Numerical approximation of BSDEs using local polynomial drivers and branching processes

Bruno Bouchard*† Xiaolu Tan*‡ Yiyi Zou*§

December 20, 2016

Abstract

We propose a new numerical scheme for Backward Stochastic Differential Equations based on branching processes. We approximate an arbitrary (Lipschitz) driver by local polynomials and then use a Picard iteration scheme. Each step of the Picard iteration can be solved by using a representation in terms of branching diffusion systems, thus avoiding the need for a fine time discretization. In contrast to the previous literature on the numerical resolution of BSDEs based on branching processes, we prove the convergence of our numerical scheme without limitation on the time horizon. Numerical simulations are provided to illustrate the performance of the algorithm.

Keywords: Bsde, Monte-Carlo methods, branching process.

MSC2010: Primary 65C05, 60J60; Secondary 60J85, 60H35.

*Université Paris-Dauphine, PSL Research University, CNRS, UMR [7534], CEREMADE, 75016 PARIS, FRANCE.

†bouchard@ceremade.dauphine.fr

‡tan@ceremade.dauphine.fr

§zou@ceremade.dauphine.fr

1 Introduction

Since the seminal paper of Pardoux and Peng [15], the theory of Backward Stochastic Differential Equations (BSDEs hereafter) has been largely developed, and has lead to many applications in optimal control, finance, etc. (see e.g. El Karoui, Peng and Quenez [8]). Different approaches have been proposed during the last decade to solve them numerically, without relying on pure PDE based resolution methods. A first family of numerical schemes, based on a time discretization technique, has been introduced by Bally and Pagès [2], Bouchard and Touzi [5] and Zhang [21], and generated a large stream of the literature. The implementation of these numerical schemes requires the estimation of a sequence of conditional expectations, which can be done by using simulations combined with either non-linear regression techniques or Malliavin integration by parts based representations of conditional expectations, or by using a quantization approach, see e.g. [6, 9] for references and error analysis.

Another type of numerical algorithms is based on a pure forward simulation of branching processes, and was introduced by Henry-Labordère [10], and Henry-Labordère, Tan and Touzi [12] (see also the recent extension by Henry-Labordère et al. [11]). The main advantage of this new algorithm is that it avoids the estimation of conditional expectations. It relies on the probabilistic representation in terms of branching processes of the so-called KPP (Kolmogorov-Petrovskii-Piskunov) equation:

$$\partial_t u(t,x) \;+\; \frac{1}{2}D^2 u(t,x) \;+\; \sum_{k\geq 0} p_k u^k(t,x) \;=\; 0,\; u(T,x) = g(x). \tag{1.1}$$

Here, D^2 is the Laplacian on $\mathbb{R}^d$, and $(p_k)_{k\geq 0}$ is a probability mass sequence, i.e. $p_k \geq 0$ and $\sum_{k\geq 0} p_k = 1$. This is a natural extension of the classical Feynmann-Kac formula, which is well known since the works of Skorokhod [17], Watanabe [20] and McKean [14], among others. The PDE (1.1) corresponds to a BSDE with a polynomial driver and terminal condition $g(W_T)$:

$$Y_\cdot = g(W_T) + \int_\cdot^T \sum_{k\geq 0} p_k (Y_t)^k dt - \int_\cdot^T Z_t dW_t,$$

in which W is a Brownian motion. Since $Y_\cdot = u(\cdot, W_\cdot)$, the Y-component of this BSDE can be estimated by making profit of the branching process based

Feynman-Kac representation of (1.1) by means of a pure forward Monte-Carlo scheme, see Section 2.3 below. The idea is not new. It was already proposed in Rasulov, Raimov and Mascagni [16], although no rigorous convergence analysis was provided. Extensions to more general drivers can be found in [10, 11, 12]. Similar algorithms have been studied by Bossy et al. [4] to solve non-linear Poisson-Boltzmann equations.

It would be tempting to use this representation to solve BSDEs with Lipschitz drivers, by approximating their drivers by polynomials. This is however not feasible in general. The reason is that PDEs (or BSDEs) with polynomial drivers, of degree bigger or equal to two, typically explode in finite time. They are only well posed on a small time interval. It is worse when the degree of the polynomial increases. Hence, no convergence can be expected for the case of general drivers.

In this paper, we propose to instead use a local polynomial approximation. Then, convergence of the sequence of approximating drivers to the original one can be ensured without explosion of the corresponding BSDEs, that can be defined on a arbitrary time interval. It requires to be combined with a Picard iteration scheme, as the choice of the polynomial form will depend on the position in space of the solution Y itself. However, unlike classical Picard iteration schemes for BSDEs, see e.g. Bender and Denk [3], we do not need to have a very precise estimation of the whole path of the solution at each Picard iteration. Indeed, if local polynomials are fixed on a partition $(A_i)_i$ of $\mathbb{R}$, then one only needs to know in which A_i the solution stays at certain branching times of the underlying branching process. If the A_i's are large enough, this does not require a very good precision in the intermediate estimations. We refer to Remark 2.3 for more details.

