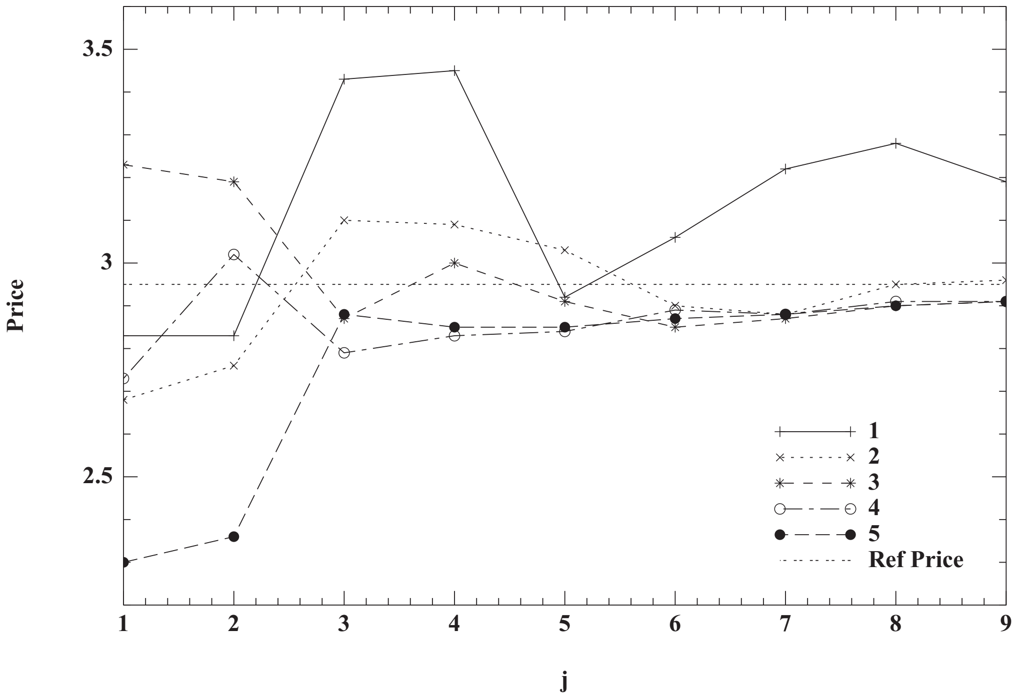

Figure 2. Basis HC, $\beta = 1$, $\alpha_M^* = 5$, $\alpha_M = 1, 2, 3, 4, 5$.

Finally, we consider an example taken from Heath *et al.* (2001). In that paper, the authors approximate via partial differential equation methods the local risk-minimization price (see Föllmer and Schweizer 1991) of a put option in the Heston stochastic volatility model. The dynamics of the asset price X and of the square of the volatility F are

$$\frac{\mathrm{d}X_t}{X_t} = \gamma F_t \mathrm{d}t + \sqrt{F_t}\,\mathrm{d}W_t, \qquad \mathrm{d}F_t = \kappa(\theta - F_t)\mathrm{d}t + \Sigma\sqrt{F_t}\,\mathrm{d}W'_t,$$

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

$$-\mathrm{d}Y_t = -\left(r_t Y_t + \frac{Z_t}{\sqrt{F_t}}(\gamma F_t - r_t)\right)\mathrm{d}t - Z_t\,\mathrm{d}W_t - \mathrm{d}L_t, \qquad Y_T = (K - X_T)_+.$$

Taking r to be 0, this gives a driver $f(t, x, F, y, z) = -\gamma z\sqrt{F}$ which is not globally Lipschitz. We nevertheless apply our algorithm for $\alpha_M = 3$ and present the results in 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 vector space of functions 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 algorithm's output still converges towards the reference price.

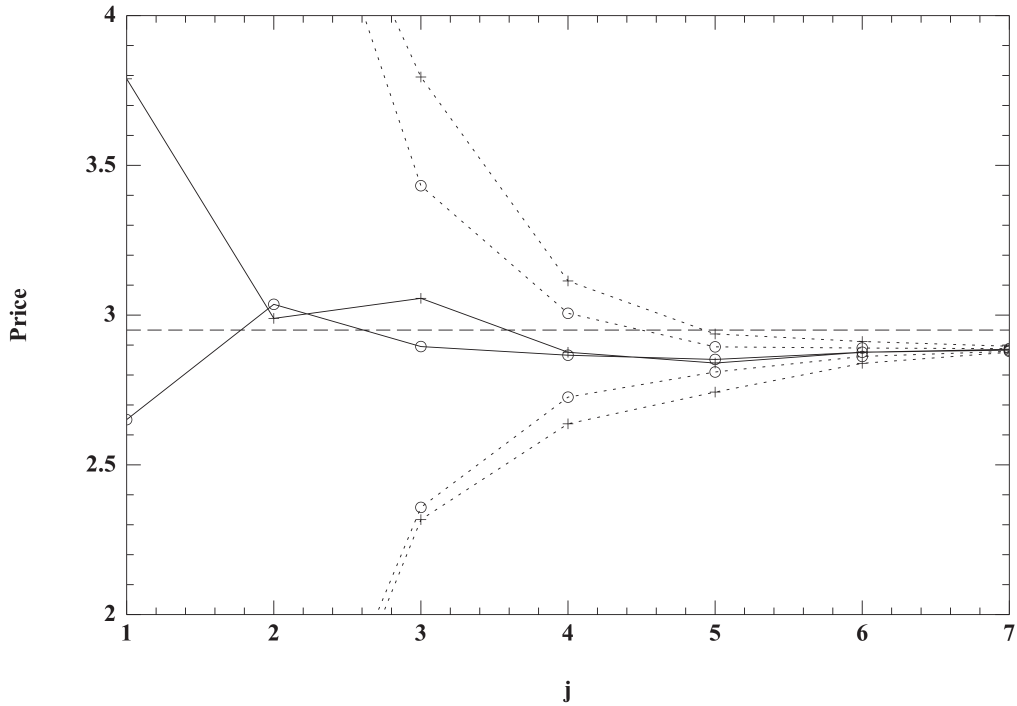

Figure 3. Basis HC, $\beta = 1$, $\alpha_M^* = 5$, $\alpha_M = 6$ (cross markers), $\alpha_M = 7$ (circle markers). Prices are shown by solid lines, and upper and lower 0.95 confidence intervals by dotted lines.

