Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 117 (2007) 803–829
stochastic processes and their applications
www.elsevier.com/locate/spa

Error expansion for the discretization of backward stochastic differential equations

Emmanuel Gobet[a,*], Céline Labart[b]

[a] ENSIMAG - INPG, IMAG - LMC, BP 53, 38041 Grenoble Cedex 9, France
[b] CMAP, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau, France

Received 23 February 2006; received in revised form 1 September 2006; accepted 16 October 2006
Available online 10 November 2006

Abstract

We study the error induced by the time discretization of decoupled forward–backward stochastic differential equations (X, Y, Z). The forward component X is the solution of a Brownian stochastic differential equation and is approximated by a Euler scheme X^N with N time steps. The backward component is approximated by a backward scheme. Firstly, we prove that the errors $(Y^N - Y, Z^N - Z)$ measured in the strong L_p-sense ($p \geq 1$) are of order $N^{-1/2}$ (this generalizes the results by Zhang [J. Zhang, A numerical scheme for BSDEs, The Annals of Applied Probability 14 (1) (2004) 459–488]). Secondly, an error expansion is derived: surprisingly, the first term is proportional to $X^N - X$ while residual terms are of order N^{-1}.

MSC: 60H07; 60F05; 60H10; 65G99

Keywords: Backward stochastic differential equation; Discretization scheme; Malliavin calculus; Semi-linear parabolic PDE

1. Introduction

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space on which is defined a q-dimensional standard Brownian motion W, whose natural filtration, augmented with $\mathbb{P}$-null sets, is denoted by $(\mathcal{F}_t)_{0 \leq t \leq T}$ (T is a fixed terminal time). We consider the solution (X, Y, Z) to a decoupled

* Corresponding author. Tel.: +33 4 76 63 56 93; fax: +33 4 76 63 12 63.
E-mail addresses: emmanuel.gobet@imag.fr (E. Gobet), labart@cmapx.polytechnique.fr (C. Labart).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2006.10.007

forward–backward stochastic differential equation (FBSDE in short). Namely, X is the $\mathbb{R}^d$-valued process solution of

$$X_t = x + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s, \tag{1}$$

and Y (resp. Z) is a real-valued adapted (resp. predictable $\mathbb{R}^q$-valued) process solution of

$$-\mathrm{d}Y_t = f(t, X_t, Y_t, Z_t)\mathrm{d}t - Z_t\mathrm{d}W_t, \qquad Y_T = \Phi(X_T). \tag{2}$$

We assume standard Lipschitz properties on the coefficients, which ensure existence and uniqueness in appropriate L_2-spaces (see Pardoux and Peng [18], or Ma and Yong [14] for numerous references). During the last decade, more and more attention has been paid to these equations, because of their natural applications in Mathematical Finance or in the probabilistic resolution of semi-linear partial differential equations (PDE in short): see El Karoui et al. [5] or Pardoux [17].

Our aim is to study the most usual time approximation of (X, Y, Z). For X, we use the Euler scheme X^N with N discretization times $(t_k = kh)_{0\le k\le N}$ ($h = \frac{T}{N}$ is the time step). For convenience, set $\Delta W_k = W_{t_{k+1}} - W_{t_k}$ (ΔW_k^l componentwise). X^N is defined by $X_0^N = x$ and

$$t \in [t_k, t_{k+1}], \quad X_t^N = X_{t_k}^N + b(t_k, X_{t_k}^N)(t - t_k) + \sigma(t_k, X_{t_k}^N)(W_t - W_{t_k}). \tag{3}$$

The backward SDE (2) is approximated by (Y^N, Z^N) defined in a backward manner by $Y_{t_N}^N = \Phi(X_{t_N}^N)$ and

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

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

where $\mathbb{E}_{t_k}$ is the conditional expectation w.r.t. $\mathcal{F}_{t_k}$ and $*$ is the transpose operator. Additional tools are needed to derive a fully implementable scheme, in particular for the computations of conditional expectations. We refer to Bouchard and Touzi [2] for Malliavin calculus techniques, or to Gobet et al. [6] and Lemor et al. [13] for empirical regression methods. In this work, we leave these further questions and we only address the error analysis between (Y, Z) and (Y^N, Z^N).

On the one hand, Zhang [20] proves (in a slightly different form) that the error $\max_{k\le N} \|Y_{t_k}^N - Y_{t_k}\|_{L_2} \le CN^{-1/2}$. This is done under rather minimal Lipschitz assumptions on b, σ, f, Φ. On the other hand, when f does not depend on z and the coefficients are smooth, one knows that $|Y_0^N - Y_0| \le CN^{-1}$ (see Chevance [3]). We aim at filling the gap regarding these two different rates of convergence. In the following, we prove that

- Chevance's results are extended to the case of f depending also on z.
- The rate N^{-1} holds true also for the difference $|Z_0^N - Z_0|$.
- More generally, for the other discretization times t_k, we expand the error as

  $$|Y_{t_k}^N - Y_{t_k} - \alpha_k \cdot (X_{t_k}^N - X_{t_k})| \le CN^{-1} \vee |X_{t_k}^N - X_{t_k}|^2$$

  (for an explicit and bounded random vector α_k).
- An analogous expansion is available for Z.

Since $|X^N_{t_k} - X_{t_k}|^2$ has the same order in L_p than N^{-1}, the error on Y is mainly due to the error $X^N_{t_k} - X_{t_k}$. Thus, Zhang's results are a consequence of this expansion, and Chevance's ones as well since $X^N_0 = X_0$. The gap is filled.

In addition, we learn from this expansion that if one could perfectly simulate X (as for Brownian motion with constant drift, geometric Brownian motion or Ornstein–Uhlenbeck process), the error on the BSDE would be of order N^{-1} and not $N^{-1/2}$ as stated by Zhang's results. Also, if one could use a discretization scheme for X of order 1 for the strong error (for instance Milshtein scheme whenever possible), the error on the BSDE would be of order N^{-1} (we would need to extend our analysis to other discretization schemes, this is straightforward for the Milshtein scheme).

The paper is organized as follows. In Section 2, we define the assumptions on the coefficients, recall the connection between BSDEs and semi-linear PDEs (which is important for our analysis). Finally, we state our main results. Firstly in Theorem 6, we extend Zhang's results to L_p-norm. Secondly in Theorem 7, we expand the error on Y. Lastly in Theorem 8, we deal with the error on Z. Naturally, stronger and stronger assumptions are required for these theorems. Proofs of the three results are postponed to Sections 3–5: we combine BSDE techniques, martingale estimates and Malliavin calculus.

Notation

- *Differentiation*. If $g : \mathbb{R}^d \mapsto \mathbb{R}^q$ is a differentiable function, its gradient $\nabla_x g(x) = (\partial_{x_1} g(x), \ldots, \partial_{x_d} g(x))$ takes values in $\mathbb{R}^q \otimes \mathbb{R}^d$. At many places, $\nabla_x g(x)$ will simply be denoted $g'(x)$. If $g : \mathbb{R}^d \mapsto \mathbb{R}$ is a twice differentiable function, its Hessian $H_x(g)$ takes values in $\mathbb{R}^d \otimes \mathbb{R}^d$: $(H_x(g))_{i,j} = \partial^2_{x_i x_j} g$. If $g : \mathbb{R}^d \times \mathbb{R}^q \mapsto \mathbb{R}$, $g''_{xy}(x, y)$ takes values in $\mathbb{R}^d \otimes \mathbb{R}^q : (g''_{xy})_{ij} = \frac{\partial^2 g}{\partial x_i \partial y_j}$, for $1 \le i \le d$, $1 \le j \le q$.
- *Function spaces*. For an integer $k \ge 1$, we denote by $C_b^{k/2,k,k,k}$ the set of continuously differentiable functions $\phi : (t, x, y, z) \in [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^q \mapsto \phi(t, x, y, z)$ such that the partial derivatives $\partial_t^{l_0} \partial_x^{l_1} \partial_y^{l_2} \partial_z^{l_3} \phi(t, x, y, z)$ exist for $2l_0 + l_1 + l_2 + l_2 \le k$ and are uniformly bounded. The analogous set of functions that do not depend on y and z is denoted by $C_b^{k/2,k}$. This set is denoted by $C_b^{(k+\alpha)/2,k+\alpha}$ ($\alpha \in]0, 1[$) if in addition the highest derivatives are Hölder continuous with index α w.r.t. x and $\alpha/2$ w.r.t. t (for a precise definition, see Ladyzenskaja et al. [12]).
- *Norm*. For a d-dimensional vector U, we set $|U|^2 = \sum_{i=1}^d U_i^2$. For a $d \times q$-dimensional matrix A, A_i denotes its ith column, and A^i its ith row. Moreover, $|A|^2 = \sum_{i,j=1}^{d,q} A^2_{i,j}$.
- *Constants*. Let C denote a generic constant which may depend on the coefficients b, σ, f, Φ and on the dimensions d and q. We will keep the same notation $K(T)$ for all finite, non-negative, and non-decreasing functions w.r.t. T: they do not depend on x and h. The generic notation $K(T, x)$ stands for any function bounded by $K(T)(1 + |x|^q)$, for some $q \ge 0$.
- *$O(U)$ and $O_k(h)$*. A random vector R is such that $R = O(U)$ for a non-negative random variable U if $|R| \le K(T, x)U$ (in particular, $R = O(h)$ means $|R| \le K(T, x)h$). The notation $R = O_k(h^p)$ means $|R| \le \lambda_k^N h^p$, where λ_k^N is $\mathcal{F}_{t_k}$-measurable, $\sup_N \mathbb{E}(\sup_k |\lambda_k^N|^q) \le K(T, x)$, for $q \ge 1$.
- *$\mathbb{E}_{t_k}$ and Var_{t_k}*. $\mathbb{E}_{t_k}$ is the conditional expectation w.r.t. $\mathcal{F}_{t_k}$ and $\mathrm{Var}_{t_k}(X) = \mathbb{E}_{t_k}(X^2) - (\mathbb{E}_{t_k}(X))^2$.
- *Malliavin calculus*. We use the notations of Nualart [16] for weak spaces $\mathbb{D}^{k,p}$.
- *Discretization*. Let $s \in [t_k, t_{k+1}[$. We define $\eta(s) = t_k$.

2. Main results

2.1. Hypotheses

The coefficients $b : [0, T]\times\mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0, T]\times\mathbb{R}^d \to \mathbb{R}^{d\times q}$, $f : [0, T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^q \to \mathbb{R}$ and $\Phi : \mathbb{R}^d \to \mathbb{R}$ satisfy one of the following set of assumptions.

Hypothesis 1. The functions b, σ, f and Φ are bounded in x, are uniformly Lipschitz continuous w.r.t. (x, y, z) and Hölder continuous of parameter $\frac{1}{2}$ w.r.t. t. In addition, Φ is of class $C_b^{2+\alpha}$ for some $\alpha \in]0, 1[$ and the matrix-valued function $a = \sigma\sigma^*$ is uniformly elliptic.

Hypothesis 2. Assume Hypothesis 1 and that the functions b, σ are in $C_b^{\frac{3}{2},3}$, f is in $C_b^{\frac{3}{2},3,3,3}$, Φ is in $C_b^{3+\alpha}$ for some $\alpha \in]0, 1[$.

Hypothesis 3. Assume Hypothesis 1 and that the functions b, σ are in $C_b^{2,4}$, f is in $C_b^{2,4,4,4}$, Φ is in $C_b^{4+\alpha}$ for some $\alpha \in]0, 1[$.

We do not assert that these smoothness and boundedness conditions are the weakest ones for our error analysis, but they are sufficient. Investigations regarding minimal assumptions would be certainly interesting but it is beyond the scope of the paper.

2.2. Connection between Markovian BSDEs and semi-linear parabolic PDEs

We recall classical results connecting (Y, Z) and the solution and its gradient of the following semi-linear PDE on $[0, T] \times \mathbb{R}^d$:

$$(\partial_t + \mathcal{L}_{(t,x)})u(t, x) + f(t, x, u(t, x), \nabla_x u(t, x)\sigma(t, x)) = 0, \qquad (6)$$
$$u(T, x) = \Phi(x),$$

where $\mathcal{L}_{(t,x)}$ is the second order differential operator

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

(see for instance Ma and Zhang [15] or Pardoux [17]).

Proposition 4. *Under Hypothesis 1, one has*

$$\forall t \in [0, T], \quad Y_t = u(t, X_t), \qquad Z_t = \nabla_x u(t, X_t)\sigma(t, X_t), \qquad (7)$$

where u is the unique classic solution $C_b^{1,2}$ of the PDE (6).

In addition under Hypothesis 2, $u \in C_b^{\frac{3}{2},3}$, and under Hypothesis 3, $u \in C_b^{2,4}$.

The first result of this Proposition corresponds to Theorem 2.1 of Delarue and Menozzi [4]. The two last regularity results can be proved in the same way. In fact for this, we would only need b, σ to be in $C_b^{1+\alpha/2,2+\alpha}$; the additional smoothness is used later for Malliavin calculus computations.

2.3. Main results

We now turn to the statement of our results. We remind the following well-known upper bound on the Euler Scheme, which is useful in the sequel.

Proposition 5. *Let σ and b be Lipschitz continuous. Then*

$$\forall p \geq 1, \quad \left[\mathbb{E}\left(\sup_{t\leq T}|X_t^N - X_t|^p\right)\right]^{\frac{1}{p}} \leq K(T,x)\frac{1}{\sqrt{N}}.$$

In fact, for all $p \geq 1$ one has

$$\left[\mathbb{E}_{t_i}\left(\sup_{t_i\leq t\leq T}|X_t^N - X_t|^p\right)\right]^{\frac{1}{p}} \leq K(T,X_{t_i})\frac{1}{\sqrt{N}} + |X_{t_i}^N - X_{t_i}|. \tag{8}$$

Our first result is an extension of the L_2-estimates in Zhang [20] to L_q-estimates (see also Gobet et al. [6]).

Theorem 6. *Let us assume Hypothesis 1. Let $q > 0$. We define the error*

$$e_q(N) = \left[\max_{0\leq k\leq N}\mathbb{E}|Y_{t_k} - Y_{t_k}^N|^q + \mathbb{E}\left(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}|Z_{t_k}^N - Z_t|^2\mathrm{d}t\right)^{\frac{q}{2}}\right]^{\frac{1}{q}},$$

where Y^N and Z^N are defined by (4) and (5). Then $|e_q(N)| \leq K(T,x)\frac{1}{\sqrt{N}}$.

By slightly strengthening the smoothness assumptions on b, σ, f and Φ, we are able to expand the error on Y.