We finally insist on the fact that our results will be presented in a Markovian context for simplification. However, all of our arguments work trivially in a non-Markovian setting too.

2 Numerical method for a class of BSDE based on branching processes

Let $T > 0$, W be a standard d-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and X be the solution of the stochastic differential equation:

$$X = X_0 + \int_0^{\cdot} \mu(X_s)\, dt + \int_0^{\cdot} \sigma(X_s)\, dW_s, \tag{2.1}$$

where X_0 is a constant, lying in a compact subset $\mathbf{X}$ of $\mathbb{R}^d$, and (μ, σ) : $[0, T] \times \mathbb{R}^d \mapsto \mathbb{R}^d \times \mathbb{M}^d$ is assumed to be Lipschitz continuous with support contained in $\mathbf{X}$. Our aim is to provide a numerical scheme for the resolution of the backward stochastic differential equation

$$Y_{\cdot} = g(X_T) + \int_{\cdot}^{T} f(X_s, Y_s)\, ds - \int_{\cdot}^{T} Z_s\, dW_s. \tag{2.2}$$

In the above, $g : \mathbb{R}^d \mapsto \mathbb{R}$ is assumed to be measurable and bounded, $f \in \mathbb{R}^d \times \mathbb{R} \mapsto \mathbb{R}$ is measurable with linear growth and Lipschitz in its second argument, uniformly in the first one. As a consequence, there exists $M \geq 1$ such that

$$|g(X_T)| \leq M \;\text{ and } |X| + |Y| \leq M \;\text{ on } [0, T]. \tag{2.3}$$

Remark 2.1. *The above conditions are imposed to easily localize the solution Y of the BSDE, which will be used in our estimates later on. One could also assume that g and f have polynomial growth in their first component and that $\mathbf{X}$ is not compact. After possibly truncating the coefficients and reducing their support, one would go back to our conditions. Then, standard estimates and stability results for SDEs and BSDEs could be used to estimate the additional error in a very standard way. See e.g. [8].*

2.1 Local polynomial approximation of the generator

A first main ingredient of our algorithm consists in approximating the driver f by a driver $f_{\ell_\circ}$ that has a local polynomial structure. Namely, let

$$f_{\ell_\circ} : (x, y, y') \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \mapsto \sum_{j=1}^{j_\circ} \sum_{\ell=0}^{\ell_\circ} a_{j,\ell}(x) y^\ell \varphi_j(y'), \tag{2.4}$$

in which $(a_{j,\ell}, \varphi_j)_{\ell \leq \ell_\circ, j \leq j_\circ}$ is a family of continuous and bounded maps satisfying

$$|a_{j,\ell}| \leq C_{\ell_\circ}\ , \ |\varphi_j(y'_1) - \varphi_j(y'_2)| \leq L_\varphi |y'_1 - y'_2| \text{ and } |\varphi_j| \leq 1, \tag{2.5}$$

for all $y'_1, y'_2 \in \mathbb{R}$, $j \le j_\circ$ and $\ell \le \ell_\circ$, for some constants $C_{\ell_\circ}, L_\varphi \ge 0$. In the following, we shall assume that $\ell_\circ \ge 2$ (without loss of generality). One can think of the $(a_{j,\ell})_{\ell \le \ell_\circ}$ as the coefficients of a polynomial approximation of f on a subset A_j, the A_j's forming a partition of $[-M, M]$. Then, the φ_j's have to be considered as smoothing kernels that allow one to pass in a Lipschitz way from one part of the partition to another one. We therefore assume that

$$\#\{j \in \{1, \cdots, j_\circ\} : \ \varphi_j(y) > 0\} \le 2 \ \text{ for all } y \in \mathbb{R}, \tag{2.6}$$

and that $y \mapsto f_{\ell_\circ}(x, y, y)$ is globally Lipschitz. In particular,

$$\bar{Y}_{\cdot} = g(X_T) + \int_{\cdot}^{T} f_{\ell_\circ}(X_s, \bar{Y}_s, \bar{Y}_s)\, ds - \int_{\cdot}^{T} \bar{Z}_s\, dW_s, \tag{2.7}$$

has a unique solution $(\bar{Y}, \bar{Z})$ such that $\mathbb{E}[\sup_{[0,T]} |\bar{Y}|^2] < \infty$. Moreover, by standard estimates, $(\bar{Y}, \bar{Z})$ provides a good approximation of (Y, Z) whenever $f_{\ell_\circ}$ is a good approximation of f:

$$\mathbb{E}\Big[\sup_{[0,T]} |Y - \bar{Y}|^2\Big] + \mathbb{E}\Big[\int_0^T |Z_t - \bar{Z}_t|^2 dt\Big] \le C\mathbb{E}\Big[\int_0^T |f - f_{\ell_\circ}|^2(X_t, Y_t, Y_t) dt\Big], \tag{2.8}$$

for some $C > 0$ that does not depend on $f_{\ell_\circ}$, see e.g. [8].
The choice of $f_{\ell_\circ}$ will obviously depend on the application at hand and does not need to be more commented. Let us just mention that our algorithm will be more efficient if the sets $\{y \in \mathbb{R} : \varphi_j(y) = 1\}$ are large and the intersection between the supports of the φ_j's are small, see Remark 2.3 below.
We also assume that

$$|\bar{Y}| \le M. \tag{2.9}$$

Since we intend to keep $f_{\ell_\circ}$ with linear growth in its first component, and bounded in the two other ones, uniformly in $\ell_\circ$, this is without loss of generality.