4. Proof of Theorem 2

To prove Theorem 2, we first need two (easy) propositions, proved in Lemor *et al.* (2005). The first states that the couple $(Y^{N,R}, Z^{N,R})$ is bounded and satisfies a Lipschitz property.

Proposition 1. *Under Assumptions 1–3 there exists a constant C such that, for all k,* $0 \leqslant k \leqslant N$,

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

where $\|\phi^R\|_\infty = \sup_x|\phi^R(x)| \leqslant C(1+|R|)$ *and* $\|f^R\|_\infty = \sup_{(t,x)}|f^R(t, x, 0, 0)| \leqslant C(1+|R|)$.

In addition under Assumption 4, 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* $|y_k^{N,R}(x) - y_k^{N,R}(x')| + \sqrt{h}|z_k^{N,R}(x) - z_k^{N,R}(x')| \leqslant C|x - x'|$ *uniformly in* k_0, *N and R.*

The second states an error bound regarding the localization.

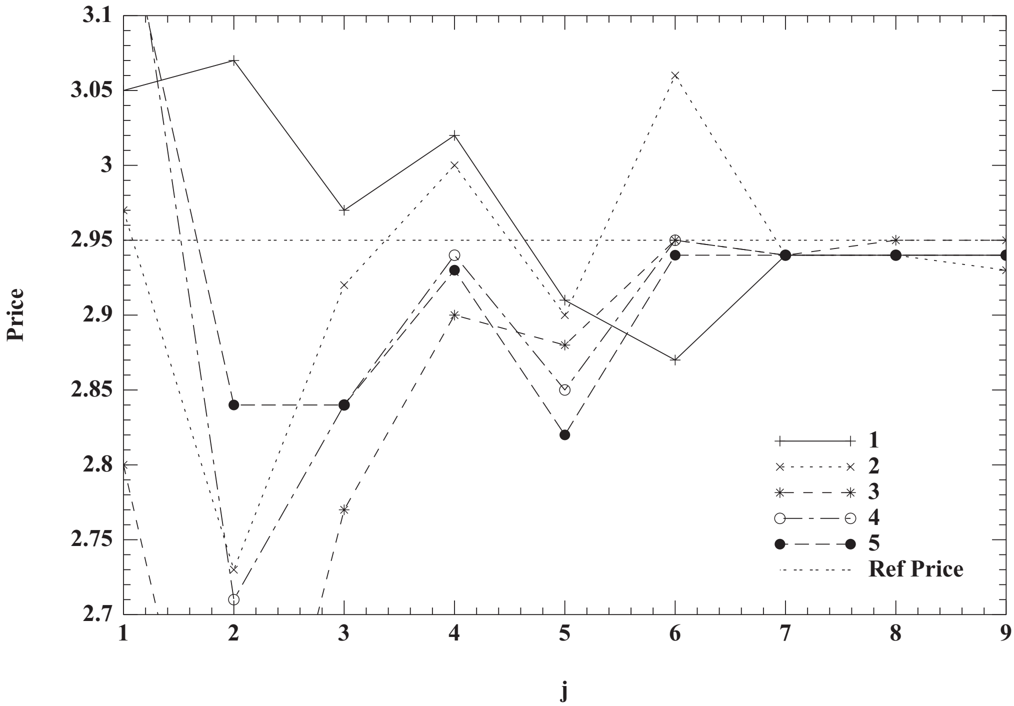

Figure 4. Basis HC(1, 0), $\beta = 0.2$, $\alpha_M^* = 3.4$, $\alpha_M = 1, 2, 3, 4, 5$.

Proposition 2. *Under Assumptions 1–4 there exists a constant C such that, for h small enough,*

$$\max_{0 \leqslant k \leqslant N} \mathbb{E}|Y_{t_k}^{N,R} - Y_{t_k}^N|^2 + h\mathbb{E} \sum_{k=0}^{N-1} |Z_{t_k}^{N,R} - Z_{t_k}^N|^2$$

$$\leqslant C\mathbb{E}|\phi(X_{t_N}^N) - \phi^R(X_{t_N}^N)|^2 + C\frac{C_y(R)^2}{h} \sum_{k=0}^{N-1} \mathbb{E}\Big(|\Delta W_k|^2 \mathbf{1}_{|\Delta W_k| \geqslant R_0\sqrt{h}}\Big)$$

$$+ Ch\mathbb{E} \sum_{k=0}^{N-1} |f(t_k, X_{t_k}^N, Y_{t_{k+1}}^N, Z_{t_k}^N) - f^R(t_k, X_{t_k}^N, Y_{t_{k+1}}^N, Z_{t_k}^N)|^2.$$

As a consequence and since $|Y_{t_k}^N| + \sqrt{h}|Z_{t_k}^N| \leqslant C(1 + |X_{t_k}^N|)$ (see Gobet *et al.* 2005), we easily obtain that

