



**HAL**  
open science

# Empirical Regression Method for Backward Doubly Stochastic Differential Equations

Achref Bachouch, Emmanuel Gobet, Anis Matoussi

► **To cite this version:**

Achref Bachouch, Emmanuel Gobet, Anis Matoussi. Empirical Regression Method for Backward Doubly Stochastic Differential Equations. SIAM/ASA Journal on Uncertainty Quantification, 2016, 4 (1), pp.358-379. 10.1137/15M1022094 . hal-01152886

**HAL Id: hal-01152886**

**<https://hal.science/hal-01152886v1>**

Submitted on 21 May 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Empirical Regression Method for Backward Doubly Stochastic Differential Equations

Achref Bachouch\*  
Humboldt-Universität  
zu Berlin

Emmanuel Gobet†  
Ecole Polytechnique

Anis MATOUSSI ‡  
University of Le Mans

May 18, 2015

## Abstract

In this paper we design a numerical scheme for approximating Backward Doubly Stochastic Differential Equations (BDSDEs for short) which represent solution to Stochastic Partial Differential Equations (SPDEs). We first use a time-discretization and then, we decompose the value function on a functions basis. The functions are deterministic and depend only on time-space variables, while decomposition coefficients depend on the external Brownian motion  $B$ . The coefficients are evaluated through a empirical regression scheme, which is performed conditionally to  $B$ . We establish non asymptotic error estimates, conditionally to  $B$ , and deduce how to tune parameters to obtain a convergence conditionally and unconditionally to  $B$ . We provide numerical experiments as well.

Keywords: Backward Doubly Stochastic Differential Equations, discrete Dynamic Programming Equations, empirical regression scheme, SPDEs.

MSC2010: Primary 60H10, Secondary 62G08

## 1 Introduction

Backward Doubly Stochastic Differential Equations (BDSDE in short) are classic tools to give Feynman-Kac representations for stochastic semilinear PDEs, see the seminal work of [PP94]. The BDSDE  $(Y^x, Z^x)$  of our interest is of the following form

$$Y_s^{t,x} = \Phi(X_T^{t,x}) + \int_s^T f(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) dr \quad (1)$$

---

\*Institute for Mathematics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. Email: [achref.bachouch@gmail.com](mailto:achref.bachouch@gmail.com).

†Centre de Mathématiques Appliquées, Ecole Polytechnique and CNRS, Route de Saclay, 91128 Palaiseau Cedex, France. Email: [emmanuel.gobet@polytechnique.edu](mailto:emmanuel.gobet@polytechnique.edu).

‡Risk and Insurance Institute and Laboratoire Manceau de Mathématiques, University of Le Mans, Avenue Olivier Messiaen, 72085 Le Mans Cedex 09, France. Email: [anis.matoussi@univ-lemans.fr](mailto:anis.matoussi@univ-lemans.fr).

$$+ \int_s^T h(r, X_r^{t,x}, Y_r^{t,x}, Z_r^{t,x}) \overleftarrow{d}B_r - \int_s^T Z_r^{t,x} dW_r,$$

where  $(X_s^{t,x})_{t \leq s \leq T}$  is a  $d$ -dimensional diffusion process starting from  $x$  at time  $t$  driven by the finite  $d$ -dimensional brownian motion  $(W_t)_{0 \leq t \leq T}$ . Here  $T > 0$  is fixed and the differential term with  $\overleftarrow{d}B_t$  refers to the backward stochastic integral with respect to a  $l$ -dimensional Brownian motion  $B$  independent from  $W$ . In addition  $W$  and  $B$  are defined on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  where we define the sigma-fields  $\mathcal{F}_{t,s}^W := \sigma\{W_r - W_t, t \leq r \leq s\}$ ,  $\mathcal{F}_{s,T}^B := \sigma\{B_r - B_s, s \leq r \leq T\}$ ,  $\mathcal{F}^W := \mathcal{F}_{0,T}^W$ ,  $\mathcal{F}^B := \mathcal{F}_{0,T}^B$ ,  $\mathcal{F} := \mathcal{F}^W \vee \mathcal{F}^B$ , all completed with the  $\mathbb{P}$ -null sets. To account for the measurability of the solution to (1), we need to define the collection of sigma-fields (for fixed  $t \in [0, T]$ )

$$\mathcal{F}_s^t := \mathcal{F}_{t,s}^W \vee \mathcal{F}_{s,T}^B,$$

and we know that the solution is such that  $Y_s^{t,x}$  is  $\mathcal{F}_s^t$ -measurable for any  $s \in [t, T]$  and  $Z_s^{t,x}$  is  $\mathcal{F}_s^t$ -measurable for a.e.  $s \in [t, T]$ .

Moreover, the random map  $(t, x) \mapsto (Y_t^{t,x}, Z_t^{t,x})$  provides the solution of the following SPDE and its gradient times  $\sigma$  at point  $(t, x)$ :

$$\begin{aligned} u(t, x) = & \Phi(x) + \int_t^T [\mathcal{L}u(s, x) + f(s, x, u(s, x), (\nabla_x u \sigma)(s, x))] ds \\ & + \int_t^T g(s, x, u(s, x), (\nabla_x u \sigma)(s, x)) \overleftarrow{d}B_s \end{aligned}$$

where  $\mathcal{L}$  is the infinitesimal generator of  $X$  (see [PP94, Theorem 3.1] for details). Such SPDEs appear in various applications like pathwise stochastic control problems, the Zakai equations in filtering and stochastic control with partial observations. Several generalizations to investigate more general nonlinear SPDEs have been developed following different approaches of the notion of weak solutions, namely, Sobolev's solutions [K99, BM01, MS02], and stochastic viscosity solutions [LS98, BuM01, LS02]. Generally, the approaches used to solve numerically SPDEs are analytic and based on time-space discretization of the equations. The discretization is achieved by different methods such as finite difference, finite element and spectral Galerkin methods [GN95, G99, W05, GK10, JK10]

Only recently some works have paid attention to the simulation and approximation of (1): See [Abo09, Ama13, BBMM13] for time discretization under various assumptions, see [Abo11] for an attempt to an implementable numerical scheme using regression methods without full convergence results, see [SYY08] for a scheme based on random walks.

In this work, we consider an empirical regression scheme (also known as regression Monte-Carlo method or least-squares Monte-Carlo method) for solving the discrete time BDSDE arising in [BBMM13]: This approach (inspired by [GLW05,

[LGW06] and more recently by [GT15b]) is increasingly popular and known to account well for high-dimensional problems, as a difference with scheme based on random walks. Our original contribution is the analysis of the regression scheme for approximating BDSDEs and its proof of convergence, with some non-asymptotic error estimates in order to have the most accurate control on the convergence w.r.t. all the parameters. Here, we adapt the tools for the regression error analysis, arising from discrete BSDE's approximation, developed recently in [GT15b] in a quite general context. These tools will allow us to analyse the regression error in the doubly stochastic framework.

We recall the different strategies of approximation using least squares algorithms, to better motivate our approach. For the sake of clarity, assume standard Lipschitz and boundedness assumptions (detailed later) and  $t = 0$ , then start with the case  $h \equiv 0$ , i.e. the usual BSDE case, and consider a time discretization scheme which takes the form (in [LGW06]) of One step forward Dynamic Programming (ODP for short) equation:  $Y_{t_N} = \Phi(X_{t_N})$  and for all  $i \in \{N - 1, \dots, 0\}$

$$\begin{aligned} Y_{t_i} &= \mathbb{E}[Y_{t_{i+1}} + f(t_i, X_{t_i}, Y_{t_{i+1}}, Z_{t_i})\Delta_i \mid \mathcal{F}_{0,t_i}^W], \\ \Delta_i Z_{t_i} &= \mathbb{E}\left[Y_{t_{i+1}} \Delta W_i^\top \mid \mathcal{F}_{0,t_i}^W\right], \end{aligned} \quad (2)$$

where  $\top$  denotes the transpose operator and where  $t_i \in \pi$ ,  $\pi := \{t_0 := 0, \dots, t_N := T\}$  being a discrete time grid of the time interval  $[0, T]$ ,  $\Delta_i := t_{i+1} - t_i$  and  $\Delta W_i := W_{t_{i+1}} - W_{t_i}$ . Since  $(X_{t_i})_i$  forms a Markov chain, there exist deterministic measurable functions  $y_i(\cdot)$  and  $z_i(\cdot)$ , but unknown, such that  $Y_{t_i} = y_i(X_{t_i})$  and  $Z_{t_i} = z_i(X_{t_i})$ . The functions  $y_i(\cdot)$  and  $z_i(\cdot)$  are solutions of least squares problems in  $L_2(\Omega, \mathbb{P}, \mathcal{F}_{t_i})$  and can be approximated on a finite dimensional subspace, which coefficients are computed using Monte-Carlo simulations. Now for the case  $h \neq 0$ , in [Abo11] a similar algorithm is proposed where the equation for  $Y_{t_i}$  is replaced by

$$Y_{t_i} = \mathbb{E}\left[Y_{t_{i+1}} + f(t_i, X_{t_i}, Y_{t_{i+1}}, Z_{t_i})\Delta_i + h(t_i, X_{t_i}, Y_{t_{i+1}}, Z_{t_{i+1}})\Delta B_i \mid \mathcal{F}_{0,t_i}^W \vee \mathcal{F}_{0,t_i}^B\right],$$

where  $\Delta B_i := B_{t_{i+1}} - B_{t_i}$  and similarly for the  $Z$ -component. Then, the author has designed an empirical least-squares algorithm by taking approximations in the space of functions w.r.t. the variables  $(X_{t_i}, B_{t_{k+1}} - B_{t_k} : i \leq k \leq N)$ : Thus, the dimension of this problem is  $\dim = d + l \times N$  and goes to infinity as the discretization parameter  $N \rightarrow +\infty$ . Since we know [LGW06, GT15b, GT15a] from the usual error analysis on BSDE that the convergence rates are of the form  $N^{-c_1/(c_2 + (\dim))}$  where  $c_1, c_2$  are positive constants and  $\dim$  is the dimension of the explanatory variables, it seems hopeless to conclude to the convergence of the above algorithm.

Our strategy of approximation is different from the above and it leads to a convergent scheme. This is inspired by the "SPDE" object seen as a PDE driven by an auxiliary independent noise (here the Brownian motion  $B$ ), i.e. we are to compute  $Y_{t_i}$  as a function  $x \mapsto y_i(\Delta B, x)$  for given Brownian increments  $\Delta B$ . As

a consequence, the dimension of the problem is still  $d$ , but the regression schemes and the error analysis are to be performed conditionally to  $B$ . This raises new difficulties, in particular because the theory of BDSDE is well posed unconditionally to  $B$ . Moreover and as a difference with [Abo11], we incorporate in our scheme an additional improvement inspired by [GT15b] in the BSDE setting, where the discrete BSDE is considered in the form of a Multi step forward Dynamic Programming (MDP for short) equation given by

$$Y_{t_i} = \mathbb{E} \left[ \Phi(X_{t_N}) + \sum_{k=i}^{N-1} f(t_k, X_{t_k}, Y_{t_{k+1}}, Z_{t_k}) \Delta_k \mid \mathcal{F}_{0,t_i}^W \right],$$

and similarly for  $Z_{t_i}$ . Using the tower property of conditional expectations, we note that ODP (based on (2)) and MDP coincide. But combined with empirical regression approximations, they are different and it is proved in [GT15b] that the MDP scheme leads to better error estimates than the ODP scheme, in particular for the  $Y$ -component. Indeed, the quadratic error is the average of local error terms rather than the sum.

