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/

A regression-based Monte-Carlo method to solve backward stochastic differential equations.

Emmanuel Gobet, Jean-Philippe Lemor and Xavier Warin

R.I. N^0 533 *May 2004*

A regression-based Monte-Carlo method to solve backward stochastic differential equations

Emmanuel Gobet[1], **Jean-Philippe Lemor**[2], **Xavier Warin**[3]

1: Corresponding author. École Polytechnique - Centre de Mathématiques Appliquées - 91128 Palaiseau Cedex - FRANCE (e-mail: emmanuel.gobet@polytechnique.fr).
2: Électricité de France - EDF R&D - Site de Clamart - 1 avenue du Général De Gaulle, 92141 Clamart - FRANCE (e-mail: lemor@cmapx.polytechnique.fr).
3: Électricité de France - EDF R&D - Site de Clamart - 1 avenue du Général De Gaulle, 92141 Clamart - FRANCE (e-mail: xavier.warin@edf.fr).

Current version: 17 May 2004

Abstract. We are concerned with the numerical resolution of backward stochastic differential equations and their applications to finance. We propose a new numerical scheme based on iterative regressions on function bases, which coefficients are evaluated using Monte Carlo simulations. A full convergence analysis is derived. Numerical experiments are included, in particular concerning option pricing with differential interest rates.

Key words: backward stochastic differential equations, regression on function bases, Monte Carlo methods.

JEL Classification: C14, C15, G1

Mathematics Subject Classification (1991): 60H10, 62G08, 65C30

1 Introduction

In this paper we are interested in numerically approximating the solution of a decoupled forward-backward stochastic differential equation (FBSDE)

$$S_{i,t} = S_{i,0} + \int_0^t S_{i,s} B_i(s, S_s) ds + \int_0^t S_{i,s} \Sigma_i(s, S_s) dW_s, \ 1 \le i \le d, \tag{1}$$

$$Y_t = \Phi(\mathbf{S}) + \int_t^T f(s, S_s, Y_s, Z_s) ds - \int_t^T Z_s dW_s. \tag{2}$$

In this representation, $\mathbf{S} = (S_t : 0 \le t \le T) = ([S_{1,t}, \cdots, S_{d,t}] : 0 \le t \le T)$ stands for the risky assets (with $S_{i,0} > 0$ for any i), whose dynamics are written under the objective probability $\mathbb{P}$. Here, W is a q-dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{0\leq t\leq T})$, where $(\mathcal{F}_t)_t$ is the completed natural filtration of W. The driver $f(\cdot,\cdot,\cdot,\cdot)$ and the terminal condition $\Phi(\cdot)$ are respectively a deterministic function and a deterministic functional of the price process $\mathbf{S}$. The assumption **(H1-H2-H3)** below ensures the existence and the unicity to such equation (1-2).

BSDEs in finance. Such equations, first studied by Peng and Pardoux [25] in a general form, are important tools in mathematical finance. We mention some applications and refer the reader to [16] for numerous references. In a complete market, for the usual valuation of a contingent claim with payoff $\Phi(\mathbf{S})$, Y is the value of the replicating portfolio and Z is related to the hedging strategy. In that case, the driver f is linear w.r.t. Y and Z. Some market imperfections can also be incorporated, such as higher interest rate for borrowing [4]: then, the driver is only Lipschitz continuous w.r.t. Y and Z. Related numerical experiments are developed in Section 4. In incomplete markets, the Föllmer-Schweizer strategy [13] is given by the solution of a BSDE. When trading constraints on some assets are imposed, the super-replication price [17] is obtained as the limit of non linear BSDEs. Connections with recursive utilities of Duffie and Epstein [12] are also available. Peng has introduced the notion of g-expectation (here g is the driver) as a non linear pricing rule [27]. Recently he has shown [26] the deep connection between BSDEs and dynamic risk measures, proving that any dynamic risk measure $(\mathcal{E}_t)_{0\leq t\leq T}$ (satisfying some axiomatic conditions) is necessarily associated to a BSDE $(Y_t)_{0\leq t\leq T}$ (the converse being known for years). The least we can say is that BSDEs are now unavoidable tools in mathematical finance. Another indirect application may concern variance reduction techniques for the Monte Carlo computations of option prices, say $\mathbb{E}(\Phi)$ (omitting here the discounting factor). Indeed, $\int_0^T Z_s\, dW_s$ is the so-called martingale control variate (see [24] among others).

The mathematical analysis of BSDE is now well understood (see [23] for recent references) and its numerical resolution has made recent progresses. However, even if several numerical methods have been proposed, they suffer of a high complexity in terms of computational time or are very costly in terms of computer memory. Thus, their uses in practice on real problems are difficult. Hence, it is still topical to devise more efficient algorithms. This article contributes in this direction, by developing a simple approach, based on Monte Carlo regression on function bases. It is in the vein of the general regression approach of Bouchard and Touzi [6], but here it is actually much simpler because only one set of paths is used to evaluate all the regression operators. Consequently, the numerical implementation is easier and more efficient. In addition, we provide a full mathematical analysis of the influence of the parameters of the method.

Numerical methods for BSDEs. In the past decade, there have been several attempts to provide approximation schemes for BSDEs. Firstly, Ma *etal.* [22] propose the *four step scheme* to solve general FBSDEs, which requires the numerical resolution of a quasilinear parabolic PDE. In [2], Bally presents a time discretization scheme based on a Poisson net: this trick avoids him to use the unknown regularity of Z and enables him to derive a rate of convergence w.r.t.

the intensity of the Poisson process. However, extra computations of very high dimensional integrals are needed and this is not handled in [2]. In a recent work [28], Zhang proves some $\mathbf{L}_2$-regularity on Z, which allows the use of a regular deterministic time mesh. Under an assumption of *constructible functionals* for Φ (which essentially means that the system can be made Markovian, by adding d' extra state variables), its approximation scheme is less consuming in terms of high dimensional integrals. If for each of the $d + d'$ state variables, one uses M points to compute the integrals, the complexity is about $M^{d+d'}$ per time step, for a global error of order M^{-1} say[1]. This approach is somewhat related to the quantization method of Bally and Pagès [3], which is an optimal space discretization of the underlying dynamic programming equation (see also the former work by Chevance [9], where the driver does not depend on Z). We should also mention the works by Ma *etal.* [21], Briand *etal.* [8], where the Brownian motion is replaced by a scaled random walk. Weak convergence results are given, without rates of approximation. The complexity becomes very large in multidimensional problems, like for the usual multinomial tree method for the pricing of Bermuda/European options [15]. Recently, in the case of path-independent terminal conditions $\Phi(\mathbf{S}) = \phi(S_T)$, Bouchard and Touzi [6] propose a Monte Carlo approach which may be more suitable for high dimensional problems. They follow the approach by Zhang [28] by approximating (1-2) by a discrete time FBSDE with N time steps (see (4-5) below), with a $\mathbf{L}_2$-error of order $N^{-1/2}$. Instead of computing the conditional expectations which appear at each discretization time by discretizing the space of each state variable, the authors use a general regression operator, which can be derived for instance from kernel estimators or from the Malliavin calculus integration by parts formulas. The regression operator at a discretization time is assumed to be build independently of the underlying process, and independently of the regression operators at the other times. For the Malliavin calculus approach for example, this means that one needs to simulate at each discrete time, M copies of the approximation of (1), which is costly. The algorithm that we propose in this paper requires only one set of paths to approximate all the regression operators at each discretization time at once. Since the regression operators are now correlated, the mathematical analysis is much more involved.

The regression operator we use in the sequel results from the $\mathbf{L}_2$-projection on a finite basis of functions, which leads in practice to solve a standard least squares problem. This approach is not new in numerical methods for financial engineering, since it has been developed by Longstaff and Schwartz [20] for the pricing of Bermuda options. See also [7] for the option pricing using simulations under the objective probability.

Organization of the paper. At the end of this section, we set the framework of our study and define some notations used throughout the paper. Then in Section 2, we describe our algorithm based on the approximation of conditional expectations by a projection on a finite basis of functions. We also state the main results about the convergence of this scheme. The rest of the paper is devoted to analyze the influence of the parameters of this scheme on the eval-

[1] actually, an analysis of the global accuracy is not provided in [28].

uation of Y and Z. Note that approximation results on Z were not previously considered in [6]. In Section 3, we provide an estimation of the time discretization error: this essentially follows from the results by Zhang [28]. Then, the impact of the function bases and the number of simulated paths is discussed, which is the major contribution of our work. Since this least squares approach is also popular to price Bermuda options [20], it is crucial to accurately estimate the propagation of errors in this type of numerical methods, i.e. to ensure that it is not explosive when the exercise frequency shrinks to 0. $\mathbf{L}_2$-estimates and a central limit theorem (see also [10] for Bermuda options) are proved. In section 4, an explicit choice of function bases is given, together with numerical examples relative to the option pricing with differential interest rates.

Standing assumptions. Throughout the paper, we consider that the following hypotheses are fulfilled.

(H1) For any $1 \leq i \leq d$, the functions $(t, x) \mapsto x_i B_i(t, x)$ and $(t, x) \mapsto x_i \Sigma_i(t, x)$ are uniformly Lipschitz continuous w.r.t. $(t, x) \in [0, T] \times \mathbb{R}^d$.

(H2) The driver f satisfies the following continuity estimate:

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

for any $(t_1, x_1, y_1, z_1), (t_2, x_2, y_2, z_2) \in [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q$. Moreover, $\sup_{0 \leq t \leq T} |f(t, 0, 0, 0)| < \infty$.

(H3) The terminal condition Φ satisfies Zhang's $\mathbf{L}^\infty$-Lipschitz condition, i.e. for any continuous functions $\mathbf{s^1}$ and $\mathbf{s^2}$ one has

$$|\Phi(\mathbf{s^1}) - \Phi(\mathbf{s^2})| \leq C \sup_{t \in [0,T]} |s_t^1 - s_t^2|.$$

In addition, $|\Phi(\mathbf{0})| < \infty$ where $\mathbf{0}$ is the function equal to 0 on $[0, T]$.

These assumptions **(H1-H2-H3)** are sufficient to ensure the existence and uniqueness of a triplet $(\mathbf{S}, \mathbf{Y}, \mathbf{Z})$ solution to (1-2) (see [23] and references therein). Note that the $\mathbf{L}^\infty$-Lipschitz condition allows a large class of exotic payoffs (see examples later).

General notations.

- The $\mathbf{L}_2(\mathbb{P})$ projection of the random variable U on a finite family $\phi = [\phi_1, \cdots, \phi_n]^*$ (considered as a random column vector) is denoted by $\mathcal{P}_\phi(U)$. We set $\mathcal{R}_\phi(U) = U - \mathcal{P}_\phi(U)$ for the projection error.
- At each time t_k some random variables $p_{0,k}, p_{1,k}, \cdots, p_{q,k}$ will be used as projection bases to approximate $Y_{t_k}, Z_{1,t_k}, \cdots, Z_{q,t_k}$ (Z_{l,t_k} is the l-th component of Z_{t_k}). The projection coefficients will be denoted $\alpha_{0,k}, \alpha_{1,k}, \cdots, \alpha_{q,k}$ (viewed as column vectors). To simplify, we write $f_k(\alpha_{0,k}, \cdots, \alpha_{q,k})$ or $f_k(\alpha_k)$ (resp. $f_k^m(\alpha_{0,k}, \cdots, \alpha_{q,k})$ or $f_k^m(\alpha_k)$) for $f(t_k, S_{t_k}^N, \alpha_{0,k} \cdot p_{0,k}, \cdots, \alpha_{q,k} \cdot p_{q,k})$ (resp. $f(t_k, S_{t_k}^{N,m}, \alpha_{0,k} \cdot p_{0,k}^m, \cdots, \alpha_{q,k} \cdot p_{q,k}^m)$ the m-th realization of the same quantity) ($S_{t_k}^N$ is an Euler approximation of S_{t_k}, see (3)).

- For convenience, we write $\mathbb{E}_k(.) = \mathbb{E}(.|\mathcal{F}_{t_k})$. We put $\Delta W_k = W_{t_{k+1}} - W_{t_k}$ (and $\Delta W_{l,k}$ component-wise). The m-th realization is denoted by $\Delta W^m_{l,k}$.
- For a vector x, $|x|$ stands as usual for its Euclidean norm. The relative dimension is still implicit. For an integer M and $x \in \mathbb{R}^M$, we put $|x|^2_M = \frac{1}{M}\sum_{m=1}^M |x_m|^2$. For a set of projection coefficients $\alpha = (\alpha_0, \cdots, \alpha_q)$, we set $|\alpha| = \max_{0\le l\le q}|\alpha_l|$ (the dimensions of the α_l may be different). For the set of basis functions at a fixed time t_k, $|p_k|$ is defined analogously.
- For a symmetric matrix A, $\|A\|$ and $\|A\|_F$ are respectively the maximum of the absolute value of its eigenvalues and its Frobenius norm.
- In the next computations, C denotes a generic constant that may change from line to line. It is still uniform in the parameters of our scheme.

Additional notations are given through the paper when needed.

2 The numerical scheme and the main approximation results

In this section, we describe our numerical scheme and give the main results about the convergence analysis. All the proofs are postponed to Section 3. The derivation of the final scheme is obtained in three main steps. The presentation below is rather expanded but it should provide an intuition of the main approximation results.