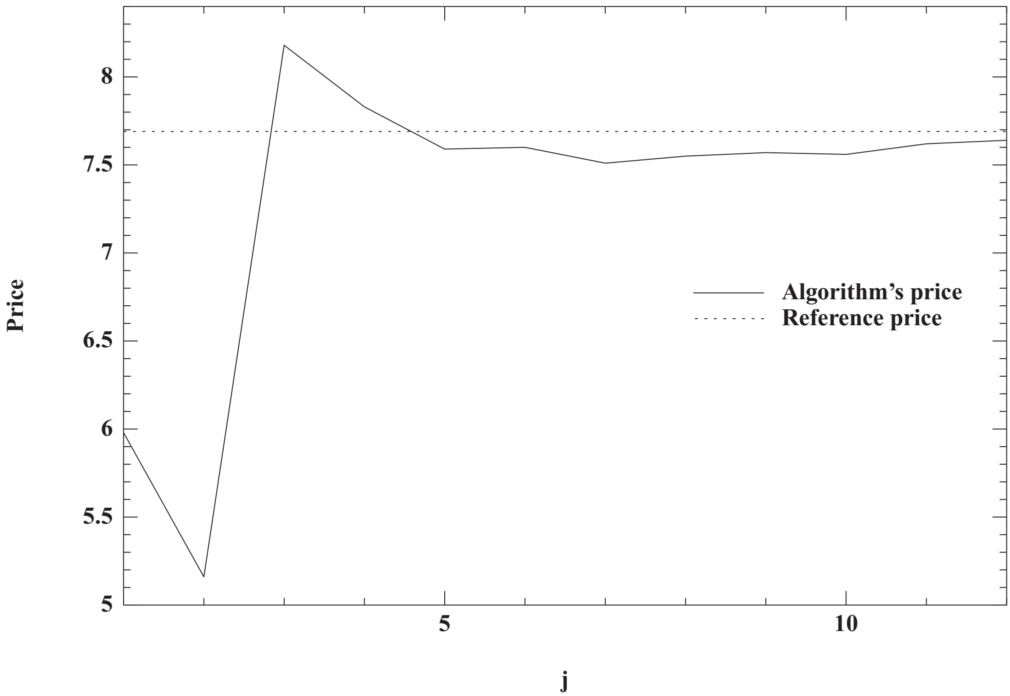

Figure 5. Basis HC, $\beta = 1$, $\alpha_M = 3$.

$$\max_{0 \leqslant k \leqslant N} \mathbb{E}|Y^{N,R}_{t_k} - Y^N_{t_k}|^2 + h\mathbb{E}\sum_{k=0}^{N-1} |Z^{N,R}_{t_k} - Z^N_{t_k}|^2$$

$$\leqslant \frac{C}{h} \max_{0 \leqslant k \leqslant N} \mathbb{E}\Big[(1 + |X^N_{t_k}|^2)1_{|X^N_{t_k}|>|R|}\Big] + C\frac{1 + |R|^2}{h} \exp\left(-\frac{R_0^2}{4}\right).$$

Hence for appropriate thresholds R going to infinity, the localization error converges to 0. Rates are available if in addition $\sup_{0 \leqslant k \leqslant N} \mathbb{E}|X^N_{t_k}|^p \leqslant C_p(1 + |x|^p)$ for some $p > 2$ (stronger moment conditions on the Lévy measure λ would lead to larger exponents p): indeed, the upper bound becomes

$$\frac{C_p}{h(1 + |R|)^{p-2}} + C\frac{1 + |R|^2}{h} \exp\left(-\frac{R_0^2}{4}\right)$$

and to obtain a contribution of order h (as in Theorem 1) it is enough to asymptotically set $R_i = h^{-2/(p-2)}$ $(i = 1, \ldots, d)$ 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 observed in the numerical tests.

We can now turn to the proof of Theorem 2. The proof is technical and we divide it into several steps. First, we introduce additional notation, closely related to non-parametric regression arguments. Second, 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.

4.1. Extra notation for the proofs

4.1.1. 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 on $X^{N,m}_{t_k}$, 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

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

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

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

4.1.2. Norms

For a function ψ, we define $(0 \leqslant k \leqslant N)$

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

4.1.3. Projection coefficients

In addition to the coefficients $(\alpha^M_{l,k})_{0\leqslant l\leqslant q}$ defined by (6) and (7), we need other coefficients in the proofs below. Thus, we define the projection coefficients $(\tilde{\alpha}^M_{l,k})_{1\leqslant l\leqslant q}$:

$$\tilde{\alpha}^M_{l,k} = \arg\inf_{\alpha} \frac{1}{M}\sum_{m=1}^{M} \left| y^{N,R,M}_{k+1}(\tilde{X}^{N,m}_{t_{k+1}}) \frac{[\Delta \tilde{W}^m_{l,k}]_w}{h} - \alpha \cdot p^m_{l,k} \right|^2, \tag{8}$$

and $\tilde{\alpha}^M_{0,k}$ is defined as the minimizer of

$$\frac{1}{M}\sum_{m=1}^{M} |y^{N,R,M}_{k+1}(\tilde{X}^{N,m}_{t_{k+1}}) + hf^R(t_k, X^{N,m}_{t_k}, y^{N,R,M}_{k+1}(\tilde{X}^{N,m}_{t_{k+1}}), [\tilde{\alpha}^M_{l,k} \cdot p^m_{l,k}]_z) - \alpha \cdot p^m_{0,k}|^2. \tag{9}$$

We also define the projection coefficients $(\tilde{\beta}^M_{l,k})_{1\leqslant l\leqslant q}$:

$$\tilde{\beta}^M_{l,k} = \arg\inf_{\alpha} \frac{1}{M} \sum_{m=1}^{M} \left| y^{N,R}_{k+1}(\tilde{X}^{N,m}_{t_{k+1}}) \frac{[\Delta \tilde{W}^m_{l,k}]_w}{h} - \alpha \cdot p^m_{l,k} \right|^2, \tag{10}$$

and $\tilde{\beta}^M_{0,k}$ is defined as the minimizer of

$$\frac{1}{M} \sum_{m=1}^{M} |y^{N,R}_{k+1}(\tilde{X}^{N,m}_{t_{k+1}}) + hf^R(t_k, X^{N,m}_{t_k}, y^{N,R}_{k+1}(\tilde{X}^{N,m}_{t_{k+1}}), z^{N,R}_{l,k}(X^{N,m}_{t_k})) - \alpha \cdot p^m_{0,k}|^2. \tag{11}$$

We emphasize the differences between these projection coefficients: for $(\alpha^M_{l,k})_{0 \leqslant l \leqslant q}$ and $(\tilde{\alpha}^M_{l,k})_{0 \leqslant l \leqslant q}$, the function $y^{N,R,M}_{k+1}$ is fixed and we estimate $\alpha^M_{l,k}$ from $(X^{N,m}_{t_k}, X^{N,m}_{t_{k+1}}, \Delta W^m_k)_{1 \leqslant m \leqslant M}$, whereas we estimate $\tilde{\alpha}^M_{l,k}$ from $(X^{N,m}_{t_k}, \tilde{X}^{N,m}_{t_{k+1}}, \Delta \tilde{W}^m_k)_{1 \leqslant m \leqslant M}$. As for $(\tilde{\beta}^M_{l,k})_{0 \leqslant l \leqslant q}$, we also estimate from $(X^{N,m}_{t_k}, \tilde{X}^{N,m}_{t_{k+1}}, \Delta \tilde{W}^m_k)_{1 \leqslant m \leqslant M}$ but knowing the true functions $y^{N,R}_{k+1}(\cdot)$ and $z^{N,R}_k(\cdot)$. We note that $\alpha^M_{0,k} = \alpha^{1,M}_{0,k} + \alpha^{2,M}_{0,k}$, where

$$\alpha^{1,M}_{0,k} = \arg\inf_{\alpha} \frac{1}{M} \sum_{m=1}^{M} |y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}}) - \alpha \cdot p^m_{0,k}|^2,$$