In this work, we specialise our analysis to the case where  $f$  and  $h$  do not depend on  $z$ , i.e. we only approximate  $Y$ . We guess that this simplification makes the reading easier for the reader (even in the "simple" BSDE case as in [GT15b, GT15a], the analysis is rather tough) and the essence of our methodology remains unchanged if  $f$  and  $h$  depend on  $z$ . This simplified setting already raises new issues (about a priori estimates and stabilities) which we partly overcome but which will still deserve deeper investigation in the future to handle more general  $f$  and  $h$ .

The organisation of the paper is as follows. In Section 2 we give preliminaries on BDSDEs and the assumptions we will use. Then, we define the discrete BDSDE to be solved, in the MDP form. After that, we establish a priori estimates that will be useful in the regression error analysis. In Section 3, we present the Least Squares MDP algorithm designed to approximate the solution of the discrete BDSDE of Section 2. Then, we give the full analysis of the regression error conditionally and unconditionally to the Brownian motion  $B$ . Section 4 is dedicated to some numerical tests.

**Usual notations.** If  $x$  is in an Euclidean space  $E$ ,  $|x|$  denotes its norm. If  $\varphi$  is a vector-valued function defined on  $E$ ,  $|\varphi|_\infty$  denotes its sup-norm. If  $\nu$  is a probability measure on  $E$ ,  $|\cdot|_\nu$  stands for the  $L_2$ -norm w.r.t. the measure  $\nu$ . If  $\mathcal{X}$  is a  $E$ -valued random variable with distribution  $\nu$ , we may write  $|\cdot|_{\mathcal{X}} := |\cdot|_\nu$ . Last, if  $A$  is a matrix,  $|A|$  stands for its Hilbert-Schmidt norm.

## 2 Preliminaries and notations

This section gathers preliminary results to be used in order to discuss the approxi-

mation of the solution and its convergence. Actually, in the sequel we consider only the solution  $(X^t, Y^t, Z^t)$  with initial condition  $t = 0$ : Extending the results to other  $t$  is rather straightforward. Thus, from now on, we omit to indicate the dependence w.r.t.  $t$  by simply write  $(X, Y, Z)$ , the starting point  $X_0$  is given.

## 2.1 Forward Backward Doubly Stochastic Differential Equation

Recall the setting related to the filtered probability space given in the introduction. Let  $x \in \mathbb{R}^d$  be given and consider  $(X_s)_{0 \leq s \leq T}$  as solution of the following SDE

$$dX_s = b(X_s)ds + \sigma(X_s)dW_s \quad \text{for } s \in [0, T], \quad X_0 = x, \quad (3)$$

where  $b$  and  $\sigma$  are two given functions on  $\mathbb{R}^d$  with values respectively in  $\mathbb{R}^d$  and  $\mathbb{R}^d \otimes \mathbb{R}^d$ , that satisfy the following standard Lipschitz assumption.

**Assumption (H1).** There exists a non-negative constant  $K$  such that

$$|b(x) - b(x')| + |\sigma(x) - \sigma(x')| \leq K|x - x'|, \quad \forall x, x' \in \mathbb{R}^d.$$

This implies the existence of a unique strong solution to (3). Besides, we consider the following BDSDE defined by

$$\begin{cases} -dY_s &= f(s, X_s, Y_s)ds + h(s, X_s, Y_s)\overleftarrow{dB}_s - Z_s dW_s, \quad s \in [0, T], \\ Y_T &= \Phi(X_T) := \xi, \end{cases} \quad (4)$$

where  $f$  and  $h$  are respectively real-valued and  $\mathbb{R}^l$ -valued functions on  $[0, T] \times \mathbb{R}^d \times \mathbb{R}$  and  $\Phi$  is a real-valued function on  $\mathbb{R}^d$  ( $h$  is considered as a row vector).

Our standing assumptions to study (4) are the following.

**Assumption (H2).** There exist non-negative constants  $C_f, C_h, C_\xi, L_f, L_h$  and  $L_\xi$  such that

- i)  $|f(s_1, x_1, y_1) - f(s_2, x_2, y_2)| \leq L_f \left( \sqrt{|s_1 - s_2|} + |x_1 - x_2| + |y_1 - y_2| \right)$  for all  $s_1, s_2 \in [0, T]$ ,  $x_1, x_2 \in \mathbb{R}^d$  and  $y_1, y_2 \in \mathbb{R}$ ,
- ii)  $|h(s_1, x_1, y_1) - h(s_2, x_2, y_2)| \leq L_h \left( \sqrt{|s_1 - s_2|} + |x_1 - x_2| + |y_1 - y_2| \right)$  for all  $s_1, s_2 \in [0, T]$ ,  $x_1, x_2 \in \mathbb{R}^d$  and  $y_1, y_2 \in \mathbb{R}$ ,
- iii)  $|f(s, x, 0)|$  and  $|h(s, x, 0)|$  are uniformly bounded on  $[0, T] \times \mathbb{R}^d$  by  $C_f$  and  $C_h$  respectively,
- iv)  $|\Phi(x_1) - \Phi(x_2)| \leq L_\xi |x_1 - x_2|$  for all  $x_1, x_2 \in \mathbb{R}^d$ ,
- v)  $\Phi$  is uniformly bounded on  $\mathbb{R}^d$  by  $C_\xi$ .

Pardoux and Peng [PP94, Theorem 1.1] proved that under the previous assumptions, there exists a unique solution  $(Y, Z) \in \mathbb{S}^2([0, T]) \times \mathbb{H}_d^2([0, T])$  to the BDSDE (4), where

- $\mathbb{H}_d^2([0, T])$  denotes the set of (classes of  $d\mathbb{P} \times dt$  a.e. equal)  $\mathbb{R}^d$ -valued jointly measurable processes  $\{\psi_s; s \in [0, T]\}$  such that  $\mathbb{E} \left[ \int_0^T |\psi_s|^2 ds \right] < +\infty$  and  $\psi_s$  is  $\mathcal{F}_s^0$ -measurable, for a.e.  $s \in [0, T]$ .
- $\mathbb{S}^2([0, T])$  denotes the set of real-valued continuous processes such that  $\mathbb{E} \left[ \sup_{0 \leq s \leq T} |\psi_s|^2 \right] < +\infty$  and  $\psi_s$  is  $\mathcal{F}_s^0$ -measurable for any  $s \in [0, T]$ .

## 2.2 Time-discretization scheme for decoupled Forward-BDSDE

In order to approximate the solution of the Forward-BDSDE (3)-(4), we introduce the following discretized version. Let

$$\pi = \{t_0 := 0 < t_1 < \dots < t_N := T\}$$

be a partition of the time interval  $[0, T]$  with time step  $\Delta_i := t_{i+1} - t_i$ ,  $0 \leq i \leq N - 1$ . Throughout this work, we will use the notations  $\Delta W_i := W_{t_{i+1}} - W_{t_i}$  and  $\Delta B_i := B_{t_{i+1}} - B_{t_i}$ , for  $i = 0, \dots, N - 1$ .

The forward component  $X$  is approximated by the classical Euler scheme:

$$\begin{cases} X_{t_0}^\pi = x, \\ X_{t_{i+1}}^\pi = X_{t_i}^\pi + b(X_{t_i}^\pi)\Delta_i + \sigma(X_{t_i}^\pi)\Delta W_i, \text{ for } i = 0, \dots, N - 1. \end{cases}$$

It is known that as  $\max_{0 \leq i \leq N-1} \Delta_i \rightarrow 0$ , one has  $\sup_{0 \leq i \leq N} \mathbb{E} [|X_{t_i} - X_{t_i}^\pi|^2] \rightarrow 0$ .

The solution  $Y$  of (4) is approximated by  $Y^\pi$  defined by the following **Multi step-forward Dynamic Programming (MDP)** equation: For  $i = N - 1, \dots, 0$ , we set

$$\begin{aligned} Y_{t_i}^\pi &= \mathbb{E}_i \left[ \Phi(X_T^\pi) + \sum_{k=i}^{N-1} \left( f(t_k, X_{t_k}^\pi, Y_{t_{k+1}}^\pi)\Delta_k + h(t_{k+1}, X_{t_{k+1}}^\pi, Y_{t_{k+1}}^\pi)\Delta B_k \right) \right] \\ &= \mathbb{E}_i \left[ Y_{t_{i+1}}^\pi + f(t_i, X_{t_i}^\pi, Y_{t_{i+1}}^\pi)\Delta_i + h(t_{i+1}, X_{t_{i+1}}^\pi, Y_{t_{i+1}}^\pi)\Delta B_i \right], \end{aligned} \quad (5)$$

where  $\mathbb{E}_i [\cdot]$  denotes the conditional expectation w.r.t.  $\mathcal{G}_i$  defined by

$$\mathcal{G}_i := \sigma(\Delta W_j, 0 \leq j \leq i - 1) \vee \mathcal{F}^{\Delta B}$$

with  $\mathcal{F}^{\Delta B} = \sigma(\Delta B_j, 0 \leq j \leq N - 1)$ . Observe that  $(\mathcal{G}_i)_{0 \leq i \leq N-1}$  is a discrete filtration associated to the time grid  $\pi$ . We recall from [BBMM13, Theorem 4.1] the following convergence result for the time discretization error. Set

$$\text{Error}_N(Y, Z) := \max_{0 \leq i \leq N-1} \sup_{t_i \leq s \leq t_{i+1}} \mathbb{E} [|Y_s - Y_{t_i}^\pi|^2] + \sum_{i=0}^{N-1} \mathbb{E} \left[ \int_{t_i}^{t_{i+1}} |Z_s - Z_{t_i}^\pi|^2 ds \right].$$

Thus, we have

**Theorem 1** *Under Assumptions (H1)-(H2) and assuming in addition that  $b$ ,  $\sigma$ ,  $\Phi$ ,  $f$  and  $h$  are of class  $C^2$  with bounded derivatives up to order 2, there exists a non-negative constant  $C_{(6)}$  (independent on  $\pi$ ) such that*

$$\text{Error}_N(Y, Z) \leq C_{(6)} \max_{0 \leq i \leq N-1} \Delta_i. \quad (6)$$

We note that in the case we are dealing with (i.e. the drivers are independent from the variable  $z$ ), we do not have to approximate the control process  $Z$ , since it does not enter in the approximation of  $Y$ .

**Extra notations.** Our aim being the Monte-Carlo approximation of the discrete BDSDE solution for a given time grid  $\pi$ , we shall alleviate the notation by simply writing  $X_i, Y_i$  for  $X_{t_i}^\pi, Y_{t_i}^\pi$ . Furthermore, we shall write

$$\Delta B := \{\Delta B_j, 0 \leq j \leq N-1\}.$$

With this notation and since the data are Lipschitz (coefficients of the BDSDE and of the Euler scheme), it is easy to check the following lemma, by combining the Equation (5) with a recursion argument.