Theorem 7. *Let us assume Hypothesis 2. Then, the following expansion holds*

$$Y_{t_k}^N - Y_{t_k} = \nabla_x u(t_k, X_{t_k})(X_{t_k}^N - X_{t_k}) + O_k\left(\frac{1}{N}\right) + O(|X_{t_k}^N - X_{t_k}|^2).$$

In view of Proposition 5, $|X_{t_k}^N - X_{t_k}|^2$ and N^{-1} have the same order (in L_p). Hence it turns out that $\nabla_x u(t_k, X_{t_k})(X_{t_k}^N - X_{t_k})$ is the first order term in the error $Y_{t_k}^N - Y_{t_k}$. Obviously, this estimate implies that of Theorem 6. As mentioned in the introduction, the evaluation of Y_0 by Y_0^N has still an accuracy of order N^{-1} since initial conditions for X^N and X coincide. Note that if there is no discretization error for the process X, $Y_{t_k}^N - Y_{t_k} = O(\frac{1}{N})$, a fact which is not clear from Eqs. (4) and (5). A nice situation corresponds to σ independent of x (this is a very specific situation where Euler and Milshtein schemes are equal): in that case $\|X_{t_k}^N - X_{t_k}\|_{L_p} = O(N^{-1})$ and one gets the order of accuracy N^{-1} for Y.

For Z which plays the role of a gradient relative to Y, we get an analogous result about the error, up to increasing by 1 the degree of smoothness of the coefficients.

Theorem 8. *Let us assume Hypothesis 3. Then, the following expansion holds*

$$Z_{t_k}^N - Z_{t_k} = \left(\nabla_x[\nabla_x u\,\sigma]^*(t_k, X_{t_k})(X_{t_k}^N - X_{t_k})\right)^* + O_k\left(\frac{1}{N}\right) + O(|X_{t_k}^N - X_{t_k}|^2).$$

Remark 9. The above results are sufficient to derive the weak convergence of the renormalized error process $[\sqrt{N}(Y_t^N - Y_t)]_{0\leq t\leq T}$ and $[\sqrt{N}(Z_t^N - Z_t)]_{0\leq t\leq T}$, except that one has to define Y^N

and Z^N between discretization times. For $t \in [t_k, t_{k+1}[$, analogously to (4) and (5) we define

$$Y_t^N = \mathbb{E}_t \left(Y_{t_{k+1}}^N + (t_{k+1} - t) f(t, X_t^N, Y_{t_{k+1}}^N, Z_t^N) \right),$$
$$Z_t^N = \frac{1}{t_{k+1} - t} \mathbb{E}_t \left(Y_{t_{k+1}}^N (W_{t_{k+1}} - W_t)^* \right).$$

Theorems 7 and 8 can be extended to all $t \in [0, T]$. We have

$$Y_t^N - Y_t = \nabla_x u(t, X_t)(X_t^N - X_t) + O_t\left(\frac{1}{N}\right) + O(|X_t^N - X_t|^2),$$
$$Z_t^N - Z_t = \left(\nabla_x [\nabla_x u\, \sigma]^*(t, X_t)(X_t^N - X_t) \right)^* + O_t\left(\frac{1}{N}\right) + O(|X_t^N - X_t|^2).$$

Theorem 3.5 of Kurtz and Protter [11] allows us to establish the weak convergence of the processes $\sqrt{N}(Y^N - Y)$, and $\sqrt{N}(Z^N - Z)$. Indeed, the process $[\sqrt{N}(X_t^N - X_t)]_{0 \le t \le T}$ weakly converges to the solution of

$$U_t = \sum_{i=1}^q \int_0^t \nabla_x \sigma_i(s, X_s) U_s \mathrm{d}W_s^i + \int_0^t \nabla_x b(s, X_s) U_s \mathrm{d}s + \frac{1}{\sqrt{2}} \sum_{i,j=1}^q \int_0^t \sum_{k=1}^d \partial_{x_k} \sigma_i(s, X_s) \sigma_{kj}(s, X_s) \mathrm{d}V_s^{ij},$$

where $(V^{ij})_{1 \le i,j \le q}$ are independent standard Brownian motions and independent of W. Furthermore, the convergence is stable (see Jacod and Protter [9]). Hence, $[\sqrt{N}(X_t^N - X_t), \sqrt{N}(Y_t^N - Y_t), \sqrt{N}(Z_t^N - Z_t), X_t]_{0 \le t \le T}$ weakly converges to $[U_t, \nabla_x u(t, X_t) U_t, ([\nabla_x [\nabla_x u\, \sigma]^*(t, X_t)] U_t)^*, X_t]_{0 \le t \le T}$.

2.4. Comments

2.4.1. Weak error

From Theorems 7 and 8 we can derive estimates related to the weak errors on Y and Z.

Theorem 10. *Let ψ be a three times continuously differentiable function with bounded derivatives. Let us assume Hypothesis 2. Then, one has*

$$\mathbb{E}(\psi(Y_{t_k}^N) - \psi(Y_{t_k})) = O\left(\frac{1}{N}\right).$$

Under Hypothesis 3, the same result applies to Z.

Proof. A Taylor expansion of ψ yields

$$\mathbb{E}(\psi(Y_{t_k}^N) - \psi(Y_{t_k})) = \mathbb{E}((Y_{t_k}^N - Y_{t_k})\psi'(Y_{t_k})) + O((Y_{t_k}^N - Y_{t_k})^2).$$

By using Theorem 7, we get

$$\begin{aligned} &\mathbb{E}((Y_{t_k}^N - Y_{t_k})\psi'(Y_{t_k})) \\ &= \mathbb{E}\left[\psi'(Y_{t_k}) \nabla_x u(t_k, X_{t_k})(X_{t_k}^N - X_{t_k}) + O_k\left(\frac{1}{N}\right) + O(|X_{t_k}^N - X_{t_k}|^2) \right], \\ &= \mathbb{E}(\psi'(u(t_k, X_{t_k})) \nabla_x u(t_k, X_{t_k})(X_{t_k}^N - X_{t_k})) + O\left(\frac{1}{N}\right). \end{aligned}$$

Hypotheses on ψ and u enable us to apply Remark 15 (see later in Section 4) to $\mathbb{E}(\psi'(u(t_k, X_{t_k}))\nabla_x u(t_k, X_{t_k})(X^N_{t_k} - X_{t_k}))$. The result follows. □

Analyzing weak errors on Y and Z is admittedly useful, but studying pathwise estimates can also be relevant. Actually, both estimates are complementary. For instance, practitioners in finance are interested in finding hedging strategies. This corresponds to solving BSDEs, where Y and Z respectively represent the value of the replicating portfolio and the hedging strategy. On the one hand, Theorems 7 and 8 are suitable tools to study these quantities for computational issues. On the other hand, Theorem 10 enables us to quantify the error on the distribution of the portfolio value, which is relevant in a risk management perspective.

2.4.2. *Global error of the numerical resolution of BSDE*

As recalled in the introduction, there exist several techniques to numerically solve BSDEs. The one we present here refers to Lemor et al. [13]; it turns out to be presumably the most efficient procedure. The authors propose a numerical scheme based on iterative regressions on function bases $p_{0,k}(\cdot), p_{1,k}(\cdot), \ldots, p_{q,k}(\cdot)$ (each being represented as a vector), whose coefficients are evaluated using M extra independent simulations of $(X^N_{t_k})_{0\le k\le N-1}$ and of the Brownian increments $(\Delta W_k)_{0\le k\le N-1}$. Let $(y^{N,M}_k(X^N_{t_k}), z^{N,M}_{1,k}(X^N_{t_k}), \ldots, z^{N,M}_{q,k}(X^N_{t_k}))_{0\le k\le N-1}$ denote the approximation of the solution of the discretized BSDE $(Y^N_{t_k}, Z^N_{1,t_k}, \ldots, Z^N_{q,t_k})_{0\le k\le N-1}$ computed in a backward manner with the following algorithm.

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

$$\alpha^M_{l,k} = \arg\inf_\alpha \frac{1}{M}\sum_{m=1}^{M}\left| y^{N,M}_{k+1}(X^{N,m}_{t_{k+1}})\frac{\Delta W^{l,m}_k}{h} - \alpha\cdot p_{l,k}(X^{N,m}_{t_k})\right|^2.$$

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

$$\frac{1}{M}\sum_{m=1}^{M}|y^{N,M}_{k+1}(X^{N,m}_{t_{k+1}}) + hf(t_k, X^{N,m}_{t_k}, y^{N,M}_{k+1}(X^{N,m}_{t_{k+1}}), \alpha^M_{l,k}\cdot p_{l,k}(X^{N,m}_{t_k}))$$
$$- \alpha\cdot p_{0,k}(X^{N,m}_{t_k})|^2.$$

Thus, define $y^{N,M}_k(\cdot)$ and $z^{N,M}_{l,k}(\cdot)$ by

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

Actually, the true algorithm requires the use of additional truncation operators that we have omitted for the sake of simplicity, see Lemor et al. [13] for details. The following error on the unknown regression functions $(y^{N,M}_k, z^{N,M}_{l,k})_{1\le l\le q, 0\le k\le N-1}$

$$\max_{0\le k\le N}\mathbb{E}|Y^N_{t_k} - y^{N,M}_k(X^N_{t_k})|^2 + h\mathbb{E}\sum_{k=0}^{N-1}|Z^N_{t_k} - z^{N,M}_k(X^N_{t_k})|^2$$

is essentially bounded by $NC_{M,p}$, where M and p respectively denote the number of simulated paths and the set of functions. For suitable choices of M and p, $C_{M,p}$ goes to 0 at a given rate. This result allows them to optimally tune the parameters to ensure a given accuracy. Hence, summing this numerical error and the discretization's one given by Theorems 7 and 8 leads to

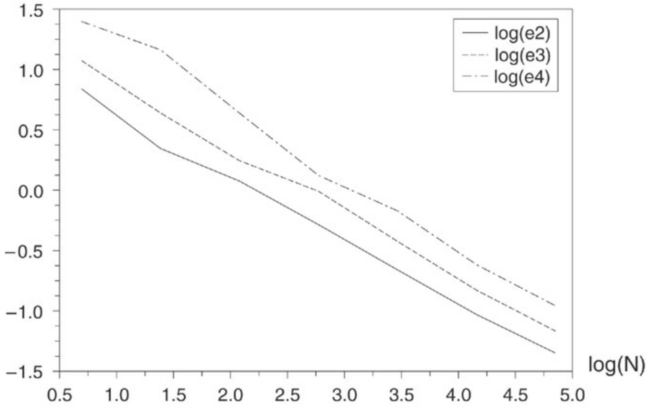

Fig. 1. Evolution of $e_2(N), e_3(N), e_4(N)$ w.r.t. $\log(N)$.

the global error. For example, assume that $\|X^N_{t_k} - X_{t_k}\|_{L_p} = O(\frac{1}{N})$. Then, from Theorem 7, we get $\mathbb{E}|Y_{t_k} - y_k^{N,M}(X^N_{t_k})|^2 \le C(\frac{1}{N^2} + NC_{M,p})$.

2.5. Numerical experiments

In this part, we draw some graphs to illustrate the results given by Theorems 6 and 7. To do so, one needs to explicitly know X and Y. Let us consider a Call option pricing problem. We assume that X follows the Black–Scholes model in dimension $d = 1$, $\frac{\mathrm{d}X_t}{X_t} = \mu \mathrm{d}t + \sigma \mathrm{d}W_t$, with $\sigma = 0.2$, $\mu = 0.1$ and $X_0 = 100$. The driver f is defined by $f(t, x, y, z) = -ry - \theta z$, where $\theta = \frac{\mu - r}{\sigma}$ and $r = 0.02$. The terminal condition $\Phi(x)$ is given by $(x - K)_+$, where $K = 100$. The maturity of the option is $T = 1$. The continuous backward equation can be solved, Y_t is the price of a standard Call option (see El Karoui et al. [5] for a detailed computation).

We compute X^N and Y^N by using (3) and (4) and get

$$X^N_{t_k} = X_0 \prod_{j=0}^{k-1} (1 + \mu h + \sigma \Delta W_j),$$

$$Y^N_{t_k} = \mathbb{E}_{t_k}\left[\Phi(X^N_T) \prod_{j=k}^{N-1} (1 - rh - \theta \Delta W_j)\right].$$

Fig. 1 refers to Theorem 6. We plot the evolution of the logarithm of $e_2(N), e_3(N)$ and $e_4(N)$ w.r.t. $\log(N)$.

We use 1000 simulations to approximate the L_p-norm $e_p(N)$ and to compute each conditional expectation $Y^N_{t_k}$, we use 1000 Monte Carlo simulations. We compute $\log(e_p(N))_{p=2,3,4}$ for $N = 2^j$, $j = 1, \ldots, 7$. Looking at the graph, we see that the evolutions of $\log(e_p(N))_{p=2,3,4}$ w.r.t. $\log(N)$ are almost linear. In view of Theorem 6, the slope should be of order $-\frac{1}{2}$. By using a linear regression method, we get the parameters a, b, std where $\log(e_p(N)) = a * \log(N) + b$, and *std* represents the standard deviation of the residuals. Table 1 sums up the values of a, b, std for $p = 2, 3, 4$. Clearly a is of order $-\frac{1}{2}$.

Fig. 2 refers to Theorem 7. We plot the evolution of $\log(a(N))$, where $a(N) = (\mathbb{E}[|Y^N_{t_k} - Y_{t_k} - \nabla_x u(t_k, X_{t_k})(X^N_{t_k} - X_{t_k})|^2])^{\frac{1}{2}}$ w.r.t. $\log(N)$, at time $t_k = \frac{T}{2}$. We use 100 simulations

Table 1
Coefficients of the linear regression of $\log(e_p(N))$, $p = 2, 3, 4$ w.r.t. $\log(N)$

	a	*b*	*std*
L_2 error	−0.5179119	1.14106	0.0384534
L_3 error	−0.5321072	1.4078415	0.0367535
L_4 error	−0.5891505	1.858531	0.0662573

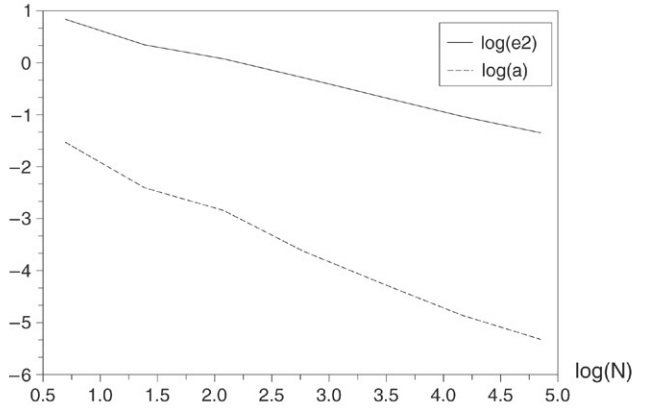

Fig. 2. Evolution of $\log(a(N))$ and $\log(e_2(N))$ w.r.t. $\log(N)$.

to approximate the L_2-norm and to compute each $Y^N_{t_k}$ we use 10^6 Monte Carlo simulations. N behaves as 2^j, $j = 1, \ldots, 7$. We note that $\log(a(N))$ actually evolves almost linearly w.r.t $\log(N)$. Regarding Theorem 7, the slope should be of order -1. If we still use a linear regression,we get the slope $a = -0.9123248$, $b = -1.0172153$ and the standard deviation of the residuals equals 0.0940069.