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

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

4.1.4. Conditional expectations

We write $\mathcal{F}^M$ for the σ-algebra generated by $((X^{N,m}_{t_k})_{0 \leqslant k \leqslant N}, (\Delta W^m_k)_{0 \leqslant k \leqslant N-1})_{1 \leqslant m \leqslant M}$ and $\mathbb{E}^M$ for the conditional expectation with respect to $\mathcal{F}^M$. We denote by $\mathbb{E}^M_k$ ($\mathbb{P}^M_k$) the conditional expectation (conditional probability) with respect to the σ-algebra generated by $((X^{N,m}_{t_i})_{0 \leqslant i \leqslant k}, (\Delta W^m_i)_{0 \leqslant i \leqslant k-1})_{1 \leqslant m \leqslant M}$.

4.1.5. Error terms

We define the following events which depend on β and the projection coefficients

$$A^M_{0,k} = \left\{ \frac{1}{M} \sum_{m=1}^{M} |p^m_{0,k} \cdot (\alpha^{1,M}_{0,k} - \tilde{\alpha}^{1,M}_{0,k})|^2 < h^{\beta+2} \right\},$$

$$A^M_{l,k} = \left\{ \frac{1}{M} \sum_{m=1}^{M} |p^m_{l,k} \cdot (\alpha^M_{l,k} - \tilde{\alpha}^M_{l,k})|^2 < h^{\beta} \right\},$$

$$A^M_k = \{\forall \psi \in [\mathcal{P}_{0,k}]_y - y^{N,R}_k : \|\psi\|_{k,\tilde{M}} - \|\psi\|_{k,M} < h^{(\beta+2)/2}\}.$$

The probabilities of these events are evaluated in Proposition 4 below. We also deal with the following quantities whose bounds are given in Proposition 4:

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

$$T_{3,l,k}^M = \mathbb{E}\|\mathbb{E}^M(\tilde\beta_{l,k}^M)\cdot p_{l,k} - z_{l,k}^{N,R}\|_{k,M}^2, \qquad T_{4,l,k}^M = \mathbb{E}\|\{\tilde\alpha_{l,k}^{M} - \mathbb{E}^M(\tilde\alpha_{l,k}^{M})\}\cdot p_{l,k}\|_{k,M}^2.$$

4.1.6. Covering numbers

In the proofs below we use random covering numbers. We refer to Györfi *et al.* (2002) for a complete description. However, for the sake of completeness, we briefly recall here that if $\mathcal{G}$ is a class of functions and $x_1^M = (x_1, \dots, x_M)$ are M points in $\mathbb{R}^d$, $\mathcal{N}_2(\varepsilon, \mathcal{G}, x_1^M)$ $(\varepsilon > 0)$ is the minimal $p \in \mathbb{N}$ such that there exist functions $g_1, \dots, g_p$, such that for all $g \in \mathcal{G}$ we can find a $j \in \{1 \dots, p\}$ with $(M^{-1}\sum_{m=1}^M |g(x_m) - g_j(x_m)|^2)^{1/2} < \epsilon$. To simplify, we adopt the notation $\mathcal{N}_2(\epsilon, k) = \mathcal{N}_2(\epsilon, [\mathcal{P}_{0,k}]_y, (X_{t_k}^{N,m}, \tilde X_{t_k}^{N,m})_{1\leqslant m\leqslant M})$.

4.2. Propagation of the error

Our main tool is the following result.

Proposition 3. *Under the assumptions of Theorem 2, for* $0 \leqslant k \leqslant N-1$,

$$\mathbb{E}\|y_k^{N,R} - y_k^{N,R,M}\|_{k,M}^2$$

$$\leqslant (1+Ch)\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}^2 + C\{T_{1,k}^M + T_{2,k}^M\} + Ch\sum_{l=1}^q \{T_{3,l,k}^M + T_{4,l,k}^M\}$$

$$+ C\frac{C_y(R)^2}{h}\left\{\mathbb{P}([A_{0,k}^M]^c) + h\sum_{l=1}^q \mathbb{P}([A_{l,k}^M]^c) + \mathbb{P}([A_{k+1}^M]^c)\right\} + Ch^{\beta+1}.$$

If we have this result and Proposition 4 which estimates each contribution, it is easy to complete the error estimation on Y in Theorem 2. Then, an easy calculation (see Lemor 2005) enables us to deduce an error estimate for Z from that for Y.

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

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

We now introduce $\tilde\beta_{0,k}^M$ (see (11)) and, noting that $\mathbb{E}^M(\tilde\beta_{0,k}^M)$ is the minimizer of $M^{-1}\sum_{m=1}^M |y_k^{N,R}(X_{t_k}^{N,m}) - \alpha\cdot p_{0,k}^m|^2$, we apply Pythagoras' theorem to obtain

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

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 apply first Young's inequality and then Jensen's inequality to obtain

$$\begin{aligned}\mathbb{E}\|\{\alpha_{0,k}^{M} - \mathbb{E}^{M}(\tilde{\beta}_{0,k}^{M})\} \cdot p_{0,k}\|_{k,M}^{2} &\leqslant (1+\gamma h)\mathbb{E}\|\{\alpha_{0,k}^{1,M} - \mathbb{E}^{M}(\tilde{\beta}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^{2} \\ &\quad + \left(1+\frac{1}{\gamma h}\right)\mathbb{E}\|\{\alpha_{0,k}^{2,M} - \mathbb{E}^{M}(\tilde{\beta}_{0,k}^{2,M})\} \cdot p_{0,k}\|_{k,M}^{2} \\ &\leqslant (1+\gamma h)\mathbb{E}\|\{\alpha_{0,k}^{1,M} - \mathbb{E}^{M}(\tilde{\beta}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^{2} \\ &\quad + \left(1+\frac{1}{\gamma h}\right)\mathbb{E}\|\{\alpha_{0,k}^{2,M} - \tilde{\beta}_{0,k}^{2,M}\} \cdot p_{0,k}\|_{k,M}^{2}. \end{aligned} \tag{12}$$