**Lemma 1** *Under Assumptions (H1)-(H2), for each  $i \in \{0, \dots, N\}$  there exists a locally Lipschitz function  $y_i : (\mathbb{R}^l)^N \times \mathbb{R}^d \mapsto \mathbb{R}$  such that*

$$Y_i = y_i(\Delta B, X_i). \quad (7)$$

### 2.3 A priori estimates

In this section, we establish a priori estimates on discrete BDSDEs. These estimates will be needed later for the regression analysis. In the case of pure BSDEs, they are rather standard (see [GT15b] among other references): on the one hand we take advantage of the driver independent of  $Z$  to provide slightly stronger estimates than usually. On the other hand, the BDSDE setting with the  $\Delta B$  contribution is a source of difficulty in the analysis.

We aim at comparing two discrete BDSDEs,  $Y_{1,\cdot}$  and  $Y_{2,\cdot}$ , defined as follows. For  $j = 1, 2$ , we set  $Y_{j,N} = \xi_j$  and for all  $i = N-1, \dots, 0$

$$Y_{j,i} = \tilde{\mathbb{E}}_i [Y_{j,i+1} + f_{j,i}(Y_{j,i+1})\Delta_i + h_{j,i+1}(Y_{j,i+1})\Delta B_i], \quad (8)$$

where

- $\tilde{\mathbb{E}}_i[\cdot]$  is the conditional expectation w.r.t.  $\mathcal{G}_i \vee \tilde{\mathcal{G}}$ , where  $\tilde{\mathcal{G}}$  is a sigma-field independent of  $W$  and  $B$ ,
- $(\omega, y) \rightarrow f_{1,i}(\omega, y) := f_{1,i}(y)$  and  $(\omega, y) \rightarrow f_{2,i}(\omega, y) := f_{2,i}(y)$  are real-valued and  $[\mathcal{G}_i \vee \tilde{\mathcal{G}}] \otimes \mathcal{B}(\mathbb{R})$ -measurable functions on  $\mathbb{R}$ ,



- $(\omega, y) \rightarrow h_{1,i+1}(\omega, y) := h_{1,i+1}(y)$  and  $(\omega, y) \rightarrow h_{2,i+1}(\omega, y) := h_{2,i+1}(y)$  are  $\mathbb{R}^l$ -valued and  $[\mathcal{G}_{i+1} \vee \tilde{\mathcal{G}}] \otimes \mathcal{B}(\mathbb{R})$ -measurable functions on  $\mathbb{R}$ .

The above choice of measurability is coherent with the MDP equation (5) and with the further feature that solutions to posterior times are built using extra independent Monte-Carlo simulations (to be associated to the sigma-field  $\tilde{\mathcal{G}}$ ). We set

$$\begin{aligned}\delta Y_i &:= Y_{1,i} - Y_{2,i}, \\ \delta \xi &:= \xi_1 - \xi_2, \\ \delta f_i &:= f_{1,i}(Y_{1,i+1}) - f_{2,i}(Y_{1,i+1}), \\ \delta h_{i+1} &:= h_{1,i+1}(Y_{1,i+1}) - h_{2,i+1}(Y_{1,i+1}).\end{aligned}$$

Now, we are in a position to state the following lemma, which gives a local estimate on the solutions of two discrete BDSDEs.

**Lemma 2** *We assume that for  $j \in \{1, 2\}$ ,  $\xi_j$  belongs to  $L_2(\mathcal{G}_N \vee \tilde{\mathcal{G}})$  and that for  $i \in \{0, \dots, N-1\}$ ,  $f_{j,i}(Y_{j,i+1})$  and  $h_{j,i+1}(Y_{j,i+1})$  belong to  $L_2(\mathcal{G}_{i+1} \vee \tilde{\mathcal{G}})$ . In addition, we assume that  $f_{2,i}$  and  $h_{2,i+1}$  are Lipschitz continuous with Lipschitz constants  $L_{f_{2,i}}$  and  $L_{h_{2,i+1}}$  (possibly  $\mathcal{G}_0 \vee \tilde{\mathcal{G}}$ -measurable). Then, we have*

$$\begin{aligned}|\delta Y_i| &\leq (1 + L_{f_{2,i}} \Delta_i + L_{h_{2,i+1}} |\Delta B_i|) \tilde{\mathbb{E}}_i [|\delta Y_{i+1}|] \\ &\quad + \tilde{\mathbb{E}}_i [|\delta f_i|] \Delta_i + \tilde{\mathbb{E}}_i [|\delta h_{i+1}|] |\Delta B_i|.\end{aligned}\tag{9}$$

**Proof.** From (8), we have

$$\begin{aligned}\delta Y_i &= \tilde{\mathbb{E}}_i [\delta Y_{i+1} + \{\delta f_i + f_{2,i}(Y_{1,i+1}) - f_{2,i}(Y_{2,i+1})\} \Delta_i] \\ &\quad + \tilde{\mathbb{E}}_i [\{\delta h_{i+1} + h_{2,i+1}(Y_{1,i+1}) - h_{2,i+1}(Y_{2,i+1})\} \Delta B_i].\end{aligned}$$

Applying the triangle and Jensen inequalities, then using the Lipschitz assumptions on  $f_{2,i}$  and  $h_{2,i}$  give the estimation (9).  $\square$

By propagating the above result, we obtain a global stability result on the solutions of two discrete BDSDEs. The proof is easy and left to the reader.

**Proposition 1** *Under the notations and assumptions of Lemma 2, the following estimation holds a.s. for all  $i \in \{1, \dots, N\}$ :*

$$\Gamma_i |\delta Y_i| \leq \Gamma_N \tilde{\mathbb{E}}_i [|\delta \xi|] + \sum_{k=i}^{N-1} \Gamma_k \left( \tilde{\mathbb{E}}_i [|\delta f_k|] \Delta_k + \tilde{\mathbb{E}}_i [|\delta h_{k+1}|] |\Delta B_k| \right),$$

where  $\Gamma_i := \prod_{j=0}^{i-1} (1 + L_{f_{2,j}} \Delta_j + L_{h_{2,j+1}} |\Delta B_j|)$  and  $\Gamma_0 := 1$ .

As an application of the above proposition, we can derive an a.s. upper bound for the solution of the discrete BDSDE (5). Such an upper bound is required in the subsequent empirical regression algorithm.

**Proposition 2** *Under Assumptions (H1) and (H2), the solution of the discrete BDSDE (5)-(7) has an a.s. upper bound, uniformly w.r.t.  $i \in \{0, \dots, N\}$ :*

$$|y_i(\Delta B, \cdot)|_\infty \leq C_y^{\Delta B} := e^{L_f T + L_h \sum_{j=0}^{N-1} |\Delta B_j|} \left( C_\xi + C_f T + C_h \sum_{j=0}^{N-1} |\Delta B_j| \right).$$

**Proof.** Apply Proposition 1 by setting  $Y_{1,i} := 0$  (with  $\xi_1 := 0$ ,  $f_{1,i} \equiv 0$ ,  $h_{1,i} \equiv 0$ ) and  $Y_{2,i} := Y_i = y_i(\Delta B, X_i)$  (with  $\xi_2 := \Phi(X_N)$ ,  $f_{2,i}(y) := f(t_i, X_i, y)$  and  $h_{2,i+1}(y) := h(t_{i+1}, X_{i+1}, y)$ ): Combined with Assumption (H2) this gives

$$\Gamma_i |y_i(\Delta B, \cdot)|_\infty \leq \Gamma_N C_\xi + \sum_{k=i}^{N-1} \Gamma_k (C_f \Delta_k + C_h |\Delta B_k|).$$

We obtain the announced result by observing that

$$\Gamma_i \leq \exp \left( \sum_{j=0}^{i-1} [L_f \Delta_j + L_h |\Delta B_j|] \right) \leq \exp \left( L_f T + L_h \sum_{j=0}^{N-1} |\Delta B_j| \right). \quad (10)$$

□

Observe that unfortunately this a.s. upper bound explodes in probability as  $N \rightarrow \infty$  because

$$\sum_{j=0}^{N-1} |\Delta B_j| \geq \left( \sum_{j=0}^{N-1} |\Delta B_j|^2 \right) / \max_{0 \leq j \leq N-1} |\Delta B_j| \sim T / \max_{0 \leq j \leq N-1} |\Delta B_j| \rightarrow +\infty$$

in probability. On the other hand, it is valid in rather great generality under the assumptions of our setting ( $f$  and  $h$  Lipschitz). Nevertheless, an easy improvement can be obtained provided that  $f$  and  $h$  are uniformly bounded, avoiding the exponential factor: Indeed, from (5) we directly have

$$|y_i(\Delta B, \cdot)|_\infty \leq C_\xi + T |f|_\infty + |h|_\infty \sum_{j=0}^{N-1} |\Delta B_j| \text{ a.s.,}$$

yielding another upper bound which still explodes as  $N \rightarrow \infty$  but at a slower rate.

Lastly, we know that  $\sup_{N \geq 1} \sup_{0 \leq i \leq N} \mathbb{E} |y_i(\Delta B, X_i)|^2 < +\infty$ , see [BBMM13], which shows the gap between a.s. and  $L_2$  estimates. This is a difficulty intrinsic to the study of pathwise property of BDSDE: To our knowledge, having good pathwise estimates is an open question.

### 3 Regression Monte-Carlo scheme

In this section, we design an algorithm to approximate conditional expectations involved in (5) using linear least squares methods (empirical regressions). We also analyse its convergence.

Since we are to consider regressions conditionally to  $\Delta B$ , it is clearer to write  $(\Omega, \mathcal{F}, \mathbb{P})$  as a product space  $(\Omega^{\Delta W} \times \Omega^{\Delta B}, \mathcal{F}^{\Delta W} \otimes \mathcal{F}^{\Delta B}, \mathbb{P}^{\Delta W} \otimes \mathbb{P}^{\Delta B})$  where  $\mathcal{F}^{\Delta W} = \sigma(\Delta W_j : 0 \leq j \leq N-1)$  and  $\mathcal{F}^{\Delta B} = \sigma(\Delta B_j : 0 \leq j \leq N-1)$ , coherently with the discrete BDSDE (5) to solve. This is our convention from now on. Then the conditional expectation w.r.t.  $\mathcal{F}^{\Delta B}$  is denoted by  $\mathbb{E}_{\Delta B} [\cdot]$ .

### 3.1 Preliminaries on Ordinary Least-Squares (OLS)

In the following, we recall the definition of the Least-Squares regression as stated in [GT15b] and specialise it to our framework. The next general probability space  $(\tilde{\Omega} \times \Omega^{\Delta B}, \tilde{\mathcal{F}} \otimes \mathcal{F}^{\Delta B}, \tilde{\mathbb{P}} \otimes \mathbb{P}^{\Delta B})$  (which will be larger than  $(\Omega, \mathcal{F}, \mathbb{P})$ ) is to account for the extra simulations used in the regression Monte-Carlo algorithm.