2.2 Picard iteration with doubly reflected BSDEs

Our next step is to introduce a Picard iteration scheme to approximate the solution $\bar{Y}$ of (2.7). Note however that, although the map $y \mapsto f(x, y, y)$ is globally Lipschitz, the map $y \mapsto f(x, y, y')$ is a polynomial, given y', and hence only locally Lipschitz in general. In order to reduce to a Lipschitz driver, we shall apply our Picard scheme to a doubly (discretely) reflected BSDE, with lower and upper barrier given by the bounds $-M$ and M for $\bar{Y}$, recall (2.9).

Let $h_\circ$ be defined by (A.1) in the Appendix. It is a lower bound for the explosion time of the BSDE with driver $y \mapsto f(x, y, y')$. Let us then fix $h \in (0, h_\circ)$ such that $N_h := T/h \in \mathbb{N}$, and define

$$t_i = ih \quad \text{and} \quad \mathbb{T} := \{t_i,\ i = 0, \cdots, N_h\}. \tag{2.10}$$

We initialize our Picard scheme by setting

$$\bar{Y}_t^0 = \mathrm{y}(t, X_t) \ \text{ for } t \in [0, T], \tag{2.11}$$

in which y is a deterministic function, bounded by M and such that $\mathrm{y}(T, \cdot) = g$. Then, given $\bar{Y}^{m-1}$, for $m \geq 1$, we define $(\bar{Y}^m, \bar{Z}^m, \bar{K}^{m,+}, \bar{K}^{m,-})$ as the solution on $[0, T]$ of

$$\begin{aligned}
\bar{Y}_t^m \ &= \ g(X_T) + \int_t^T f_{\ell_\circ}(X_s, \bar{Y}_s^m, \bar{Y}_s^{m-1})\, ds - \int_t^T \bar{Z}_s^m \, dW_s \\
&\qquad + \int_{[t,T]\cap\mathbb{T}} d(\bar{K}^{m,+} - \bar{K}^{m,-})_s, \\
-M &\leq \bar{Y}_t^m \leq M, \ \forall t \in \mathbb{T}, \ a.s. \\
\int_{\mathbb{T}} &(\bar{Y}_s^m + M) d\bar{K}_s^{m,+} = \int_{\mathbb{T}} (\bar{Y}_s^m - M) d\bar{K}_s^{m,-} = 0,
\end{aligned} \tag{2.12}$$

where $\bar{K}^{m,+}$ and $\bar{K}^{m,-}$ are non-decreasing processes.

Remark 2.1. *Since the solution $\bar{Y}$ of* (2.7) *is bounded by M, the quadruple of processes $(\bar{Y}, \bar{Z}, \bar{K}^+, \bar{K}^-)$ (with $\bar{K}^+ \equiv \bar{K}^- \equiv 0$) is in fact the unique solution of the same reflected BSDE as in* (2.12) *but with $f_{\ell_\circ}(X, \bar{Y}, \bar{Y})$ in place of $f_{\ell_\circ}(X_s, \bar{Y}^m, \bar{Y}^{m-1})$.*

Remark 2.2. *One can equivalently define the process $\bar{Y}^m$ in a recursive way. Let $\bar{Y}_T^m := g(X_T)$ be the terminal condition, and define, on each interval $[t_i, t_{i+1}]$, $(Y_\cdot^m, Z_\cdot^m)$ as the solution on $[t_i, t_{i+1}]$ of*

$$Y_\cdot^m \ = \ \bar{Y}_{t_{i+1}}^m + \int_\cdot^{t_{i+1}} f_{\ell_\circ}(X_s, Y_s^m, \bar{Y}_s^{m-1}) ds - \int_\cdot^{t_{i+1}} Z_s^m dW_s. \tag{2.13}$$

Then, $\bar{Y}^m := Y^m$ on $(t_i, t_{i+1}]$, and $\bar{Y}_{t_i}^m := (-M) \vee Y_{t_i}^m \wedge M$.

The error due to our Picard iteration scheme is handled in a standard way. It depends on the constants

$$L_1 := 2C_{\ell_\circ}\sum_{\ell=1}^{\ell_\circ}\ell(M_{h_\circ})^{\ell-1}, \quad M_{h_\circ}L_2 := L_\varphi\sum_{\ell=0}^{\ell_\circ}2C_{\ell_\circ}(M_{h_\circ})^{\ell},$$

where $M_{h_\circ}$ is defined by (A.2).