2.1 Step 1

We first consider a time approximation of equations (1) and (2). Let h be a time step (say smaller than 1), associated to equidistant discretization times $(t_k = kh = kT/N)_{0\le k\le N}$.

For the forward component (1), we use a standard Euler scheme, not on $\mathbf{S}$ but on $\mathbf{log}(\mathbf{S})$. This modification does not change the further rates of convergence but it is more satisfactory since it leads to non negative prices for the approximation of $\mathbf{S}$ as it is expected. This is defined by $(S^N_{t_k})_{0\le k\le N}$, which writes component-wise $S^N_{i,0} = S_{i,0}$ and

$$S^N_{i,t_{k+1}} = S^N_{i,t_k}\exp\big([B_i(t_k, S^N_{t_k}) - \frac{1}{2}|\Sigma_i|^2(t_k, S^N_{t_k})]h + \Sigma_i(t_k, S^N_{t_k})\Delta W_k\big). \quad (3)$$

To ensure the convergence of this modified Euler scheme towards $\mathbf{S}$, we slightly reinforce **(H1)**.

(H1') The assumption **(H1)** is satisfied. In addition, the functions $b_i(t,x) = B_i(t, \exp(x_1), \cdots, \exp(x_d))$ and $\sigma_i(t,x) = \Sigma_i(t, \exp(x_1), \cdots, \exp(x_d))$ are bounded and uniformly Lipschitz continuous w.r.t. $(t,x) \in [0,T]\times\mathbb{R}^d$.

The terminal condition $\Phi(\mathbf{S})$ is approximated by $\Phi^N(P^N_{t_N})$, where Φ^N is a deterministic function and $(P^N_{t_k})_{0\leq k\leq N}$ is a Markov chain, whose first components are given by those of $(S^N_{t_k})_{0\leq k\leq N}$. In other words, we eventually add extra state variables to make Markovian the implicit dynamics of the terminal condition. We also assume that $P^N_{t_k}$ is $\mathcal{F}_{t_k}$-measurable and that $\mathbb{E}[\Phi^N(P^N_{t_N})]^2 < \infty$. Of course, this approximation strongly depends on the terminal condition type and its impact is measured in Theorem 1 by the error $\mathbb{E}|\Phi(\mathbf{S}) - \Phi^N(P^N_{t_N})|^2$. Let us give some important examples with $d = 1$ and $q = 1$.

- Vanilla payoff: $\Phi(\mathbf{S}) = \phi(S_T)$. Set $P^N_{t_k} = S^N_{t_k}$ and $\Phi^N(P^N_{t_N}) = \phi(P^N_{t_N})$. Under the $\mathbf{L}^\infty$-condition **(H3)**, it gives $\mathbb{E}|\Phi^N(P^N_{t_N}) - \Phi(\mathbf{S})|^2 \leq Ch$.

- Asian payoff: $\Phi(\mathbf{S}) = \phi(S_T, \int_0^T S_t dt)$. Set $P^N_{t_k} = (S^N_{t_k}, h\sum_{i=0}^{k-1} S^N_{t_i})$ and $\Phi^N(P^N_{t_N}) = \phi(P^N_{t_N})$. For usual functions ϕ, the $\mathbf{L}_2$-error is of order $1/2$ w.r.t. h. More accurate approximations of the average of $\mathbf{S}$ could be incorporated [18].

- Lookback payoff: $\Phi(\mathbf{S}) = \phi(S_T, \min_{t\in[0,T]} S_t, \max_{t\in[0,T]} S_t)$. Set $\Phi^N(P^N_{t_N}) = \phi(P^N_{t_N})$ with $P^N_{t_k} = (S^N_{t_k}, \min_{i\leq k} S^N_{t_i}, \max_{i\leq k} S^N_{t_i})$. In general, this induces an $\mathbf{L}_2$-error of magnitude $\sqrt{h\log(1/h)}$ [28]. The rate $\sqrt{h}$ can be achieved by considering the exact extrema of the continuous Euler scheme [1].

The backward component (2) is approximated in a backward manner. First, we set $Y^N_{t_N} = \Phi^N(P^N_{t_N})$. Then, $(Y^N_{t_k}, Z^N_{t_k})_{0\leq k\leq N-1}$ are defined by

$$Z^N_{l,t_k} = \frac{1}{h}\mathbb{E}_k(Y^N_{t_{k+1}}\Delta W_{l,k}), \tag{4}$$

$$Y^N_{t_k} = \mathbb{E}_k(Y^N_{t_{k+1}}) + hf(t_k, S^N_{t_k}, Y^N_{t_k}, Z^N_{t_k}). \tag{5}$$

Using in particular the inequality $|Z^N_{l,t_k}| \leq \frac{1}{\sqrt{h}}\sqrt{\mathbb{E}_k(Y^N_{t_{k+1}})^2}$, it is easy to see by a recurrence argument that $Y^N_{t_k}$ and $Z^N_{t_k}$ belong[2] to $\mathbf{L}_2(\mathcal{F}_{t_k})$. It is also equivalent to assert that they minimize the quantity

$$\mathbb{E}(Y^N_{t_{k+1}} - Y + hf(t_k, S^N_{t_k}, Y, Z) - Z\Delta W_k)^2 \tag{6}$$

over $\mathbf{L}_2(\mathcal{F}_{t_k})$ random variables (Y, Z). Note that $Y^N_{t_k}$ is well defined in (5), because the application $Y \mapsto \mathbb{E}_k(Y^N_{t_{k+1}}) + hf(t_k, S^N_{t_k}, Y, Z^N_{t_k})$ is a contraction in $\mathbf{L}_2(\mathcal{F}_{t_k})$, for h small enough. The following result provides an estimate of the error induced by this first step.

Theorem 1 *Assume* **(H1'-H2-H3)**. *For h small enough, we have*

$$\max_{0\leq k\leq N} \mathbb{E}|Y_{t_k} - Y^N_{t_k}|^2 + \sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} \mathbb{E}|Z_t - Z^N_{t_k}|^2 dt$$
$$\leq C\big((1 + |S_0|^2)h + \mathbb{E}|\Phi(\mathbf{S}) - \Phi^N(P^N_{t_N})|^2\big).$$

[2] as usual, $\mathbf{L}_2(\mathcal{F}_{t_k})$ stands for the square integrable and $\mathcal{F}_{t_k}$-measurable (possibly multidimensional) random variables.

Owing to the Markov chain $(P^N_{t_k})_{0\le k\le N}$, the independent increments $(\Delta W_k)_{0\le k\le N-1}$ and the equations (4-5), we easily get the following result.

Proposition 1 *Assume* **(H1'-H2-H3)**. *For h small enough, we have*

$$Y^N_{t_k} = y^N_k(P^N_{t_k}),\ Z^N_{l,t_k} = z^N_{l,k}(P^N_{t_k}) \quad \text{for } \ 0 \le k \le N \text{ and } \ 1 \le l \le q, \tag{7}$$

where $(y^N_k(\cdot))_k$ and $(z^N_{l,k}(\cdot))_{k,l}$ are measurable functions.

More interesting is the derivation of Lipschitz continuity property for these functions. This is achieved under a Lipschitz property of the terminal condition and the Markov chain $P^N_{t_k}$ w.r.t. its initial condition. To state a relevant condition, we use the usual representation of a Markov chain as a random iterative sequence of the form $P^N_{t_k} = F^N_k(U_k, P^N_{t_{k-1}})$ where $(F^N_k)_k$ are measurable functions and $(U_k)_k$ are i.i.d. random variables. This representation allows to deal with flow properties of $P^N_{t_k}$. Consequently, for any initial condition (k_0, x) (such that $P^N_{t_{k_0}} = x$), we can associate the relative solutions $(Y^{N,k_0,x}_{t_k})_{k\ge k_0}$, $(Z^{N,k_0,x}_{t_k})_{k\ge k_0}$, $(P^{N,k_0,x}_{t_k})_{k\ge k_0}$. Our Lipschitz assumption takes the following form.

(H4) The function $\Phi^N(\cdot)$ is Lipschitz continuous (uniformly in N) and $\sup_N |\Phi^N(\mathbf{0})| < \infty$. In addition, $\mathbb{E}|P^{N,k_0,x}_{t_N} - P^{N,k_0,x'}_{t_N}|^2 + \mathbb{E}|P^{N,k_0,x}_{t_{k_0+1}} - P^{N,k_0,x'}_{t_{k_0+1}}|^2 \le C|x - x'|^2$ uniformly in k_0 and N.

Proposition 2 *Assume* **(H1'-H2-H3-H4)**. *For h small enough, we have*

$$|y^N_{k_0}(x) - y^N_{k_0}(x')| + \sqrt{h}|z^N_{k_0}(x) - z^N_{k_0}(x')| \le C|x - x'| \tag{8}$$

uniformly in $k_0 \le N - 1$.

It is easy to check that for the previous examples of vanilla, Asian or lookback options, the assumption **(H4)** is verified and thus we have to approximate, at each time t_k, Lipschitz continuous functions.

2.2 Step 2

Here, the conditional expectations which appear in the definition of $Y^N_{t_k}$ and $Z^N_{t_k}$ are replaced by a $\mathbf{L}_2(\mathbb{P})$ projection on function bases.

Definition 1 *To approximate $Y^N_{t_k}$ and Z^N_{l,t_k} $(1 \le l \le q)$, we use respectively finite-dimensional function bases $p_{0,k}(P^N_{t_k})$ and $p_{l,k}(P^N_{t_k})$ $(1 \le l \le q)$, which may be also written $p_{0,k}$ and $p_{l,k}$ $(1 \le l \le q)$ to simplify. We assume that $\mathbb{E}|p_{l,k}|^2 < \infty$ $(0 \le l \le q)$ and w.l.o.g. that each function basis is free, which ensures the uniqueness of the coefficients of the projection $\mathcal{P}_{p_{l,k}}$ $(0 \le l \le q)$.*

A numerical difficulty still remains in the approximation of $Y^N_{t_k}$ in (5), which is usually obtained as a fixed point. To circumvent this problem, we propose a solution combining the projection on the function basis and I Picard iterations. The integer I is a fixed parameter of our scheme (the analysis below shows that the value $I = 3$ is relevant).

Definition 2 *We denote[3] by* $Y_{t_k}^{N,i,I}$ *the approximation of* $Y_{t_k}^N$*, where* i *Picard iterations with projections have been performed at time* t_k *and* I *Picard iterations with projections at any time after* t_k*. Analogous notations stand for* $Z_{l,t_k}^{N,i,I}$*. We associate to* $Y_{t_k}^{N,i,I}$ *and* $Z_{l,t_k}^{N,i,I}$ *their respective projection coefficients* $\alpha_{0,k}^{i,I}$ *and* $\alpha_{l,k}^{i,I}$*, on the function bases* $p_{0,k}$ *and* $p_{l,k}$ $(1 \le l \le q)$*.*

We now turn to a precise definition of the above quantities. We set $Y_{t_N}^{N,i,I} = \Phi^N(P_{t_N}^N)$, independently of i and I. Assume that $Y_{t_{k+1}}^{N,I,I}$ is obtained and let us define $Y_{t_k}^{N,i,I}, Z_{l,t_k}^{N,i,I}$ for $i = 0, \cdots, I$. We begin with $Y_{t_k}^{N,0,I} = 0$ and $Z_{t_k}^{N,0,I} = 0$, corresponding to $\alpha_{l,k}^{0,I} = 0$ $(0 \le l \le q)$. By analogy with (6), we set $\alpha_k^{i,I} = (\alpha_{l,k}^{i,I})_{0 \le l \le q}$ as the argmin in $(\alpha_0, \cdots, \alpha_q)$ of the quantity

$$\mathbb{E}\big(Y_{t_{k+1}}^{N,I,I} - \alpha_0 \cdot p_{0,k} + h f_k(\alpha_k^{i-1,I}) - \sum_{l=1}^{q} \alpha_l \cdot p_{l,k}\ \Delta W_{l,k}\big)^2. \tag{9}$$

Iterating with $i = 1, \cdots, I$, at the end we get $(\alpha_{l,k}^{I,I})_{0 \le l \le q}$, thus $Y_{t_k}^{N,I,I} = \alpha_{0,k}^{I,I} \cdot p_{0,k}$ and $Z_{l,t_k}^{N,I,I} = \alpha_{l,k}^{I,I} \cdot p_{l,k}$ $(1 \le l \le q)$. The least squares problem (9) can be formulated in different ways but this one is more convenient for the next step of the scheme. The error induced by this second step is analyzed by the following result.