**Definition 1** *Let  $n \geq 1$ . We consider the two probability spaces  $(\tilde{\Omega} \times \Omega^{\Delta B}, \tilde{\mathcal{F}} \otimes \mathcal{F}^{\Delta B}, \tilde{\mathbb{P}} \otimes \mathbb{P}^{\Delta B})$  and  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \nu)$ . Let*

- *$S$  be a  $\tilde{\mathcal{F}} \otimes \mathcal{F}^{\Delta B} \otimes \mathcal{B}(\mathbb{R}^n)$ -measurable  $\mathbb{R}$ -valued function such that  $S(\tilde{\omega}, \Delta B, \cdot) \in L_2(\mathcal{B}(\mathbb{R}^n), \nu)$  for  $\tilde{\mathbb{P}} \otimes \mathbb{P}^{\Delta B}$ -a.e.  $(\tilde{\omega}, \Delta B) \in \tilde{\Omega} \otimes \Omega^{\Delta B}$ ,*
- *$\mathcal{K}$  be the linear vector subspace of  $L_2(\mathcal{B}(\mathbb{R}^n), \nu)$  spanned by  $\mathcal{F}^{\Delta B} \otimes \mathcal{B}(\mathbb{R}^n)$ -measurable  $\mathbb{R}$ -valued functions  $\{p^j(\Delta B, \cdot), j \geq 1\}$ .*

*The least squares approximation of  $S$  in the space  $\mathcal{K}$  with respect to  $\nu$  is the  $\tilde{\mathbb{P}} \otimes \mathbb{P}^{\Delta B} \otimes \nu$ -a.e. unique and  $\tilde{\mathcal{F}} \otimes \mathcal{F}^{\Delta B} \otimes \mathcal{B}(\mathbb{R}^n)$ -measurable function  $S^*$  given by:*

$$S^*(\tilde{\omega}, \Delta B, \cdot) := \operatorname{arginf}_{\phi \in \mathcal{K}} \int |S(\tilde{\omega}, \Delta B, x) - \phi(x)|^2 \nu(dx).$$

*Then, we say that  $S^*$  solves OLS( $S, \mathcal{K}, \nu$ ).*

*In the same manner, let  $M := M^{\Delta B}$  be a positive integer-valued  $\mathcal{F}^{\Delta B}$ -random variable and  $\nu_M := \frac{1}{M^{\Delta B}} \sum_{m=1}^{M^{\Delta B}} \delta_{\mathcal{X}^{(m)}}$  be a discrete probability measure on  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ , where  $\delta_x$  is the Dirac measure at  $x$  and  $(\mathcal{X}^{(m)} : \tilde{\Omega} \rightarrow \mathbb{R}^n, m \geq 1)$  is an infinite sequence of i.i.d. random variables. For an  $\tilde{\mathcal{F}} \otimes \mathcal{F}^{\Delta B} \otimes \mathcal{B}(\mathbb{R}^n)$ -measurable real-valued function  $S$  such that  $|S(\tilde{\omega}, \Delta B, \mathcal{X}^{(m)}(\tilde{\omega}))| < +\infty$  for all  $m$  and  $\tilde{\mathbb{P}} \otimes \mathbb{P}^{\Delta B}$ -a.e.  $(\tilde{\omega}, \Delta B) \in \tilde{\Omega} \times \Omega^{\Delta B}$ , the least squares approximation of  $S$  in the space  $\mathcal{K}$  with respect to  $\nu_M$  is the  $(\tilde{\mathbb{P}} \otimes \mathbb{P}^{\Delta B}$ -a.e.) unique,  $\tilde{\mathcal{F}} \otimes \mathcal{F}^{\Delta B} \otimes \mathcal{B}(\mathbb{R}^n)$ -measurable function  $S^*$  given by*

$$S^*(\tilde{\omega}, \Delta B, \cdot) := \operatorname{arginf}_{\phi \in \mathcal{K}} \frac{1}{M^{\Delta B}} \sum_{m=1}^{M^{\Delta B}} |S(\tilde{\omega}, \Delta B, \mathcal{X}^{(m)}(\tilde{\omega})) - \phi(\mathcal{X}^{(m)}(\tilde{\omega}))|^2. \quad (11)$$

*Then, we say that  $S^*$  solves OLS( $S, \mathcal{K}, \nu_M$ ).*

Due to (7), the MDP equation (5) is interpreted in terms of Definition 1 as follows:

For all  $i \in \{0, \dots, N-1\}$ ,  $y_i(\Delta B, \cdot)$  is the measurable function given by:

$$y_i(\Delta B, \cdot) = \text{solution of OLS} \left( \mathcal{Y}_i(\Delta B, \cdot), \mathcal{K}_i, \nu_i \right), \quad (12)$$

where  $\nu_i := \mathbb{P} \circ (X_i, \dots, X_N)^{-1}$ ,  $\mathcal{K}_i$  is any dense subset in the real-valued functions belonging to  $L_2(\mathcal{B}(\mathbb{R}^d), \mathbb{P} \circ (X_i)^{-1})$  and

$$\begin{aligned} \mathcal{Y}_i(\Delta B, x_{i:N}) := & \Phi(x_N) + \sum_{k=i}^{N-1} \left( f(t_k, x_k, y_{k+1}(\Delta B, x_{k+1})) \Delta B_k \right. \\ & \left. + h(t_{k+1}, x_{k+1}, y_{k+1}(\Delta B, x_{k+1})) \Delta B_k \right) \end{aligned} \quad (13)$$

with  $x_{i:N} := (x_i, \dots, x_N) \in (\mathbb{R}^d)^{N-i+1}$ . To make the algorithm implementable, the infinite-dimensional space  $\mathcal{K}_i$  and the exact measure  $\nu_i$  in (12) are replaced respectively by a finite-dimensional space and an empirical measure.

### 3.2 Notations and algorithm

The solution  $y_i(\Delta B, \cdot)$  of (12) will be approximated in a finite dimensional functional linear space, defined hereafter. Because the algorithm and the regression analysis are performed conditionally on  $\Delta B$ , it is important that the number  $M$  of data to be used and the functions space  $\mathcal{K}$  depend on  $\Delta B$  in Definitions 1 and 2. This is a significant difference with [GT15b] where  $\mathcal{K}$  and  $M$  are not stochastic.

**Definition 2 (Finite dimensional approximation spaces)** *For each  $i \in \{0, \dots, N-1\}$ , the finite dimensional approximation space  $\mathcal{K}_{Y,i}^{\Delta B}$  (of cardinality  $K_{Y,i}^{\Delta B}$  which is a finite  $\mathcal{F}^{\Delta B}$ -random variable) is given by:*

$$\mathcal{K}_{Y,i}^{\Delta B} := \text{span} \{ p_i^j(\Delta B, \cdot), j = 1, \dots, K_{Y,i}^{\Delta B} \}$$

where for all  $j$ , the  $\mathcal{F}^{\Delta B} \otimes \mathcal{B}(\mathbb{R}^d)$ -measurable function  $p_i^j : \Omega^{\Delta B} \otimes \mathbb{R}^d \rightarrow \mathbb{R}$  satisfies the condition  $\mathbb{E}_{\Delta B} \left[ |p_i^j(\Delta B, X_i)|^2 \right] < +\infty$ .

The best approximation error of  $y_i(\Delta B, \cdot)$  on the linear space  $\mathcal{K}_{Y,i}^{\Delta B}$  is given by

$$\tau_{1,i}^{\Delta B, Y} := \inf_{\phi \in \mathcal{K}_{Y,i}^{\Delta B}} \mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, X_i) - \phi(X_i) \right|^2 \right].$$

The computation of the OLS (12) involves the law of  $X_i, \dots, X_N$ , which is replaced by the empirical measure, defined as follows.

**Definition 3 (Simulations and empirical measure)** *For any  $i \in \{0, \dots, N-1\}$ , let  $M_i^{\Delta B}$  be<sup>1</sup> the number of Monte-Carlo simulations used for the regression at*

---

<sup>1</sup> $M_i^{\Delta B}$  may depend on  $\Delta B$  to allow an optimal tuning of parameters as a function of  $\Delta B$ , see Corollary 1. To avoid overfitting, we assume w.l.o.g.  $M_i^{\Delta B} \geq K_{Y,i}^{\Delta B}$ .

time  $t_i$ : namely, we sample independent copies of  $X_{i:N} := (X_i, \dots, X_N)$ , that we denote by

$$\mathcal{C}_i := \{X_{i:N}^{(i,m)}, m \geq 1\}$$

and that we call cloud of simulations at time  $t_i$ . For the algorithm we will use only the first  $M_i^{\Delta B}$  simulations of  $\mathcal{C}_i$ , however for the sake of clarity in the analysis error it is more convenient to write  $\mathcal{C}_i$  with an infinite sequence.

In addition, we assume that the clouds  $\{\mathcal{C}_i; i = 0, \dots, N-1\}$  are sampled independently. The random variables  $(X_{i:N}^{(i,m)} : 0 \leq i \leq N-1, m \geq 1)$  are supported by a probability space  $(\Omega^{(M)}, \mathcal{F}^{(M)}, \mathbb{P}^{(M)})$  and we define the empirical probability measure associated to the cloud  $\mathcal{C}_i$ :

$$\nu_{i,M} := \frac{1}{M_i^{\Delta B}} \sum_{m=1}^{M_i^{\Delta B}} \delta_{(X_i^{(i,m)}, \dots, X_N^{(i,m)})}.$$

The  $L_2$ -norm w.r.t  $\nu_{i,M}$  will be denoted as usually as  $|\cdot|_{\nu_{i,M}}$  (and  $|\cdot|_{\nu}$  for another measure  $\nu$ ).

Then, the full probability space used to analyse the following algorithm is  $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}}) = (\Omega, \mathcal{F}, \mathbb{P}) \otimes (\Omega^{(M)}, \mathcal{F}^{(M)}, \mathbb{P}^{(M)})$ . Within this extended probability space, we keep the same notation for probability and expectation, whenever unambiguous, for the sake of simplicity.

The algorithm is defined as follows.

**Algorithm 1 (Least-Squares MDP (LSMDP) Algorithm)** We define  $y_i^{(M)}(\Delta B, \cdot)$ , for all  $i$ , by a backward induction. We start with

$$y_N^{(M)}(\Delta B, \cdot) := \Phi(\cdot)$$

and for  $i = N-1, \dots, 0$ , we define

$$\psi_i^{(M)}(\Delta B, \cdot) \text{ as the solution of OLS} \left( \mathcal{Y}_i^{(M)}(\Delta B, \cdot), \mathcal{K}_{Y,i}^{\Delta B}, \nu_{i,M} \right) \quad (14)$$

where for any  $x_{i:N} := (x_i, \dots, x_N) \in (\mathbb{R}^d)^{N-i+1}$ , we set

$$\begin{aligned} \mathcal{Y}_i^{(M)}(\Delta B, x_{i:N}) := & \Phi(x_N) + \sum_{k=i}^{N-1} \left( f(t_k, x_k, y_{k+1}^{(M)}(\Delta B, x_{k+1})) \Delta_k \right. \\ & \left. + h(t_{k+1}, x_{k+1}, y_{k+1}^{(M)}(\Delta B, x_{k+1})) \Delta B_k \right). \end{aligned} \quad (15)$$

After that, we set

$$y_i^{(M)}(\Delta B, \cdot) := \left[ \psi_i^{(M)}(\Delta B, \cdot) \right]_i, \quad (16)$$

where  $[\cdot]_i$  is the soft thresholding operator defined by

$$[y]_i := -C_y^{\Delta B} \vee y \wedge C_y^{\Delta B}, \quad (17)$$

$C_y^{\Delta B}$  being the bound computed in Proposition 2. Any other (and better) upper bound on  $|y_i(\Delta B, \cdot)|_{\infty}$  could advantageously replace  $C_y^{\Delta B}$ .