Theorem 2.3. *The system* (2.12) *of doubly reflected BSDEs admits a unique solution* $(\bar{Y}^m, \bar{Z}^m, \bar{K}^{m,+}, \bar{K}^{m,-})_{m\geq 0}$ *such that* $\bar{Y}^m$ *is uniformly bounded for each* $m \geq 0$*. Moreover, for all* $m \geq 0$*,* $|\bar{Y}^m|$ *is uniformly bounded by the constant* $M_{h_\circ}$*, and*

$$|\bar{Y}_t^m - \bar{Y}_t|^2 \;\leq\; \frac{L_2}{\lambda^2}\Big(\frac{L_2(T-t)}{\lambda^2}\Big)^m(2M)^2\frac{e^{\beta T}}{\beta},$$

for all $t \leq T$*, and all constants* $\lambda > 0$*,* $\beta > 2L_1 + L_2\lambda^2$*.*

Proof. i) First, when $\bar{Y}^m$ is uniformly bounded, $f_{\ell_\circ}(X_s, \bar{Y}_s^m, \bar{Y}_s^{m-1})$ can be considered to be uniformly Lipschitz in $\bar{Y}^m$, then (2.12) has at most one bounded solution. Next, in view of Lemma A.1 and Remark 2.2, it is easy to see that (2.13) has a unique solution Y^m, bounded by $M_{h_\circ}$ (defined by (A.2)) on each interval $[t_i, t_{i+1}]$. It follows the existence of the solution to (2.12). Moreover, $\bar{Y}^m$ is also bounded by $M_{h_\circ}$ on $[0, T]$, and more precisely bounded by M on the discrete grid $\mathbb{T}$, by construction.

ii) Consequently, the generator $f_{\ell_\circ}(x, y, y')$ can be considered to be uniformly Lipschitz in y and y'. Moreover, using (2.5) and (2.6), one can identify the corresponding Lipschitz constants as L_1 and L_2.

Let us denote $\Delta\bar{Y}^m := \bar{Y}^m - \bar{Y}$ for all $m \geq 1$. We notice that, in Remark 2.2, the truncation operation $\bar{Y}_{t_i}^m := (-M) \vee Y_{t_i}^m \wedge M$ can only make the value $(\Delta\bar{Y}_{t_i}^m)^2$ smaller than $(Y_{t_i}^m - \bar{Y}_{t_i})^2$, since $|\bar{Y}| \leq M$. Thus we can apply Itô's formula to $(e^{\beta t}(\Delta\bar{Y}_t^{m+1})^2)_{t\geq 0}$ on each interval $[t_i, t_{i+1}]$, and then take expectation to obtain

$$\begin{aligned} &\mathbb{E}\big[e^{\beta t}(\Delta\bar{Y}_t^{m+1})^2\big] + \beta\mathbb{E}\Big[\int_t^T e^{\beta s}|\Delta\bar{Y}_s^{m+1}|^2 ds + \int_t^T e^{\beta s}|\Delta\bar{Z}_s^{m+1}|^2 ds\Big] \\ \leq\;& 2\mathbb{E}\Big[\int_t^T e^{\beta s}\Delta\bar{Y}_s^{m+1}\big(f_{\ell_\circ}(X_s, \bar{Y}_s^{m+1}, \bar{Y}_s^m) - f_{\ell_\circ}(X_s, \bar{Y}_s, \bar{Y}_s)\big)ds\Big]. \end{aligned}$$

Using the Lipschitz property of $f_{\ell_\circ}$ and the inequality $\lambda^2 + \frac{1}{\lambda^2} \geq 2$, it follows that the r.h.s. of the above inequality is bounded by

$$(2L_1 + L_2\lambda^2)\mathbb{E}\Big[\int_t^T e^{\beta s}(\Delta\bar{Y}_s^{m+1})^2 ds\Big] + \frac{L_2}{\lambda^2}\mathbb{E}\Big[\int_t^T e^{\beta s}(\Delta\bar{Y}_s^m)^2 ds\Big].$$

Since $\beta \geq 2L_1 + L_2\lambda^2$, the above implies

$$\mathbb{E}\Big[e^{\beta t}(\Delta\bar{Y}_t^{m+1})^2\Big] \;\leq\; \frac{L_2}{\lambda^2}\mathbb{E}\Big[\int_t^T e^{\beta s}(\Delta\bar{Y}_s^m)^2 ds\Big], \tag{2.14}$$

and hence

$$\mathbb{E}\Big[\int_0^T e^{\beta t}(\Delta\bar{Y}_t^{m+1})^2 dt\Big] \;\leq\; \frac{L_2}{\lambda^2}T\mathbb{E}\Big[\int_0^T e^{\beta s}(\Delta\bar{Y}_s^m)^2 ds\Big].$$