Theorem 2 *Assume* **(H1'-H2-H3)***. For* h *small enough, we have*

$$\begin{aligned}
&\max_{0 \le k \le N} \mathbb{E}|Y_{t_k}^{N,I,I} - Y_{t_k}^N|^2 + h \sum_{k=0}^{N-1} \mathbb{E}|Z_{t_k}^{N,I,I} - Z_{t_k}^N|^2 \\
&\le C h^{2I-2}\big[1 + |S_0|^2 + \mathbb{E}|\Phi^N(P_{t_N}^N)|^2\big] \\
&\quad + C \sum_{k=0}^{N-1} \mathbb{E}|\mathcal{R}_{p_{0,k}}(Y_{t_k}^N)|^2 + Ch \sum_{k=0}^{N-1} \sum_{l=1}^{q} \mathbb{E}|\mathcal{R}_{p_{l,k}}(Z_{l,t_k}^N)|^2.
\end{aligned}$$

The above result shows how projection errors cumulate along the backward iteration. The key point is to note that they only sum up, with a factor C which does not explode as $N \to \infty$. These estimates improve those of Theorem 4.1 in [6] for two reasons. Firstly, error estimates on Z^N are provided here. Secondly, in the cited theorem, the error is analyzed in terms of $\mathbb{E}|\mathcal{R}_{p_{0,k}}(Y_{t_k}^{N,I,I})|^2$ and $\mathbb{E}|\mathcal{R}_{p_{l,k}}(Z_{l,t_k}^{N,I,I})|^2$ say: hence the influence of function bases is still questionable, since it is hidden in the projection residuals $\mathcal{R}_{p_k}$ and also in the random variables $Y_{t_k}^{N,I,I}$ and $Z_{l,t_k}^{N,I,I}$. Our estimates are relevant to directly analyze the influence of function bases (see Section 4 for explicit computations, where the Lipschitz property from Proposition 2 is used). This feature is crucial in our opinion. Regarding the influence of I, it is enough here to have $I = 2$ to get an error of the same order than in Theorem 1. At the third step, $I = 3$ is needed.

[3] The notation $Y_{t_k}^{N,i,\overbrace{I,\cdots,I}^{N-k-1 \text{ times}}}$ would be clearer but certainly not convenient at all.

2.3 Step 3

This step is very analogous to Step 2, except that in the sequence of iterative least squares problems (9), the expectation $\mathbb{E}$ is replaced by an empirical mean built on M independent simulations of $(P^N_{t_k})_{0\le k\le N}, (\Delta W_k)_{0\le k\le N-1}$. We denote them $((P^{N,m}_{t_k})_{0\le k\le N}, (\Delta W^m_k)_{0\le k\le N-1})_{1\le m\le M}$. The values of basis functions along these simulations are denoted $(p^m_{l,k} = p_l(P^{N,m}_{t_k}))_{0\le l\le q, 0\le k\le N-1, 1\le m\le M}$. A subtlety remains: it is useful to take advantage of a priori estimates on $Y^{N,i,I}_{t_k}, Z^{N,i,I}_{l,t_k}$, in order to force their simulation-based evaluations to satisfy the same estimates. These a priori estimates are given by the following result.

Proposition 3 *Under* **(H1'-H2-H3)**, *for a sequence of explicit functions $(\rho^N_{l,k})_{0\le l\le q, 0\le k\le N-1}$ bounded from below by 1, one has*

$$|Y^{N,i,I}_{t_k}| \le \rho^N_{0,k}(P^N_{t_k}), \quad \sqrt{h}|Z^{N,i,I}_{l,t_k}| \le \rho^N_{l,k}(P^N_{t_k}), a.s.,$$

for any $i \ge 0$, $I \ge 0$ and $0 \le k \le N-1$, with $\mathbb{E}|\rho^N_{l,k}(P^N_{t_k})|^2 < \infty$. We can take $\rho^N_{l,k}(x) = \max(1, C_0|p_{l,k}(x)|)$ for a constant C_0 large enough.

In the sequel, we set $\rho^N_k(P^N_{t_k}) = [\rho^N_{0,k}(P^N_{t_k}), \cdots, \rho^N_{q,k}(P^N_{t_k})]^*$.

Definition 3 *Associated to these a priori estimates, we define (random) truncation functions $\hat\rho^N_{l,k}$ (resp. $\hat\rho^{N,m}_{l,k}$) such that:*

- *they leave invariant $\alpha^{i,I}_{0,k} \cdot p_{0,k} = Y^{N,i,I}_{t_k}$ if $l = 0$ or $\sqrt{h}\alpha^{i,I}_{l,k} \cdot p_{l,k} = \sqrt{h}Z^{N,i,I}_{l,t_k}$ if $l \ge 1$ (resp. $\alpha^{i,I}_{0,k} \cdot p^m_{0,k}$ if $l = 0$ or $\sqrt{h}\alpha^{i,I}_{l,k} \cdot p^m_{l,k}$ if $l \ge 1$);*
- *they are bounded by $2\rho^N_{l,k}(P^N_{t_k})$ (resp. $2\rho^N_{l,k}(P^{N,m}_{t_k})$);*
- *their first derivative is bounded by 1;*
- *their second derivative is uniformly bounded in N, l, k, m.*

A possible construction may be as follows. Take a C^2_b-function $\xi : \mathbb{R} \mapsto \mathbb{R}$, such that $\xi(x) = x$ for $|x| \le 3/2$, $|\xi|_\infty \le 2$ and $|\xi'|_\infty \le 1$. Then, set $\hat\rho^N_{l,k}(x) = \rho^N_{l,k}(P^N_{t_k})\xi(x/\rho^N_{l,k}(P^N_{t_k}))$ and $\hat\rho^{N,m}_{l,k}(x) = \rho^N_{l,k}(P^{N,m}_{t_k})\xi(x/\rho^N_{l,k}(P^{N,m}_{t_k}))$.

We are now in a position to define the simulation-based approximations of $Y^{N,i,I}_{t_k}, Z^{N,i,I}_{l,t_k}$. The backward in time iteration starts with $Y^{N,i,I,M}_{t_N} = \Phi^N(P^N_{t_N})$ independently of i and I. At a given discretization time t_k, the Picard iterations are initialized with $Y^{N,0,I,M}_{t_k} = 0$ and $Z^{N,0,I,M}_{t_k} = 0$, i.e. $\alpha^{0,I,M}_{l,k} = 0$ $(0 \le l \le q)$. Given some projection coefficients $(\alpha^{i,I,M}_{l,k})_{0\le l\le q}$, we define the approximation candidates $Y^{N,i,I,M}_{t_k} = \hat\rho^N_{0,k}(\alpha^{i,I,M}_{0,k} \cdot p_{0,k})$, $\sqrt{h}\ Z^{N,i,I,M}_{l,t_k} = \hat\rho^N_{l,k}(\sqrt{h}\ \alpha^{i,I,M}_{l,k} \cdot p_{l,k})$, and their realizations along the simulations $Y^{N,i,I,M,m}_{t_k} = \hat\rho^{N,m}_{0,k}(\alpha^{i,I,M}_{0,k} \cdot p^m_{0,k})$, $\sqrt{h}\ Z^{N,i,I,M,m}_{l,t_k} = \hat\rho^{N,m}_{l,k}(\sqrt{h}\ \alpha^{i,I,M}_{l,k} \cdot p^m_{l,k})$. These coefficients $\alpha^{i,I,M}_k = (\alpha^{i,I,M}_{l,k})_{0\le l\le q}$ are iteratively obtained as the argmin in $(\alpha_0, \cdots, \alpha_q)$ of the quantity

$$\frac{1}{M}\sum_{m=1}^{M}\left(Y^{N,I,I,M,m}_{t_{k+1}} - \alpha_0 \cdot p^m_{0,k} + hf^m_k(\alpha^{i-1,I,M}_k) - \sum_{l=1}^{q}\alpha_l \cdot p^m_{l,k}\ \Delta W^m_{l,k}\right)^2. \quad (10)$$

If the above least squares problem has multiple solutions (i.e. the empirical regression matrix is not invertible, which occurs with small probability when M becomes large), we may choose for instance the (unique) solution of minimal norm. Actually, this choice is arbitrary and has no incidence on the further analysis.

Now, we aim at quantifying the error between $(Y_{t_k}^{N,I,I,M}, \sqrt{h}Z_{l,t_k}^{N,I,I,M})_{l,k}$ and $(Y_{t_k}^{N,I,I}, \sqrt{h}Z_{l,t_k}^{N,I,I})_{l,k}$, in terms of the number of simulations M, the function bases and the time step h. The analysis here is more involved than in [6] since all the regression operators are correlated by the same set of simulated paths. To obtain more tractable theoretical estimates, we shall assume that each function basis $p_{l,k}$ is orthonormal. Of course, this hypothesis does not affect the numerical scheme, since the projection on a function basis is unchanged by any linear transformation of the basis.

Extra notations. Define

- v_k (resp v_k^m), the (column) vector given by $[v_k]^* = (p_{0,k}{}^*, p_{1,k}{}^* \frac{\Delta W_{1,k}}{\sqrt{h}}, \cdots, p_{q,k}{}^* \frac{\Delta W_{q,k}}{\sqrt{h}})$ (resp. $[v_k^m]^* = (p_{0,k}^m{}^*, p_{1,k}^m{}^* \frac{\Delta W_{1,k}^m}{\sqrt{h}}, \cdots, p_{q,k}^m{}^* \frac{\Delta W_{q,k}^m}{\sqrt{h}})$;
- V_k^M, the matrix given by $V_k^M = \frac{1}{M}\sum_{m=1}^M v_k^m [v_k^m]^*$;
- $P_{l,k}^M$, the matrix given by $P_{l,k}^M = \frac{1}{M}\sum_{m=1}^M p_{l,k}^m [p_{l,k}^m]^*$ $(0 \le l \le q)$;
- the event $\mathbf{A}_k^M = \{\forall\ j \in \{k, \cdots, N-1\} : \|V_j^M - \mathrm{Id}\| \le h, \|P_{0,j}^M - \mathrm{Id}\| \le h$ and $\|P_{l,j}^M - \mathrm{Id}\| \le 1$ for $1 \le l \le q\}$.

Under the orthonormality assumption for each basis $p_{l,k}$, the matrices $(V_k^M)_{0\le k\le N-1}$, $(P_{l,k}^M)_{0\le l\le q, 0\le k\le N-1}$ converge to the identity with probability 1 as $M \to \infty$. Thus, we have $\lim_{M\to\infty} \mathbb{P}(\mathbf{A}_k^M) = 1$. We now state our main result about the influence of the number of simulations.

Theorem 3 *Assume* **(H1'-H2-H3)**, *$I \ge 3$, that each function basis $p_{l,k}$ is orthonormal and that $\mathbb{E}|p_{l,k}|^4 < \infty$ for any k, l. For h small enough, we have for any $0 \le k \le N-1$*

$$\mathbb{E}|Y_{t_k}^{N,I,I} - Y_{t_k}^{N,I,I,M}|^2 + h\sum_{j=k}^{N-1} \mathbb{E}|Z_{t_j}^{I,I} - Z_{t_j}^{I,I,M}|^2$$
$$\le 9\sum_{j=k}^{N-1} \mathbb{E}(|\rho_j^N(P_{t_j}^N)|^2 \mathbf{1}_{[\mathbf{A}_k^M]^c}) + Ch^{I-1}\sum_{j=k}^{N-1}\big[1 + |S_0|^2 + \mathbb{E}|\rho_j^N(P_{t_j}^N)|^2\big]$$
$$+ \frac{C}{hM}\sum_{j=k}^{N-1}\Big(\mathbb{E}\|v_j v_j^* - \mathrm{Id}\|_F^2\ \mathbb{E}|\rho_j^N(P_{t_j}^N)|^2 + \mathbb{E}(|v_j|^2 |p_{0,j+1}|^2)\mathbb{E}|\rho_{0,j}^N(P_{t_j}^N)|^2$$
$$+ h^2\mathbb{E}\big[|v_j|^2(1 + |S_{t_j}^N|^2 + |p_{0,j}|^2\mathbb{E}|\rho_{0,j}^N(P_{t_j}^N)|^2 + \frac{1}{h}\sum_{l=1}^q |p_{l,j}|^2\mathbb{E}|\rho_{l,j}^N(P_{t_j}^N)|^2)\big]\Big).$$

The term with $[\mathbf{A}_k^M]^c$ readily converges to 0 as $M \to \infty$ but we have not made estimations more explicit because the derivation of an optimal upper bound essentially depends on extra moment assumptions that may be available. For instance, if $\rho_j^N(P_{t_j}^N)$ has moments of order higher than 2, we are reduced via an Hölder inequality to estimate the probability $\mathbb{P}([\mathbf{A}_k^M]^c) \le \sum_{j=k}^{N-1} \big[\mathbb{P}(\|V_j^M - \mathrm{Id}\| > h) + \mathbb{P}(\|P_{0,j}^M - \mathrm{Id}\| > h) + \sum_{l=1}^q \mathbb{P}(\|P_{l,j}^M - \mathrm{Id}\| > 1)\big]$. We have $\mathbb{P}(\|V_k^M - \mathrm{Id}\| > h) \le h^{-2}\mathbb{E}\|V_k^M - \mathrm{Id}\|^2 \le h^{-2}\mathbb{E}\|V_k^M - \mathrm{Id}\|_F^2 = (Mh^2)^{-1}\mathbb{E}\|v_k v_k^* - \mathrm{Id}\|_F^2$. This simple calculus illustrates the possible computations, other terms can be handled analogously.

The previous theorem is really informative since it provides a non asymptotic error estimation. With Theorems 1 and 2, it enables to see how to optimally choose the time step h, the function bases and the number of simulations to achieve a given accuracy. We do not report this analysis which seems to be hard to derive for general function bases. This will be addressed in further researches. However, our next numerical experiments give an idea of this optimal choice.

We conclude our theoretical analysis by stating a central limit theorem on the coefficients $\alpha_k^{i,I,M}$ as M goes to ∞. This is less informative than Theorem 3 since this is an asymptotic result. Thus, we remain vague about the asymptotic variance. Explicit expressions can be derived from the proof.