The further statements will be made in terms of the squared approximation error of  $y_i$  in the linear space  $\mathcal{K}_{Y,i}^{\Delta B}$  with respect to the empirical measure  $\nu_{i,M}$ , defined by

$$\tau_{1,i,M}^{\Delta B,Y} := \mathbb{E}_{\Delta B} \left[ \inf_{\phi \in \mathcal{K}_{Y,i}^{\Delta B}} \left| y_i(\Delta B, \cdot) - \phi \right|_{\nu_{i,M}}^2 \right]. \quad (18)$$

A simple argument based on the inversion of  $\mathbb{E}_{\Delta B}$  and inf and on the independence between simulations  $(X_{i:N}^{i,m}, m \geq 1)$  and  $\Delta B$  yields the following bound.

**Lemma 3** *For all  $i \in \{0, \dots, N-1\}$ , we have*

$$\tau_{1,i,M}^{\Delta B,Y} \leq \tau_{1,i}^{\Delta B,Y}.$$

### 3.3 Main result: non-asymptotic error estimates for the regression scheme

The following theorem gives the conditional regression error of Algorithm 1 for approximating solutions of (5): It is measured in terms of

$$\eta_{i,M}^{Y,\Delta B} := \sqrt{\mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, \cdot) - y_i^{(M)}(\Delta B, \cdot) \right|_{\nu_{i,M}}^2 \right]}, \quad (19)$$

$$\eta_i^{Y,\Delta B} := \sqrt{\mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, X_i) - y_i^{(M)}(\Delta B, X_i) \right|^2 \right]}. \quad (20)$$

Actually by using uniform concentration-of-measure estimates, we can switch from one error to the other, up to a small error term; see later Proposition 4 in our specific setting or more generally Proposition 5.

**Theorem 2** *Under Assumptions (H1-H2), for any  $i \in \{0, \dots, N-1\}$  we have*

$$\eta_{i,M}^{Y,\Delta B} \leq \delta_i + \sqrt{2} \exp(\sqrt{2}L_f T + \sqrt{2}L_h \sum_{k=i}^{N-1} |\Delta B_k|) \sum_{k=i}^{N-2} (L_f \Delta_k + L_h |\Delta B_k|) \delta_{k+1}, \quad (21)$$

where for all  $k$  in  $\{0, \dots, N-1\}$

$$\begin{aligned} \delta_k &:= \left( \tau_{1,k,M}^{\Delta B,Y} \right)^{\frac{1}{2}} + \left( \frac{\text{card}(\mathcal{K}_{Y,k}^{\Delta B})}{M_k^{\Delta B}} \right)^{\frac{1}{2}} \sigma_{\mathcal{Y}_k}(\Delta B) \\ &+ \sqrt{2028} C_y^{\Delta B} \sum_{j=k}^{N-2} (L_f \Delta_j + L_h |\Delta B_j|) \sqrt{\frac{(\text{card}(\mathcal{K}_{Y,j+1}^{\Delta B}) + 1) \log(3M_{j+1}^{\Delta B})}{M_{j+1}^{\Delta B}}}, \end{aligned} \quad (22)$$

with

$$\sigma_{\mathcal{Y}_k}(\Delta B) := C_\xi + T(L_f C_y^{\Delta B} + C_f) + (L_h C_y^{\Delta B} + C_h) \sum_{j=k}^{N-1} |\Delta B_j|. \quad (23)$$

Theorem 2 gives explicit non asymptotic error estimates for the algorithm, since the constants of the error upper bound depend explicitly on the time grid  $\pi$  and on the path  $(\Delta B_k)_{0 \leq k \leq N-1}$ . As in [GT15b], it allows an easy tuning of the convergence parameters  $\mathcal{K}_{Y,i}^{\Delta B}$  and  $M_i^{\Delta B}$  to obtain an a.s. convergence given the external noise  $B$  as in the spirit of SPDEs. The subsequent convergence result (Corollary 1) is made possible owing to the Lipschitz regularity of the unknown solution  $x \mapsto y_i(\Delta B, x)$ , which is stated as follows.

**Proposition 3** *Under Assumptions (H1-H2), for any  $x, x' \in \mathbb{R}^d$  and  $i_0 \in \{0, \dots, N-1\}$  we have*

$$|y_{i_0}(\Delta B, x) - y_{i_0}(\Delta B, x')| \leq C_{(24)}^{\Delta B} |x - x'| \quad (24)$$

where  $C_{(24)}^{\Delta B} := C_{(25)} e^{L_f T + L_h \sum_{j=0}^{N-1} |\Delta B_j|} \left( L_\xi + L_f T + L_h \sum_{j=0}^{N-1} |\Delta B_j| \right)$ .

**Proof.** Set  $X_i^{i_0, x} = x$  for  $i \leq i_0$  and let  $(X_i^{i_0, x})_{i \geq i_0}$  be the Euler scheme starting from  $x$  at time  $i_0$ . We apply Proposition 1 by setting  $Y_{1,i} := y(\Delta B, X_i^{i_0, x})$  (with  $\xi_1 := \Phi(X_N^{i_0, x})$ ,  $f_{1,i}(\cdot) := f(t_i, X_i^{i_0, x}, \cdot)$ ,  $h_{1,i+1}(\cdot) := h(t_{i+1}, X_{i+1}^{i_0, x}, \cdot)$ ) and  $Y_{2,i} := y_i(\Delta B, X_i^{i_0, x'})$  (with  $\xi_2 := \Phi(X_N^{i_0, x'})$ ,  $f_{2,i}(\cdot) := f(t_i, X_i^{i_0, x'}, \cdot)$  and  $h_{2,i+1}(\cdot) := h(t_{i+1}, X_{i+1}^{i_0, x'}, \cdot)$ ). With Assumption (H2), we get

$$\begin{aligned} \Gamma_i |y_i(\Delta B, X_i^{i_0, x}) - y_i(\Delta B, X_i^{i_0, x'})| &\leq \Gamma_N L_\xi \mathbb{E}_i \left[ |X_N^{i_0, x} - X_N^{i_0, x'}| \right] \\ &+ \sum_{k=i}^{N-1} \Gamma_k \left( L_f \mathbb{E}_i \left[ |X_k^{i_0, x} - X_k^{i_0, x'}| \right] \Delta_k + L_h \mathbb{E}_i \left[ |X_{k+1}^{i_0, x} - X_{k+1}^{i_0, x'}| \right] |\Delta B_k| \right). \end{aligned}$$

Using (10) and taking  $i = i_0$  in the above inequality, we get

$$\begin{aligned} |y_{i_0}(\Delta B, x) - y_{i_0}(\Delta B, x')| &\leq e^{L_f T + L_h \sum_{j=0}^{N-1} |\Delta B_j|} \left( L_\xi \mathbb{E}_{i_0} \left[ |X_N^{i_0, x} - X_N^{i_0, x'}| \right] \right. \\ &\left. + \sum_{k=i_0}^{N-1} \{ L_f \mathbb{E}_{i_0} \left[ |X_k^{i_0, x} - X_k^{i_0, x'}| \right] \Delta_k + L_h \mathbb{E}_{i_0} \left[ |X_{k+1}^{i_0, x} - X_{k+1}^{i_0, x'}| \right] |\Delta B_k| \} \right). \end{aligned}$$

Since the coefficients  $b$  and  $\sigma$  are globally Lipschitz, there exists a non negative constant  $C_{(25)}$  such that

$$\sup_{i_0 \leq k \leq N} \mathbb{E}_i \left[ |X_k^{i_0, x} - X_k^{i_0, x'}| \right] \leq C_{(25)} |x - x'|. \quad (25)$$

By the last estimation, we conclude immediately.  $\square$

### 3.4 Convergence of the algorithm and complexity

We are now in a position to study in details the convergence of the Algorithm 1, by choosing appropriately the approximation space  $\mathcal{K}_{Y,i}^{\Delta B}$  and the number of simulations  $M_i^{\Delta B}$ . We are to handle the analysis conditionally to  $\Delta B$ . To simplify the presentation,  $\Delta_i = \text{Cst} = T/N$  and  $\mathcal{K}_{Y,i}^{\Delta B}$  will not depend on  $i$ ; nevertheless less restrictive investigations are possible in light of the general estimates of Theorem 2.

**Approximation spaces.** Since the unknown function  $y_i(\Delta B, \cdot)$  is Lipschitz (Proposition 3), it is enough to consider piecewise approximations. Let  $D^{\Delta B}$  be a large hypercube of  $\mathbb{R}^d$  centered on  $X_0 = x$ , that is  $D^{\Delta B} = \prod_{k=1}^d (x_k - H^{\Delta B}, x_k + H^{\Delta B}]$  for some parameter  $H^{\Delta B}$  large enough. Then,  $D^{\Delta B}$  can be partitioned in a finite number of small hypercubes  $C_{j_1, \dots, j_d}^{\Delta B}$  of edge  $\rho^{\Delta B} > 0$  i.e.

$$D^{\Delta B} = \bigcup_{j_1, \dots, j_d} C_{j_1, \dots, j_d}^{\Delta B}$$

where

$$C_{j_1, \dots, j_d}^{\Delta B} = \prod_{k=1}^d (x_k - H^{\Delta B} + j_k \rho^{\Delta B}, x_k - H^{\Delta B} + (j_k + 1) \rho^{\Delta B}]$$

and  $j_k \in \{0, \dots, \frac{2H^{\Delta B}}{\rho^{\Delta B}} - 1\}$ . To simplify the exposition we neglect the rounding effect by assuming  $\frac{2H^{\Delta B}}{\rho^{\Delta B}}$  is integer. The number of hypercubes is  $(2H^{\Delta B}/\rho^{\Delta B})^d$ , which equals  $\text{card}(\mathcal{K}_{Y,i}^{\Delta B})$  since on each hypercube, the approximation is piecewise constant.