3. Proof of Theorem 6

Extra notations for all the proofs. For any process U (except the Brownian increments ΔW_k), we define $\Delta U_k = U^N_{t_k} - U_{t_k}$. Let θ_s denote (s, X_s, Y_s, Z_s) and $f^N_{t_k}$ denote $f(t_k, X^N_{t_k}, Y^N_{t_{k+1}}, Z^N_{t_k})$.

$\overline{Z}_{t_k}$ is defined as $h\overline{Z}_{t_k} := \mathbb{E}_{t_k} \int_{t_k}^{t_{k+1}} Z_s \mathrm{d}s$ and we put $\Delta\overline{Z}_k = Z^N_{t_k} - \overline{Z}_{t_k}$.

If $q = 2$, the result has already been proved in Gobet et al. [6], under Lipschitz conditions on b, σ, f, Φ. Thanks to the inequality $\mathbb{E}|U|^q \leq (\mathbb{E}|U|^{2p})^{\frac{q}{2p}}$ for $2p \geq q$, we only need to prove the theorem for $q = 2p$, where $p \in \mathbb{N}^*$.

First, we give some estimates which can be easily established. We have, under Hypothesis 1, $\forall s \in [t_k, t_{k+1}]$,

$$\mathbb{E}_{t_k}(|X_s - X_{t_k}|^{2p} + |Y_s - Y_{t_k}|^{2p} + |Z_s - \overline{Z}_{t_k}|^{2p}) \leq Ch^p. \tag{9}$$

In the following computations, these estimates are repeatedly used.

3.1. *Proof of* $\max_{0 \leq k \leq N} \mathbb{E}|Y_{t_k} - Y^N_{t_k}|^{2p} = O(h^p)$

We prove the following result, which is a bit more general.

Proposition 11. $\max_{i\leq k\leq N}\mathbb{E}_{t_i}|Y_{t_k}-Y^N_{t_k}|^{2p}=O_i(h^p)+|\Delta X_i|^{2p}$.

By taking $i=0$, we get $\max_{0\leq k\leq N}\mathbb{E}|Y_{t_k}-Y^N_{t_k}|^{2p}=O(h^p)$.
Assume that we have

$$|\Delta Y_k|^2\leq(1+Ch)\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2+Ch|\Delta X_k|^2+Ch^2. \tag{10}$$

Then, using the inequality $(a+b)^p\leq a^p(1+\epsilon(2^{p-1}-1))+b^p(1+\frac{2^{p-1}-1}{\epsilon^{p-1}})$ for $0<\epsilon<1$, we deduce

$$|\Delta Y_k|^{2p}\leq(1+Ch)^{p+1}\mathbb{E}_{t_k}|\Delta Y_{k+1}|^{2p}+C^ph^p(|\Delta X_k|^2+Ch)^p\left(1+\frac{C}{h^{p-1}}\right).$$

Take the conditional expectation w.r.t. $\mathcal{F}_{t_i}$ to get $\mathbb{E}_{t_i}|\Delta Y_k|^{2p}\leq(1+Ch)\mathbb{E}_{t_i}|\Delta Y_{k+1}|^{2p}+h(h^p+\mathbb{E}_{t_i}|\Delta X_k|^{2p})$. Using (8) for $|\Delta X_k|$ and Gronwall's lemma yields $\max_{i\leq k\leq N}\mathbb{E}_{t_i}|Y_{t_k}-Y^N_{t_k}|^{2p}=O_i(h^p)+|\Delta X_i|^{2p}$. □
Now we prove the inequality (10). From (2) and (4) we obtain

$$\Delta Y_k=\mathbb{E}_{t_k}(\Delta Y_{k+1})+\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}(f^N_{t_k}-f(\theta_s))\mathrm{d}s. \tag{11}$$

By applying Young's inequality, that is $(a+b)^2\leq(1+\gamma h)a^2+(1+\frac{1}{\gamma h})b^2$, where γ will be fixed later, and using the Lipschitz property of f, we get

$$\begin{aligned}|\Delta Y_k|^2\leq{}&(1+\gamma h)(\mathbb{E}_{t_k}(\Delta Y_{k+1}))^2+C\left(h+\frac{1}{\gamma}\right)\left[h^2+\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|X_s-X^N_{t_k}|^2\mathrm{d}s\right]\\&+C\left(h+\frac{1}{\gamma}\right)\left[\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|Y_s-Y^N_{t_{k+1}}|^2\mathrm{d}s+\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|Z_s-Z^N_{t_k}|^2\mathrm{d}s\right].\end{aligned} \tag{12}$$

Let us introduce $\overline{Z}_{t_k}$ (see extra notations at the beginning of Section 3):

$$\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|Z_s-Z^N_{t_k}|^2\mathrm{d}s=\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|Z_s-\overline{Z}_{t_k}|^2\mathrm{d}s+h\mathbb{E}_{t_k}|\overline{Z}_{t_k}-Z^N_{t_k}|^2. \tag{13}$$

Thanks to the Cauchy–Schwarz inequality we have

$$|\mathbb{E}_{t_k}(\Delta Y_{k+1}\Delta W^l_k)|^2\leq h\{\mathbb{E}_{t_k}(|\Delta Y_{k+1}|^2)-|\mathbb{E}_{t_k}(\Delta Y_{k+1})|^2\}.$$

Hence, as $h\overline{Z}_{t_k}=\mathbb{E}_{t_k}(\{Y_{t_{k+1}}+\int_{t_k}^{t_{k+1}}f(\theta_s)\mathrm{d}s\}\Delta W^*_k)$, with a bounded f, it follows that

$$h^2|\overline{Z}_{t_k}-Z^N_{t_k}|^2\leq dh\left(\mathbb{E}_{t_k}(|\Delta Y_{k+1}|^2)-|\mathbb{E}_{t_k}(\Delta Y_{k+1})|^2\right)+Ch^3. \tag{14}$$

By plugging (13) and (14) into (12), we get

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

We can write $\mathbb{E}_{t_k}|Y_s - Y^N_{t_{k+1}}|^2 \le 2\mathbb{E}_{t_k}|Y_s - Y_{t_{k+1}}|^2 + 2\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2$. By doing the same for $X_s - X^N_{t_{k+1}}$, and taking $\gamma = C$, we obtain

$$\begin{aligned}|\Delta Y_k|^2 \le{}& (1+Ch)\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2 + Ch|\Delta X_k|^2 + Ch\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2\\ &+ C\left[h^2 + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|X_s - X_{t_k}|^2\mathrm{d}s + \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|Y_s - Y_{t_{k+1}}|^2\mathrm{d}s\right]\\ &+ C\left[\mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}|Z_s - \overline{Z}_{t_k}|^2\mathrm{d}s\right].\end{aligned}$$

Using (9) yields $|\Delta Y_k|^2 \le (1+Ch)\mathbb{E}_{t_k}|\Delta Y_{k+1}|^2 + Ch|\Delta X_k|^2 + Ch^2$. □

3.2. *Proof of* $\mathbb{E}(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}|Z^N_{t_k} - Z_t|^2\mathrm{d}t)^{\frac{p}{2}} = O(h^p)$

First of all, we can split this summation into two terms

$$\begin{aligned}\mathbb{E}\left(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}|Z^N_{t_k} - Z_t|^2\mathrm{d}t\right)^p \le{}& C\mathbb{E}\left(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}|\overline{Z}_{t_k} - Z_t|^2\mathrm{d}t\right)^p\\ &+ C\mathbb{E}\left(h\sum_{k=0}^{N-1}|\Delta\overline{Z}_k|^2\right)^p.\end{aligned}$$

Thanks to (9), we have $\mathbb{E}(\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}|\overline{Z}_{t_k} - Z_t|^2\mathrm{d}t)^p \le T^{p-1}\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}\mathbb{E}|\overline{Z}_{t_k} - Z_t|^{2p}\mathrm{d}t = O(h^p)$.

Scheme of the proof of $\mathbb{E}(h\sum_{k=0}^{N-1}|\Delta\overline{Z}_k|^2)^p = O(h^p)$. The first key point is to slice the summation into small intervals and show that the result is true for small time intervals. The second key point is to use Rosenthal's inequality, see Theorem 2.12, page 23 of Hall and Heyde [8]. By using (14) and taking the expectation, we can write:

$$\mathbb{E}\left(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2\right)^p \le C\mathbb{E}\left(\sum_{k=0}^{k_1}\mathrm{Var}_{t_k}\Delta Y_{k+1}\right)^p + Ch^p. \tag{15}$$

We use Rosenthal's inequality to upper bound

$$\begin{aligned}\mathbb{E}\left(\sum_{k=0}^{k_1}\mathrm{Var}_{t_k}\Delta Y_{k+1}\right)^p &\le C\mathbb{E}\left(\sum_{k=0}^{k_1}\Delta Y_{k+1} - \mathbb{E}_{t_k}\Delta Y_{k+1}\right)^{2p},\\ &\le C3^{2p-1}\left[\mathbb{E}\Delta Y_{k_1+1}^{2p} + \mathbb{E}\Delta Y_0^{2p}\right.\\ &\quad\left. + \mathbb{E}\left(\sum_{k=0}^{k_1}(\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1})\right)^{2p}\right].\end{aligned}$$

By plugging this inequality into (15) and using the previous estimate on $|\Delta Y_k|$, we get

$$\mathbb{E}\left(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2\right)^p \le O(h^p) + C\mathbb{E}\left(\sum_{k=0}^{k_1}(\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1})\right)^{2p}. \tag{16}$$

We now tackle the term $\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1}$. Using (11), we have $\sum_{k=0}^{k_1}(\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1}) = \sum_{k=0}^{k_1}\int_{t_k}^{t_{k+1}}(\mathbb{E}_{t_k}(f_{t_k}^N - f(\theta_s)))\mathrm{d}s$. By doing the same kind of proof as before, that is using the fact that f is Lipschitz and the results on $\mathbb{E}|\Delta X_k|^{2p}$ and $\mathbb{E}|\Delta Y_k|^{2p}$, we find

$$\mathbb{E}\left(\sum_{k=0}^{k_1}(\Delta Y_k - \mathbb{E}_{t_k}\Delta Y_{k+1})\right)^{2p} \leq O(h^p) + C(hk_1)^p\mathbb{E}\left(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2\right)^p.$$

By plugging this term back into (16), we can write $(1-C(hk_1)^p)\mathbb{E}(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2)^p = O(h^p)$. Consequently, if we choose $k_1 \leq \frac{1}{(2C)^{\frac{1}{p}}h}$ we come up with $\mathbb{E}(h\sum_{k=0}^{k_1}|\Delta\overline{Z}_k|^2)^p = O(h^p)$. This result can be extended to any summation involving at most Δk terms, where $\Delta k \leq \frac{1}{(2C)^{\frac{1}{p}}h}$. We can cover the interval $\{0,\ldots,N-1\}$ with a finite number of elementary intervals of size Δk and we get $\mathbb{E}(h\sum_{k=0}^{N-1}|\Delta\overline{Z}_k|^2)^p = O(h^p)$, which completes our proof. □

From this result and (9), we also deduce

$$\mathbb{E}\left(h\sum_{k=0}^{N-1}|\Delta Z_k|^2\right)^p = O(h^p), \tag{17}$$

which is very useful in the following.

4. Proof of Theorem 7

To expand the error, we use the usual techniques of stochastic analysis, combining martingale estimates and Malliavin calculus tools.

4.1. Preliminary estimates

Sections 4 and 5 contain proofs with similar calculations, which are quite technical. In order to be as clear as possible, we state two results really useful in the sequel, which are related to Malliavin calculus (see Nualart [16]). The results give sufficient conditions for expectations and conditional expectations to be small w.r.t. the time step h. They are based on ideas from Kohatsu-Higa and Pettersson [10] and Gobet and Munos [7].

Proposition 12. *Let $F \in \mathbb{D}^{1,2}$ with $\mathbb{E}_{t_k}|F|^2 + \sup_{t_k\leq s\leq T}\mathbb{E}_{t_k}|D_sF|^2 < \infty$ and let U be an Itô process of the form $U_t = U_0 + \int_0^t \alpha_s\mathrm{d}s + \int_0^t \beta_s\mathrm{d}W_s$, with $\sup_{t_k\leq s\leq T}\mathbb{E}_{t_k}|\alpha_s|^2 + \sup_{t_k\leq s\leq T}\mathbb{E}_{t_k}|\beta_s|^2 < \infty$. Then, $\forall(t,t')$ such that $t_k \leq t \leq t' \leq t_{k+1}$,*

$$|\mathbb{E}_{t_k}[F(U_t - U_{t'})]| \leq (t'-t)\left[(\mathbb{E}_{t_k}|F|^2)^{\frac{1}{2}}\left(\sup_{t\leq s\leq t'}\mathbb{E}_{t_k}|\alpha_s|^2\right)^{\frac{1}{2}} + \left(\sup_{t\leq s\leq t'}\mathbb{E}_{t_k}|D_sF|^2\right)^{\frac{1}{2}}\left(\sup_{t\leq s\leq t'}\mathbb{E}_{t_k}|\beta_s|^2\right)^{\frac{1}{2}}\right].$$

This proposition can be easily proved. Assume without loss of generality that F and U are one-dimensional. From the duality formula, we have $\mathbb{E}_{t_k}[F(\int_t^{t'}\alpha_s\mathrm{d}s + \int_t^{t'}\beta_s\mathrm{d}W_s)] = \mathbb{E}_{t_k}[\int_t^{t'}(F\alpha_s +$

$D_sF \cdot \beta_s)\mathrm{d}s]$. Thanks to Cauchy–Schwarz inequality and hypotheses on α and β, we get the result.