Theorem 4 *Assume* **(H1'-H2-H3)***, that the driver is continuously differentiable w.r.t.* (y,z) *with a bounded and uniformly Hölder continuous derivatives and that* $\mathbb{E}|p_{l,k}|^{2+\varepsilon} < \infty$ *for any* k,l $(\varepsilon > 0)$*. Then, the vector* $[\sqrt{M}(\alpha_k^{i,I,M} - \alpha_k^{i,I})]_{i\le I, k\le N-1}$ *weakly converges to a centered Gaussian vector as* M *goes to* ∞*.*

3 Proofs of the approximation results

3.1 Proof of Theorem 1

This follows more or less directly from Theorem 5.3 in [28] and from its proof. We can not apply this theorem to our case since we use an Euler scheme on $\log(S)$ instead of S, so the proof has to be slightly adapted. We briefly give the key points which make Theorem 1 valid. Firstly, under **(H1-H2-H3)**, Z is càdlàg (see Remark 2.6.ii in [28]). Secondly, our representation formula for Z seems to be different from that of (5.2) in [28], but it actually coincides if we use the isometry property of Itô's integral. At last, examining the proof of the cited theorem, to obtain our result it just remains to prove

$$\mathbb{E}(\sup_{0\le t\le T} |S_t^N|^4) \le C\big(1 + |S_0|^4\big), \tag{11}$$

$$\mathbb{E}(\sup_{0\le t\le T} |S_t - S_t^N|^2) \le C\big(1 + |S_0|^2\big)h, \tag{12}$$

where $(S_t^N)_t$ is the continuous Euler scheme defined for $t \in [t_k, t_{k+1}[$ by

$$S_{i,t}^N = S_{i,t_k}^N \exp\big([B_i(t_k, S_{t_k}^N) - \frac{1}{2}|\Sigma_i|^2(t_k, S_{t_k}^N)](t - t_k) + \Sigma_i(t_k, S_{t_k}^N)(W_t - W_{t_k})\big).$$

To derive estimates (11) and (12), note that $X^N_{i,t} = \log(S^N_{i,t})$ is the *usual* Euler scheme associated to the drift and diffusion coefficients $b_i(t,x) - \frac{1}{2}|\sigma_i(t,x)|^2$ and $\sigma_i(t,x)$. Thus, under **(H1')** the following classic estimates hold: $\mathbb{E}(\sup_{0\le t\le T}|X_t - X^N_t|^p) \le C_p h^{p/2}$, $\mathbb{E}(\sup_{0\le t\le T} e^{4X_{i,t}} + \sup_{0\le t\le T} e^{4X^N_{i,t}}) \le CS^4_{i,0}$. Now, the inequality (11) is clear. In addition, writing $S_{i,t} - S^N_{i,t} = (X_{i,t} - X^N_{i,t})\int_0^1 e^{\lambda X_{i,t}+(1-\lambda)X^N_{i,t}} d\lambda$, we deduce (12) using the previous estimates on X and X^N. □

3.2 Proof of Theorem 2

For convenience, we denote $\mathcal{A}^N(S_0) = 1 + |S_0|^2 + \mathbb{E}|\Phi^N(P^N_{t_N})|^2$. In the following computations, we repeatedly use three standard inequalities.

1. The contraction property of the $\mathbf{L}_2$-projection operator: for any random variable $X \in \mathbf{L}_2$, we have $\mathbb{E}|\mathcal{P}_{p_{l,k}}(X)|^2 \le \mathbb{E}|X|^2$.
2. The Young inequality: $\forall\ \gamma > 0, \forall\ (a,b) \in \mathbb{R}^2, \quad (a+b)^2 \le (1+\gamma h)a^2 + (1+\frac{1}{\gamma h})b^2$.
3. The discrete Gronwall lemma: for any non-negative sequences $(a_k)_{0\le k\le N}$, $(b_k)_{0\le k\le N}$ and $(c_k)_{0\le k\le N}$ satisfying $a_{k-1}+c_{k-1} \le (1+\gamma h)a_k + b_{k-1}$ (with $\gamma > 0$), we have $a_k + \sum_{i=k}^{N-1} c_i \le e^{\gamma(T-t_k)}\big[a_N + \sum_{i=k}^{N-1} b_i\big]$. Most of the time, it will be used with $c_i = 0$.

Because ΔW_k is centered and independent of $(p_{l,k})_{0\le l\le q}$, it is straightforward to see that the solution of the least squares problem (9) is given for $i \ge 1$ by

$$Z^{N,i,I}_{l,t_k} = \frac{1}{h}\mathcal{P}_{p_{l,k}}\big(Y^{N,I,I}_{t_{k+1}}\Delta W_{l,k}\big), \tag{13}$$

$$Y^{N,i,I}_{t_k} = \mathcal{P}_{p_{0,k}}\big(Y^{N,I,I}_{t_{k+1}} + hf(t_k, S^N_{t_k}, Y^{N,i-1,I}_{t_k}, Z^{N,i-1,I}_{t_k})\big). \tag{14}$$

The proof of Theorem 2 may be divided in several steps.
Step 1: a (tight) preliminary upper bound for $\mathbb{E}|Z^{N,i,I}_{l,t_k}|^2$. First note that $Z^{N,i,I}_{l,t_k}$ is constant for $i \ge 1$. Moreover, the Cauchy-Schwarz inequality yields $|\mathbb{E}_k(Y^{N,I,I}_{t_{k+1}}\Delta W_{l,k})|^2 = |\mathbb{E}_k([Y^{N,I,I}_{t_{k+1}} - \mathbb{E}_k(Y^{N,I,I}_{t_{k+1}})]\Delta W_{l,k})|^2 \le h\big(\mathbb{E}_k[Y^{N,I,I}_{t_{k+1}}]^2 - [\mathbb{E}_k(Y^{N,I,I}_{t_{k+1}})]^2\big)$. Since $(p_{l,k})_l$ is $\mathcal{F}_{t_k}$-measurable and owing to the contraction of the projection operator, it follows that

$$\begin{aligned}\mathbb{E}|Z^{N,i,I}_{l,t_k}|^2 = \frac{1}{h^2}\mathbb{E}\big[\mathcal{P}_{p_{l,k}}\big(\mathbb{E}_k[Y^{N,I,I}_{t_{k+1}}\Delta W_{l,k}]\big)\big]^2 &\le \frac{1}{h^2}\mathbb{E}\big(\mathbb{E}_k[Y^{N,I,I}_{t_{k+1}}\Delta W_{l,k}]\big)^2\\ &\le \frac{1}{h}\big(\mathbb{E}[Y^{N,I,I}_{t_{k+1}}]^2 - \mathbb{E}[\mathbb{E}_k(Y^{N,I,I}_{t_{k+1}})]^2\big).\end{aligned} \tag{15}$$

As it may be seen in the computations below, the term $\mathbb{E}[\mathbb{E}_k(Y^{N,I,I}_{t_{k+1}})]^2$ in (15) plays a crucial role to make further estimates not explosive w.r.t. h.
Step 2: $\mathbf{L}_2$ bounds for $Y^{N,i,I}_{t_k}$ and $\sqrt{h}Z^{N,i,I}_{l,t_k}$. Actually, it is an easy exercise to check that the random variables $Y^{N,i,I}_{t_k}$ and $\sqrt{h}Z^{N,i,I}_{l,t_k}$ are square integrable.

We aim at proving that uniform $\mathbf{L}_2$ bounds w.r.t. i, I, k are available. Denote $\chi_k^{N,I} : Y \in \mathbf{L}_2(\mathcal{F}_{t_k}) \mapsto \mathcal{P}_{p_{0,k}}\big(Y_{t_{k+1}}^{N,I,I} + hf(t_k, S_{t_k}^N, Y, Z_{t_k}^{N,i-1,I})\big) \in \mathbf{L}_2(\mathcal{F}_{t_k})$. Clearly, $\mathbb{E}|\chi_k^{N,I}(Y_2) - \chi_k^{N,I}(Y_1)|^2 \le (C_f h)^2 \mathbb{E}|Y_2 - Y_1|^2$ where C_f is the Lipschitz constant of f. Consequently for h small enough, the application $\chi_k^{N,I}$ is contracting and has an unique fixed point $Y_{t_k}^{N,\infty,I} \in \mathbf{L}_2(\mathcal{F}_{t_k})$ (remind that $Z_{l,t_k}^{N,i,I}$ does not depend on $i \ge 1$). One has

$$Y_{t_k}^{N,\infty,I} = \mathcal{P}_{p_{0,k}}\big(Y_{t_{k+1}}^{N,I,I} + hf(t_k, S_{t_k}^N, Y_{t_k}^{N,\infty,I}, Z_{t_k}^{N,I,I})\big), \tag{16}$$

$$\mathbb{E}|Y_{t_k}^{N,\infty,I} - Y_{t_k}^{N,i,I}|^2 \le (C_f h)^{2i} \mathbb{E}|Y_{t_k}^{N,\infty,I}|^2 \tag{17}$$

since $Y_{t_k}^{N,0,I} = 0$. Thus, Young's inequality yields for $i \ge 1$

$$\begin{aligned}\mathbb{E}|Y_{t_k}^{N,i,I}|^2 \le& (1 + \frac{1}{h})\mathbb{E}|Y_{t_k}^{N,\infty,I} - Y_{t_k}^{N,i,I}|^2 + (1+h)\mathbb{E}|Y_{t_k}^{N,\infty,I}|^2 \\ \le& (1 + Ch)\mathbb{E}|Y_{t_k}^{N,\infty,I}|^2. \end{aligned}\tag{18}$$

The above inequality is also true for $i = 0$ because $Y_{t_k}^{N,0,I} = 0$. We now estimate $\mathbb{E}|Y_{t_k}^{N,\infty,I}|^2$ from the identity (16). Combining Young's inequality (with γ to be chosen later), the identity $\mathcal{P}_{p_{0,k}}(Y_{t_{k+1}}^{N,I,I}) = \mathcal{P}_{p_{0,k}}(\mathbb{E}_k[Y_{t_{k+1}}^{N,I,I}])$, the contraction of $\mathcal{P}_{p_{0,k}}$, the Lipschitz property of f, we get

$$\begin{aligned}\mathbb{E}|Y_{t_k}^{N,\infty,I}|^2 \le& (1 + \gamma h)\mathbb{E}|\mathbb{E}_k[Y_{t_{k+1}}^{N,I,I}]|^2 \\ &+ Ch(h + \frac{1}{\gamma})\big[\mathbb{E}f_k^2(0, \cdots, 0) + \mathbb{E}|Y_{t_k}^{N,\infty,I}|^2 + \mathbb{E}|Z_{t_k}^{N,I,I}|^2\big]. \end{aligned}\tag{19}$$

Bringing together terms $\mathbb{E}|Y_{t_k}^{N,\infty,I}|^2$, then using (15) and the upper bound $\mathbb{E}f_k^2(0, \cdots, 0) \le C(1 + |S_0|^2)$ (see (11)), it readily follows that

$$\begin{aligned}\mathbb{E}|Y_{t_k}^{N,\infty,I}|^2 \le& \frac{(1+\gamma h)}{1 - Ch(h + \frac{1}{\gamma})}\mathbb{E}|\mathbb{E}_k[Y_{t_{k+1}}^{N,I,I}]|^2 + \frac{Ch(h + \frac{1}{\gamma})}{1 - Ch(h + \frac{1}{\gamma})}\big[1 + |S_0|^2\big] \\ &+ \frac{C(h + \frac{1}{\gamma})}{1 - Ch(h + \frac{1}{\gamma})}\big(\mathbb{E}|Y_{t_{k+1}}^{N,I,I}|^2 - \mathbb{E}|\mathbb{E}_k[Y_{t_{k+1}}^{N,I,I}]|^2\big) \end{aligned}\tag{20}$$

provided that h is small enough. Take $\gamma = C$ to get

$$\begin{aligned}\mathbb{E}|Y_{t_k}^{N,\infty,I}|^2 &\le Ch\big[1 + |S_0|^2\big] + (1 + Ch)\mathbb{E}|Y_{t_{k+1}}^{N,I,I}|^2 + Ch\mathbb{E}|\mathbb{E}_k[Y_{t_{k+1}}^{N,I,I}]|^2 \\ &\le Ch\big[1 + |S_0|^2\big] + (1 + 2Ch)\mathbb{E}|Y_{t_{k+1}}^{N,I,I}|^2 \end{aligned}\tag{21}$$

with a new constant C. Plugging this estimate into (18) with $i = I$, we get $\mathbb{E}|Y_{t_k}^{N,I,I}|^2 \le Ch\big[1 + |S_0|^2\big] + (1 + Ch)\mathbb{E}|Y_{t_{k+1}}^{N,I,I}|^2$ and thus, by Gronwall's lemma, $\sup_{0 \le k \le N} \mathbb{E}|Y_{t_k}^{N,I,I}|^2 \le C\mathcal{A}^N(S_0)$. This upper bound combined with (21), (18) and (15) finally provides the required uniform estimates for $\mathbb{E}|Y_{t_k}^{N,i,I}|^2$ and $\mathbb{E}|Z_{l,t_k}^{N,i,I}|^2$:

$$\sup_{I \ge 1}\sup_{i \ge 0}\sup_{0 \le k \le N}(\mathbb{E}|Y_{t_k}^{N,i,I}|^2 + h\mathbb{E}|Z_{l,t_k}^{N,i,I}|^2) \le C\mathcal{A}^N(S_0). \tag{22}$$