Recall that under **(H1)**,  $X_i$  has finite moments at any order, i.e. for any  $q > 0$

$$\sup_{0 \leq i \leq N} \mathbb{E}[|X_i - x|_\infty^q] \leq C_{q,(26)} \quad (26)$$

for a constant independent of  $N$ . With Proposition 3 at hand, we easily upper bound the squared approximation error as follows:

$$\begin{aligned} \tau_{1,i,M}^{\Delta B,Y} &\leq \tau_{1,i}^{\Delta B,Y} := \inf_{\phi \in \mathcal{K}_{Y,i}^{\Delta B}} \mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, X_i) - \phi(X_i) \right|^2 \right] \\ &\leq \mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, X_i) \right|^2 1_{\{X_i \notin D^{\Delta B}\}} \right] \\ &\quad + \sum_{j_1, \dots, j_d} \mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, X_i) - y_i(\Delta B, x_{j_1, \dots, j_d}) \right|^2 1_{\{X_i \in C_{j_1, \dots, j_d}^{\Delta B}\}} \right] \end{aligned}$$

for an arbitrary point  $x_{j_1, \dots, j_d}$  in the hypercube  $C_{j_1, \dots, j_d}^{\Delta B}$

$$\leq (C_y^{\Delta B})^2 C_{2,(26)} (H^{\Delta B})^{-2} + (C_{(24)}^{\Delta B} \rho^{\Delta B})^2$$



using the Markov inequality with  $q = 2$  for the first term and the Lipschitz property of  $y_i(\Delta B, \cdot)$  for the second. To get a squared approximation error  $\tau_{1,i,M}^{\Delta B, Y}$  of order  $N^{-1}$  (in coherence with Theorem 1), it is enough to choose

$$H^{\Delta B} = C_y^{\Delta B} \sqrt{N}, \quad \rho^{\Delta B} = \frac{1}{C_{(24)}^{\Delta B} \sqrt{N}}.$$

However, this is not sufficient to contribute in  $\eta_{i,M}^{Y, \Delta B}$  with an error of magnitude  $N^{-1/2}$  because of the summation over  $k$  in (21). An appropriate choice is

$$H^{\Delta B} = e^{c \sum_{k=0}^{N-1} |\Delta B_k|} N^{3/2}, \quad \rho^{\Delta B} = e^{-c \sum_{k=0}^{N-1} |\Delta B_k|} N^{-3/2}.$$

For  $c$  large enough (and explicit w.r.t. model data), this shows that the  $\tau_{1,k,M}^{\Delta B, Y}$ -errors contribute in  $\eta_{i,M}^{Y, \Delta B}$  as  $CN^{-1/2}$  for a deterministic constant  $C$ . With the above choice, we have

$$\text{card}(\mathcal{K}_{Y,i}^{\Delta B}) = (2e^{2c \sum_{k=0}^{N-1} |\Delta B_k|} N^3)^d.$$

**Number of simulations.** A careful analysis of the upper bound (21) shows that

$$M_i^{\Delta B} := N^{3d+5} \exp\left(c' \sum_{k=0}^{N-1} |\Delta B_k|\right) \quad (27)$$

for  $c'$  large enough implies

$$\eta_{i,M}^{Y, \Delta B} \leq C_{(28)} \sqrt{\frac{\log(N+1)}{N}}, \quad a.s., \quad (28)$$

for some deterministic constant  $C_{(28)} > 0$ . We have proved the first part of the following result.

**Corollary 1 (Convergence of the algorithm)** *For the uniform time grid with  $N$  time steps and for a.s. any discrete path  $(\Delta B_k)_{0 \leq k \leq N-1}$ , the empirical regression algorithm with appropriate choices of  $\mathcal{K}_{Y,k}^{\Delta B}$  and  $M_k^{\Delta B}$  yields an error in  $L_2$  conditionally to  $\Delta B$  bounded by  $C_{(28)} \sqrt{\frac{\log(N+1)}{N}}$  where  $C_{(28)}$  is deterministic.*

*Furthermore, the complexity of the algorithm is  $\mathcal{C} \sim N \sum_{i=0}^{N-1} M_i^{\Delta B}$  up to a deterministic constant. Thus,*

- *conditionally on  $\Delta B$ , the complexity is of order*

$$\mathcal{C} = O(N^{3d+7});$$

- *in expectation, the complexity is of order*

$$\mathbb{E}[\mathcal{C}] \approx \tilde{c} N^{3d+7} \exp(\tilde{c} \sqrt{N})$$

*for some  $\tilde{c} > 0$ .*

It remains to justify the second part regarding the complexity  $\mathcal{C}$ . The latter is directly evaluated by counting the elementary operations as in [GT15b]. Then, its evaluation conditionally and unconditionally on  $\Delta B$  are easily obtained in view of (27).

As for simple BSDEs, the curse of dimensionality occurs. But here, the effect of external Brownian motion is seemingly much more determinant, it is responsible for the factors  $\exp\left(c' \sum_{k=0}^{N-1} |\Delta B_k|\right)$  and  $\exp(\tilde{c}\sqrt{N})$ . In other words, the convergence holds, but in average at a logarithmic speed w.r.t. the accuracy. In a pathwise sense, the approximation may be more or less accurate depending on the realization of  $\Delta B$ , which is intuitively meaningful.

### 3.5 Proof of Theorem 2

The following proposition is useful to interchange the errors (19) and (20).

**Proposition 4** *For all  $i \in \{0, \dots, N-1\}$ , we have*

$$\left(\eta_i^{Y, \Delta B}\right)^2 \leq 2 \left(\eta_{i, M}^{Y, \Delta B}\right)^2 + (C_y^{\Delta B})^2 \frac{2028(\text{card}(\mathcal{K}_{Y, i}^{\Delta B}) + 1) \log(3M_i^{\Delta B})}{M_i^{\Delta B}}.$$

**Proof.** Let  $i \in \{0, \dots, N-1\}$ . Write

$$\begin{aligned} & \mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, X_i) - y_i^{(M)}(\Delta B, X_i) \right|^2 \right] \leq 2 \mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, \cdot) - y_i^{(M)}(\Delta B, \cdot) \right|_{\nu_{i, M}}^2 \right] \\ & + \mathbb{E}_{\Delta B} \left[ \left( \mathbb{E} \left[ \left| y_i(\Delta B, X_i) - y_i^{(M)}(\Delta B, X_i) \right|^2 \mid \Delta B, \{\mathcal{C}_k : k \geq i\} \right] \right. \right. \\ & \quad \left. \left. - 2 \left| y_i(\Delta B, \cdot) - y_i^{(M)}(\Delta B, \cdot) \right|_{\nu_{i, M}}^2 \right)_+ \right], \end{aligned}$$

and apply Proposition 5, with  $p = 2$ ,  $\lambda = C_y^{\Delta B}$  (owing to Proposition 2),  $\mathcal{K} = \mathcal{K}_{Y, i}^{\Delta B}$ ,  $M = M_i^{\Delta B}$ .  $\square$

To prove Theorem 2, we need few extra notations.

1. We define the following  $\sigma$ -fields  $\mathcal{G}_i^* := \mathcal{F}^{\Delta B} \vee \sigma(\mathcal{C}_{i+1}, \dots, \mathcal{C}_{N-1})$  and  $\mathcal{G}_i^{i, 1:M} := \mathcal{G}_i^* \vee \sigma(X_i^{(i, m)} : m \geq 1)$  for all  $i = 0, \dots, N-1$ .
2. We define

$$\psi_i(\Delta B, \cdot) \text{ as the solution of OLS} \left( \mathcal{Y}_i(\Delta B, \cdot), \mathcal{K}_{Y, i}^{\Delta B}, \nu_{i, M} \right),$$

for all  $i = 0, \dots, N-1$ . This is the OLS solution when the functions  $f$  and  $h$  are computed with the right solution, as opposed to the definition (14) of  $\psi_i^{(M)}(\Delta B, \cdot)$ .

We now turn to the proof of Theorem 2. In view of Proposition 2 and the definition (17) of the thresholding operator  $[\cdot]_i$  which is 1-Lipschitz, we have

$$\begin{aligned}\eta_{i,M}^{Y,\Delta B} &= \sqrt{\mathbb{E}_{\Delta B} \left[ \left| \left[ y_i(\Delta B, \cdot) \right]_i - \left[ \psi_i^{(M)}(\Delta B, \cdot) \right]_i \right|_{\nu_{i,M}}^2 \right]} \\ &\leq \sqrt{\mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, \cdot) - \psi_i^{(M)}(\Delta B, \cdot) \right|_{\nu_{i,M}}^2 \right]}.\end{aligned}$$

Then, inserting the  $\mathcal{G}_i^{i,1:M}$ -conditional expectations  $\mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i(\Delta B, \cdot)]$  and  $\mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i^{(M)}(\Delta B, \cdot)]$  and using the triangle inequality, we get

$$\begin{aligned}\eta_i^{Y,\Delta B} &\leq \sqrt{\mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, \cdot) - \mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i(\Delta B, \cdot)] \right|_{\nu_{i,M}}^2 \right]} \quad (29) \\ &\quad + \sqrt{\mathbb{E}_{\Delta B} \left[ \left| \mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i^{(M)}(\Delta B, \cdot)] - \psi_i^{(M)}(\Delta B, \cdot) \right|_{\nu_{i,M}}^2 \right]} \\ &\quad + \sqrt{\mathbb{E}_{\Delta B} \left[ \left| \mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i(\Delta B, \cdot)] - \psi_i^{(M)}(\Delta B, \cdot) \right|_{\nu_{i,M}}^2 \right]}.\end{aligned}$$

We separately deal with each term of the previous inequality.

**a) Term  $\sqrt{\mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, \cdot) - \mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i(\Delta B, \cdot)] \right|_{\nu_{i,M}}^2 \right]}$  in (29).** We note that for all  $m$  in  $\{1, \dots, M_i^{\Delta B}\}$ ,  $\mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\mathcal{Y}_i(\Delta B, X_i^{(i,m)})] = y_i(\Delta B, X_i^{(i,m)})$ . It follows from Proposition 6-(iii) that  $\mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i(\Delta B, \cdot)]$  solves OLS  $(y_i(\Delta B, \cdot), \mathcal{K}_{Y,i}^{\Delta B}, \nu_{i,M})$ , that is

$$\mathbb{E}_{\Delta B} \left[ \left| y_i(\Delta B, \cdot) - \mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i(\Delta B, \cdot)] \right|_{\nu_{i,M}}^2 \right] = \tau_{1,i,M}^{\Delta B, Y} \quad (30)$$

where  $\tau_{1,i,M}^{\Delta B, Y}$  is defined in (18).

**b) Term  $\sqrt{\mathbb{E}_{\Delta B} \left[ \left| \mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i^{(M)}(\Delta B, \cdot)] - \psi_i^{(M)}(\Delta B, \cdot) \right|_{\nu_{i,M}}^2 \right]}$  in (29).** The terms  $\mathcal{Y}_i^{(M)}(\Delta B, \cdot)$  are computed only using the clouds  $\{\mathcal{C}_k, k \geq i+1\}$ . Thus, we obtain by Proposition 6-(iv)

$$\mathbb{E}_{\Delta B} \left[ \left| \psi_i^{(M)}(\Delta B, \cdot) - \mathbb{E}_{\mathcal{G}_i^{i,1:M}} [\psi_i^{(M)}(\Delta B, \cdot)] \right|_{\nu_{i,M}}^2 \right] \leq \frac{\text{card}(\mathcal{K}_{Y,i}^{\Delta B})}{M_i^{\Delta B}} \sigma_{\mathcal{Y}_i}^2(\Delta B), \quad (31)$$

where  $\sigma_{\mathcal{Y}_i}^2(\Delta B)$  is a  $\mathcal{F}^{\Delta B}$ -measurable random variable bounding the conditional variance  $\text{Var}(\mathcal{Y}_i^{(M)}(\Delta B, X_{i:N}) \mid \Delta B, \sigma(\mathcal{C}_{i+1}, \dots, \mathcal{C}_{N-1}), X_i = x_i)$  uniformly in  $x_i$ . A simple bound of the latter is  $\sup_{x_{i:N}} |\mathcal{Y}_i^{(M)}(\Delta B, x_{i:N})|^2$  which

we now make explicit: starting from the definition (15), since  $y_i^{(M)}(\Delta B, \cdot)$  is bounded by  $C_y^{\Delta B}$  (see (16) and (17)) and in view of the **Assumption (H2)**, we easily derive that we can take  $\sigma_{\mathcal{Y}_i}(\Delta B)$  as defined in (23).