Definition 13. F satisfies the condition R_k if $F \in \mathbb{D}^{k,\infty}$ and if $\mathcal{C}_{k,p}(F) := \|F\|_{L_p} + \sum_{j\le k} \sup_{0\le s_1,\ldots,s_j\le T} \|D_{s_1,\ldots,s_j}F\|_{L_p} < \infty$.

Proposition 14. *Let F satisfy the condition R_2. For simplicity we set $\mathrm{d}W_s^0 = \mathrm{d}s$. Assume that $U_t \in \mathbb{R}^d$ satisfies the following stochastic expansion property*

$$U_t = \sum_{i,j=0}^{q} c_{i,j}^{U,0}(t) \int_0^t c_{i,j}^{U,1}(s) \left(\int_{\eta(s)}^s c_{i,j}^{U,2}(r)\mathrm{d}W_r^i \right) \mathrm{d}W_s^j, \tag{$\mathcal{P}$}$$

where $\{(c_{i,j}^{U,i_1}(t))_{t\ge 0} : 0 \le i, j \le q, 0 \le i_1 \le 2\}$ are adapted processes satisfying

- $\forall (i,j), 1 \le i,j \le q, \forall t \in [0,T]$, *$c_{i,j}^{U,0}(t)$ satisfies R_2, and $\mathcal{C}_{2,p}^U := \sup_{0\le t\le T}\sup_{1\le i,j\le q} \mathcal{C}_{2,p}(c_{i,j}^{U,0}(t)) < \infty$, $p \ge 1$.*
- $\forall (i,j), 1 \le i,j \le q, \forall t \in [0,T]$, *$c_{i,j}^{U,1}(t), c_{0,j}^{U,0}(t), c_{i,0}^{U,0}(t), c_{i,0}^{U,1}(t)$ satisfy R_1, and $\mathcal{C}_{1,p}^U := \sup_{0\le t\le T}\sup_{1\le i,j\le q}\{\mathcal{C}_{1,p}(c_{i,j}^{U,1}(t)) + \mathcal{C}_{1,p}(c_{0,j}^{U,0}(t)) + \mathcal{C}_{1,p}(c_{i,0}^{U,0}(t)) + \mathcal{C}_{1,p}(c_{i,0}^{U,1}(t))\} < \infty$, $p \ge 1$.*
- $\forall (i,j), 0 \le i,j \le q, \forall t \in [0,T]$, *$c_{i,j}^{U,2}(t), c_{0,j}^{U,1}(t), c_{0,0}^{U,0}(t)$ satisfy R_0, and $\mathcal{C}_{0,p}^U := \sup_{0\le t\le T}\sup_{0\le i,j\le q}\{\mathcal{C}_{0,p}(c_{i,j}^{U,2}(t)) + \mathcal{C}_{0,p}(c_{0,j}^{U,1}(t)) + \mathcal{C}_{0,p}(c_{0,0}^{U,0}(t))\} < \infty$, $p \ge 1$.*

Thus, there is a constant $K(T)$ which depends polynomially on $\mathcal{C}_{2,p}(F), \mathcal{C}_{2,p}^U, \mathcal{C}_{1,p}^U, \mathcal{C}_{0,p}^U$ (for some $p \ge 1$) such that $|\mathbb{E}[FU_t]| \le K(T)h$.

Indeed, we have

$$\begin{aligned}
\mathbb{E}(FU_t) &= \sum_{i,j=0}^{q} \mathbb{E}\left(F c_{i,j}^{U,0}(t) \int_0^t c_{i,j}^{U,1}(s) \left(\int_{\eta(s)}^s c_{i,j}^{U,2}(r)\mathrm{d}W_r^i \right) \mathrm{d}W_s^j \right) \\
&= \sum_{i,j=1}^{q} \int_0^t \int_{\eta(s)}^s \mathbb{E}\left(D_r^i \left[D_s^j \{F c_{i,j}^{U,0}(t)\} c_{i,j}^{U,1}(s) \right] c_{i,j}^{U,2}(r) \right) \mathrm{d}r\mathrm{d}s \\
&\quad + \sum_{j=1}^{q} \int_0^t \int_{\eta(s)}^s \mathbb{E}\left[D_s^j \{F c_{0,j}^{U,0}(t)\} c_{0,j}^{U,1}(s) c_{0,j}^{U,2}(r) \right] \mathrm{d}r\mathrm{d}s \\
&\quad + \sum_{i=1}^{q} \int_0^t \int_{\eta(s)}^s \mathbb{E}\left[D_r^i \{F c_{i,0}^{U,0}(t) c_{i,0}^{U,1}(s)\} c_{i,0}^{U,2}(r) \right] \mathrm{d}r\mathrm{d}s \\
&\quad + \int_0^t \int_{\eta(s)}^s \mathbb{E}\left(F c_{0,0}^{U,0}(t) c_{0,0}^{U,1}(s) c_{0,0}^{U,2}(r) \right) \mathrm{d}r\mathrm{d}s.
\end{aligned}$$

Then, the result readily follows.

Remark 15. Under Hypothesis 2, we can show (see later the proof of (41)) that for each t, $X_t^N - X_t$ satisfies the expansion $(\mathcal{P})$. Hence, if F satisfies R_2, Proposition 14 yields

$$|\mathbb{E}[F(X_t^N - X_t)]| = O(h)$$

uniformly in $t \in [0,T]$, which is a very useful result for the sequel.

4.2. Expansion of $Y^N_{t_k} - Y_{t_k}$

In the following, we assume that Hypothesis 2 is in force. This implies in particular that u is bounded, of class $C_b^{3/2,3}$ (see Proposition 4). We also easily prove that $\forall p \geq 1, \forall k \in \{0, \ldots, N-1\}$ (see Nualart [16] e.g.)

- $$\mathbb{E}_{t_k}\left(\sup_{t_k \le t \le T} |X_t|^{2p}\right) < K(T)(1 + |X_{t_k}|^{2p}), \qquad \sup_{t_k \le s \le T} \mathbb{E}_{t_k}\left(\sup_{t_k \le t \le T} |D_s X_t|^p\right) \le C,$$
$$\sup_{t_k \le s,r \le T} \mathbb{E}_{t_k}\left(\sup_{t_k \le t \le T} |D_r D_s X_t|^p\right) + \sup_{t_k \le s,r,v \le T} \mathbb{E}_{t_k}\left(\sup_{t_k \le t \le T} |D_v D_r D_s X_t|^p\right) \le C, \quad (18)$$
- $$\mathbb{E}_{t_k}\left(\sup_{t_k \le t \le T} |X^N_t|^{2p}\right) < K(T)(1 + |X^N_{t_k}|^{2p}), \qquad \sup_{N, t_k \le s \le T} \mathbb{E}_{t_k}\left(\sup_{t_k \le t \le T} |D_s X^N_t|^p\right) \le C,$$
$$\sup_{N, t_k \le s,r \le T} \mathbb{E}_{t_k}\left(\sup_{t_k \le t \le T} |D_r D_s X^N_t|^p\right) + \sup_{N, t_k \le s,r,v \le T} \mathbb{E}_{t_k}\left(\sup_{t_k \le t \le T} |D_v D_r D_s X^N_t|^p\right) \le C. \quad (19)$$

Due to the Markov property of $(X^N_{t_k})_k$, one has $Y^N_{t_k} = u^N(t_k, X^N_{t_k})$ for some Lipschitz function $u^N(t_k, \cdot)$ (see Gobet et al. [6]) with an obvious definition of u^N. Actually, under our assumptions, this function is even three times differentiable w.r.t. x. Thus, the difference ΔY_k can be written as follows:

$$\Delta Y_k = (u^N(t_k, X^N_{t_k}) - u(t_k, X^N_{t_k})) + (u(t_k, X^N_{t_k}) - u(t_k, X_{t_k})).$$

Since u is of class $C_b^{3/2,3}$, the last term of the previous inequality becomes

$$u(t_k, X^N_{t_k}) - u(t_k, X_{t_k}) = \nabla_x u(t_k, X_{t_k}) \Delta X_k + O(|\Delta X_k|^2). \quad (20)$$

To complete the proof, we apply the following lemma.

Lemma 16. *Under Hypothesis 2,* $|u^N(t_k, x) - u(t_k, x)| \le K(T, x)h$.

The result above is new but not so surprising. Indeed, if f is identically zero, the difference is only related to the weak approximation of $\Phi(X_T)$ by $\Phi(X^N_T)$: from Bally and Talay [1], one knows that this is of order h.

The rest of this section is devoted to the proof of the lemma. We only give the proof for $t_k = 0$. We want to find an upper bound for $|u^N(0, x) - u(0, x)| = |\Delta Y_0|$.

For the sake of clarity, we split the proof into several steps.

Step 1: *Linearization of the error*. We show that

$$\Delta Y_k = \mathbb{E}_{t_k}(\Delta Y_{k+1} \xi_k + h f'_x(\theta_{t_k}) \Delta X_k + h \chi_k), \quad (21)$$

with

$$\xi_k = (1 + h f'_y(\theta_{t_k}) + f'_z(\theta_{t_k}) \Delta W_k), \quad (22)$$

$$\begin{aligned}\chi_k &= \int_{t_k}^{t_{k+1}} (G_0(s, X_s) + f'_y(\theta_{t_k}) G_y(s, X_s) + f'_z(\theta_{t_k}) G_z(s, X_s))\mathrm{d}s \\ &\quad + \int_0^1 (1-\lambda) \Big[\Delta X_k^* f''_{xx}(\theta_{t_k}^\lambda) \Delta X_k + f''_{yy}(\theta_{t_k}^\lambda)(Y_{t_{k+1}}^N - Y_{t_k})^2 + \Delta Z_k f''_{zz}(\theta_{t_k}^\lambda) \Delta Z_k^* \\ &\quad + 2\Delta X_k^* f''_{xy}(\theta_{t_k}^\lambda)(Y_{t_{k+1}}^N - Y_{t_k}) + 2\Delta X_k^* f''_{xz}(\theta_{t_k}^\lambda) \Delta Z_k^* \\ &\quad + 2(Y_{t_{k+1}}^N - Y_{t_k}) f''_{yz}(\theta_{t_k}^\lambda) \Delta Z_k^* \Big] \mathrm{d}\lambda, \end{aligned} \tag{23}$$

where $\theta_{t_k}^\lambda = \lambda(t_k, X_{t_k}^N, Y_{t_{k+1}}^N, Z_{t_k}^N) + (1-\lambda)\theta_{t_k}$ and G_0, G_y, G_z are bounded functions. From (11) and by introducing $f(\theta_{t_k})$, we have

$$\Delta Y_k = \mathbb{E}_{t_k} \left(\Delta Y_{k+1} + h(f_{t_k}^N - f(\theta_{t_k})) + \int_{t_k}^{t_{k+1}} (f(\theta_{t_k}) - f(\theta_s))\mathrm{d}s \right). \tag{24}$$

By applying Itô's formula to $f(\theta_u)$ between t_k and s we show that, under Hypothesis 2, $\int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k}(f(\theta_{t_k}) - f(\theta_s))\mathrm{d}s = h \int_{t_k}^{t_{k+1}} E_{t_k}(G_0(s, X_s))\mathrm{d}s$, where G_0 is a bounded function. In the second term, perform a second order expansion of f around θ_{t_k} to get

$$\begin{aligned} f_{t_k}^N - f(\theta_{t_k}) &= f'_x(\theta_{t_k})\Delta X_k + f'_y(\theta_{t_k})\Delta Y_{k+1} + f'_z(\theta_{t_k})\Delta Z_k^* + f'_y(\theta_{t_k})(Y_{t_{k+1}} - Y_{t_k}) \\ &\quad + \int_0^1 (1-\lambda) \Big[\Delta X_k^* f''_{xx}(\theta_{t_k}^\lambda)\Delta X_k + f''_{yy}(\theta_{t_k}^\lambda)(Y_{t_{k+1}}^N - Y_{t_k})^2 \\ &\quad + \Delta Z_k f''_{zz}(\theta_{t_k}^\lambda)\Delta Z_k^* + 2\Delta X_k^* f''_{xy}(\theta_{t_k}^\lambda)(Y_{t_{k+1}}^N - Y_{t_k}) + 2\Delta X_k^* f''_{xz}(\theta_{t_k}^\lambda)\Delta Z_k^* \\ &\quad + 2(Y_{t_{k+1}}^N - Y_{t_k}) f''_{yz}(\theta_{t_k}^\lambda)\Delta Z_k^* \Big] \mathrm{d}\lambda. \end{aligned} \tag{25}$$

Note that $\mathbb{E}_{t_k}(Y_{t_{k+1}} - Y_{t_k}) = \mathbb{E}_{t_k} \int_{t_k}^{t_{k+1}} G_y(s, X_s)\mathrm{d}s$. If we closely look at (25), we can see that we need to develop ΔZ_k. By using (5), we can write

$$Z_{t_k}^N = \frac{1}{h}\mathbb{E}_{t_k}(\Delta Y_{k+1}\Delta W_k^*) + \frac{1}{h}\mathbb{E}_{t_k}(u(t_{k+1}, X_{t_{k+1}})\Delta W_k^*).$$

Introducing the weak derivative of $X_{t_{k+1}}$ (see Nualart [16, p. 109]), the second term of this summation equals $\frac{1}{h}\mathbb{E}_{t_k} \int_{t_k}^{t_{k+1}} \nabla_x u(t_{k+1}, X_{t_{k+1}}) D_t X_{t_{k+1}} \mathrm{d}t$, where $D_t X_{t_{k+1}} = \nabla_x X_{t_{k+1}} (\nabla_x X_t)^{-1} \sigma(t, X_t)$. Since $Z_{t_k} = \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k})$, one gets

$$\begin{aligned}\Delta Z_k &= \frac{1}{h}\mathbb{E}_{t_k}(\Delta Y_{k+1}\Delta W_k^*) + \frac{1}{h}\int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k}\Big(\nabla_x u(t_{k+1}, X_{t_{k+1}})\nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1}\sigma(t, X_t) \\ &\quad - \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k})\Big)\mathrm{d}t.\end{aligned}$$

The term in the second conditional expectation is equal to $\nabla_x u(t_{k+1}, X_{t_{k+1}})\nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1}$ $\sigma(t, X_t) \pm \nabla_x u(t, X_t)\sigma(t, X_t) - \nabla_x u(t_k, X_{t_k})\sigma(t_k, X_{t_k})$: hence, two applications of Itô's formula (for the first contribution between t and t_{k+1}, for the second one between t_k and t) prove that

$$\Delta Z_k^* = \int_{t_k}^{t_{k+1}} \mathbb{E}_{t_k}(G_z(s, X_s))\mathrm{d}s + \frac{1}{h}\mathbb{E}_{t_k}(\Delta Y_{k+1}\Delta W_k), \tag{26}$$

for a bounded function G_z. Plugging this equality and (25) into (24) yields (21).