Step 3: upper bounds for $\eta_k^{N,I} = \mathbb{E}|Y_{t_k}^{N,I,I} - Y_{t_k}^N|^2$. Note that $\eta_N^{N,I} = 0$. Our purpose is to prove the following relation for $0 \le k < N$:

$$\begin{aligned}\eta_k^{N,I} \le &(1+Ch)\eta_{k+1}^{N,I} + Ch^{2I-1}\mathcal{A}^N(S_0)\\ &+ C\mathbb{E}|\mathcal{R}_{p_{0,k}}(Y_{t_k}^N)|^2 + Ch\sum_{l=1}^{q}\mathbb{E}|\mathcal{R}_{p_{l,k}}(Z_{l,t_k}^N)|^2.\end{aligned} \tag{23}$$

Note that the estimate on $\max_{0\le k\le N}\mathbb{E}|Y_{t_k}^{N,I,I} - Y_{t_k}^N|^2$ given in Theorem 2 directly follows from the relation above. With the arguments used to derive (18) and using the estimate (22), we easily get

$$\begin{aligned}&\eta_k^{N,I} \le Ch^{2I-1}\mathcal{A}^N(S_0) + (1+h)\mathbb{E}|Y_{t_k}^{N,\infty,I} - Y_{t_k}^N|^2\\ &= Ch^{2I-1}\mathcal{A}^N(S_0) + (1+h)\mathbb{E}|\mathcal{R}_{p_{0,k}}(Y_{t_k}^N)|^2 + (1+h)\mathbb{E}|Y_{t_k}^{N,\infty,I} - \mathcal{P}_{p_{0,k}}(Y_{t_k}^N)|^2\end{aligned} \tag{24}$$

where we used at the last equality the orthogonality property relative to $\mathcal{P}_{p_{0,k}}$:

$$\mathbb{E}|Y_{t_k}^{N,\infty,I} - Y_{t_k}^N|^2 = \mathbb{E}|\mathcal{R}_{p_{0,k}}(Y_{t_k}^N)|^2 + \mathbb{E}|Y_{t_k}^{N,\infty,I} - \mathcal{P}_{p_{0,k}}(Y_{t_k}^N)|^2. \tag{25}$$

Furthermore, with the same techniques than for (15) and (19), we can prove

$$\begin{aligned}&\mathbb{E}|Z_{t_k}^{N,I,I} - Z_{t_k}^N|^2 = \sum_{l=1}^{q}\mathbb{E}|\mathcal{R}_{p_{l,k}}(Z_{l,t_k}^N)|^2 + \sum_{l=1}^{q}\mathbb{E}|Z_{l,t_k}^{N,I,I} - \mathcal{P}_{p_{l,k}}(Z_{l,t_k}^N)|^2\\ &\le \sum_{l=1}^{q}\mathbb{E}|\mathcal{R}_{p_{l,k}}(Z_{l,t_k}^N)|^2 + \frac{d}{h}\big(\mathbb{E}[Y_{t_{k+1}}^{N,I,I} - Y_{t_{k+1}}^N]^2 - \mathbb{E}[\mathbb{E}_k(Y_{t_{k+1}}^{N,I,I} - Y_{t_{k+1}}^N)]^2\big), \quad (26)\\ &\mathbb{E}|Y_{t_k}^{N,\infty,I} - \mathcal{P}_{p_{0,k}}(Y_{t_k}^N)|^2 \le (1+\gamma h)\mathbb{E}|\mathbb{E}_k[Y_{t_{k+1}}^{N,I,I} - Y_{t_{k+1}}^N]|^2\\ &\qquad + Ch(h+\frac{1}{\gamma})\big[\mathbb{E}|Y_{t_k}^{N,\infty,I} - Y_{t_k}^N|^2 + \mathbb{E}|Z_{t_k}^{N,I,I} - Z_{t_k}^N|^2\big]. \quad (27)\end{aligned}$$

Replacing the estimate (26) in (27), choosing $\gamma = Cd$ and using (25) directly leads to

$$\begin{aligned}(1-Ch)\mathbb{E}|Y_{t_k}^{N,\infty,I} - &\mathcal{P}_{p_{0,k}}(Y_{t_k}^N)|^2 \le (1+Ch)\eta_{k+1}^{N,I}\\ &+ Ch\sum_{l=1}^{q}\mathbb{E}|\mathcal{R}_{p_{l,k}}(Z_{l,t_k}^N)|^2 + Ch\mathbb{E}|\mathcal{R}_{p_{0,k}}(Y_{t_k}^N)|^2.\end{aligned} \tag{28}$$

Plugging this estimate into (24) completes the proof of (23).

Step 4: upper bounds for $\zeta^N = h\sum_{k=0}^{N-1}\mathbb{E}|Z_{t_k}^{N,I,I} - Z_{t_k}^N|^2$. We aim at showing

$$\begin{aligned}\zeta^N \le &Ch^{2I-2}\mathcal{A}^N(S_0) + Ch\sum_{k=0}^{N-1}\sum_{l=1}^{q}\mathbb{E}|\mathcal{R}_{p_{l,k}}(Z_{l,t_k}^N)|^2\\ &+ C\sum_{k=0}^{N-1}\mathbb{E}|\mathcal{R}_{p_{0,k}}(Y_{t_k}^N)|^2 + C\max_{0\le k\le N-1}\eta_k^{N,I}.\end{aligned} \tag{29}$$

In view of (26), we have $\zeta^N \leq h\sum_{k=0}^{N-1}\sum_{l=1}^{q}\mathbb{E}|\mathcal{R}_{p_{l,k}}(Z^N_{l,t_k})|^2 + d\sum_{k=0}^{N-1}\big(\mathbb{E}[Y^{N,I,I}_{t_k}-Y^N_{t_k}]^2-\mathbb{E}[\mathbb{E}_k(Y^{N,I,I}_{t_{k+1}}-Y^N_{t_{k+1}})]^2\big)$. Owing to (24) and (27), we obtain

$$\begin{aligned}\mathbb{E}|Y^{N,I,I}_{t_k}-&Y^N_{t_k}|^2-\mathbb{E}[\mathbb{E}_k(Y^{N,I,I}_{t_{k+1}}-Y^N_{t_{k+1}})]^2 \leq Ch^{2I-1}\mathcal{A}^N(S_0)\\ &+C\mathbb{E}|\mathcal{R}_{p_{0,k}}(Y^N_{t_k})|^2+[(1+h)(1+\gamma h)-1]\mathbb{E}|\mathbb{E}_k[Y^{N,I,I}_{t_{k+1}}-Y^N_{t_{k+1}}]|^2\\ &+Ch(h+\frac{1}{\gamma})\big[\mathbb{E}|Y^{N,\infty,I}_{t_k}-Y^N_{t_k}|^2+\mathbb{E}|Z^{N,I,I}_{t_k}-Z^N_{t_k}|^2\big].\end{aligned}$$

Taking $\gamma = 4Cd$ and h small enough such that $dC(h+\frac{1}{\gamma})\leq\frac{1}{2}$, we have proved

$$\begin{aligned}\zeta^N \leq Ch^{2I-2}\mathcal{A}^N(S_0)+Ch\sum_{k=0}^{N-1}\sum_{l=1}^{q}\mathbb{E}|\mathcal{R}_{p_{l,k}}(Z^N_{l,t_k})|^2+C\sum_{k=0}^{N-1}\mathbb{E}|\mathcal{R}_{p_{0,k}}(Y^N_{t_k})|^2\\ +C\max_{0\leq k\leq N-1}\eta^{N,I}_k+\frac{1}{2}h\sum_{k=0}^{N-1}\mathbb{E}|Y^{N,\infty,I}_{t_k}-Y^N_{t_k}|^2+\frac{1}{2}\zeta^N.\end{aligned}$$

But taking into account (25) and (28) to estimate $\mathbb{E}|Y^{N,\infty,I}_{t_k}-Y^N_{t_k}|^2$, we clearly obtain (29). This easily completes the proof of Theorem 2.

3.3 Proof of Proposition 2

As for (20), we can obtain

$$\begin{aligned}\mathbb{E}|Y^{N,k_0,x}_{t_k}-Y^{N,k_0,x'}_{t_k}|^2 \leq \frac{(1+\gamma h)}{1-Ch(h+\frac{1}{\gamma})}\mathbb{E}|\mathbb{E}_k(Y^{N,k_0,x}_{t_{k+1}}-Y^{N,k_0,x'}_{t_{k+1}})|^2\\ +\frac{Ch(h+\frac{1}{\gamma})}{1-Ch(h+\frac{1}{\gamma})}\mathbb{E}|S^{N,k_0,x}_{t_k}-S^{N,k_0,x'}_{t_k}|^2\\ +\frac{C(h+\frac{1}{\gamma})}{1-Ch(h+\frac{1}{\gamma})}\big(\mathbb{E}|Y^{N,k_0,x}_{t_{k+1}}-Y^{N,k_0,x'}_{t_{k+1}}|^2-\mathbb{E}|\mathbb{E}_k(Y^{N,k_0,x}_{t_{k+1}}-Y^{N,k_0,x'}_{t_{k+1}})|^2\big).\end{aligned}$$

Choosing $\gamma = C$ and h small enough, we get (for another constant C):

$$\mathbb{E}|Y^{N,k_0,x}_{t_k}-Y^{N,k_0,x'}_{t_k}|^2\leq(1+Ch)\mathbb{E}|Y^{N,k_0,x}_{t_{k+1}}-Y^{N,k_0,x'}_{t_{k+1}}|^2+Ch\mathbb{E}|S^{N,k_0,x}_{t_k}-S^{N,k_0,x'}_{t_k}|^2.$$

The last term above is bounded by $C|x-x'|^2$ under assumption **(H1')**. Thus, using Gronwall's lemma and assumption **(H4)**, we get the result for $y^N_{k_0}(\cdot)$. The result for $\sqrt{h}z^N_{k_0}(\cdot)$ follows by considering (4).

3.4 Proof of Proposition 3

In view of Proposition 1, it is tempting to apply a Markov property argument and to assert that Proposition 3 results from (22) written with conditional expectations $\mathbb{E}_k$. But this argumentation fails because the law used for the projection is not the conditional law $\mathbb{E}_k$ but $\mathbb{E}_0$. The right argument may be the following one. Write $Y_{t_k}^{N,i,I} = \alpha_{0,k}^{i,I} \cdot p_{0,k}(P_{t_k}^N)$. On the one hand, by (22) we have $C\mathcal{A}^N(S_0) \geq \mathbb{E}|Y_{t_k}^{N,i,I}|^2 = \alpha_{0,k}^{i,I} \cdot \mathbb{E}[p_{0,k}p_{0,k}^*]\alpha_{0,k}^{i,I} \geq |\alpha_{0,k}^{i,I}|^2\lambda_{\min}(\mathbb{E}[p_{0,k}p_{0,k}^*])$. On the other hand, $|Y_{t_k}^{N,i,I}| \leq |\alpha_{0,k}^{i,I}||p_{0,k}(P_{t_k}^N)| \leq |p_{0,k}|\sqrt{C\mathcal{A}^N(S_0)/\lambda_{\min}(\mathbb{E}[p_{0,k}p_{0,k}^*])}$. Thus, we can take $\rho_{0,k}^N(x) = \max(1, |p_{0,k}(x)|\sqrt{C\mathcal{A}^N(S_0)/\lambda_{\min}(\mathbb{E}[p_{0,k}p_{0,k}^*])})$. Analogously for $\sqrt{h}|Z_{l,t_k}^{N,i,I}|$, we have $\rho_{l,k}^N(x) = \max(1, |p_{l,k}(x)|\sqrt{C\mathcal{A}^N(S_0)/\lambda_{\min}(\mathbb{E}[p_{l,k}p_{l,k}^*])})$. Note that if $p_{l,k}$ is an orthonormal function basis, we have $\lambda_{\min}(\mathbb{E}[p_{l,k}p_{l,k}^*]) = 1$ and previous upper bounds have simpler expressions.