Since $|\Delta\bar{Y}^0| = |\mathrm{y}(\cdot, X) - \bar{Y}| \leq 2M$ by (2.9) and our assumption $|\mathrm{y}| \leq M$, this shows that

$$\mathbb{E}\Big[\int_0^T e^{\beta t}(\Delta\bar{Y}_t^m)^2 dt\Big] \;\leq\; \big(\frac{L_2}{\lambda^2}T\big)^m (2M)^2 e^{\beta T}/\beta.$$

Plugging this in (2.14) leads to the required result at $t = 0$. It is then clear that the above estimation does not depend on the initial condition $(0, X_0)$, so that the same result holds true for every $t \in [0, T]$. □

2.3 A branching diffusion representation for $\bar{Y}^m$

We now explain how the solution of (2.13) on $[t_i, t_{i+1})$ can be represented by means of a branching diffusion system. More precisely, let us consider an element $(p_\ell)_{0\leq\ell\leq\ell_\circ} \in \mathbb{R}_+^{\ell_\circ+1}$ such that $\sum_{\ell=0}^{\ell_\circ} p_\ell = 1$, set $K_n := \{(1, k_2, \ldots, k_n) : (k_2, \ldots, k_n) \in \{0, \ldots, \ell_\circ\}^n\}$ for $n \geq 1$, and $K := \cup_{n\geq 1} K_n$. Let $(W^k)_{k\in K}$ be a sequence of independent d-dimensional Brownian motions, $(\xi_k)_{k\in K}$ and $(\delta_k)_{k\in K}$ be two sequences of independent random variables, such that

$$\mathbb{P}[\xi_k = \ell] = p_\ell,\;\; \ell \leq \ell_\circ, k \in K,$$

and

$$\bar{F}(t) := \mathbb{P}[\delta_k > t] = \int_t^\infty \rho(s)ds,\; t \geq 0,\; k \in K, \tag{2.15}$$

for some continuous strictly positive map $\rho : \mathbb{R}_+ \to \mathbb{R}_+$. We assume that

$$(W^k)_{k\in K},\ (\xi_k)_{k\in K},\ (\delta_k)_{k\in K} \text{ and } W \text{ are independent.} \tag{2.16}$$

Given the above, we construct particles $X^{(k)}$ that have the dynamics (2.1) up to a killing time T_k at which they split in ξ_k different (conditionally) independent particles with dynamics (2.1) up to their own killing time. The construction is done as follows. First, we set $T_{(1)} := \delta_1$, and, given $k = (1, k_2, \ldots, k_n) \in K_n$ with $n \geq 2$, we let $T_k := \delta_k + T_{k-}$ in which $k- := (1, k_2, \ldots, k_{n-1}) \in K_{n-1}$. We can then define the Brownian particles $(W^{(k)})_{k\in K}$ by using the following induction: we first set

$$W^{((1))} := W^1 \mathbf{1}_{[0,T_{(1)}]}\ ,\ \mathcal{K}^1_t := \{(1)\}\mathbf{1}_{[0,T_{(1)}]}(t) + \emptyset \mathbf{1}_{[0,T_{(1)}]^c}(t),\ t \geq 0,$$

then, given $n \geq 2$ and $k \in \bar{\mathcal{K}}^{n-1}_T := \cup_{t\leq T}\mathcal{K}^{n-1}_t$, we let

$$W^{(k\oplus j)} := \left(W^{(k)}_{\cdot\wedge T_k} + W^{k\oplus j}_{\cdot\vee T_k} - W^{k\oplus j}_{T_k}\right)\mathbf{1}_{[0,T_{k\oplus j}]},\ 1 \leq j \leq \xi_k,$$

and

$$\bar{\mathcal{K}}^n_t := \{k \oplus j : k \in \bar{\mathcal{K}}^{n-1}_T, 1 \leq j \leq \xi_k \text{ s.t. } t \in (0, T_{k\oplus j}]\},\ \ \bar{\mathcal{K}}_t := \cup_{n\geq 1}\bar{\mathcal{K}}^n_t,$$

$$\mathcal{K}^n_t := \{k \oplus j : k \in \bar{\mathcal{K}}^{n-1}_T, 1 \leq j \leq \xi_k \text{ s.t. } t \in (T_k, T_{k\oplus j}]\},\ \ \mathcal{K}_t := \cup_{n\geq 1}\mathcal{K}^n_t,$$

in which we use the notation $(1, k_1, \ldots, k_{n-1}) \oplus j = (1, k_1, \ldots, k_{n-1}, j)$.
Now observe that the solution X^x of (2.1) on $[0, T]$ with initial condition $X^x_0 = x \in \mathbb{R}^d$ can be identified in law on the canonical space as a process of the form $\Phi[x](\cdot, W)$ in which the deterministic map $(x, s, \omega) \mapsto \Phi[x](s, \omega)$ is $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{P}$-measurable, where $\mathcal{P}$ is the predictable σ-filed on $[0, T] \times \Omega$. We then define the corresponding particles $(X^{x,(k)})_{k\in K}$ by $X^{x,(k)} := \Phi[x](\cdot, W^{(k)})$.