**c) Term**  $\sqrt{\mathbb{E}_{\Delta B} \left[ \left| \mathbb{E}_{\mathcal{G}_i^{i,1:M}} \left[ \psi_i(\Delta B, \cdot) - \psi_i^{(M)}(\Delta B, \cdot) \right] \right|_{\nu_{i,M}}^2 \right]}$  in (29).

We set  $\mathcal{E}_{Y,i}^M(\Delta B, x_i) := \mathbb{E} \left[ \mathcal{Y}_i(\Delta B, X_{i:N}) - \mathcal{Y}_i^{(M)}(\Delta B, X_{i:N}) \mid X_i = x_i, \mathcal{G}_i^* \right]$ . As  $\mathcal{Y}_i(\Delta B, X_{i:N}) - \mathcal{Y}_i^{(M)}(\Delta B, X_{i:N})$  are computed only with the clouds  $\{\mathcal{C}_k, k \geq i+1\}$ , we have for all  $m$

$$\mathbb{E}_{\mathcal{G}_i^{i,1:M}} \left[ \mathcal{Y}_i(\Delta B, X_{i:N}^{(i,m)}) - \mathcal{Y}_i^{(M)}(\Delta B, X_{i:N}^{(i,m)}) \right] = \mathcal{E}_{Y,i}^M(\Delta B, X_i^{(i,m)}).$$

Thus, by Proposition 6-(i-iii),  $\mathbb{E}_{\mathcal{G}_i^{i,1:M}} \left[ \psi_i(\Delta B, \cdot) - \psi_i^{(M)}(\Delta B, \cdot) \right]$  solves OLS  $\left( \mathcal{E}_{Y,i}^M(\Delta B, \cdot), \mathcal{K}_{Y,i}^{\Delta B}, \nu_{i,M} \right)$ . Using Proposition 6-(ii) (the norm stability property of the OLS operator), we get

$$\begin{aligned} \mathbb{E}_{\Delta B} \left[ \left| \mathbb{E}_{\mathcal{G}_i^{i,1:M}} \left[ \psi_i(\Delta B, \cdot) - \psi_i^{(M)}(\Delta B, \cdot) \right] \right|_{\nu_{i,M}}^2 \right] &\leq \mathbb{E}_{\Delta B} \left[ \left| \mathcal{E}_{Y,i}^M(\Delta B, \cdot) \right|_{\nu_{i,M}}^2 \right] \\ &= \mathbb{E}_{\Delta B} \left[ \left| \mathcal{E}_{Y,i}^M(\Delta B, X_i) \right|^2 \right] \end{aligned}$$

where we use at the last equality the independence (conditionally on  $\mathcal{F}^{\Delta B}$ ) between clouds of simulations. Using the triangle inequality on the conditional  $L_2$ -norm and the definitions of (13) and (15), we get

$$\begin{aligned} &\sqrt{\mathbb{E}_{\Delta B} \left[ \left| \mathcal{E}_{Y,i}^M(\Delta B, X_i) \right|^2 \right]} \\ &\leq \sum_{k=i}^{N-2} (L_f \Delta_k + L_h |\Delta B_k|) \sqrt{\mathbb{E}_{\Delta B} \left[ \left| y_{k+1}(\Delta B, X_{k+1}) - y_{k+1}^{(M)}(\Delta B, X_{k+1}) \right|^2 \right]}. \end{aligned}$$

To summarise, by plugging the last estimation and the estimations (30) and (31) in the inequality (29), we have

$$\eta_{i,M}^{Y,\Delta B} \leq \left( \tau_{1,i,M}^{\Delta B,Y} \right)^{\frac{1}{2}} + \left( \frac{\text{card}(\mathcal{K}_{Y,i}^{\Delta B})}{M_i^{\Delta B}} \right)^{\frac{1}{2}} \sigma_{\mathcal{Y}_i}(\Delta B) + \sum_{k=i}^{N-2} (L_f \Delta_k + L_h |\Delta B_k|) \eta_{k+1}^{Y,\Delta B}.$$

Thanks to Proposition 4, the above conditional  $L_2$ -norms  $\eta_{k+1}^{Y,\Delta B}$  can be expressed in terms of  $\eta_{k+1,M}^{Y,\Delta B}$ : namely

$$\eta_{k+1}^{Y,\Delta B} \leq \sqrt{2} \eta_{k+1,M}^{Y,\Delta B} + C_y^{\Delta B} \sqrt{\frac{2028(\text{card}(\mathcal{K}_{Y,k+1}^{\Delta B}) + 1) \log(3M_{k+1}^{\Delta B})}{M_{k+1}^{\Delta B}}}.$$

Therefore, we have proved

$$\eta_{i,M}^{Y,\Delta B} \leq \delta_i + \sum_{k=i}^{N-2} \sqrt{2}(L_f \Delta_k + L_h |\Delta B_k|) \eta_{k+1,M}^{Y,\Delta B}, \quad (32)$$

where  $\delta_i$  is defined in (22). By setting  $C_{i+1} := \sqrt{2}(L_f \Delta_i + L_h |\Delta B_i|)$ , (32) writes under the form

$$\eta_{i,M}^{Y,\Delta B} \leq \delta_i + \sum_{k=i+1}^{N-1} C_k \eta_{k,M}^{Y,\Delta B}, \quad (33)$$

i.e. , the error  $\eta_{i,M}^{Y,\Delta B}$  is bounded by a local error term  $\delta_i$  and the sum of the errors  $\eta_{k,M}^{Y,\Delta B}$  arising before the steps  $k > i$ . This inequality is a discrete Gronwall type inequality, which leads to the following upper bound for all  $i$  in  $\{0, \dots, N-1\}$

$$\eta_{i,M}^{Y,\Delta B} \leq \delta_i + \sum_{k=i+1}^{N-1} \Gamma_{i,k} C_k \delta_k, \quad (34)$$

where

$$\Gamma_{i,k} := \begin{cases} \prod_{i < j < k} (1 + C_j) \leq \exp \left( \sqrt{2} L_f T + \sqrt{2} L_h \sum_{k=i}^{N-1} |\Delta B_k| \right), & \text{for } i+1 < k, \\ 1, & \text{otherwise.} \end{cases} \quad (35)$$

The proof of (34) is postponed to Appendix A.3, while the proof of the upper bound in (35) can be undertaken similarly to (10). Making explicit the terms in (34) readily leads to (21), we are done.  $\square$

## 4 Numerical tests

In this section, we aim to test the algorithm on an example where we know the exact solution. In any case, since the solution is random (through  $B$ ), checking the accuracy is challenging and our theoretical results of Theorem 2 are useful to assert the convergence in general.

We take a linear BDSDE of the form

$$\begin{cases} dX_t = X_t(\mu dt + \sigma dW_t), & X_0 = x, \\ \Phi(x) \text{ given, } f(y) = a_0 y, \quad h(y) = b_0 y, \end{cases}$$

which is written in a one-dimensional setting ( $d = l = 1$ ). From [PP94, Remark 3.4], we know that the solution is given by

$$Y_t^{exact} = \mathbb{E} \left[ \Phi(X_T) e^{a_0(T-t) + b_0(B_T - B_t) - \frac{1}{2} b_0^2 (T-t)} \mid \mathcal{F}_t^0 \right]$$

$$= \mathbb{E} [\Phi(X_T) \mid X_t] e^{a_0(T-t)+b_0(B_T-B_t)-\frac{1}{2}b_0^2(T-t)}.$$

To allow an explicit computation of the above expectation, we take  $\Phi(x) = -x + K$ . The parameters are set to  $K = 115$ ,  $x = 100$ ,  $\mu = 0.05$ ,  $\sigma = 0.2$ ,  $T = 0.25$ ,  $a_0 = b_0 = 0.5$ .

The domain of resolution is  $D^{\Delta B} = [60, 200]$ . We use piecewise approximations on hypercubes as explained in Subsection 3.4. Below we report the values of the error  $y_0^{(M)}(\Delta B, x) - Y_0^{\text{exact}}$  at time  $t = 0$  and additionally, we provide the standard deviation  $\sigma^M$  of  $y_0^{(M)}(\Delta B, x)$  (conditionally to  $\Delta B$ ) by performing 50 independent macro-runs of the empirical regression algorithms (with the same  $B$ ). The value  $\bar{y}_0^{(M)}(\Delta B, x)$  stands for the mean over the macro-runs.

The following tables present the results for different choices of  $\rho^{\Delta B}$  and  $M_i^{\Delta B}$ . The relative error is  $\left| \frac{\bar{y}_0^{(M)}(\Delta B, x) - Y_0^{\text{exact}}}{Y_0^{\text{exact}}} \right|$ .

Let  $N = 20$ : a first simulation of  $B$  gives  $Y_0^{\text{exact}} = 13.724$ . Then, by taking  $\rho^{\Delta B} = 1$ , we obtain

| $M_i^{\Delta B}$ | $\bar{y}_0^{(M)}(\Delta B, x)(\sigma^M)$ | relative error |
|------------------|--|----------------|
| 100              | 13.910(1.178)                            | 1.3%           |
| 1000             | 13.792(0.309)                            | 0.4%           |
| 5000             | 13.847(0.117)                            | 0.8%           |

Now take  $N = 30$  and another simulation of  $B$  yields  $Y_0^{\text{exact}} = 14.115$ . By setting  $\rho^{\Delta B} = 0.5$  we obtain the following results.

| $M_i^{\Delta B}$ | $\bar{y}_0^{(M)}(\Delta B, x)(\sigma^M)$ | relative error |
|------------------|--|----------------|
| 100              | 14.246(1.045)                            | 0.9%           |
| 1000             | 14.195(0.337)                            | 0.5%           |
| 5000             | 14.236(0.129)                            | 0.8%           |

As predicted by our convergence analysis, we observe an improvement of accuracy when  $\rho^{\Delta B} \rightarrow 0$  and  $M_i^{\Delta B} \rightarrow +\infty$ . Additional numerical experiments are available in [Bac14], also including  $z$  in the driver.

## A Appendix

### A.1 Concentration-of-measure inequalities

We state an upper bound result, for a sample deviation, uniformly on the function spaces.