3.5 Proof of Theorem 3

In the sequel, set

$$\mathcal{A}_k^{N,M} = \frac{1}{M}\sum_{m=1}^{M} |\rho_{0,k}^N(P_{t_k}^{N,m})|^2, \quad \mathcal{B}_k^{N,M} = \frac{1}{M}\sum_{m=1}^{M} |f_k^m(0, \cdots, 0)|^2.$$

Obviously, we have $\mathbb{E}(\mathcal{A}_k^{N,M}) = \mathbb{E}|\rho_{0,k}^N(P_{t_k}^N)|^2$ and $\mathbb{E}(\mathcal{B}_k^{N,M}) \leq C(1+|S_0|^2)$. Now, we remind the standard contraction property in the case of least squares problems in $\mathbb{R}^M$, analogously to the case $\mathbf{L}_2(\mathbb{P})$. Consider a sequence of real numbers $(x^m)_{1\leq m\leq M}$ and a sequence $(v^m)_{1\leq m\leq M}$ of vectors in $\mathbb{R}^n$, associated to the matrix $V^M = \frac{1}{M}\sum_{m=1}^M v^m[v^m]^*$ which is supposed to be invertible ($\lambda_{\min}(V^M) > 0$). Then, the (unique) $\mathbb{R}^n$-valued vector $\theta_x = \arg\inf_\theta |x - \theta \cdot v|_M^2$ is given by

$$\theta_x = \frac{[V^M]^{-1}}{M}\sum_{m=1}^{M} v^m x^m. \tag{30}$$

The application $x \mapsto \theta_x$ is linear and moreover, we have the inequality

$$\lambda_{\min}(V^M)|\theta_x|^2 \leq |\theta_x.v|_M^2 \leq |x|_M^2. \tag{31}$$

For the further computations, it is more convenient to deal with

$$(\theta_k^{i,I,M})^* = \big(\alpha_{0,k}^{i,I,M*}, \sqrt{h}\alpha_{1,k}^{i,I,M*}, \cdots, \sqrt{h}\alpha_{q,k}^{i,I,M*}\big)$$

instead of $\alpha_k^{i,I,M}$. Then, the Picard iterations given in (10) can be rewritten

$$\theta_k^{i+1,I,M} = \arg\inf_\theta \frac{1}{M}\sum_{m=1}^{M}\big(\hat\rho_{0,k+1}^{N,m}(\alpha_{0,k+1}^{I,I,M}.p_{0,k+1}^m) + hf_k^m(\alpha_k^{i,I,M}) - \theta.v_k^m\big)^2. \tag{32}$$

Introducing the event $\mathbf{A}_k^M$, taking into account the Lipschitz property of the functions $\hat\rho_{l,k}^N$ and using the orthonormality of $p_{l,k}$, we get

$$\begin{aligned}\mathbb{E}|Y_{t_k}^{N,I,I} - Y_{t_k}^{N,I,I,M}|^2 + h\sum_{j=k}^{N-1}\mathbb{E}|Z_{t_j}^{N,I,I} - Z_{t_j}^{N,I,I,M}|^2 \leq 9\sum_{j=k}^{N-1}\mathbb{E}(|\rho_j^N(P_{t_j}^N)|^2\mathbf{1}_{[\mathbf{A}_k^M]^c})\\ + \mathbb{E}(\mathbf{1}_{\mathbf{A}_k^M}|\alpha_{0,k}^{I,I,M} - \alpha_{0,k}^{I,I}|^2) + h\sum_{j=k}^{N-1}\sum_{l=1}^{q}\mathbb{E}(\mathbf{1}_{\mathbf{A}_k^M}|\alpha_{l,j}^{I,I,M} - \alpha_{l,j}^{I,I}|^2).\end{aligned} \tag{33}$$

To obtain Theorem 3, we estimate $|\theta_k^{I,I,M} - \theta_k^{I,I}|^2$ on the event $\mathbf{A}_k^M$. This is achieved in several steps.
Step 1: contraction properties relative to the sequence $(\theta_k^{i,I,M})_{i\geq 0}$. They are summed up in the following lemma.

Lemma 1 *For h small enough, on $\mathbf{A}_k^M$ the following properties hold.*

a) $|\theta_k^{i+1,I,M} - \theta_k^{i,I,M}|^2 \leq Ch|\theta_k^{i,I,M} - \theta_k^{i-1,I,M}|^2$.

b) There is an unique vector $\theta_k^{\infty,I,M}$ such that

$$\theta_k^{\infty,I,M} = \arg\inf_\theta \frac{1}{M}\sum_{m=1}^{M}\big(\hat\rho_{0,k+1}^{N,m}(\alpha_{0,k+1}^{I,I,M}.p_{0,k+1}^m) + hf_k^m(\alpha_k^{\infty,I,M}) - \theta.v_k^m\big)^2.$$

c) We have $|\theta_k^{\infty,I,M} - \theta_k^{I,I,M}|^2 \leq [Ch]^I|\theta_k^{\infty,I,M}|^2$.

Proof. We prove a). Since $1-h \leq \lambda_{\min}(V_k^M)$ and $\lambda_{\max}(P_{l,k}^M) \leq 2$ $(0\leq l\leq q)$ on $\mathbf{A}_k^M$, in view of (31) we obtain that $(1-h)|\theta_k^{i+1,I,M} - \theta_k^{i,I,M}|^2$ is bounded by

$$\begin{aligned}&\frac{h^2}{M}\sum_{m=1}^{M}\big(f_k^m(\alpha_k^{i,I,M}) - f_k^m(\alpha_k^{i-1,I,M})\big)^2\\ \leq &Ch^2\sum_{l=0}^{q}|\alpha_{l,k}^{i,I,M} - \alpha_{l,k}^{i-1,I,M}|^2\lambda_{\max}(P_{l,k}^M) \leq Ch|\theta_k^{i,I,M} - \theta_k^{i-1,I,M}|^2.\end{aligned}$$

Now, statements a) and b) are clear. For c), apply a) reminding that $\theta_k^{0,I,M} = 0$. □

Step 2: bounds for $|\theta_k^{i,I,M}|$ on the event $\mathbf{A}_k^M$. Namely, we aim at showing that

$$|\theta_k^{i,I,M}|^2 \leq C\big(\mathcal{A}_{k+1}^{N,M} + h\mathcal{B}_k^{N,M}\big) \text{ on } \mathbf{A}_k^M. \tag{34}$$

We first consider $i=\infty$. As in the proof of Lemma 1, we get

$$\begin{aligned}(1-h)|\theta_k^{\infty,I,M}|^2 &\leq \frac{1}{M}\sum_{m=1}^{M}\big[\hat\rho_{0,k+1}^{N,m}(\alpha_{0,k+1}^{I,I,M}.p_{0,k+1}^m) + hf_k^m(\alpha_k^{\infty,I,M})\big]^2\\ &\leq (1+\gamma h)\mathcal{A}_{k+1}^{N,M} + Ch(h+\frac{1}{\gamma})\big(\mathcal{B}_k^{N,M} + \sum_{l=0}^{q}|\alpha_{l,k}^{\infty,I,M}|^2\lambda_{\max}(P_{l,k}^M)\big).\end{aligned}$$

Take $\gamma = 8C$ and h small enough to ensure $2C(h+\frac{1}{\gamma})(1+h) \leq \frac{1}{2}(1-h)$. It readily follows $|\theta_k^{\infty,I,M}|^2 \leq C(\mathcal{A}_{k+1}^{N,M} + h\mathcal{B}_k^{N,M})$, proving that (34) holds for $i=\infty$. Lemma 1-c) leads to expected bounds for other values of i.
Step 3: we remind bounds for $\theta^{i,I}$. Using Proposition 3 and in view of (13-17), we have for $i\geq 1$

$$|\theta_{l,k}^{i,I}|^2 \leq \mathbb{E}|\rho_{l,k}^N(P_{t_k}^N)|^2,\ 0\leq l\leq q; \quad |\theta_k^{\infty,I} - \theta_k^{i,I}|^2 \leq (C_fh)^{2i}\mathbb{E}|\rho_{0,k}^N(P_{t_k}^N)|^2. \tag{35}$$

Remind also the following expression of $\theta_k^{\infty,I}$, derived from (13-16) and the orthonormality of each basis $p_{l,k}$:

$$\theta_k^{\infty,I} = \mathbb{E}\big(v_k[\alpha_{0,k+1}^{I,I}.p_{0,k+1} + hf_k(\alpha_k^{\infty,I})]\big). \tag{36}$$

Step 4: decomposition of the quantity $\mathbb{E}(\mathbf{1}_{\mathbf{A}_k^M}|\theta_k^{I,I,M} - \theta_k^{I,I}|^2)$. Due to lemma 1, on $\mathbf{A}_k^M$ we get $|\theta_k^{\infty,I,M} - \theta_k^{I,I,M}|^2 \le Ch^I|\theta_k^{\infty,I,M}|^2 \le Ch^I|\theta_k^{\infty,I}|^2 + Ch^I|\theta_k^{\infty,I,M} - \theta_k^{\infty,I}|^2$. Thus, using (35), it readily follows that $\mathbb{E}(1_{\mathbf{A}_k^M}|\theta_k^{I,I,M} - \theta_k^{I,I}|^2)$ is bounded by

$$\begin{aligned}
&(1+h)\mathbb{E}(1_{\mathbf{A}_k^M}|\theta_k^{\infty,I,M} - \theta_k^{\infty,I}|^2)\\
&\quad + 2(1+\frac{1}{h})\{\mathbb{E}(1_{\mathbf{A}_k^M}|\theta_k^{I,I,M} - \theta_k^{\infty,I,M}|^2) + |\theta_k^{I,I} - \theta_k^{\infty,I}|^2\}\\
&\le (1+Ch)\mathbb{E}(1_{\mathbf{A}_k^M}|\theta_k^{\infty,I,M} - \theta_k^{\infty,I}|^2) + Ch^{I-1}\mathbb{E}|\rho_k^N(P_{t_k}^N)|^2
\end{aligned} \tag{37}$$

taking account that $I \ge 3$. On $\mathbf{A}_k^M$, V_k^M is invertible and we can set

$$\begin{aligned}
B_1 &= (\mathrm{Id} - (V_k^M)^{-1})\theta_k^{\infty,I}\\
B_2 &= (V_k^M)^{-1}\big[\mathbb{E}(v_k\hat\rho_{0,k+1}^N(\alpha_{0,k+1}^{I,I}\cdot p_{0,k+1})) - \frac{1}{M}\sum_{m=1}^M v_k^m\hat\rho_{0,k+1}^{N,m}(\alpha_{0,k+1}^{I,I}\cdot p_{0,k+1}^m)\big],\\
B_3 &= (V_k^M)^{-1}h\big[\mathbb{E}(v_kf_k(\alpha_k^{\infty,I})) - \frac{1}{M}\sum_{m=1}^M v_k^m f_k^m(\alpha_k^{\infty,I})\big],\\
B_4 &= \frac{(V_k^M)^{-1}}{M}\sum_{m=1}^M v_k^m\big[\hat\rho_{0,k+1}^{N,m}(\alpha_{0,k+1}^{I,I}\cdot p_{0,k+1}^m) - \hat\rho_{0,k+1}^{N,m}(\alpha_{0,k+1}^{I,I,M}\cdot p_{0,k+1}^m)\\
&\qquad\qquad + h(f_k^m(\alpha_k^{\infty,I}) - f_k^m(\alpha_k^{\infty,I,M}))\big].
\end{aligned}$$

Thus, by (30-36) and Definition 3 we can write $\theta_k^{\infty,I} - \theta_k^{\infty,I,M} = B_1 + B_2 + B_3 + B_4$, which gives on $\mathbf{A}_k^M$

$$|\theta_k^{\infty,I} - \theta_k^{\infty,I,M}|^2 \le 3(1+\frac{1}{h})(|B_1|^2 + |B_2|^2 + |B_3|^2) + (1+h)|B_4|^2. \tag{38}$$

Step 5: individual estimation of B_1, B_2, B_3, B_4 on $\mathbf{A}_k^M$. Remind the classic result [14]: if $\|\mathrm{Id} - F\| < 1$, $F^{-1} - \mathrm{Id} = \sum_{k=1}^\infty[\mathrm{Id} - F]^k$ and $\|\mathrm{Id} - F^{-1}\| \le \frac{\|F-\mathrm{Id}\|}{1-\|F-\mathrm{Id}\|}$. Consequently, for $F = V_k^M$ we get $\mathbb{E}(\mathbf{1}_{\mathbf{A}_k^M}\|\mathrm{Id} - (V_k^M)^{-1}\|^2) \le (1-h)^{-2}\mathbb{E}\|\mathrm{Id} - V_k^M\|^2 \le (1-h)^{-2}\mathbb{E}\|V_k^M - \mathrm{Id}\|_F^2 = (M(1-h)^2)^{-1}\mathbb{E}\|v_kv_k^* - \mathrm{Id}\|_F^2$. Thus, we have

$$\mathbb{E}(|B_1|^2 1_{\mathbf{A}_k^M}) \le \frac{C}{M}\mathbb{E}\|v_kv_k^* - \mathrm{Id}\|_F^2\ \mathbb{E}|\rho_k^N(P_{t_k}^N)|^2.$$

Since on $\mathbf{A}_k^M$ one has $\|(V_k^M)^{-1}\| \le 2$, it readily follows

$$\mathbb{E}(|B_2|^2\mathbf{1}_{\mathbf{A}_k^M}) \le \frac{C}{M}\mathbb{E}(|v_k|^2|p_{0,k+1}|^2)\mathbb{E}|\rho_{0,k}^N(P_{t_k}^N)|^2,$$

$$\mathbb{E}(|B_3|^2 1_{\mathbf{A}_k^M}) \leq \frac{Ch^2}{M}\mathbb{E}\big[|v_k|^2(1+|S_{t_k}^N|^2+|p_{0,k}|^2\mathbb{E}|\rho_{0,k}^N(P_{t_k}^N)|^2 + \frac{1}{h}\sum_{l=1}^q |p_{l,k}|^2\mathbb{E}|\rho_{l,k}^N(P_{t_k}^N)|^2)\big].$$

As in the proof of Lemma 1 and using $\|P_{0,k+1}^M\| \leq 1+h$ on $\mathbf{A}_k^M$, we easily obtain

$$(1-h)|B_4|^2 \leq (1+h)(1+\gamma h)|\alpha_{0,k+1}^{I,I} - \alpha_{0,k+1}^{I,I,M}|^2 + Ch(h+\frac{1}{\gamma})\sum_{l=0}^q |\alpha_{l,k}^{\infty,I} - \alpha_{l,k}^{\infty,I,M}|^2.$$