We deal separately with the two terms on the right-hand side of (12). For the first term, we introduce $\tilde{\alpha}_{0,k}^{1,M}$ (see (9)) and use the definition of $A_{0,k}^{M}$ and $T_{2,k}^{M}$ to obtain

$$\begin{aligned}&\mathbb{E}\|\{\alpha_{0,k}^{1,M} - \mathbb{E}^{M}(\tilde{\beta}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^{2} \\ &\quad\leqslant (1+h^{-1})\mathbb{E}\|\{\alpha_{0,k}^{1,M} - \tilde{\alpha}_{0,k}^{1,M}\} \cdot p_{0,k}\|_{k,M}^{2} + (1+h)\mathbb{E}\|\{\tilde{\alpha}_{0,k}^{1,M} - \mathbb{E}^{M}(\tilde{\beta}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^{2} \\ &\quad\leqslant Ch^{\beta+1} + \frac{C}{h}C_y(R)^2\mathbb{P}\Big([A_{0,k}^{M}]^c\Big) + (1+h)(\mathbb{E}\|\{\tilde{\alpha}_{0,k}^{1,M} - \mathbb{E}^{M}(\tilde{\alpha}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^{2} \\ &\qquad + \mathbb{E}\|\{\mathbb{E}^{M}(\tilde{\alpha}_{0,k}^{1,M}) - \mathbb{E}^{M}(\tilde{\beta}_{0,k}^{1,M})\} \cdot p_{0,k}\|_{k,M}^{2}) \\ &\quad\leqslant Ch^{\beta+1} + \frac{C}{h}C_y(R)^2\mathbb{P}\Big([A_{0,k}^{M}]^c\Big) + (1+h)T_{2,k}^{M} \\ &\qquad + (1+h)\mathbb{E}\frac{1}{M}\sum_{m=1}^{M}|\mathbb{E}^{M}\{y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}) - y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})\}|^2. \end{aligned} \tag{13}$$

In the second and third inequalities, we have used the contraction property of the projection on $((p_{0,k}^{m})^*)_{1\leqslant m\leqslant M}$. Using this contraction property once again, we obtain for the second term on the right-hand side of (12),

$$\begin{aligned}&\mathbb{E}\|\{\alpha_{0,k}^{2,M} - \tilde{\beta}_{0,k}^{2,M}\} \cdot p_{0,k}\|_{k,M}^{2} \\ &\quad\leqslant \frac{Ch^2}{M}\,\mathbb{E}\sum_{m=1}^{M}|y_{k+1}^{N,R,M}(X_{t_{k+1}}^{N,m}) - y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m})|^2 + Ch^2\mathbb{E}\|z_k^{N,R,M} - z_k^{N,R}\|_{k,M}^{2}. \end{aligned} \tag{14}$$

Let us deal with the last term of the right-hand side of (14). For $1 \leqslant l \leqslant q$, we have that

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

as $[z_{l,k}^{N,R}]_z = z_{l,k}^{N,R}$. Next, we introduce $\tilde{\beta}_{l,k}^M$ (see (10)) and as $\mathbb{E}^M(\tilde{\beta}_{l,k}^M)$ is the minimizer of $M^{-1}\sum_{m=1}^M |z_{l,k}^{N,R}(X_{t_k}^{N,m}) - \alpha \cdot p_{l,k}^m|^2$, we obtain

$$\begin{aligned}
&\mathbb{E}\|\alpha_{l,k}^M \cdot p_{l,k} - z_{l,k}^{N,R}\|_{k,M}^2 \\
&\quad = \mathbb{E}\|\mathbb{E}^M(\tilde{\beta}_{l,k}^M).p_{l,k} - z_{l,k}^{N,R}\|_{k,M}^2 + \mathbb{E}\|\{\mathbb{E}^M(\tilde{\beta}_{l,k}^M) - \alpha_{l,k}^M\} \cdot p_{l,k}\|_{k,M}^2 \\
&\quad \leqslant T_{3,l,k}^M + 3\mathbb{E}\|\{\mathbb{E}^M(\tilde{\beta}_{l,k}^M) - \mathbb{E}^M(\tilde{\alpha}_{l,k}^M)\} \cdot p_{l,k}\|_{k,M}^2 + 3\mathbb{E}\|\{\mathbb{E}^M(\tilde{\alpha}_{l,k}^M) - \tilde{\alpha}_{l,k}^M\} \cdot p_{l,k}\|_{k,M}^2 \\
&\qquad + 3\mathbb{E}\|\{\tilde{\alpha}_{l,k}^M - \alpha_{l,k}^M\} \cdot p_{l,k}\|_{k,M}^2 \\
&\quad \leqslant T_{3,l,k}^M + 3\mathbb{E}\|\{\mathbb{E}^M(\tilde{\beta}_{l,k}^M) - \mathbb{E}^M(\tilde{\alpha}_{l,k}^M)\} \cdot p_{l,k}\|_{k,M}^2 + 3T_{4,l,k}^M + Ch^\beta \\
&\qquad + C\,\frac{C_y(R)^2}{h}\,\mathbb{P}([A_{l,k}^M]^c).
\end{aligned} \tag{15}$$

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

$$\begin{aligned}
&\frac{3}{M}\,\mathbb{E}\sum_{m=1}^M |\{\mathbb{E}^M(\tilde{\beta}_{l,k}^M) - \mathbb{E}^M(\tilde{\alpha}_{l,k}^M)\} \cdot p_{l,k}^m|^2 \\
&\quad \leqslant \frac{3}{M}\,\mathbb{E}\sum_{m=1}^M \left|\mathbb{E}^M\left\{(y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m}) - y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}))\frac{[\Delta\tilde{W}_{l,k}^m]_w}{h}\right\}\right|^2 \\
&\quad \leqslant \frac{3}{Mh}\,\mathbb{E}\sum_{m=1}^M \{|y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m}) - y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m})|^2 \\
&\qquad - |\mathbb{E}^M(y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m}) - y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m}))|^2\}.
\end{aligned} \tag{16}$$