Step 2: *Another formula of* ΔY_0. First of all, we replace $Y_{t_{k+1}}^N - Y_{t_k}$ by $\Delta Y_{k+1} + Y_{t_{k+1}} - Y_{t_k}$ in the expression of χ_k. Then, easy computations combining Proposition 11 and estimates (9) show that

$$\tilde{\chi}_k = \mathbb{E}_{t_k}(\chi_k) = O_k(h) + O(|\Delta X_k|^2 + |\Delta Z_k|^2). \tag{27}$$

From (21), we deduce the following equality

$$\Delta Y_0 = \mathbb{E}\left(\Delta Y_N \xi_0 \cdots \xi_{N-1} + h \sum_{i=0}^{N-1} (f'_x(\theta_{t_i})\Delta X_i + \tilde{\chi}_i)\xi_0 \cdots \xi_{i-1}\right). \tag{28}$$

Now it is enough to show that all terms of this summation are $O(h)$. In the following, $\eta_0 = 1$ and $\eta_i = \xi_0 \cdots \xi_{i-1}$ for $i \le N$.

Step 3: *Some results on* $\eta_N = \xi_0 \cdots \xi_{N-1}$.

We establish the following results on η_N:

$$\eta_k \text{ satisfies the condition } R_2 \text{ uniformly in } k, \text{ i.e. } \forall k, \eta_k \in \mathbb{D}^{2,\infty} \text{ and } \max_{k \le N} \mathcal{C}_{2,p}(\eta_k) < \infty, \quad \forall p \ge 1, \tag{29}$$

$$\mathbb{E}\left(\max_{0\le k\le N} |\eta_k|^p\right) + \sup_{r \le T} \mathbb{E}\left(\max_{0\le k\le N} |D_r\eta_k|^p\right) + \sup_{r,s\le T} \mathbb{E}\left(\max_{0\le k\le N} |D_r D_s \eta_k|^p\right) < \infty. \tag{30}$$

Proof of (29). We have $\eta_0 = 1$, and for $i \ge 1$

$$\eta_i = \eta_{i-1}(1 + hf'_y(\theta_{t_{i-1}}) + f'_z(\theta_{t_{i-1}})\Delta W_{i-1}). \tag{31}$$

We begin to show that $\max_{k\le N} \|\eta_k\|_{L_p} = O(1)$ for $p \ge 1$. Since f'_y and f'_z are bounded, we easily prove that $\mathbb{E}_{t_{i-1}}(1 + hf'_y(\theta_{t_{i-1}}) + f'_z(\theta_{t_{i-1}})\Delta W_{i-1})^{2p} \le (1 + Ch)$, whence $\mathbb{E}|\eta_i|^{2p} \le (1 + Ch)\mathbb{E}|\eta_{i-1}|^{2p}$. We deduce that $\max_{k\le N} \|\eta_k\|_{L_p} = O(1)$.

Now, let us show that $\max_{k\le N} \mathbb{E}|D_r\eta_k|^p = O(1)$, uniformly in r. Let r be such that $t_{k-1} < r \le t_k$. $\forall i \le k-1$, $D_r\eta_i = 0$. We note that $D_r\eta_k = \eta_{k-1} f'_z(\theta_{t_{k-1}})$. For $i \ge k+1$, we have

$$\begin{aligned} D_r\eta_i &= D_r\eta_{i-1} + hD_r(\eta_{i-1} f'_y(\theta_{t_{i-1}})) + \sum_{l=1}^{q} D_r(\eta_{i-1} f'_{z_l}(\theta_{t_{i-1}}))\Delta W^l_{i-1}, \\ &= \eta_{k-1} f'_z(\theta_{t_{k-1}}) + h\sum_{j=k}^{i-1} D_r(\eta_j f'_y(\theta_{t_j})) + \sum_{l=1}^{q}\sum_{j=k}^{i-1} D_r(\eta_j f'_{z_l}(\theta_{t_j}))\Delta W^l_j. \end{aligned} \tag{32}$$

Applying Burkholder–Davis–Gundy's inequality to the martingale $\sum_{j=k}^{i-1} D_r(\eta_j f'_{z_l}(\theta_{t_j}))\Delta W^l_j$ yields

$$\begin{aligned} \mathbb{E}|D_r\eta_i|^p &\le C\mathbb{E}|\eta_{k-1}|^p + C_p h \sum_{j=k}^{i-1} \mathbb{E}|D_r(\eta_j f'_y(\theta_{t_j}))|^p \\ &\quad + C\sum_{l=1}^{q} \mathbb{E}\left|h\sum_{j=k}^{i-1} |D_r(\eta_j f'_{z_l}(\theta_{t_j}))|^2\right|^{\frac{p}{2}} \\ &\le C\mathbb{E}|\eta_{k-1}|^p + Ch\sum_{j=k}^{i-1} \mathbb{E}|D_r(\eta_j f'_y(\theta_{t_j}))|^p + C\sum_{l=1}^{q} h\sum_{j=k}^{i-1} \mathbb{E}|D_r(\eta_j f'_{z_l}(\theta_{t_j}))|^p \end{aligned}$$

$$\leq C(1+\mathbb{E}|\eta_{k-1}|^p) + Ch\sum_{j=k+1}^{i-1}\mathbb{E}|D_r\eta_j|^p,$$

using the boundedness of the derivatives of f, $\max_{j\leq N}\|\eta_j\|_q = O(1)$, identity (7), $u,\sigma\in C_b^{1,2}$, and estimates (18). By applying Gronwall's lemma, we get $\max_{k\leq i\leq N}\mathbb{E}|D_r\eta_i|^p \leq C(1+\mathbb{E}|\eta_{k-1}|^p)$, $t_{k-1}<r\leq t_k$.

Then, $\max_{k\leq N}\mathbb{E}|D_r\eta_k|^p = O(1)$, uniformly in $r\in[0,T]$. The proof concerning the derivative of order 2 can be done following the same scheme. □

Proof of (30). We begin to show that $\mathbb{E}(\max_{k\leq N}|\eta_k|^p) < \infty$. The idea is to use a martingale property in order to apply Doob's inequality. Since $\eta_i = \eta_{i-1} + h\eta_{i-1}f'_y(\theta_{t_{i-1}}) + \eta_{i-1}f'_z(\theta_{t_{i-1}})\Delta W_{i-1}$, one has $\eta_k = 1+\sum_{i=1}^k(h\eta_{i-1}f'_y(\theta_{t_{i-1}}) + \eta_{i-1}f'_z(\theta_{t_{i-1}})\Delta W_{i-1})$. Thus,

$$\mathbb{E}\left(\max_{k\leq N}|\eta_k|^p\right) \leq C\left(1+\mathbb{E}\left(\sum_{i=1}^N h|\eta_{i-1}||f'_y(\theta_{t_{i-1}})|\right)^p + \mathbb{E}\left(\max_{k\leq N}\left|\sum_{i=1}^k \eta_{i-1}f'_z(\theta_{t_{i-1}})\Delta W_{i-1}\right|^p\right)\right).$$

The last term is upper bounded by $C\mathbb{E}(h\sum_{i=1}^N|\eta_{i-1}f'_z(\theta_{t_{i-1}})|^2)^{\frac{p}{2}} \leq Ch\sum_{i=1}^N\mathbb{E}|\eta_{i-1}f'_z(\theta_{t_{i-1}})|^p$. Using the estimate (29), we get $\mathbb{E}(\max_{k\leq N}|\eta_k|^p)<\infty$.

To prove that $\sup_{r\leq T}\mathbb{E}(\max_{k\leq N}|D_r\eta_k|^p) < \infty$, we proceed in the same way, by starting from (32). For the second derivative, this is analogous.

Step 4: *We prove that* $\mathbb{E}(\Delta Y_N\eta_N) = O(h)$.

If η_N were equal to 1, the results of Bally and Talay [1] would directly apply. Here the approach has to be different and we use techniques of Malliavin calculus. We have $\mathbb{E}(\Delta Y_N\eta_N) = \mathbb{E}(\eta_N\Phi(X_T^N) - \eta_N\Phi(X_T))$. Let us introduce $X_t^{N,\lambda} = (1-\lambda)X_t + \lambda X_t^N$. Thus, we have

$$\mathbb{E}(\Delta Y_N\eta_N) = \int_0^1 \mathbb{E}\left(\eta_N\Phi'_x(X_T^{N,\lambda})(X_T^N - X_T)\right)\mathrm{d}\lambda.$$

As $\Phi\in C^{3+\alpha}$, by using (29), (18) and (19), we note that $\eta_N\Phi'_x(X_T^{N,\lambda})$ satisfies R_2. By applying Remark 15, we deduce that $\mathbb{E}(\Delta Y_N\eta_N) = O(h)$.

Step 5: *We prove that* $\mathbb{E}(f'_x(\theta_{t_i})\Delta X_i\eta_i) = O(h)$. This is a very similar proof to Step 4, in a case where $\Phi(x) = x$.

Conclusion. We now work on $h\mathbb{E}(\sum_{i=0}^{N-1}\tilde{\chi}_i\eta_i)$, where $|\tilde{\chi}_k| \leq \lambda_k^N h + K(T,x)|\Delta X_k|^2 + K(T,x)|\Delta Z_k|^2$. Hence,

$$\begin{aligned}\left|h\sum_{i=0}^{N-1}\mathbb{E}(\tilde{\chi}_i\eta_i)\right| &\leq C\sum_{i=0}^{N-1}\mathbb{E}(\lambda_i^N|\eta_i|)h^2 + K(T,x)\sum_{i=0}^{N-1}h\mathbb{E}\left(|\eta_i|(|\Delta X_i|^2+|\Delta Z_i|^2)\right)\\ &\leq K(T,x)h + K(T,x)\sum_{i=0}^{N-1}h\mathbb{E}(|\eta_i||\Delta Z_i|^2)\\ &\leq K(T,x)h + K(T,x)\left(\mathbb{E}\Big(\max_{0\leq i\leq N-1}|\eta_i|\Big)^2\right)^{\frac{1}{2}}\left(\mathbb{E}\Big(h\sum_{i=0}^{N-1}|\Delta Z_i|^2\Big)^2\right)^{\frac{1}{2}}.\end{aligned}$$

By using (30) on $(\eta_i)_i$ and the upper bound (17) we get that $|h\mathbb{E}(\sum_{i=0}^{N-1}\tilde{\chi}_i\eta_i)| \le K(T,x)h$. By combining this result and the results of Step 4 and Step 5, (28) shows that $|\Delta Y_0| \le K(T,x)h$. Lemma 16 is proved. □

5. Proof of Theorem 8

As could be expected, its proof is more difficult. The main extra ingredient is the convergence of the weak derivative of the discrete BSDE (Y^N, Z^N), with the rate of convergence $N^{-1/2}$. The next paragraph is aimed at proving this result. In the following, Hypothesis 3 is in force.

5.1. *Proof of an intermediate result*

Proposition 17. *Let* $r \in]0, t_1[$. *Under Hypothesis 3, we have* $\max_{1\le i\le N}\mathbb{E}|D_r\Delta Y_i|^2 + h\mathbb{E}(\sum_{i=1}^{N-1}|D_r\Delta Z_i^*|^2) = O(h)$, *uniformly in* r.

This proposition is analogous to Theorem 6, where $q = 2$, and the scheme of its proof as well. However, there is a significant difference: the BSDE solved by the weak derivatives (see (33)–(35)) has a non-Lipschitz driver, which requires extra technicalities that we detail. In what follows, we fix $r \in]0, t_1[$ and introduce some specific notations. $\widehat{X_t}$ stands for D_rX_t. In the case of Z_t, which is a row vector, $\widehat{Z_t}$ is a matrix whose ith column is $D_r^iZ_t^*$. It is well-known (Proposition 5.3 of El Karoui et al. [5]) that $(\widehat{Y_t}, \widehat{Z_t})_{r\le t\le T}$ solves

$$\widehat{Y_t} = \Phi_x'(X_T)\widehat{X_T} + \int_t^T (f_x'(\theta_s)\widehat{X_s} + f_y'(\theta_s)\widehat{Y_s} + f_z'(\theta_s)\widehat{Z_s})\mathrm{d}s - \left(\int_t^T \widehat{Z_s}^*\mathrm{d}W_s\right)^*. \quad (33)$$

Regarding $(\widehat{Y^N}, \widehat{Z^N})$, one obtains

$$\widehat{Y_{t_k}^N} = \mathbb{E}_{t_k}[\widehat{Y_{t_{k+1}}^N} + h\nabla_x f_{t_k}^N\widehat{X_{t_k}^N} + h\nabla_y f_{t_k}^N\widehat{Y_{t_{k+1}}^N} + h\nabla_z f_{t_k}^N\widehat{Z_{t_k}^N}], \quad (34)$$

$$\widehat{Z_{t_k}^N} = \frac{1}{h}\mathbb{E}_{t_k}[\Delta W_k\widehat{Y_{t_{k+1}}^N}], \quad (35)$$

where we set $\nabla_x f_{t_k}^N = \nabla_x f(t_k, X_{t_k}^N, Y_{t_{k+1}}^N, Z_{t_k}^N)$ and analogously for $\nabla_y f_{t_k}^N$ and $\nabla_z f_{t_k}^N$. Indeed, we can start from (4) and (5) and interchange conditional expectations and weak derivatives (see Proposition 1.2.4 in Nualart [16]). Another way to get (34) and (35) is to take advantage of the Markov structure of $(X_{t_k}^N)_k$ to write $Y_{t_k}^N = y^N(t_k, X_{t_k}^N)$, where the function y^N is the solution of a dynamic programming equation, and then apply the chain rule. We omit further details.

From (7), we also have

$$\widehat{Y_t} = \nabla_x u(t, X_t)\widehat{X_t}, \qquad \widehat{Z_t} = \nabla_x(\nabla_x u\sigma)^*(t, X_t)\widehat{X_t}. \quad (36)$$

For the sake of clarity, let us write, for any process V, $\widehat{\Delta V_k} = D_rV_{t_k}^N - D_rV_{t_k}$. In particular, we have $\widehat{\Delta \overline{Z}_k} = D_r(Z_{t_k}^{N*} - \overline{Z}_{t_k}^*) = \widehat{Z_{t_k}^N} - \widehat{\overline{Z}_{t_k}}$, where $\widehat{\overline{Z}_{t_k}}$ is defined as $h\widehat{\overline{Z}_{t_k}} = \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}\widehat{Z_s}\mathrm{d}s$ (see the beginning of Section 3).