**Proposition 5** *Let  $n \geq 1$ ,  $p \geq 1$  and  $\mathcal{K}$  be a finite dimensional vector space of measurable functions from  $\mathbb{R}^n$  to  $\mathbb{R}$ . Let  $\lambda \in (0, +\infty)$  be a threshold,  $\vartheta$  a measurable function from  $\mathbb{R}^n$  to  $[-\lambda, \lambda]$  and set*

$$\mathcal{K}_\lambda := \{|\tau_\lambda \phi(\cdot) - \vartheta(\cdot)|^p : \phi(\cdot) \in \mathcal{K}\},$$

where  $\tau_\lambda \phi(\cdot) := -\lambda \vee \phi(\cdot) \wedge \lambda$ . Then, for any  $M \geq 1$  and any  $\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(M)}$  i.i.d. random variables distributed as a  $\mathbb{R}^n$ -valued random variable  $\mathcal{X}$ , we have

$$\begin{aligned} \mathbb{E} \left( \sup_{g \in \mathcal{K}_\lambda} \left( \int_{\mathbb{R}^n} g(x) \mathbb{P} \circ \mathcal{X}^{-1}(\mathrm{d}x) - \frac{2}{M} \sum_{m=1}^M g(\mathcal{X}^{(m)}) \right)_+ \right) \\ \leq 507p2^{p-1}\lambda^p \frac{(\mathrm{card}(\mathcal{K}) + 1) \log(3M)}{M}. \end{aligned}$$

**Proof.** Apply [GT15b, Proposition 4.9] with  $\psi(x) = (|x| \wedge 2\lambda)^p$  which is Lipschitz continuous with Lipschitz constant equal to  $p(2\lambda)^{p-1}$ .  $\square$

## A.2 OLS: deterministic and probabilistic properties

For the sake of completeness, we recall a result proved in [GT15b, Proposition 4.12], gathering the important properties of the least-squares operator, and adapted to our setting.

**Proposition 6** *Assume the notation of Definition 1 with a finite dimensional vector subspace  $\mathcal{K}$  and consider the second case with discrete probability measure  $\nu_M$ . Let  $S^*$  solve  $\mathrm{OLS}(S, \mathcal{K}, \nu_M)$ , according to (11). The following properties are satisfied.*

- (i) **LINEARITY:** *The mapping  $S \rightarrow S^*$  is linear.*
- (ii) **EMPIRICAL  $L_2$ -NORM STABILITY:**  $|S^*(\tilde{\omega}, \Delta B, \cdot)|_{\nu_M} \leq |S(\tilde{\omega}, \Delta B, \cdot)|_{\nu_M}$ .
- (iii) **CONDITIONAL EXPECTATION SOLUTION:** *Assume additionally that  $\mathcal{Q}$  is a  $\sigma$ -field such that  $\{p^j(\Delta B, \mathcal{X}^{(m)}), j \geq 1, m \geq 1\}$  is  $\mathcal{F}^{\Delta B} \vee \mathcal{Q}$ -measurable. Setting  $S_{\mathcal{Q}}(\Delta B, \mathcal{X}^{(m)}) = \mathbb{E}[S(\tilde{\omega}, \Delta B, \mathcal{X}^{(m)}) \mid \mathcal{F}^{\Delta B} \vee \mathcal{Q}]$  for each  $m \in \{1, \dots, M^{\Delta B}\}$ , then  $\mathbb{E}[S^* \mid \mathcal{F}^{\Delta B} \vee \mathcal{Q}]$  solves  $\mathrm{OLS}(S_{\mathcal{Q}}, \mathcal{K}, \nu_M)$ .*
- (iv) **BOUNDED CONDITIONAL VARIANCE:** *Suppose first that  $S(\tilde{\omega}, \Delta B, x)$  is  $\tilde{\mathcal{G}} \otimes \mathcal{F}^{\Delta B} \otimes \mathcal{B}(\mathbb{R}^n)$ -measurable for a  $\sigma$ -field  $\tilde{\mathcal{G}} \subset \tilde{\mathcal{F}}$  independent of  $\sigma(\mathcal{X}^{(m)}, m \geq 1)$ , second that there exists a Borel measurable function  $\kappa : \mathbb{R}^n \rightarrow E$ , for some Euclidean space  $E$ , such that the random variables  $\{p^j(\Delta B, \mathcal{X}^{(m)}), m \geq 1, j \geq 1\}$  are  $\mathcal{F}^{\Delta B} \vee \mathcal{H}$ -measurable with  $\mathcal{H} := \sigma(\kappa(\mathcal{X}^{(m)}), m \geq 1)$ , and third that there exists a finite  $\mathcal{F}^{\Delta B}$ -random variable  $\zeta(\Delta B) \geq 0$  that uniformly bounds the conditional variances*

$$\mathrm{Var}(S(\tilde{\omega}, \Delta B, \mathcal{X}^{(m)}) \mid \mathcal{F}^{\Delta B} \vee \tilde{\mathcal{G}} \vee \mathcal{H}) \leq \zeta(\Delta B) \quad \tilde{\mathbb{P}} \otimes \mathbb{P}^{\Delta B} - a.s.$$

for all  $m \geq 1$ . Then

$$\mathbb{E} \left[ |S^*(\cdot) - \mathbb{E}[S^*(\cdot) \mid \mathcal{F}^{\Delta B} \vee \tilde{\mathcal{G}} \vee \mathcal{H}]|_{\nu_M}^2 \mid \mathcal{F}^{\Delta B} \vee \tilde{\mathcal{G}} \vee \mathcal{H} \right] \leq \frac{\mathrm{card}(\mathcal{K})}{M} \zeta(\Delta B).$$

### A.3 Proof of inequality (34)

The proof is made by backward induction on  $i \in \{0, \dots, N-1\}$ .

- For  $i = N-1$ , by inequality (33) we check that  $\eta_{N-1,M}^{Y,\Delta B} \leq \delta_{N-1}$ . Thus the estimation (34) holds.
- Now, let  $i \in \{0, \dots, N-1\}$ , assume that (34) is true for  $\{i, \dots, N-1\}$  and let us prove that it holds for  $\{i-1, \dots, N-1\}$ . Using (33) and (34), we get

$$\begin{aligned}
 \eta_{i-1,M}^{Y,\Delta B} &\leq \delta_{i-1} + \sum_{k=i}^{N-1} C_k \eta_{k,M}^{Y,\Delta B} \\
 &\leq \delta_{i-1} + \sum_{k=i}^{N-1} C_k \left( \delta_k + \sum_{s=k+1}^{N-1} \Gamma_{k,s} C_s \delta_s \right) \\
 &= \delta_{i-1} + \sum_{k=i}^{N-1} C_k \delta_k + \sum_{s=i+1}^{N-1} \sum_{k=i}^{s-1} C_k \Gamma_{k,s} C_s \delta_s \\
 &= \delta_{i-1} + C_i \delta_i + \sum_{k=i+1}^{N-1} C_k \delta_k \left\{ 1 + \sum_{s=i}^{k-1} C_s \Gamma_{s,k} \right\}.
 \end{aligned}$$

Owing to  $\Gamma_{i-1,k} + \Gamma_{i-1,k} C_{i-1} = \Gamma_{i-2,k}$  and  $\Gamma_{k-1,k} = 1$ , observe that

$$1 + \sum_{s=i}^{k-1} C_s \Gamma_{s,k} = \Gamma_{i-1,k},$$

and therefore

$$\eta_{i-1,M}^{Y,\Delta B} \leq \delta_{i-1} + C_i \delta_i + \sum_{k=i+1}^{N-1} C_k \delta_k \Gamma_{i-1,k}.$$

Recalling that  $\Gamma_{i-1,i} = 1$ , we conclude that (34) holds for  $i-1$ .

## References

- [Abo09] O. Aboura. On the discretization of backward doubly stochastic differential equations. *Arxiv:0907.1406v1*, 2009.
- [Abo11] O. Aboura. A regression Monte-Carlo method for Backward Doubly Stochastic Differential Equations. *HAL preprint hal-00607274*, 2011.
- [Ama13] A. Aman. A numerical scheme for backward doubly stochastic differential equations. *Bernoulli*, 19(1):93–114, 2013.
- [Bac14] A. Bachouch. *Numerical computations for Backward Doubly Stochastic Differential Equations and Nonlinear Stochastic PDEs*. PhD thesis, University Le Mans, 2014.



- [BBMM13] A. Bachouch, M.A. Ben Lasmer, A. Matoussi, and M. Mnif. Numerical scheme for semilinear stochastic PDEs via Backward Doubly Stochastic Differential Equations. *Arxiv: 1302.0440 v3. In revision*, 2013.
- [BM01] V. Bally, V. and A. Matoussi. Weak solutions for SPDEs and backward doubly stochastic differential equations. *J. Theoret. Probab.*, 14(1):125-164, 2001.
- [BuM01] R. Buckdahn and J. Ma. Stochastic viscosity solutions for nonlinear stochastic partial differential equations. I. *Stochastic Process. Appl.*, 93(2):181-204, 2001.
- [GLW05] E. Gobet, J-P. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Annals of Applied Probability*, 15(3):2172–2202, 2005.
- [GT15a] E. Gobet and P. Turkedjiev. Approximation of BSDEs using Malliavin weights and least-squares regression. *To appear in Bernoulli*, 2015.
- [GT15b] E. Gobet and P. Turkedjiev. Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions. *To appear in Mathematics of Computation*, 2015.
- [GK10] I. Gyongy and N. Krylov. Accelerated finite difference schemes for linear stochastic partial differential equations in the whole space. *SIAM J.Math.Anal.*,42:2275-2296, 2010.
- [GN95] I. Gyongy and D. Nualart. Implicit scheme for quasi-linear parabolic partial differential equations perturbed by space-time white noise. *Stochastic Processes and Their applications*, 58:57-72, 1995.
- [G99] I. Gyongy. Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by a space-time white noise I. *Potential Anal.*, 11:1-37, 1999.
- [JK10] A. Jentzen and P. Kloeden. Taylor Expansions of solutions of stochastic partial differential equation with additive noise. *The Annals of Probability*, 38(2): 532-569, 2010.
- [K99] N.V. Krylov. An analytic approach to SPDEs. In Stochastic partial differential equations: six perspectives, *Math. Surveys Monogr. Amer. Math. Soc., Providence, RI*, vol. 64:185-242, 1999.
- [LGW06] J-P. Lemor, E. Gobet, and X. Warin. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5):889–916, 2006.

- [LS98] P.-L. Lions and P.E. Souganidis. Fully nonlinear viscosity stochastic partial differential equations: non-smooth equations and applications. *C.R. Acad. Sci. Paris*, 327(1):735-741, 1998.
- [LS02] P.-L. Lions and P.E. Souganidis. Viscosity solutions of fully nonlinear stochastic partial differential equations. *Sūrikaisekikenkyūsho Kōkyūroku*, 1287: 58-65, 2002.
- [MS02] A. Matoussi and M. Scheutzw. Stochastic PDEs driven by nonlinear noise and backward doubly SDEs. *J. Theoret. Probab.*, 15(1):1-39, 2002.
- [PP94] E. Pardoux and S.G. Peng. Backward doubly stochastic differential equations and systems of quasilinear SPDEs. *Probability Theory and Related Fields*, 98(2):209–227, 1994.
- [SYY08] Y. Shi, W. Yang, and J. Yuan. Numerical computations for Backward Doubly SDEs and SPDEs. *arXiv preprint arXiv:0805.4662*, 2008.
- [W05] J.B. Walsh. Finite Elements Methods for Parabolic Stochastic PDE's. *Potential Analysis*, 23:1-43, 2005.