Given the above construction, we can now introduce a sequence of deterministic map associated to $(\bar{Y}^m)_{m\geq 0}$. First, we set

$$v^0 := \mathrm{y}\ , \tag{2.17}$$

recall (2.11). Then, given v^{m-1}, we define

$$\begin{aligned}
V^m_{t,x} &:= \Big(\prod_{k\in\mathcal{K}_{t_{i+1}-t}} G^m_{t,x}(k)\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_{t_{i+1}-t}\setminus\mathcal{K}_{t_{i+1}-t}} A^m_{t,x}(k)\Big),\\
G^m_{t,x}(k) &:= \frac{v^m\big(t_{i+1}, X^{x,(k)}_{t_{i+1}-t}\big)}{\bar{F}(t_{i+1} - t - T_{k-})},\\
A^m_{t,x}(k) &:= \frac{\sum_{j=1}^{j_\circ} a_{j,\xi_k}(X^{x,(k)}_{T_k})\varphi_j(v^{m-1}(t + T_k, X^{x,(k)}_{T_k}))}{p_{\xi_k}\,\rho(\delta_k)},\quad \forall (t, x) \in [t_i, t_{i+1}) \times \mathbf{X}.
\end{aligned}$$

We finally set, whenever $V^m_{t,x}$ is integrable,

$$v^m(t,x) := \mathbb{E}\left[V^m_{t,x}\right], \ \ (t,x) \in (t_i, t_{i+1}) \times \mathbf{X}, \ m \geq 1,$$

and

$$v^m(t_i,x) := (-M) \vee \mathbb{E}\left[V^m_{t_i,x}\right] \wedge M, \ \ x \in \mathbf{X}, \ m \geq 1. \tag{2.18}$$

Proposition 2.1. *For all $m \geq 1$ and $(t,x) \in [0,T] \times \mathbf{X}$, the random variable $V^m_{t,x}$ is integrable. Moreover, one has $\bar{Y}^m_\cdot = v^m(\cdot, X)$ on $[0,T]$.*

This follows from Proposition A.2 proved in the Appendix, which is in spirit of [11]. The main use of this representation result here is that it provides a numerical scheme for the approximation of the component $\bar{Y}$ of (2.7), as explained in the next section.

2.4 The numerical algorithm

The representation result in Proposition 2.1 suggests to use a simple Monte-Carlo estimation of the expectation in the definition of v^m based on the simulation of the corresponding particle system. However, it requires the knowledge of v^{m-1} in the Picard scheme which is used to localize our approximating polynomials. We therefore need to approximate the corresponding (conditional) expectations at each step of the Picard iteration scheme. In practice, we shall replace the expectation operator $\mathbb{E}$ in the definition of v^m by an operator $\hat{\mathbb{E}}$ that can be computed explicitly, see Remark 2.2 below.

In order to perform a general (abstract) analysis, let us first recall that $v^m(t,x) = \mathbb{E}[V_{t,x}(v^m(t_{i+1},\cdot), v^{m-1}(\cdot)]$ for all $t \in (t_i, t_{i+1})$ and $v^m(t_i,x) = (-M) \vee \mathbb{E}[V_{t_i,x}(v^m(t_{i+1},\cdot), v^{m-1}(\cdot)] \wedge M$, where, given two functions $\phi, \phi' : (t_i, t_{i+1}] \times \mathbb{R}^d \to \mathbb{R}$,

$$\begin{aligned}
V_{t,x}(\phi,\phi') &:= \Big(\prod_{k\in\mathcal{K}_{t_{i+1}-t}} G_{t,x}(\phi,k)\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_{t_{i+1}-t}\setminus\mathcal{K}_{t_{i+1}-t}} A_{t,x}(\phi',k)\Big),\\
G_{t,x}(\phi,k) &:= \frac{\phi(t_{i+1}, X^{x,(k)}_{t_{i+1}-t})}{\bar{F}(t_{i+1}-t-T_{k-})},\\
A_{t,x}(\phi',k) &:= \frac{\sum_{j=1}^{j_\circ} a_{j,\xi_k}(X^{x,(k)}_{T_k})\varphi_j(\phi'(t+T_k, X^{x,(k)}_{T_k})}{p_{\xi_k}\,\rho(\delta_k)}.
\end{aligned}$$

Let us then denote by $\mathbf{L}^\infty_{M_{h_\circ}}$ the class of all Borel measurable functions $\phi : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ that are bounded by $M_{h_\circ}$, and let $\mathbf{L}^\infty_{M_{h_\circ},0} \subset \mathbf{L}^\infty_{M_{h_\circ}}$ be a

subspace, generated by a finite number of basis functions. Besides, let us consider a sequence $(U_i)_{i\geq 1}$ of i.i.d. random variables of uniform distribution on $[0,1]$, independent of $(W^k)_{k\in K}$, $(\xi_k)_{k\in K}$, $(\delta_k)_{k\in K}$ and W introduced in (2.16). Denote $\hat{\mathcal{F}} := \sigma(U_i, i\geq 1)$.
From now on, we use the notations

$$\|\phi\|_{t_i} := \sup_{(t,x)\in[t_i,t_{i+1})\times\mathbb{R}^d} |\phi(t,x)| \text{ and } \|\phi\|_\infty := \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} |\phi(t,x)|$$

for all functions $\phi : [0,T]\times\mathbb{R}^d \to \mathbb{R}$.