5.1.1. *Preparatory estimates*

In this part we give some L_p-estimates ($p \ge 1$), which are repeatedly used in the following calculations.

- $$\sup_{i\le j\le N}(\mathbb{E}_{t_i}|\widehat{X^N_{t_j}}|^{2p}) \le C|\widehat{X^N_i}|^{2p}, \tag{37}$$
- $$\mathbb{E}\left(\max_{0\le j\le N}|\widehat{X^N_{t_j}}|^{2p}\right) = O(1), \tag{38}$$
- $$\forall j\in 0\cdots N-1,\quad |\widehat{Y^N_{t_j}}|^2 \le C|\widehat{X^N_{t_j}}|^2,\qquad \mathbb{E}\left(\max_{0\le j\le N}|\widehat{Y^N_{t_j}}|^{2p}\right) = O(1), \tag{39}$$
- $$\mathbb{E}\left(\sup_{0\le t\le T}|\widehat{X_t}|^{2p} + \sup_{0\le t\le T}|\widehat{Y_t}|^{2p} + \sup_{0\le t\le T}|\widehat{Z_t}|^{2p}\right) = O(1). \tag{40}$$
- Let F satisfy R_3. Then, $|\mathbb{E}(F(\widehat{X^N_t} - \widehat{X_t}))| = O(h)$. Furthermore,
$$\sup_{0\le k\le N}\mathbb{E}|\widehat{\Delta X_k}|^{2p} = O(h^p). \tag{41}$$
- Analogously to (9), $\forall s\in[t_k,t_{k+1}]$, we have
$$\mathbb{E}_{t_k}\left(|\widehat{X_s}-\widehat{X_{t_k}}|^{2p} + |\widehat{Y_s}-\widehat{Y_{t_k}}|^{2p} + |\widehat{Z_s}-\widehat{\overline{Z}_{t_k}}|^{2p}\right) = O_k(h^p). \tag{42}$$

Note that $\widehat{X^N_{t_1}} = \sigma(0,x)$, and $\widehat{X^N_{t_{k+1}}} = (1 + hb'_x(t_k, X^N_{t_k}) + \sum_{i=1}^q(\sigma_i)'_x(t_k, X^N_{t_k})\Delta W^i_k)\widehat{X^N_{t_k}}$ for $1\le k\le N$. Thus, we easily get $\mathbb{E}_{t_i}|\widehat{X^N_{t_j}}|^{2p} \le (1+Ch)\mathbb{E}_{t_i}|\widehat{X^N_{t_{j-1}}}|^{2p}$, and (37) follows. The proof of (38) can be done as the proof of (30).

Proof of (39). From (34), we use Young's inequality and boundedness of ∇f to get

$$|\widehat{Y^N_{t_i}}|^2 \le (1+\gamma h)|\mathbb{E}_{t_i}\widehat{Y^N_{t_{i+1}}}|^2 + Ch\left(h+\frac{1}{\gamma}\right)\left(|\widehat{X^N_{t_i}}|^2 + \mathbb{E}_{t_i}|\widehat{Y^N_{t_{i+1}}}|^2 + |\widehat{Z^N_{t_i}}|^2\right). \tag{43}$$

From (35) and the Cauchy–Schwarz inequality, we obtain $h|\widehat{Z^N_{t_i}}|^2 \le C(\mathbb{E}_{t_i}|\widehat{Y^N_{t_{i+1}}}|^2 - |\mathbb{E}_{t_i}\widehat{Y^N_{t_{i+1}}}|^2)$. Hence, with an appropriate choice of γ, (43) is reduced to $|\widehat{Y^N_{t_i}}|^2 \le (1+Ch)\mathbb{E}_{t_i}|\widehat{Y^N_{t_{i+1}}}|^2 + Ch|\widehat{X^N_{t_i}}|^2$, and thus Gronwall's lemma yields

$$|\widehat{Y^N_{t_i}}|^2 \le C\mathbb{E}_{t_i}\left(|\widehat{Y^N_{t_N}}|^2 + h\sum_{j=i}^{N-1}|\widehat{X^N_{t_j}}|^2\right) \le C\sup_{i\le j\le N-1}\mathbb{E}_{t_i}|\widehat{X^N_{t_j}}|^2.$$

Finally, estimates (37) and (38) complete the proof.

Proof of (40). $\mathbb{E}(\sup_{0\le t\le T}|\widehat{X_t}|^{2p}) = O(1)$ follows from (18). The other estimates come from this result and (36).

Proof of (41). Let us introduce $X'_t = \nabla_x X_t(\nabla_x X_r)^{-1}\sigma(0,x)$ and write $\widehat{X^N_t} - \widehat{X_t} = \widehat{X^N_t} - X'_t + X'_t - \widehat{X_t}$.

Since $\widehat{X_t} = \nabla_x X_t (\nabla_x X_r)^{-1} \sigma(r, X_r)$, a direct application of Proposition 12 with $U_t = \sigma(t, X_t)$ gives $\mathbb{E}(F(X'_t - \widehat{X_t})) = O(h)$ for F satisfying R_2. Moreover, simple increment estimates yield $\sup_{t \le T} \mathbb{E}|X'_t - \widehat{X_t}|^{2p} = O(h^p)$.

It remains to study the impact of the difference $\widehat{X^N_t} - X'_t$. $(\widehat{X^N_t})_{t \ge r}$ and $(X'_t)_{t \ge r}$ are solutions of

$$\widehat{X^N_t} = \sigma(0, x) + \int_r^t b'_x(\eta(s), X^N_{\eta(s)}) \widehat{X^N_{\eta(s)}} \mathrm{d}s + \sum_{i=1}^q \int_r^t (\sigma_i)'_x(\eta(s), X^N_{\eta(s)}) \widehat{X^N_{\eta(s)}} \mathrm{d}W^i_s,$$

$$X'_t = \sigma(0, x) + \int_r^t b'_x(s, X_s) X'_s \mathrm{d}s + \sum_{i=1}^q \int_r^t (\sigma_i)'_x(s, X_s) X'_s \mathrm{d}W^i_s. \tag{44}$$

For the sake of simplicity, we take $b \equiv 0$ and $d = q = 1$. If we set $\sigma'(s) = \int_0^1 \sigma'_x(s, X_s + \lambda(X^N_s - X_s))\mathrm{d}\lambda$, we observe that ΔX_t solves the linear equation $\Delta X_t = \int_0^t [\sigma(\eta(s), X^N_{\eta(s)}) - \sigma(s, X^N_s)]\mathrm{d}W_s + \int_0^t \sigma'(s) \Delta X_s \mathrm{d}W_s$, whose solution is given by (see Theorem 56, p. 271 in Protter [19])

$$\begin{aligned} \Delta X_t &= \epsilon_t \int_0^t \epsilon_s^{-1} [\sigma(\eta(s), X^N_{\eta(s)}) - \sigma(s, X^N_s)](\mathrm{d}W_s - \sigma'(s)\mathrm{d}s) \\ &= -\epsilon_t \int_0^t \epsilon_s^{-1} \left[\int_{\eta(s)}^s \sigma'_x(v, X^N_v) \sigma(\eta(v), X^N_{\eta(v)}) \mathrm{d}W_v \right. \\ &\quad \left. + (\sigma'_t(v, X^N_v) + \frac{1}{2} \sigma''_{xx}(v, X^N_v) \sigma^2(\eta(v), X^N_{\eta(v)}))\mathrm{d}v \right] (\mathrm{d}W_s - \sigma'(s)\mathrm{d}s) \end{aligned}$$

where $\epsilon_t = 1 + \int_0^t \sigma'(s) \epsilon_s \mathrm{d}W_s$. This proves that ΔX_t satisfies the property $(\mathcal{P})$. Analogously, if we define $\sigma''(s) = \int_0^1 \sigma''_{xx}(s, X_s + \lambda(X^N_s - X_s))\mathrm{d}\lambda$ and $\epsilon^N_t = 1 + \int_r^t \sigma'_x(s, X^N_s) \epsilon^N_s \mathrm{d}W_s$, simple computations lead to

$$\begin{aligned} \widehat{X^N_t} - X'_t = \epsilon^N_t \int_r^t &(\epsilon^N_s)^{-1} ([\sigma'_x(\eta(s), X^N_{\eta(s)}) \widehat{X^N_{\eta(s)}} - \sigma'_x(s, X^N_s) \widehat{X^N_s}] + \sigma''(s) X'_s \Delta X_s) \\ &\times (\mathrm{d}W_s - \sigma'_x(s, X^N_s)\mathrm{d}s). \end{aligned}$$

From the above representation, it is straightforward to conclude $\sup_{t \le T} \mathbb{E}|\widehat{\Delta X_t}|^{2p} = O(h^p)$. Now, let us upper bound $\mathbb{E}(F(\widehat{X^N_t} - X'_t))$ which can be decomposed into several terms.

- The contribution associated to $\epsilon^N_t \int_r^t (\epsilon^N_s)^{-1} [\sigma'_x(\eta(s), X^N_{\eta(s)}) \widehat{X^N_{\eta(s)}} - \sigma'_x(s, X^N_s) \widehat{X^N_s}](\mathrm{d}W_s - \sigma'_x(s, X^N_s)\mathrm{d}s)$ satisfies property $(\mathcal{P})$, thus Proposition 14 yields the expected result.
- The contribution $\mathbb{E}(F \epsilon^N_t \int_r^t (\epsilon^N_s)^{-1} \sigma''(s) X'_s \Delta X_s \sigma'_x(s, X^N_s)\mathrm{d}s)$ is equal to $\int_r^t \mathbb{E}(F \epsilon^N_t (\epsilon^N_s)^{-1} \sigma''(s) X'_s \Delta X_s \sigma'_x(s, X^N_s))\mathrm{d}s = O(h)$ in view of Remark 15.
- In the same way, the duality relationship ensures that the last contribution $\mathbb{E}(F \epsilon^N_t \int_r^t (\epsilon^N_s)^{-1} \sigma''(s) X'_s \Delta X_s \mathrm{d}W_s) = \int_r^t \mathbb{E}(D_s(F \epsilon^N_t)(\epsilon^N_s)^{-1} \sigma''(s) X'_s \Delta X_s)\mathrm{d}s$ is a $O(h)$ (using here that F satisfies R_3).

Proof of (42). In view of $\widehat{X_t} = D_r X_t = \nabla_x X_t (\nabla_x X_r)^{-1} \sigma(r, X_r)$, the estimate on the increments of $\widehat{X_t}$ becomes clear. The other ones easily follow. □

5.1.2. Proof of $\max_{1\le i\le N}\mathbb{E}|\widehat{\Delta Y_i}|^2 = O(h)$

Assume that for some non-negative random variable $\Lambda_k = O_k(h) + |\Delta X_k|^2 + |\Delta Z_k|^2$, one has

$$|\widehat{\Delta Y_k}|^2 \le (1+Ch)\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2 + h|\widehat{\Delta X_k}|^2 + h\Lambda_k O_k(1). \tag{45}$$

Take the expectation on both sides, use estimates (41) and those of Proposition 5 to get

$$\mathbb{E}|\widehat{\Delta Y_k}|^2 \le C\mathbb{E}|\widehat{\Delta Y_N}|^2 + O(h) + Ch\sum_{k=0}^{N-1}\mathbb{E}(|\Delta Z_k|^2 O_k(1)).$$

On the one hand, as $\widehat{\Delta Y_N} = \Phi'(X^N_{t_N})\widehat{X^N_{t_N}} - \Phi'(X_{t_N})\widehat{X_{t_N}}$, clearly $\mathbb{E}|\widehat{\Delta Y_N}|^2 = O(h)$. On the other hand, in view of (17) with $p = 2$, the summation above is a $O(h)$. This proves $\max_{1\le k\le N}\mathbb{E}|\widehat{\Delta Y_k}|^2 = O(h)$.

Proof of (45). From (33) and (34), we obtain

$$\widehat{\Delta Y_k} = \mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}}) + \mathbb{E}_{t_k}\left(\int_{t_k}^{t_{k+1}}[\nabla_x f^N_{t_k}\widehat{X^N_{t_k}} - f'_x(\theta_s)\widehat{X_s} + \nabla_y f^N_{t_k}\widehat{Y^N_{t_{k+1}}} - f'_y(\theta_s)\widehat{Y_s} + \nabla_z f^N_{t_k}\widehat{Z^N_{t_k}} - f'_z(\theta_s)\widehat{Z_s}]\mathrm{d}s\right).$$

Since $f\in C_b^{2,4,4,4}$, it follows that for any $\gamma > 0$ (to be fixed later)

$$\begin{aligned}|\widehat{\Delta Y_k}|^2 &\le (1+\gamma h)|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 + C\left(h+\frac{1}{\gamma}\right)\mathbb{E}_{t_k}\left(\int_{t_k}^{t_{k+1}}[|\nabla_x f^N_{t_k}\widehat{X^N_{t_k}} - f'_x(\theta_s)\widehat{X_s}|^2\right.\\ &\qquad\left. + |\nabla_y f^N_{t_k}\widehat{Y^N_{t_{k+1}}} - f'_y(\theta_s)\widehat{Y_s}|^2 + |\nabla_z f^N_{t_k}\widehat{Z^N_{t_k}} - f'_z(\theta_s)\widehat{Z_s}|^2]\mathrm{d}s\right) &(46)\\ &\le (1+\gamma h)|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 + C\left(h+\frac{1}{\gamma}\right)(T^1_k + T^2_k), &(47)\end{aligned}$$

where we put $T^1_k = \mathbb{E}_{t_k}(\int_{t_k}^{t_{k+1}}[|\widehat{X^N_{t_k}} - \widehat{X_s}|^2 + |\widehat{Y^N_{t_{k+1}}} - \widehat{Y_s}|^2 + |\widehat{Z^N_{t_k}} - \widehat{Z_s}|^2]\mathrm{d}s)$, $T^2_k = \mathbb{E}_{t_k}\int_{t_k}^{t_{k+1}}(h + |X_s - X^N_{t_k}|^2 + |Y_s - Y^N_{t_{k+1}}|^2 + |Z_s - Z^N_{t_k}|^2)(|\widehat{X_s}|^2 + |\widehat{Y_s}|^2 + |\widehat{Z_s}|^2)\mathrm{d}s$. To get (45), we need to simplify (47), by estimating T^1_k and T^2_k.