Step 6: final estimations. Put $\epsilon_k = \mathbb{E}\|v_k v_k^* - \mathrm{Id}\|_F^2\ \mathbb{E}|\rho_k^N(P_{t_k}^N)|^2 + \mathbb{E}(|v_k|^2|p_{0,k+1}|^2)\mathbb{E}|\rho_{0,k}^N(P_{t_k}^N)|^2 + h^2\mathbb{E}\big[|v_k|^2(1+|S_{t_k}^N|^2+|p_{0,k}|^2\mathbb{E}|\rho_{0,k}^N(P_{t_k}^N)|^2 + \frac{1}{h}\sum_{l=1}^q |p_{l,k}|^2\mathbb{E}|\rho_{l,k}^N(P_{t_k}^N)|^2)\big]$. Plug the above estimates on B_1, B_2, B_3, B_4 into (38), choose $\gamma = 3C$ and h close to 0 to ensure $Ch + \frac{C}{\gamma} \leq \frac{1}{2}$; after simplifications, we get

$$\mathbb{E}(1_{\mathbf{A}_k^M}|\theta_k^{\infty,I,M} - \theta_k^{\infty,I}|^2) \leq C\frac{\epsilon_k}{hM} + (1+Ch)\mathbb{E}(1_{\mathbf{A}_k^M}|\alpha_{0,k+1}^{I,I} - \alpha_{0,k+1}^{I,I,M}|^2).$$

But in view of Lemma 1-c) and estimates (35-34), we have $\mathbb{E}(1_{\mathbf{A}_k^M}|\alpha_{0,k+1}^{I,I} - \alpha_{0,k+1}^{I,I,M}|^2) \leq (1+h)\mathbb{E}(1_{\mathbf{A}_k^M}|\alpha_{0,k+1}^{\infty,I} - \alpha_{0,k+1}^{\infty,I,M}|^2) + Ch^{I-1}\big(1+|S_0|^2 + \mathbb{E}|\rho_{0,k+1}^N(P_{t_{k+1}}^N)|^2 + \mathbb{E}|\rho_{0,k+2}^N(P_{t_{k+2}}^N)|^2\big)$. Finally, we have proved

$$\mathbb{E}(1_{\mathbf{A}_k^M}|\theta_k^{\infty,I,M} - \theta_k^{\infty,I}|^2) \leq C\frac{\epsilon_k}{hM} + Ch^{I-1}\big(1+|S_0|^2 + \mathbb{E}|\rho_{0,k+1}^N(P_{t_{k+1}}^N)|^2 + \mathbb{E}|\rho_{0,k+2}^N(P_{t_{k+2}}^N)|^2\big) + (1+Ch)\mathbb{E}(1_{\mathbf{A}_k^M}|\alpha_{0,k+1}^{\infty,I,M} - \alpha_{0,k+1}^{\infty,I}|^2).$$

Using a contraction argument as in (37), the index ∞ can be replaced by I, without changing the inequality (with a possibly different constant C). This can be written

$$\begin{aligned}&\mathbb{E}(1_{\mathbf{A}_k^M}|\alpha_{0,k}^{I,I,M} - \alpha_{0,k}^{I,I}|^2) + h\sum_{l=1}^q \mathbb{E}(1_{\mathbf{A}_k^M}|\alpha_{l,k}^{I,I,M} - \alpha_{l,k}^{I,I}|^2)\\ &\leq C\frac{\epsilon_k}{hM} + Ch^{I-1}\big(1+|S_0|^2 + \mathbb{E}|\rho_{0,k+1}^N(P_{t_{k+1}}^N)|^2 + \mathbb{E}|\rho_{0,k+2}^N(P_{t_{k+2}}^N)|^2\big)\\ &\quad + (1+Ch)\mathbb{E}(1_{\mathbf{A}_k^M}|\alpha_{0,k+1}^{I,I,M} - \alpha_{0,k+1}^{I,I}|^2).\end{aligned}$$

Using Gronwall's lemma, the proof is complete. □

The attentive reader may have noted that powers of h are smaller here than in Theorem 2, which leads to take $I \geq 3$ instead of $I \geq 2$ before. Indeed, we can not take advantage of conditional expectations on the simulations as we did in (15) for instance. This degradation seems to be unavoidable.

Note that in the proof above, we only use the Lipschitz property of the truncation functions $\hat{\rho}_{l,k}^N$ and $\hat{\rho}_{l,k}^{N,m}$.

3.6 Proof of Theorem 4

The arguments are standard and there are essentially notational difficulties. The first partial derivatives of f w.r.t. y and z_l are respectively denoted $\partial_0 f$ and $\partial_l f$. The parameter $\beta \in]0,1]$ stands for their Hölder continuity index. Suppose w.l.o.g. that $\varepsilon < \beta$ and that each function basis $p_{l,k}$ is orthonormal. For $k < N-1$, define the quantities

$$\begin{aligned}
A^M_{l,k}(\alpha) =& \frac{1}{M}\sum_{m=1}^{M} v_k^m \partial_l f(t_k, S^{N,m}_{t_k}, \alpha_0 \cdot p^m_{0,k}, \cdots, \alpha_q \cdot p^m_{q,k})[p^m_{l,k}]^*,\\
B^M_k =& \frac{1}{M}\sum_{m=1}^{M} v_k^m [p^m_{0,k+1}]^*, \qquad D^M_k = \sqrt{M}(\mathrm{Id} - V^M_k),\\
C^M_k(\alpha) =& \sum_{m=1}^{M} \frac{\{v_k^m[\alpha^{I,I}_{0,k+1}\cdot p^m_{0,k+1} + hf^m_k(\alpha)] - \mathbb{E}\big(v_k[\alpha^{I,I}_{0,k+1}\cdot p_{0,k+1} + hf_k(\alpha)]\big)\}}{\sqrt{M}}.
\end{aligned}$$

For $k = N-1$, we set $B^M_k = 0$ and in $C^M_k(\alpha)$, the terms $\alpha^{I,I}_{0,k+1}\cdot p^m_{0,k+1}$ and $\alpha^{I,I}_{0,k+1}\cdot p_{0,k+1}$ have to be replaced respectively by $\Phi^N(P^{N,m}_{t_N})$ and $\Phi^N(P^N_{t_N})$. The definitions of $A^M_{l,k}(\alpha)$ and D^M_k are still valid. For convenience, we write $X^M \stackrel{w}{\rightarrow}$ if the (possibly vector or matrix valued) sequence $(X^M)_M$ weakly converges to a centered Gaussian variable, as M goes to infinity. For the convergence in probability to a constant, we denote $X^M \stackrel{\mathbb{P}}{\rightarrow}$. Since simulations are independent, observe that the following convergences hold:

$$\begin{aligned}
(A^M_{l,k}(\alpha^{i,I}_k), B^M_k, V^M_k)_{i\le I-1, l\le q, k\le N-1} &\stackrel{\mathbb{P}}{\rightarrow},\\
\mathcal{G}^M = (C^M_k(\alpha^{i,I}_k), D^M_k)_{i\le I-1, l\le q, k\le N-1} &\stackrel{w}{\rightarrow}. \qquad (39)
\end{aligned}$$

Note that $\lim_{M\to\infty} V^M_k \stackrel{a.s.}{=} \mathrm{Id}$ is invertible. Linearizing the functions f and $\hat\rho^{N,m}_{0,k+1}$ in the expressions of $\theta^{i,I}_k = \mathbb{E}(v_k[\alpha^{I,I}_{0,k+1}.p_{0,k+1} + hf_k(\alpha^{i-1,I}_{0,k}, \cdots, \alpha^{i-1,I}_{q,k})])$ and $\theta^{i,I,M}_k$ given by (30) leads to

$$\begin{aligned}
&|V^M_k\sqrt{M}(\theta^{i,I,M}_k - \theta^{i,I}_k) - D^M_k\theta^{i,I}_k - C^M_k(\alpha^{i-1,I}_k)\\
&\quad - B^M_k\sqrt{M}(\alpha^{I,I,M}_{0,k+1} - \alpha^{I,I}_{0,k+1}) - h\sum_{l=0}^{q} A^M_{l,k}(\alpha^{i-1,I}_k)\sqrt{M}(\alpha^{i-1,I,M}_{l,k} - \alpha^{i-1,I}_{l,k})|\\
\le{}& \mathbf{1}_{k<N-1}\frac{C}{\sqrt{M}}|\alpha^{I,I,M}_{0,k+1} - \alpha^{I,I}_{0,k+1}|^2\sum_{m=1}^{M}|v_k^m||p^m_{0,k+1}|^2\\
&\quad + \frac{C}{\sqrt{M}}|\alpha^{i-1,I,M}_k - \alpha^{i-1,I}_k|^{1+\beta}\sum_{m=1}^{M}|v_k^m||p_k^m|^{1+\beta}. \qquad (40)
\end{aligned}$$

To get Theorem 4, we prove by induction on k that $([\sqrt{M}(\theta^{i,I,M}_j - \theta^{i,I}_j)]_{j\ge k, i\le I}, \mathcal{G}^M) \stackrel{w}{\rightarrow}$. Remind that $\theta^{0,I,M}_j = \theta^{0,I}_j = 0$ for any j. Consider first $k = N-1$, for which $B^M_k = 0$, and $i = 1$. In view of (39-40),

clearly $([\sqrt{M}(\theta^{i,I,M}_{N-1} - \theta^{i,I}_{N-1})]_{i\leq \mathbf{1}}, \mathcal{G}^M) \xrightarrow{w}$. For $i = 2$, we may invoke the same argument using (39-40) and obtain $([\sqrt{M}(\theta^{i,I,M}_{N-1} - \theta^{i,I}_{N-1})]_{i\leq \mathbf{2}}, \mathcal{G}^M) \xrightarrow{w}$ provided that the upper bound in (40) converge to 0 in probability. To prove this, put $\mathcal{M}^M = M^{-1-\beta/2}\sum_{m=1}^M |v^m_{N-1}||p^m_{N-1}|^{1+\beta}$ and write $\frac{1}{\sqrt{M}}|\alpha^{1,I,M}_{N-1} - \alpha^{1,I}_{N-1}|^{1+\beta}\sum_{m=1}^M |v^m_{N-1}||p^m_{N-1}|^{1+\beta} = |\sqrt{M}(\alpha^{1,I,M}_{N-1} - \alpha^{1,I}_{N-1})|^{1+\beta}\mathcal{M}^M$. Since $[\sqrt{M}(\alpha^{1,I,M}_{N-1} - \alpha^{1,I}_{N-1})]_M$ is tight, our assertion holds if $\mathcal{M}^M$ converges to 0 as $M \to \infty$. Note that $|v_{N-1}||p_{N-1}|^{1+\beta} \in \mathbf{L}_{\frac{2+\varepsilon}{2+\beta}}(\mathbb{P})$. Thus, the strong Law of Large Numbers, in the case of i.i.d. random variables with infinite mean, leads to $\sum_{m=1}^M |v^m_{N-1}||p^m_{N-1}|^{1+\beta} = O(M^{\frac{2+\beta}{2+\varepsilon}+r})$ a.s. for any $r > 0$. Consequently, from the choice of r small enough it follows $\mathcal{M}^M \to 0$ a.s..
Iterating this argumentation readily leads to $([\sqrt{M}(\theta^{i,I,M}_{N-1} - \theta^{i,I}_{N-1})]_{i\leq \mathbf{I}}, \mathcal{G}^M) \xrightarrow{w}$. For the induction for $k < N-1$, we apply the techniques above. There is an additional contribution due to B^M_k, which can be handled as before.