Assumption 2.4. *There exists an operator $\hat{\mathbb{E}}[\hat{V}_{t,x}(\phi,\phi')](\omega)$, defined for all $\phi,\phi'\in\mathbf{L}^\infty_{M_{h_\circ},0}$, such that $(t,x,\omega)\mapsto\hat{\mathbb{E}}[\hat{V}_{t,x}(\phi,\phi')](\omega)$ is $\mathcal{B}([0,T]\times\mathbb{R}^d)\otimes\hat{\mathcal{F}}$-measurable, and such that the function $(t,x)\in[0,T]\times\mathbb{R}^d\mapsto\hat{\mathbb{E}}[\hat{V}_{t,x}(\phi,\phi')](\omega)$ belongs to $\mathbf{L}^\infty_{M_{h_\circ},0}$ for every fixed $\omega\in\Omega$. Moreover, one has*

$$\mathcal{E}(\hat{\mathbb{E}}) := \| \sup_{\phi,\phi'\in\mathbf{L}^\infty_{M_{h_\circ},0}} \mathbb{E}\Big[\big|\mathbb{E}\left[V_\cdot(\phi,\phi')\right] - \hat{\mathbb{E}}[\hat{V}_\cdot(\phi,\phi')]\big|\Big] \|_\infty < \infty.$$

Then, one can construct a numerical algorithm by first setting $\hat{v}^0 \equiv \mathrm{y}$, $\hat{v}^m(T,\cdot) = g$, $m\geq 1$, and then by defining by induction over $m\geq 1$

$$\hat{v}^m(t,x) := (-M_{h_\circ}) \vee \hat{\mathbb{E}}\Big[\hat{V}_{t,x}(\hat{v}^m(t_{i+1},\cdot),\hat{v}^{m-1})\Big] \wedge M_{h_\circ},\ t\in(t_i,t_{i+1}),$$

and

$$\hat{v}^m(t_i,x) := (-M) \vee \hat{\mathbb{E}}\Big[\hat{V}_{t_i,x}(\hat{v}^m(t_{i+1},\cdot),\hat{v}^{m-1})\Big] \wedge M. \tag{2.19}$$

In order to analyse the error due to the approximation of the expectation error, let us set

$$\bar{q}_t := \#\bar{\mathcal{K}}_t\ \ ,\ \ q_t := \#\mathcal{K}_t,$$

and denote

$$V^M_t := \Big(\prod_{k\in\mathcal{K}_t} \frac{M}{\bar{F}(t-T_{k-})}\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_t\setminus\mathcal{K}_t} \frac{2C_{\ell_\circ}}{p_{\xi_k}\rho(\delta_k)}\Big).$$

Recall that $h < h_\circ$ that is defined by (A.1) in the Appendix.

Lemma 2.5. *The two constants*

$$M_h^1 := \sup_{0\le t\le h} \mathbb{E}\big[q_t V_t^M\big] \quad \textit{and} \;\; M_h^2 := \sup_{0\le t\le h} \mathbb{E}\big[\bar{q}_t V_t^M\big]$$

are finite.

Proof. Notice that for any constant $\varepsilon > 0$, there is some constant $C_\varepsilon > 0$ such that $n \le C_\varepsilon(1+\varepsilon)^n$ for all $n \ge 1$. Then

$$M_h^1 \le C_\varepsilon \mathbb{E}\big[\sup_{0\le t\le h}(1+\varepsilon)^{q_t} V_t^M\big] \le C_\varepsilon \mathbb{E}\Big[\prod_{k\in\mathcal{K}_h} \frac{(1+\varepsilon)M}{\bar{F}(h-T_{k-})} \prod_{k\in\bar{\mathcal{K}}_h\setminus\mathcal{K}_h} \frac{2(1+\varepsilon)C_{\ell_\circ}}{p_{\xi_k}\rho(\delta_k)}\Big],$$

where the latter expectation is finite for ε small enough. This follows from the fact that $h < h_\circ$ for $h_\circ$ defined by (A.1) and from the same arguments as in Lemma A.1 in the Appendix. One can similarly obtain that M_h^2 is also finite.

□

Proposition 2.6. *Let Assumption 2.4 hold true. Then*

$$\|\mathbb{E}\left[|v^m - \hat{v}^m|\right]\|_\infty \le \mathcal{E}(\hat{\mathbb{E}})\big(1+N_h\big)\frac{(m+N_h)^{N_h}}{N_h!}\Big((2L_\varphi M_h^2) \vee \frac{M_h^1}{M} \vee 1\Big)^{m+N_h}.$$

Before turning to the proof of the above, let us comment on the use of this numerical scheme.