Term T^1_k. Firstly, we write $\mathbb{E}_{t_k}|\widehat{Y^N_{t_{k+1}}} - \widehat{Y_s}|^2 \le 2\mathbb{E}_{t_k}|\widehat{Y_{t_{k+1}}} - \widehat{Y_s}|^2 + 2\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2$. We do the same for $\widehat{X^N_{t_k}} - \widehat{X_s}$. Then, the usual increment estimates yield

$$\mathbb{E}_{t_k}|\widehat{Y^N_{t_{k+1}}} - \widehat{Y_s}|^2 + \mathbb{E}_{t_k}|\widehat{X^N_{t_k}} - \widehat{X_s}|^2 \le O_k(h) + 2|\widehat{\Delta X_k}|^2 + 2\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2.$$

Secondly, analogously to (13), we have

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

Finally, we obtain $T^1_k \le Ch(O_k(h) + |\widehat{\Delta X_k}|^2 + \mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2 + |\widehat{\Delta \overline{Z}_k}|^2)$.

Term T^2_k. Easy calculations combining (9), Proposition 11 and (40) give $T^2_k \le (O_k(h^2) + h|\Delta X_k|^2 + h|\Delta Z_k|^2)O_k(1) = h\Lambda_k O_k(1)$.

Conclusion. Plugging the estimates on T_k^1 and T_k^2 into (47), we get

$$|\widehat{\Delta Y_k}|^2 \leq (1+\gamma h)|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 + Ch\left(h+\frac{1}{\gamma}\right)|\widehat{\Delta \overline{Z}_k}|^2 + Ch\left(h+\frac{1}{\gamma}\right)(|\widehat{\Delta X_k}|^2 + \mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2 + \Lambda_k O_k(1)). \tag{48}$$

Note that $h\widehat{\overline{Z}_{t_k}} = \mathbb{E}_{t_k}(\Delta W_k(\widehat{Y_{t_{k+1}}} + \int_{t_k}^{t_{k+1}}[f'_x(\theta_s)\widehat{X_s} + f'_y(\theta_s)\widehat{Y_s} + f'_z(\theta_s)\widehat{Z_s}]\mathrm{d}s))$, whence $h\widehat{\Delta\overline{Z}_{t_k}} = \mathbb{E}_{t_k}(\Delta W_k(\widehat{\Delta Y_{k+1}} + \int_{t_k}^{t_{k+1}}[f'_x(\theta_s)\widehat{X_s} + f'_y(\theta_s)\widehat{Y_s} + f'_z(\theta_s)\widehat{Z_s}]\mathrm{d}s))$. By proceeding as before, we easily prove

$$h|\widehat{\Delta\overline{Z}_{t_k}}|^2 \leq C(\mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2 - |\mathbb{E}_{t_k}\widehat{\Delta Y_{k+1}}|^2) + O_k(h^2). \tag{49}$$

Combining this upper bound with (48) for a good choice of γ gives (45). □

5.1.3. Proof of $h\mathbb{E}(\sum_{k=1}^{N-1}|\widehat{\Delta Z_k}|^2) = O(h)$

In view of (42), this is equivalent to proving $h\mathbb{E}(\sum_{k=1}^{N-1}|\widehat{\Delta\overline{Z}_k}|^2) = O(h)$. To establish this estimate, we start from (49) to get

$$h\sum_{k=1}^{N-1}\mathbb{E}|\widehat{\Delta\overline{Z}_k}|^2 \leq C\sum_{k=1}^{N-1}(\mathbb{E}|\widehat{\Delta Y_k}|^2 - \mathbb{E}|\mathbb{E}_{t_k}\widehat{\Delta Y_{k+1}}|^2) + C\mathbb{E}|\widehat{\Delta Y_N}|^2 + O(h). \tag{50}$$

Now, we work on $|\widehat{\Delta Y_k}|^2 - |\mathbb{E}_{t_k}\widehat{\Delta Y_{k+1}}|^2$. The choice $\gamma = 2C^2$ in (48) leads to

$$|\widehat{\Delta Y_k}|^2 - |\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 \leq \gamma h|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 + h\left(\frac{1}{2C}+Ch\right)|\widehat{\Delta\overline{Z}_k}|^2 + h\left(Ch+\frac{1}{2C}\right)(|\widehat{\Delta X_k}|^2 + \mathbb{E}_{t_k}|\widehat{\Delta Y_{k+1}}|^2 + \Lambda_k O_k(1)).$$

From (41) and the result from Section 5.1.2, we have $\max_{1\leq k\leq N}\mathbb{E}(|\widehat{\Delta X_k}|^2 + |\widehat{\Delta Y_k}|^2) = O(h)$. We also have $\mathbb{E}(\Lambda_k O_k(1)) = O(h) + \mathbb{E}(|\Delta Z_k|^2 O_k(1))$. Consequently, for h small enough, one has $\mathbb{E}|\widehat{\Delta Y_k}|^2 - \mathbb{E}|\mathbb{E}_{t_k}(\widehat{\Delta Y_{k+1}})|^2 \leq \frac{2h}{3C}\mathbb{E}|\widehat{\Delta\overline{Z}_k}|^2 + O(h^2) + Ch\mathbb{E}(|\Delta Z_k|^2 O_k(1))$. Putting this estimate into (50) yields

$$\frac{1}{3}h\sum_{k=1}^{N-1}\mathbb{E}|\widehat{\Delta\overline{Z}_k}|^2 \leq O(h) + Ch\sum_{k=1}^{N-1}\mathbb{E}(|\Delta Z_k|^2 O_k(1)).$$

Inequality (17) with $p = 2$ directly shows that the sum above is a $O(h)$. □

5.2. Expansion of $Z_{t_k}^N - Z_{t_k}$

We recall that $u \in C_b^{2,4}$ owing to Hypothesis 3. From (26), we have $\Delta Z_k = O(h) + \frac{1}{h}\mathbb{E}_{t_k}[(u^N(t_{k+1}, X_{t_{k+1}}^N) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*]$. Let $(X_t^{s,\overline{x}})_{t\geq s}$ denote the solution of the SDE (1) starting at time s from $\overline{x}$. We write X_t for $X_t^{0,x}$. Note that $X_{t_{k+1}} = X_{t_{k+1}}^{t_k, X_{t_k}}$. In the same way, the Euler scheme starting at time t_k at $\overline{x}$ is denoted by $(X_{t_j}^{N,t_k,\overline{x}})_{j\geq k}$. With this notation we can rewrite ΔZ_k

$$\Delta Z_k = \frac{1}{h}\mathbb{E}_{t_k}\left[(u^N(t_{k+1}, X_{t_{k+1}}^{N,t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*\right] + O(h),$$
$$= \frac{1}{h}\mathbb{E}_{t_k}\left[(u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*\right]$$
$$+ \frac{1}{h}\mathbb{E}_{t_k}\left[(u^N(t_{k+1}, X_{t_{k+1}}^{N,t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}))\Delta W_k^*\right] + O(h). \quad (51)$$

We work on the first two terms separately by proving

Lemma 18. $\frac{1}{h}\mathbb{E}_{t_k}[(u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*] = O(|\Delta X_k|^2) + O(h) + [\nabla_x(\nabla_x u\, \sigma)^*(t_k, X_{t_k})\Delta X_k]^*$.

Lemma 19. $\frac{1}{h}|\mathbb{E}_{t_k}[(u^N(t_{k+1}, X_{t_{k+1}}^{N,t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}))\Delta W_k^*]| = O_k(h)$.

The combination of these lemmas completes the proof of Theorem 8.

5.2.1. *Proof of Lemma 18*

For the sake of simplicity, let $\Delta_N X_{k+1}$ denote $X_{t_{k+1}}^{t_k,X_{t_k}^N} - X_{t_{k+1}}$ (which is different from $\Delta X_{k+1} = X_{t_{k+1}}^{N,t_k,X_{t_k}^N} - X_{t_{k+1}}$). From a Taylor–Lagrange formula, we obtain

$$u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}) = u'_x(t_{k+1}, X_{t_{k+1}})\Delta_N X_{k+1}$$
$$+ \int_0^1 (1-\lambda)(\Delta_N X_{k+1})^* H_x(u)\left(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1}\right)\Delta_N X_{k+1}\mathrm{d}\lambda.$$

Thus, using the duality relationship, one has

$$\mathbb{E}_{t_k}[(u(t_{k+1}, X_{t_{k+1}}^{t_k,X_{t_k}^N}) - u(t_{k+1}, X_{t_{k+1}}))\Delta W_k^*]$$
$$= \int_{t_k}^{t_{k+1}} R_k^1(t)\mathrm{d}t + \int_{t_k}^{t_{k+1}} R_k^2(t)\mathrm{d}t + \int_0^1 (1-\lambda)R_k^3(\lambda)\mathrm{d}\lambda,$$

with

$$R_k^1(t) = \mathbb{E}_{t_k}[(\Delta_N X_{k+1})^* H_x(u)(t_{k+1}, X_{t_{k+1}})D_t X_{t_{k+1}}],$$
$$R_k^2(t) = \mathbb{E}_{t_k}[u'_x(t_{k+1}, X_{t_{k+1}})D_t(\Delta_N X_{k+1})],$$
$$R_k^3(\lambda) = \mathbb{E}_{t_k}[(\Delta_N X_{k+1})^* H_x(u)(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1})\Delta_N X_{k+1}\Delta W_k^*].$$

Expansion of $R_k^1(t)$. Clearly $\Delta_N X_{k+1} = \Delta X_k + U_{t_{k+1}} - U_{t_k}$, where U is an Itô process with drift term $\alpha_s = b(s, X_s^{t_k,X_{t_k}^N}) - b(s, X_s)$ and diffusion term $\beta_s = \sigma(s, X_s^{t_k,X_{t_k}^N}) - \sigma(s, X_s)$, both being bounded. Thus, we can apply Proposition 12, letting $F = H_x(u)(t_{k+1}, X_{t_{k+1}})D_t X_{t_{k+1}}$. Because $u \in C_b^{2,4}$ and in view of (18), we get

$$R_k^1(t) = O(h) + (\Delta X_k)^*\mathbb{E}_{t_k}[H_x(u)(t_{k+1}, X_{t_{k+1}})D_t X_{t_{k+1}}].$$

We expand the latter factor. As $D_t X_{t_{k+1}} = \nabla_x X_{t_{k+1}} (\nabla_x X_t)^{-1} \sigma(t, X_t)$, we have

$$\begin{aligned} H_x(u)(t_{k+1}, X_{t_{k+1}}) D_t X_{t_{k+1}} = {} & (H_x(u)(t_{k+1}, X_{t_{k+1}})\sigma(t, X_t) - H_x(u)(t, X_t)\sigma(t, X_t)) \\ & + (H_x(u)(t, X_t)\sigma(t, X_t) - H_x(u)(t_k, X_{t_k})\sigma(t_k, X_{t_k})) \\ & + (H_x(u)(t_{k+1}, X_{t_{k+1}})[\nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1} - I]\sigma(t, X_t)) \\ & + H_x(u)(t_k, X_{t_k})\sigma(t_k, X_{t_k}). \end{aligned}$$

The first three contributions in the r.h.s. above can be handled in the same way and we give a detailed proof only for the first one. It is enough to apply Proposition 12 with $F = \sigma(t, X_t)$ and $U_s = H_x(u)(s, X_s)$. Then, $\mathbb{E}_{t_k}[F(U_{t_{k+1}} - U_t)]$ is of order h with a constant involving b, σ, u and its derivatives up to order 4. Finally, this gives

$$R_k^1(t) = O(h) + (\Delta X_k)^* H_x(u)(t_k, X_{t_k})\sigma(t_k, X_{t_k}),$$

uniformly in $t \in [t_k, t_{k+1}]$.

Expansion of $R_k^2(t)$. For $t_k \le t \le t_{k+1}$, we have

$$\begin{aligned} D_t(\Delta_N X_{k+1}) = {} & [\nabla_x X_{t_{k+1}}^{X_{t_k}^N, t_k} (\nabla_x X_t^{X_{t_k}^N, t_k})^{-1} - I]\sigma(t, X_t^{X_{t_k}^N, t_k}) \\ & - [\nabla_x X_{t_{k+1}}(\nabla_x X_t)^{-1} - I]\sigma(t, X_t) - (\sigma(t, X_t) - \sigma(t_k, X_{t_k})) \\ & + \sigma(t, X_t^{X_{t_k}^N, t_k}) - \sigma(t_k, X_{t_k}^N) + \sigma(t_k, X_{t_k}^N) - \sigma(t_k, X_{t_k}). \end{aligned}$$

As before, apply Proposition 12 to each of these terms but the last one, with $F = u'_x(t_{k+1}, X_{t_{k+1}})$, using $u, b, \sigma \in C_b^{2,4}$ and (18). It follows that $R_k^2(t) = O(h) + \mathbb{E}_{t_k}[u'_x(t_{k+1}, X_{t_{k+1}})](\sigma(t_k, X_{t_k}^N) - \sigma(t_k, X_{t_k}))$. An application of Itô's formula yields

$$\begin{aligned} R_k^2(t) &= O(h) + \sum_{i=1}^d u'_{x_i}(t_k, X_{t_k})(\sigma^i(t_k, X_{t_k}^N) - \sigma^i(t_k, X_{t_k})) \\ &= O(h + |\Delta X_k|^2) + \sum_{i=1}^d u'_{x_i} \nabla_x([\sigma^i]^*)(t_k, X_{t_k})\Delta X_k, \end{aligned}$$

uniformly in $t \in [t_k, t_{k+1}]$. Finally, simple matrix computations lead to

$$R_k^1(t) + R_k^2(t) = O(h + |\Delta X_k|^2) + [\nabla_x(\nabla_x u\sigma)^*(t_k, X_{t_k})\Delta X_k]^*.$$

Upper bound for $R_k^3(\lambda)$. To complete the proof of Lemma 18, note that it remains to justify that $R_k^3(\lambda) = hO(h + |\Delta X_k|^2)$ uniformly in λ. The duality formula gives

$$R_k^3(\lambda) = \mathbb{E}_{t_k}\left[\int_{t_k}^{t_{k+1}} D_t[(\Delta_N X_{k+1})^* H_x(u)(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1})\Delta_N X_{k+1}]\mathrm{d}t\right].$$

The term in the integral equals $\sum_{i,j=1}^d [2D_t(\Delta_N X_{k+1,i})\Delta_N X_{k+1,j}\partial^2_{x_i,x_j} u(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1}) + \Delta_N X_{k+1,i}\Delta_N X_{k+1,j} D_t(\partial^2_{x_i,x_j} u(t_{k+1}, X_{t_{k+1}} + \lambda\Delta_N X_{k+1}))]$. Thanks to (18) and (19) and successive applications of Proposition 12, we finally prove our assertion. We omit further details. □