Substituting (13)–(16) into (12) gives

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

Letting $\gamma = C$, we obtain the following simplification:

$$
\begin{aligned}
&\mathbb{E}\|y_k^{N,R} - y_k^{N,R,M}\|_{k,M}^2 \qquad (17)\\
&\quad \leqslant (1+Ch)\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,\tilde{M}}^2 + C\,\frac{C_y(R)^2}{h}\left(\mathbb{P}([A_{0,k}^M]^c) + h\sum_{l=1}^q \mathbb{P}([A_{l,k}^M]^c)\right) \\
&\qquad + T_{1,k}^M + CT_{2,k}^M + Ch\sum_{l=1}^q \{T_{3,l,k}^M + T_{4,l,k}^M\} + Ch^{\beta+1} \\
&\qquad + Ch\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}^2 + C\,\frac{h}{M}\,\mathbb{E}\sum_{m=1}^M |y_{k+1}^{N,R}(\tilde{X}_{t_{k+1}}^{N,m}) - y_{k+1}^{N,R}(X_{t_{k+1}}^{N,m})|^2.
\end{aligned}
$$

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 Assumption 4(b)). To obtain the result of Proposition 3, the first term of the right-hand side should be changed to $\mathbb{E}\|y_{k+1}^{N,R} - y_{k+1}^{N,R,M}\|_{k+1,M}^2$. Thus, we use the definition of A_{k+1}^M to write:

$$
\begin{aligned}
&\mathbb{E}\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,\tilde{M}}^2\\
&\quad=\mathbb{E}\Big(\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,\tilde{M}}-\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,M}+\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,M}\Big)^2\\
&\quad\leqslant(1+h^{-1})\mathbb{E}\Big(\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,\tilde{M}}-\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,M}\Big)_+^2\\
&\qquad+(1+h)\mathbb{E}\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,M}^2\\
&\quad\leqslant Ch^{\beta+1}+C\,\frac{C_y(R)^2}{h}\,\mathbb{P}([A_{k+1}^M]^c)+(1+h)\mathbb{E}\|y_{k+1}^{N,R}-y_{k+1}^{N,R,M}\|_{k+1,M}^2. \qquad (18)
\end{aligned}
$$

Plugging (18) into (17) gives the result. □

4.3. Other estimates

Proposition 4. *Under the assumptions of Theorem 2, for* $0\leqslant k\leqslant N-1$,

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

Proof of the bound for $\mathbb{P}([A_{0,k}^M]^c)$**.** As already mentioned in Section 2.3, we suppose without loss of generality that $(B^M_{0,k})^* B^M_{0,k}/M = \mathrm{Id}$ and that $B^M_{0,k}$ is a matrix of dimension $M \times K^M_{0,k}$. Under this assumption, we can write (see our notation for $\mathbb{P}^M_k$)

$$\mathbb{P}([A^M_{0,k}]^c) = \mathbb{E}\Big(\mathbb{P}^M_k\Big(|\alpha^{1,M}_{0,k} - \tilde{\alpha}^{1,M}_{0,k}|_2^2 \geqslant h^{\beta+2}\Big)\Big). \tag{19}$$

By making the norm $|\cdot|_2$ in (19) explicit, we obtain

$$\begin{aligned}
&\mathbb{P}^M_k(|\alpha^{1,M}_{0,k} - \tilde{\alpha}^{1,M}_{0,k}|_2^2 \geqslant h^{\beta+2}) \\
&\quad = \mathbb{P}^M_k\left(\sum_{i=1}^{K^M_{0,k}} \left|\frac{1}{M}\sum_{m=1}^{M} p^m_{0,k,i}\{y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}}) - y^{N,R,M}_{k+1}(\tilde{X}^{N,m}_{t_{k+1}})\}\right|^2 \geqslant h^{\beta+2}\right) \\
&\quad \leqslant \sum_{i=1}^{K^M_{0,k}} \mathbb{P}^M_k\left(\left|\frac{1}{M}\sum_{m=1}^{M} p^m_{0,k,i}\{y^{N,R,M}_{k+1}(X^{N,m}_{t_{k+1}}) - y^{N,R,M}_{k+1}(\tilde{X}^{N,m}_{t_{k+1}})\}\right|^2 \geqslant \frac{h^{\beta+2}}{K^M_{0,k}}\right) \\
&\quad \leqslant \sum_{i=1}^{K^M_{0,k}} \mathbb{P}^M_k\left(\exists \psi \in [\mathcal{P}_{0,k+1}]_y : \left|\frac{1}{M}\sum_{m=1}^{M} p^m_{0,k,i}\{\psi(\tilde{X}^{N,m}_{t_{k+1}}) - \psi(X^{N,m}_{t_{k+1}})\}\right|^2 \geqslant \frac{h^{\beta+2}}{K^M_{0,k}}\right) \\
&\quad = \sum_{i=1}^{K^M_{0,k}} \mathbb{P}^M_k\left(\exists \psi \in [\mathcal{P}_{0,k+1}]_y : \left|\frac{1}{M}\sum_{m=1}^{M} p^m_{0,k,i} U_m\{\psi(\tilde{X}^{N,m}_{t_{k+1}}) - \psi(X^{N,m}_{t_{k+1}})\}\right| \geqslant \sqrt{\frac{h^{\beta+2}}{K^M_{0,k}}}\right),
\end{aligned}$$

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 $\tilde{X}^{N,m}_{t_{k+1}}$ and $X^{N,m}_{t_{k+1}}$ have the same law. We now introduce a covering $\mathcal{G}$ of $[\mathcal{P}_{0,k+1}]_y$ such that for all $\psi \in [\mathcal{P}_{0,k+1}]_y$, there exists $g \in \mathcal{G}$ such that

$$\frac{1}{2M}\sum_{m=1}^{M}\{|\psi(X^{N,m}_{t_{k+1}}) - g(X^{N,m}_{t_{k+1}})|^2 + |\psi(\tilde{X}^{N,m}_{t_{k+1}}) - g(\tilde{X}^{N,m}_{t_{k+1}})|^2\} \leqslant \frac{h^{\beta+2}}{18K^M_{0,k}}.$$