4 Numerical experiments

To use the algorithm, we need to specify the basis functions that we choose at each time t_k. In all the cases described below, assumption **(H4)** is fulfilled. Thus, we can take advantage of the Lipschitz continuity of $y^N_k(\cdot)$ and $\sqrt{h}z^N_k(\cdot)$ (cf. Proposition 2) to choose the basis functions. Firstly, as $y^N_k(\cdot)$ and $\sqrt{h}z^N_k(\cdot)$ have the same regularity, we take the same basis $p_{l,k}$ for $0 \leq l \leq q$.
Suppose that $P^N_{t_k}$ takes its value in $\mathbb{R}^{d'}$. To define the *finite* basis $p_{0,k}$, we consider a *bounded* domain $D_k = \{x \in \mathbb{R}^{d'} : \forall i, 1 \leq i \leq d', |x_i - \bar{x}_{i,k}| \leq R_k\}$ centered[4] in $\bar{x}_k$, that we partition into small hypercubes of edge δ (of course, when some components of $P^N_{t_k}$ are known to take their values in particular sub-domains, these sub-domains are considered). We denote this partition $(D_{i,k})_i$. As basis functions we consider the indicator functions of the small hypercubes. To analyze the error of projections on this particular basis, observe that for any arbitrary point x_i in $D_{i,k}$, we have

$$\begin{aligned}\mathbb{E}\big(\mathcal{R}_{p_{0,k}}(Y^N_{t_k})^2\big) &\leq \mathbb{E}\big(|Y^N_{t_k}|^2 1_{D^c_k}(P^N_{t_k})\big) + \sum_i \mathbb{E}\big(1_{D_{i,k}}(P^N_{t_k})|y^N_k(P^N_{t_k}) - y^N_k(x_i)|^2\big)\\ &\leq C\delta^2 + \mathbb{E}(|Y^N_{t_k}|^2 1_{D^c_k}(P^N_{t_k})),\end{aligned}$$

using (8) for the second inequality. To evaluate $\mathbb{E}(|Y^N_{t_k}|^2 1_{D^c_k}(P^N_{t_k}))$, note that, by adapting the proof of Proposition 2, we have $|Y^N_{t_k}|^2 \leq C(1 + |S^N_{t_k}|^2 + \mathbb{E}_k|P^N_{t_N}|^2)$. Thus, if $\mathbb{E}|P^N_{t_k}|^\alpha < \infty$ for $\alpha > 2$, we have $\mathbb{E}(|Y^N_k|^2 1_{D^c}(P^N_{t_k})) \leq \frac{C_{\alpha,k}}{R_k^{\alpha-2}}$, with an explicit constant $C_{\alpha,k}$. The choice $R_k \approx h^{-\frac{2}{\alpha-2}}$ and $\delta = h$ leads to

$$\mathbb{E}|\mathcal{R}_{p_{0,k}}(Y^N_{t_k})|^2 \leq Ch^2.$$

The same estimates hold for $\mathbb{E}|\mathcal{R}_{p_{0,k}}(\sqrt{h}Z^N_{l,t_k})|^2$. Thus we obtain the same accuracy than in Theorem 1. Observe that quantization techniques [3] could help

[4] the center $\bar{x}$ should be chosen approximately equal to $\mathbb{E}(P^N_{t_k})$.

in defining a better partition of the domain D_k at each time t_k.
Moreover, we may expect that the conditional expectations (4) and (5) define very smooth functions $y_k^N(\cdot)$ and $\sqrt{h}z_k^N(\cdot)$, which would justify to take more regular basis functions. But it seems to be difficult to control uniformly (in h) the high order derivatives of $y_k^N(\cdot)$ and $\sqrt{h}z_k^N(\cdot)$. Thus, we only exploit their Lipschitz continuity property.
Concerning the Picard iterations, note that performing I iterations is as costly as only one iteration, because the regression matrix V_k^M (which requires most of the computational efforts) is computed once for all the Picard iterations. Note also that computing each V_k^M (and its inverse) on parallel processors is feasible and certainly speeds up the method.

Call option in a perfect market model. Here we consider a call option with maturity T, strike K on an underlying asset ($d = 1$) whose dynamics is given by the Black-Scholes model with drift μ and volatility σ. We take $T = 0.25$, $K = S_0 = 100$, $\mu = 5\%$, $\sigma = 20\%$ and $r = 2\%$ for the interest rate. The associated Black-Scholes price (BSP) equals $C(K, r) = 4.23$. The goal in considering such a simple case is to compare the results of our algorithm with a reference value. As basis functions, we take a slightly different basis compared to the description above. To avoid the restriction to the domain D_k, we consider at time t_k the interval defined by the extremes of $S_{t_k}^{N,m}$ over m. Then we divide it into smaller intervals of size $\delta = h$. This idea is relative to histograms with fixed or random bandwidth as described, for example, in [5].

To test the algorithm, we compare $Y_0^{N,I,I,M}$ ($I = 3$) with the BSP. We test different values of N and M and report the CPU time (in seconds) : the results have been obtained with a 2 GhZ processor. As $Y_0^{N,I,I,M}$ satisfies a central limit theorem (Theorem 4), we estimate its accuracy by calculating the bias and the standard deviation of the price given by the algorithm. More precisely, we launch 100 times the algorithm : from the collected values, we estimate the empirical mean and standard deviation (in parenthesis below). We obtain the following results:

N	M	100 ($< 1s$)	400 ($< 1s$)	1600 ($< 1s$)	6400 ($< 1s$)
2	Price	4.32 (0.37)	4.28 (0.13)	4.29 (0.07)	4.27 (0.03)

N	M	100 ($< 1s$)	400 ($< 1s$)	1600 ($< 1s$)	6400 ($< 1s$)
4	Price	4.14 (0.25)	4.22 (0.10)	4.25 (0.06)	4.25 (0.03)

N	M	100 ($< 1s$)	400 ($< 1s$)	1600 ($< 1s$)	6400 ($< 1s$)
8	Price	3.73 (0.54)	4.12 (0.10)	4.20 (0.03)	4.23 (0.02)

N	M	100 ($< 1s$)	400 ($< 1s$)	1600 ($1s$)	6400 ($6s$)
16	Price	2.71(2.01)	3.83 (0.24)	4.15 (0.07)	4.19 (0.02)

These results highlight several features, which are coherent with previous theoretical estimates. Firstly, the convergence w.r.t. h holds, but it is surprisingly fast. It is presumably due to the simplicity of the example (constant coefficients

and linear driver). Secondly, for any given value of N, the standard deviation decreases as $\frac{1}{\sqrt{M}}$ as it can be expected from Theorems 3 and 4. Thirdly, for N large, we need more and more simulations to obtain an accurate price : this is coherent with Theorem 3, where the upper bound of the error explodes as N tends to infinity, M and the basis functions being fixed.
Note that for a given value of M, the standard deviation is smaller than for the direct Monte-Carlo method. For example, for $N = 8$ and $M = 6400$, the standard deviation of the price given by our algorithm is 0.02 and 0.08 for the usual Monte-Carlo method under the risk neutral probability. This is not surprising because at each time t_k, the term $Z_{t_k}^{N,i,I,M} \Delta W_k$ plays the role of a control variate.

Different interest rates [4]. We now consider the same option, with the same dynamics for S_t, but the seller of the option have two different interest rates: R for borrowing and r for lending with $R > r$. Here, the driver f in (2) is no more linear and takes the form $f(t, x, y, z) = -\{yr + z\theta - (y - \frac{z}{\sigma})^-(R - r)\}$ where $\theta = \frac{\mu - r}{\sigma}$. To test the algorithm, we take $r = 0.02$ and $R = 0.04$. As mentionned in [11], the price for the option must be the Black-Scholes price $C(K, R) = 4.48$ since one has to permanently borrow money to replicate the option. We make the same tests as before and report the mean price and the standard deviation. We obtain:

N	M	100 ($< 1s$)	400 ($< 1s$)	1600 ($< 1s$)	6400 ($< 1s$)
4	Price	4.30 (0.27)	4.43 (0.11)	4.46 (0.04)	4.47 (0.02)

For other values of N, the results are quite similar to those obtained for the previous case of $R = r$: the linearity or non linearity of f seem not to modify the accuracy of the algorithm.
Call spread. Now, we test the algorithm in the case of a call spread option with payoff $(S_T - K_1)_+ - (S_T - K_2)_+$. The dynamics of S_t, T, r and R are still the same and we take $K_1 = 95$ and $K_2 = 105$. The results are:

N	M	100 ($< 1s$)	400 ($< 1s$)	1600 ($< 1s$)	6400 ($< 1s$)
4	Price	5.09 (0.17)	5.13 (0.08)	5.14 (0.04)	5.14 (0.02)

We note that the price of the algorithm converges to the difference of the Black-Scholes prices $5.15 = C(K_1, R) - C(K_2, R)$. It means that in this case, the seller of the option always has to borrow money to replicate the option. It is thus of interest to see what happens when the seller sometimes borrows and sometimes lends money to replicate the option. We consider a call spread option with payoff $(S_T - K_1)_+ - 2(S_T - K_2)_+$. We put $r = 0.01$ and $R = 0.06$ to increase the impact of different interest rates. We obtain:

N	M	100 ($< 1s$)	400 ($< 1s$)	1600 ($< 1s$)	6400 ($< 1s$)
4	Price	3.05 (0.3)	2.97 (0.18)	2.92 (0.07)	2.91 (0.03)

N	M	100 ($< 1s$)	400 ($< 1s$)	1600 ($< 1s$)	6400 ($< 1s$)
8	Price	3.29 (0.3)	3.02 (0.11)	2.96 (0.05)	2.95 (0.03)

N	M	100 $(< 1s)$	400 $(< 1s)$	1600 $(< 1s)$	6400 $(< 1s)$
12	Price	2.48 (9.36)	3.12 (0.14)	3.00 (0.06)	2.96 (0.02)

Here the impact of different interest rates can not be neglected. Indeed, the price is higher than either $2.76 = C(K_1, r) - 2C(K_2, r)$ (i.e. one always lends money to replicate the payoff) or $2.75 = C(K_1, R) - 2C(K_2, R)$ (i.e. one always borrows money to replicate the payoff).

Asian option. Here, we consider a discrete Asian option, with payoff $(\frac{1}{5}\sum_{i=0}^{4} S_{t_i} - K)_+$ where $t_i = i\frac{T}{4}$. The model parameters are set to those of the first example. The risk-neutral price is calculated via a standard Monte-Carlo method with $M' = 1000000$ paths. The price is 2.30 (0.003). The basis functions are still chosen as described above, except that now we define squares in $\mathbb{R}^2$ instead of intervals in $\mathbb{R}^1$.

N	M	100 $(< 1s)$	400 $(< 1s)$	1600 $(< 1s)$	6400 $(< 1s)$
4	Price	2.57 (7.53)	2.12 (0.15)	2.25 (0.04)	2.29 (0.01)

N	M	100 $(< 1s)$	400 $(< 1s)$	1600 $(< 1s)$	6400 $(8s)$
8	Price	0.8 (11.3)	1.43 (2.69)	2.33 (0.76)	2.26 (0.03)

As expected, the algorithm price converges towards the reference price. In this two-dimensional case, the previous remarks are valid. In particular, the greater N is, the slower the convergence with M is, according to Theorem 3.

5 Conclusion

In this paper, we design a new algorithm for the numerical resolution of BSDEs. At each discretization time, it combines a finite number of Picard iterations (3 seems to be relevant) and regressions on function bases. These regressions are evaluated at once with one set of simulated paths, unlike [6] where one needs as many sets of paths as discretization times. We mainly focus on the theoretical justification of this scheme. We prove $\mathbf{L}_2$ estimates and a central limit theorem as the number of simulations goes to infinity. To confirm the accuracy of the method, we only present few convincing tests and we refer to [19] for a more detailed numerical analysis. Even if no related results have been presented here, an extension to reflected BSDEs is straightforward (as in [6]) and allows to deal with American options. At last, we mention that our results prove the convergence of the *Hedged Monte Carlo* method of Bouchaud *etal.* [7], which can be expressed in terms of BSDEs with a linear driver.

References

[1] L. Andersen and R. Brotherton-Ratcliffe. Exact exotics. *Risk*, 9:85–89, 1996.

[2] V. Bally. Approximation scheme for solutions of BSDE. In *Backward stochastic differential equations (Paris, 1995–1996)*, volume 364 of *Pitman Res. Notes Math. Ser.*, pages 177–191. Longman, Harlow, 1997.

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

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

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

[6] B. Bouchard and N. Touzi. Discrete time approximation and Monte-Carlo simulation of backward stochastic differential equations. *to appear in Stochastic Processes and their Applications*, 2004.

[7] J.P. Bouchaud, M. Potters, and D. Sestovic. Hedged Monte-Carlo: low variance derivative pricing with objective probabilities. *Physica A*, 289:517–525, 2001.

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

[9] D. Chevance. Numerical methods for backward stochastic differential equations. In *Numerical methods in finance*, Publ. Newton Inst., pages 232–244. Cambridge Univ. Press, Cambridge, 1997.

[10] E. Clément, D. Lamberton, and P. Protter. An analysis of a least squares regression method for American option pricing. *Finance Stoch.*, 6(4):449–471, 2002.

[11] J. Cvitanic, J. Ma Hedging options for a large investor and Forward-Backward SDE's. *Annals of Applied Probability*, 6:370–398, 1996.

[12] D. Duffie and L. Epstein. Stochastic differential utility. *Econometrica*, 60(2):353–394, 1992.

[13] H. Föllmer and M. Schweizer. Hedging of contingent claims under incomplete information. In M.H.A. Davis and R.J. Elliot, editors, *Applied stochastic analysis*. Gordon and Breach, London, 1990.

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

[15] B. Kamrad and P. Ritchken. Multinomial approximating models for options with k state variables. *Manage. Sci.*, 37(12):1640–1652, 1991.

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

[17] N. El Karoui and M.C. Quenez. Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM J. Control Optim.*, 33(1):29–66, 1995.

[18] B. Lapeyre and E. Temam. Competitive Monte Carlo methods for the pricing of Asian options. *Journal of Computational Finance*, 5(1):39–59, 2001.

[19] J.P. Lemor. PhD Thesis. *In progress.* 2004.

[20] F. Longstaff and E.S. Schwartz. Valuing American options by simulation: a simple least squares approach. *The Review of Financial Studies*, 14:113–147, 2001.

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

[22] J. Ma, P. Protter, and J.M. Yong. Solving forward-backward stochastic differential equations explicitly—a four step scheme. *Probab. Theory Related Fields*, 98(3):339–359, 1994.

[23] J. Ma and J. Zhang. Path regularity for solutions of backward stochastic differential equations. *Probab. Theory Related Fields*, 122(2):163–190, 2002.

[24] N.J. Newton. Variance reduction for simulated diffusions. *SIAM Journal on Applied Mathematics*, 54(6):1780–1805, 1994.

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

[26] S. Peng. Dynamically consistent evaluations and expectations. Technical report, Institute of Mathematics, Shandong University, 2003.

[27] S. Peng. Nonlinear expectations, noninear evaluations and risk measures. In *Lectures presented in CIME EMS Bressanone, Italy July 6 13, 2003*, 2003.

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