5.2.2. *Proof of Lemma 19*

As for Lemma 16, we only do the proof for $t_k = 0$, i.e. we have to show $|\mathbb{E}_{t_k}[(u^N(t_1, X_{t_1}^{N,0,x}) - u(t_1, X_{t_1}^{0,x}))\Delta W_0^*]| \le K(T, x)h^2$. We have $\mathbb{E}[(u^N(t_1, X_{t_1}^{N,0,x}) - u(t_1, X_{t_1}^{0,x}))\Delta W_0^*] =$

$\mathbb{E}[\Delta Y_1 \Delta W_0^*]$. By using (21), we come up with

$$\mathbb{E}[\Delta Y_1 \Delta W_0^*] = \mathbb{E}[\xi_1 \cdots \xi_{N-1} \Delta Y_N \Delta W_0^*] + \mathbb{E}\left[h \sum_{i=1}^{N-1} (f_x'(\theta_{t_i}) \Delta X_i + \tilde{\chi}_i) \xi_1 \cdots \xi_{i-1} \Delta W_0^*\right],$$

where $\tilde{\chi}_i = \mathbb{E}_{t_i}(\chi_i)$ (ξ_i and χ_i are defined in (22) and (23)). In the following $\tilde{\eta}_i$ denotes $\xi_1 \cdots \xi_{i-1}$ and $\tilde{\eta}_1 = 1$. We easily prove that $(\tilde{\eta}_i)_{1 \le i \le N}$ has analogous properties to $(\eta_i)_{0 \le i \le N}$. Estimates (29) and (30) remain valid for $\tilde{\eta}$ and under Hypothesis 3, the estimate (29) becomes

$$\tilde{\eta}_k \text{ satisfies } R_3 \text{ uniformly in } k. \tag{52}$$

Step 1: *Proof of* $\mathbb{E}[\xi_1 \cdots \xi_{N-1} \Delta Y_N \Delta W_0^*] = \mathbb{E}[\tilde{\eta}_N \Delta Y_N \Delta W_0^*] = O(h^2)$. As before, we use the duality formula:

$$\mathbb{E}[\tilde{\eta}_N \Delta Y_N \Delta W_0^*] = \mathbb{E} \int_0^{t_1} (D_t[\tilde{\eta}_N] \Delta Y_N + \tilde{\eta}_N D_t[\Delta Y_N]) \mathrm{d}t.$$

Since $\tilde{\eta}_N$ satisfies (52), we proceed as in *Step* 4 of Lemma 16 and we get $\mathbb{E}(D_t[\tilde{\eta}_N] \Delta Y_N) = O(h)$. Furthermore, we have

$$D_t[\Delta Y_N] = (\Phi'(X_T^N) - \Phi'(X_T)) D_t X_T^N + \Phi'(X_T)(D_t X_T^N - D_t X_T).$$

On the one hand, analogously to previous computations, we establish $\mathbb{E}(\tilde{\eta}_N (\Phi'(X_T^N) - \Phi'(X_T)) D_t X_T^N) = O(h)$.

On the other hand, we prove $\mathbb{E}(\tilde{\eta}_N \Phi'(X_T)(D_t X_T^N - D_t X_T)) = O(h)$. Thanks to (18) and (29), $\tilde{\eta}_N \Phi'(X_T)$ satisfies condition R_3. Then, by applying (41), we get the result.

Step 2: *Proof of* $\mathbb{E}[h \sum_{i=1}^{N-1} f_x'(\theta_{t_i}) \Delta X_i \xi_1 \cdots \xi_{i-1} \Delta W_0^*] = O(h^2)$. This is a similar proof to the one done at *Step* 1, with $\Phi(x) = x$.

Step 3: *Proof of* $\mathbb{E}[h \sum_{i=1}^{N-1} \tilde{\chi}_i \tilde{\eta}_i \Delta W_0^*] = O(h^2)$. A careful inspection of the definition of G_0, G_y and G_z appearing in (23) shows that under Hypothesis 3, these functions are continuously differentiable w.r.t. the variable x (with a bounded derivative). Hence, if we write $\chi_i = \chi_i^1 + \int_0^1 (1 - \lambda) \chi_i^2(\lambda) \mathrm{d}\lambda$ with (see (23))

$$\chi_i^1 = \int_{t_i}^{t_{i+1}} (G_0(s, X_s) + f_y'(\theta_{t_i}) G_y(s, X_s) + f_z'(\theta_{t_i}) G_z(s, X_s)) \mathrm{d}s,$$

$$\begin{aligned}\chi_i^2(\lambda) = {} & \Delta X_i^* f_{xx}''(\theta_{t_i}^\lambda) \Delta X_i + f_{yy}''(\theta_{t_i}^\lambda)(Y_{t_{i+1}}^N - Y_{t_i})^2 + \Delta Z_i f_{zz}''(\theta_{t_i}^\lambda) \Delta Z_i^* \\ & + 2 \Delta X_i^* f_{xy}''(\theta_{t_i}^\lambda)(Y_{t_{i+1}}^N - Y_{t_i}) + 2 \Delta X_i^* f_{xz}''(\theta_{t_i}^\lambda) \Delta Z_i^* \\ & + 2 (Y_{t_{i+1}}^N - Y_{t_i}) f_{yz}''(\theta_{t_i}^\lambda) \Delta Z_i^*,\end{aligned}$$

we note that the random variable χ_i is in $\mathbb{D}^{1,\infty}$. Thus and because $\tilde{\chi}_i = \mathbb{E}_{t_i}(\chi_i)$, one has $\mathbb{E}[\tilde{\chi}_i \tilde{\eta}_i \Delta W_0^*] = \mathbb{E}[\chi_i \tilde{\eta}_i \Delta W_0^*] = \mathbb{E}[\int_0^{t_1} (\chi_i D_t \tilde{\eta}_i + \tilde{\eta}_i D_t \chi_i) \mathrm{d}t]$.

The upper bound $\tilde{\chi}_i = \mathbb{E}_{t_i}(\chi_i) = O_i(h) + O(|\Delta X_i|^2 + |\Delta Z_i|^2)$ (see (27)) is sufficient to show $\mathbb{E}[\sum_{i=1}^{N-1} \chi_i D_t \tilde{\eta}_i] = O(1)$ uniformly in t (follow the arguments of the conclusion of the proof of Lemma 16 and use (30) with $\tilde{\eta}$).

Now, it remains to establish $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i D_t \chi_i] = O(1)$. On the one hand, clearly $\mathbb{E}_{t_i}[D_t \chi_i^1] = O_i(h)$ and we conclude $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i D_t \chi_i^1] = O(1)$ uniformly in t. On the other hand, χ_i^2 can be decomposed into several contributions, which can be analyzed with the same arguments. Let us detail how to handle one of them, for instance $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i D_t(\Delta X_i^* f''_{xz}(\theta_{t_i}^\lambda)\Delta Z_i^*)]$ which has to be a $O(1)$. We do the proof for $d = q = 1$. Write $D_t(\Delta X_i f''_{xz}(\theta_{t_i}^\lambda)\Delta Z_i) = \Delta X_i f''_{xz}(\theta_{t_i}^\lambda) D_t(\Delta Z_i) + D_t(\Delta X_i) f''_{xz}(\theta_{t_i}^\lambda)\Delta Z_i + \Delta X_i D_t(f''_{xz}(\theta_{t_i}^\lambda))\Delta Z_i$. As f'' is bounded, we have

$$\begin{aligned}\left|\mathbb{E}\left[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i f''_{xz}(\theta_{t_i}^\lambda) D_t(\Delta Z_i)\right]\right| &\leq \mathbb{E}\left[\sum_{i=1}^{N-1} |\tilde{\eta}_i||\Delta X_i||f''_{xz}(\theta_{t_i}^\lambda)||D_t(\Delta Z_i)|\right] \\ &\leq C\left(\mathbb{E}\left(\sum_{i=1}^{N-1}(|\tilde{\eta}_i|^2|\Delta X_i|^2)\right)\right)^{\frac{1}{2}} \\ &\quad\times\left(\mathbb{E}\left(\sum_{i=1}^{N-1}|D_t(\Delta Z_i)|^2\right)\right)^{\frac{1}{2}}.\end{aligned}$$

Thanks to Proposition 17, (52) and Proposition 5, we get that $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i f''_{xz}(\theta_{t_i}^\lambda)(D_t \Delta Z_i)] = O(1)$. Analogously, using (52), (17) and (41), we obtain $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i (D_t \Delta X_i) f''_{xz}(\theta_{t_i}^\lambda)\Delta Z_i] = O(1)$. It remains to demonstrate that $|\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i D_t(f''_{xz}(\theta_{t_i}^\lambda))\Delta Z_i]| = O(1)$. We have

$$\begin{aligned}D_t(f''_{xz}(\theta_{t_i}^\lambda)) = {}& f'''_{xzx}(\theta_{t_i}^\lambda)(\lambda D_t X_{t_i}^N + (1-\lambda)D_t X_{t_i}) + f'''_{xzy}(\theta_{t_i}^\lambda)(\lambda D_t Y_{t_{i+1}}^N + (1-\lambda)D_t Y_{t_i}) \\ &+ f'''_{xzz}(\theta_{t_i}^\lambda)(\lambda D_t Z_{t_i}^N + (1-\lambda)D_t Z_{t_i}).\end{aligned}$$

The most difficult term to bound among these three ones is the one which contains $D_t Z_{t_i}^N$. If we write $\lambda D_t Z_{t_i}^N + (1-\lambda)D_t Z_{t_i} = \lambda D_t(\Delta Z_i) + D_t Z_{t_i}$, we obtain

$$\begin{aligned}&\left|\mathbb{E}\left[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i f'''_{xzz}(\theta_{t_i}^\lambda)\lambda D_t(\Delta Z_i)\Delta Z_i\right]\right| \\ &\quad\leq C\left(\mathbb{E}\left(\sum_{i=1}^{N-1}|D_t(\Delta Z_i)|^2\right)\right)^{\frac{1}{2}}\left(\mathbb{E}\left(\sum_{i=1}^{N-1}(|\Delta X_i|^2|\tilde{\eta}_i|^2|\Delta Z_i|^2)\right)\right)^{\frac{1}{2}}, \\ &\quad\leq C\left(\mathbb{E}\left(\sum_{i=1}^{N-1}|D_t(\Delta Z_i)|^2\right)\right)^{\frac{1}{2}}\left(\mathbb{E}\left(\sum_{i=1}^{N-1}|\Delta Z_i|^2\right)^2\right)^{\frac{1}{4}} \\ &\qquad\times\left(\mathbb{E}\left(\max_{0\leq i\leq N}|\tilde{\eta}_i|^4 \max_{0\leq i\leq N}|\Delta X_i|^4\right)\right)^{\frac{1}{4}}.\end{aligned}$$

Applying Proposition 17, (17), Proposition 5 and (30) (with $\tilde{\eta}$) lead to $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i f'''_{xzz}(\theta_{t_i}^\lambda)\lambda D_t(\Delta Z_i)\Delta Z_i] = O(1)$. Proposition 5, (17), (38)–(40) and (52) enable us to prove that the other terms of $\mathbb{E}[\sum_{i=1}^{N-1} \tilde{\eta}_i \Delta X_i D_t(f''_{xz}(\theta_{t_i}^\lambda))\Delta Z_i]$ are $O(1)$. □

References

[1] V. Bally, D. Talay, The law of Euler scheme for stochastic differential equations, Probability Theory and related Fields 104 (1996) 43–60.

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

[3] D. Chevance, Numerical methods for backward stochastic differential equations, in: L.C.G. Rogers, D. Talay (Eds.), Numerical Methods in Financial Mathematics, Newton Institute, Cambridge University Press, 1997, pp. 232–244.
[4] F. Delarue, S. Menozzi, A forward–backward stochastic algorithm for quasi-linear PDEs, Annals of Applied Probability 1 (1) (2005) 1–49.
[5] N. El Karoui, S. Peng, M. Quenez, Backward stochastic differential equations in finance, Mathematical Finance 7 (1) (1997) 1–71.
[6] E. Gobet, J. Lemor, X. Warin, A regression-based Monte Carlo method to solve backward stochastic differential equations, Annals of Applied Probability 15 (3) (2005) 2172–2202.
[7] E. Gobet, R. Munos, Sensitivity analysis using Itô-Malliavin calculus and martingales, and application to stochastic optimal control, SIAM Journal on control and optimization 43 (5) (2004) 1676–1713.
[8] P. Hall, C. Heyde, Martingale Limit Theory and its Application, Academic Press, 1980.
[9] J. Jacod, P. Protter, Asymptotic error distributions for the Euler method for stochastic differential equations, Annals of Probability 26 (1) (1998) 267–307.
[10] A. Kohatsu-Higa, R. Pettersson, Variance reduction methods for simulation of densities on Wiener space, SIAM Journal on Numerical Analysis 40 (2) (2002) 431–450.
[11] T. Kurtz, P. Protter, Wong–Zakai corrections, random evolutions, and simulation schemes for SDEs, in: E. Mayer-Wolf, E. Merzbach, A. Schwartz (Eds.), Stochastic Analysis, Academic Press, 1991, pp. 331–346.
[12] O. Ladyzenskaja, V. Solonnikov, N. Ural'ceva, Linear and quasi-linear equations of parabolic type, in: Translations of Mathematical Monographs, vol. 23, American Mathematical Society, Providence, 1968.
[13] J. Lemor, E. Gobet, X. Warin, Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations, Bernoulli 12 (5) (2006).
[14] J. Ma, J. Yong, Forward–Backward Stochastic Differential Equations and Their Applications, in: Lecture Notes in Math., vol. 1702, Springer, 1999.
[15] J. Ma, J. Zhang, Representation theorems for backward stochastic differential equations, The Annals of Applied Probability 12 (4) (2002) 1390–1418.
[16] D. Nualart, Malliavin Calculus and Related Topics, Springer-Verlag, New York, 1995.
[17] E. Pardoux, Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order, in: Stochastic Analysis and Related Topics, VI (Geilo, 1996), in: Progr. Probab., vol. 42, Birkhäuser, Boston, MA, 1998, pp. 79–127.
[18] E. Pardoux, S. Peng, Adapted solution of a backward stochastic differential equation, Systems & Control Letters 14 (1) (1990) 55–61.
[19] P. Protter, Stochastic Integration and Differential Equations, Springer Verlag, 1990.
[20] J. Zhang, A numerical scheme for BSDEs, The Annals of Applied Probability 14 (1) (2004) 459–488.