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

$$\mathbb{P}_k^M\left(\exists\psi\in[\mathcal{P}_{0,k+1}]_y:\left|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^m U_m\{\psi(\tilde{X}_{t_{k+1}}^{N,m})-\psi(X_{t_{k+1}}^{N,m})\}\right|\geqslant\sqrt{\frac{h^{\beta+2}}{K_{0,k}^M}}\right)$$

$$\leqslant\mathbb{P}_k^M\left(\exists g\in\mathcal{G}:\left|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^m U_m\{g(\tilde{X}_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})\}\right|\geqslant\frac{1}{3}\sqrt{\frac{h^{\beta+2}}{K_{0,k}^M}}\right)$$

$$\leqslant\mathcal{N}_2\left(\sqrt{\frac{h^{\beta+2}}{18K_{0,k}^M}},\,k+1\right)\max_{g\in\mathcal{G}}\mathbb{P}_k^M\left(\left|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^m U_m\{g(\tilde{X}_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})\}\right|\geqslant\frac{1}{3}\sqrt{\frac{h^{\beta+2}}{K_{0,k}^M}}\right).$$

To bound this last probability, we additionally condition on $(X_{t_{k+1}}^{N,m},\tilde{X}_{t_{k+1}}^{N,m})_{1\leqslant m\leqslant M}$ and denote by $\tilde{\mathbb{P}}_{k,k+1}^M$ the resulting conditional probability. We note that, if $H_m=p_{0,k,i}^m U_m\{g(\tilde{X}_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})\}$, $\tilde{\mathbb{E}}_{k,k+1}^M(H_m)=0$ and $|H_m|\leqslant 2C_y(R)|p_{0,k,i}^m|$. A combination of Hoeffding's inequality and $M^{-1}\sum_{m=1}^M|p_{0,k,i}^m|^2=1$ gives

$$\tilde{\mathbb{P}}_{k,k+1}^M\left(\left|\frac{1}{M}\sum_{m=1}^{M}p_{0,k,i}^m U_m\{g(\tilde{X}_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})\}\right|\geqslant\frac{1}{3}\sqrt{\frac{h^{\beta+2}}{K_{0,k}^M}}\right)$$

$$\leqslant 2\exp\left(-\frac{2Mh^{\beta+2}}{144C_y(R)^2K_{0,k}^M M^{-1}\sum_{m=1}^M|p_{0,k,i}^m|^2}\right)=2\exp\left(-\frac{Mh^{\beta+2}}{72C_y(R)^2K_{0,k}^M}\right).$$

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

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

$$\frac{1}{2M}\sum_{m=1}^{M}\{|\psi(\tilde{X}_{t_{k+1}}^{N,m})-g(\tilde{X}_{t_{k+1}}^{N,m})|^2+|\psi(X_{t_{k+1}}^{N,m})-g(X_{t_{k+1}}^{N,m})|^2\}\leqslant\frac{h^{\beta+1}}{18K_{l,k}^M\mathrm{R}_0^2}.$$

□

Proof of the bound for $\mathbb{P}([A_k^M]^c)$**.** We partially follow the proof of Theorem 11.2 in Györfi *et al.* (2002) and define the vector $(Z_m)_{1\leqslant m\leqslant 2M}$ by $(Z_m,Z_{M+m})=(\tilde{X}_{t_k}^{N,m},X_{t_k}^{N,m})$ if $U_m=1$ and $(Z_m,Z_{M+m})=(X_{t_k}^{N,m},\tilde{X}_{t_k}^{N,m})$ if $U_m=-1$, where $(U_m)_{1\leqslant m\leqslant M}$ 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 $\mathbb{P}([A_{0,k}^M]^c)$, we introduce a covering $\mathcal{G}$ (whose elements are bounded by $2C_y(R)$) of $[\mathcal{P}_{0,k}]_y-y_k^{N,R}$ such that for all $\psi\in[\mathcal{P}_{0,k}]_y-y_k^{N,R}$, there exists a $g\in\mathcal{G}$ such that

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

Thus, we can write that

$$\mathbb{P}([A_k^M]^c)$$

$$= \mathbb{P}\left(\exists \psi \in [\mathcal{P}_{0,k}]_y - y_k^{N,R} : \left\{ \frac{1}{M} \sum_{m=1}^{M} |\psi(Z_m)|^2 \right\}^{1/2} - \left\{ \frac{1}{M} \sum_{m=1}^{M} |\psi(Z_{M+m})|^2 \right\}^{1/2} \geqslant h^{(\beta+2)/2} \right)$$

$$\leqslant \mathbb{P}\left(\exists g \in \mathcal{G} : \left\{ \frac{1}{M} \sum_{m=1}^{M} |g(Z_m)|^2 \right\}^{1/2} - \left\{ \frac{1}{M} \sum_{m=1}^{M} |g(Z_{M+m})|^2 \right\}^{1/2} \geqslant \frac{h^{(\beta+2)/2}}{3} \right).$$

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

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

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

The above exponential is bounded by $\exp\big(-Mh^{\beta+2}/72C_y(R)^2\big)$ because

$$\sum_{m=1}^{M} \{ |g(\tilde{X}_{t_k}^{N,m})|^2 - |g(X_{t_k}^{N,m})|^2 \}^2 \leqslant \sum_{m=1}^{M} \Big(|g(\tilde{X}_{t_k}^{N,m})|^4 + |g(X_{t_k}^{N,m})|^4 \Big)$$

$$\leqslant 4C_y(R)^2 \sum_{m=1}^{M} \Big(|g(\tilde{X}_{t_k}^{N,m})|^2 + |g(X_{t_k}^{N,m})|^2 \Big). \qquad (20)$$

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 $M^{-1}\sum_{m=1}^M |y_k^{N,R}(X_{t_k}^{N,m}) - \alpha \cdot p_{0,k}^m|^2 = \|y_k^{N,R} - \alpha \cdot p_{0,k}\|_{k,M}^2$, that is, $T_{1,k}^M = \inf_\alpha \|y_k^{N,R} - \alpha \cdot p_{0,k}\|_{k,M}^2$. 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 $\mathbb{E}T_{2,k}^M$, the technique 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 $(B_{0,k}^M)^* B_{0,k}^M / M = \mathrm{Id}$ as before. We can thus write, for $1 \leqslant m \leqslant M$,

$$\mathbb{E}^M\Big(|p_{0,k}^m.\{\tilde{\alpha}_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\alpha}_{0,k}^{1,M})\}|^2\Big) \tag{21}$$

$$= \mathbb{E}^M\left((p_{0,k}^m)^* \frac{(B_{0,k}^M)^*}{M}\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^* \frac{B_{0,k}^M}{M} p_{0,k}^m\right)$$

$$= (p_{0,k}^m)^* \frac{(B_{0,k}^M)^*}{M}\mathbb{E}^M\Big(\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^*\Big) \frac{B_{0,k}^M}{M} p_{0,k}^m,$$

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 $\|\mathbb{E}^M\ (\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^*)\|_2$. As $\alpha_{0,k+1}^M$ is $\mathcal{F}^M$-measurable, we obtain

$$\mathbb{E}^M\Big(y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m})y_{k+1}^{N,R,M}(\tilde{X}_{t_{k+1}}^{N,m'})\Big)$$

$$= \mathbb{E}^M\Big([\alpha_{0,k+1}^M \cdot p_{0,k+1}(\tilde{X}_{t_{k+1}}^{N,m})]_y[\alpha_{0,k+1}^M \cdot p_{0,k+1}(\tilde{X}_{t_{k+1}}^{N,m'})]_y\Big)$$

$$= \mathbb{E}^M\Big([\alpha_{0,k+1}^M \cdot p_{0,k+1}(\tilde{X}_{t_{k+1}}^{N,m})]_y\Big)\mathbb{E}^M\Big([\alpha_{0,k+1}^M \cdot p_{0,k+1}(\tilde{X}_{t_{k+1}}^{N,m'})]_y\Big),$$

from which we deduce that the non-diagonal terms of the matrix $\mathbb{E}^M(\{V-\mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^*)$ are equal to 0. The introduction of the projection coefficients $\tilde{\alpha}_{0,k}^M$ ensures this crucial property which is not true for the projection coefficients $\alpha_{0,k}^M$. As for the diagonal terms, they are bounded by $C_y(R)^2$. Thus $\|\mathbb{E}^M(\{V - \mathbb{E}^M(V)\}\{V - \mathbb{E}^M(V)\}^*)\|_2 \leqslant C_y(R)^2$. Finally, in view of (21), we obtain

$$\mathbb{E}^M\left(\frac{1}{M}\sum_{m=1}^M |p_{0,k}^m.\{\tilde{\alpha}_{0,k}^{1,M} - \mathbb{E}^M(\tilde{\alpha}_{0,k}^{1,M})\}|^2\right)$$

$$\leqslant \frac{1}{M}\sum_{m=1}^M \left|\frac{B_{0,k}^M}{M}\ p_{0,k}^m\right|_2^2 C_y(R)^2$$

$$\leqslant \frac{C_y(R)^2}{M^2}\sum_{m=1}^M |p_{0,k}^m|^2 = \frac{C_y(R)^2}{M^2}\mathrm{tr}[(B_{0,k}^M)^* B_{0,k}^M] = \frac{C_y(R)^2}{M} K_{0,k}^M.$$

The estimate for $\mathbb{E}T_{2,k}^M$ 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 gives

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

whence our result. □

5. Conclusion

We have proposed a simple algorithm to solve GBSDEs. The dynamic programming equation resulting from the time discretization of equation (1) is solved 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 Gobet *et al.* 2005). We have derived explicit error bounds which allow us to optimally choose the parameters of the method to achieve a given accuracy. This is a significant improvement compared to previous work. However, our numerical tests reveal that the convergence can be faster than our theoretical estimates predict. The explanation of this phenomenon is a matter for future research. Additional work is also necessary to consider in (1) a martingale other than W and to allow the driver to depend on L.

Acknowledgements

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

References

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

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

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

Bosq, D. and Lecoutre, J.P. (1987) *Théorie de l'estimation fonctionnelle*. Paris: Économica.

Bouchard, B. and Touzi, N. (2004) Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Process. Appl.*, **111**, 175–206.

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

Egloff, D. (2005) Monte Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Probab.*, **15**, 1396–1432.

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

Föllmer, H. and Schweizer, M. (1991) Hedging of contingent claims under incomplete information. In M.H.A. Davis and R.J. Elliott (eds), *Applied Stochastic Analysis*, pp. 389–414. London: Gordon and Breach.

Gobet, E. and Lemor, J.-P. (2006) Numerical simulation of BSDEs using empirical regression methods: theory and practice. In *Proceedings of the Fifth Colloquium on BSDEs* (Shanghai, May 2005). To appear.

Gobet, E., Lemor, J.-P. and Warin, X. (2005) A regression-based Monte-Carlo method for backward stochastic differential equations. *Ann. Appl. Probab.*, **15**, 2172–2202.

Golub, G. and Van Loan, C.F. (1996) *Matrix Computations*, 3rd edn. Baltimore, MD: Johns Hopkins University Press.

Györfi, L., Kohler, M., Krzyzak, A. and Walk, H. (2002) A distribution-free theory of nonparametric regression. New York: Springer-Verlag.

Heath, D., Platen, E. and Schweizer, M. (2001) Comparison of two quadratic approaches to hedging in incomplete markets. *Math. Finance*, **11**, 385–413.

Jacod, J. (2004) The Euler scheme for Lévy driven stochastic differential equations: limit theorems. *Ann. Probab.*, **32**(3), 1830–1872.

Lemor, J.-P. (2005) Approximation par projections et simulations de Monte-Carlo des équations différentielles stochastiques rétrogrades. Doctoral thesis, École Polytechnique. http://www.imprimerie.polytechnique.fr/Theses/Files/lemor.pdf (accessed 3 February 2006).

Lemor, J.-P., Gobet, E. and Warin, X. (2005) Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. Report 574, CMAP, École Polytechnique http://www.cmap.polytechnique.fr/preprint/ (accessed 3 February 2006).

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

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

Peng, S.G. (2004) Nonlinear expectations, nonlinear evaluations and risk measures. In M. Frittelli and W.J. Runggaldier (eds) *Stochastic Methods in Finance*, Lecture Notes in Math. 1856, pp. 165–253. Berlin: Springer-Verlag.

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

Received May 2005 and revised January 2006