Remark 2.2. *In practice, the approximation of the expectation operator can be simply constructed by using pure forward simulations of the branching process. Let us explain this first in the case $h_\circ = T$. Given that $\hat{v}^m$ has already been computed, one takes it as a given function, one draws some independent copies of the branching process (independently of $\hat{v}^m$) and computes $\hat{v}^{m+1}(t,x)$ as the Monte-Carlo counterpart of $\mathbb{E}[V_{t,x}(\hat{v}^{m+1}(T,\cdot),\hat{v}^m)]$, and truncates it with the a-priori bound $M_{h_\circ}$ for $(\bar{Y}^m)_{m\ge1}$. This corresponds to the operator $\hat{\mathbb{E}}[\hat{V}_{t,x}(\hat{v}^{m+1}(T,\cdot),\hat{v}^m)]$. If $h_\circ < T$, one needs to iterate backward over the periods $[t_i,t_{i+1}]$. Obviously one cannot in practice compute the whole map $(t,x)\mapsto \hat{v}^{m+1}(t,x)$ and this requires an additional discretization on a suitable time-space grid. Then, the additional error analysis can be handled for instance by using the continuity property of v^m in Proposition A.5 in the Appendix. This is in particular the case if one just computes $\hat{v}^{m+1}(t,x)$ by replacing (t,x) by its projection on a discrete time-space grid.*

Remark 2.3. i). *In the classical time discretization schemes of BSDEs, such as those in [5, 9, 21], one needs to let the time step go to* 0 *to reduce the discretization error. Here, the representation formula in Proposition 2.1 has no discretization error related to the BSDE itself (assuming the solution of the previous Picard iteration is known perfectly), we only need to use a fixed discrete time grid* $(t_i)_{0\le i\le N_h}$ *for* $t_i = ih$ *with* h *small enough.*
ii). *Let* $A'_j := \{y \in \mathbb{R} \,:\, \varphi_j(y) = 1\} \subset A_j$ *for* $j \le j_\circ$*, and assume that the* A'_j*'s are disjoint. If the* A'_j *are large enough, we do not need to be very precise on* $\hat{v}^m$ *to obtain a good approximation of* $\mathbb{E}[V_{t,x}(g, v^m)]$ *by* $\mathbb{E}[V_{t,x}(g, \hat{v}^m)]$ *for* $t \in [t_{N_h-1}, t_{N_h})$*. One just needs to ensure that* $\hat{v}^m$ *and* v^m *belong to the same set* A'_j *at the different branching times and at the corresponding* X*-positions. We can therefore use a rather rough time-space grid on this interval (i.e.* $[t_{N_h-1}, t_{N_h}]$*). Further, only a precise value of* $\hat{v}^m(t_{N_h-1}, \cdot)$ *will be required for the estimation of* $\hat{v}^{m+1}$ *on* $[t_{N_h-2}, t_{N_h-1})$ *and this is where a fine space grid should be used. Iterating this argument, one can use rather rough time-space grid on each* (t_i, t_{i+1}) *and concentrate on each* t_i *at which a finer space grid is required. This is the main difference with the usual backward Euler schemes of [5, 9, 21] and the forward Picard schemes of [3].*

Proof of Proposition 2.6. Define

$$\tilde{v}^m(\cdot) := (-M_{h_\circ}) \vee \mathbb{E}\left[V_\cdot(\hat{v}^m(t_{i+1}, \cdot), \hat{v}^{m-1}) \Big| \hat{\mathcal{F}}\right] \wedge M_{h_\circ}.$$

Then, Lemma A.3 below combined with the inequality $|\varphi| \le 1$ implies that for $(t, x) \in [t_i, t_{i+1}) \times \mathbf{X}$

$$\begin{aligned}
&|\tilde{v}^m(t,x) - v^m(t,x)| \\
&\le \mathbb{E}\Big[\sum_{k \in \mathcal{K}_{t_{i+1}-t}} \frac{1}{M} V^M_{t_{i+1}-t} \big|\hat{v}^m(t_{i+1}, X^{x,(k)}_{t_{i+1}}) - v^m(t_{i+1}, X^{x,(k)}_{t_{i+1}})\big| \Big| \hat{\mathcal{F}}\Big] \\
&+ \mathbb{E}\Big[\sum_{k \in \bar{\mathcal{K}}_{t_{i+1}-t} \setminus \mathcal{K}_{t_{i+1}-t}} 2L_\varphi V^M_{t_{i+1}-t} \big|\hat{v}^{m-1}(T_k, X^{x,(k)}_{T_k}) - v^{m-1}(T_k, X^{x,(k)}_{T_k})\big| \Big| \hat{\mathcal{F}}\Big].
\end{aligned}$$

Let us compute the expectation of the first term. Denoting by $\bar{\mathcal{F}}$ the σ-field generated by the branching processes, we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{k\in\mathcal{K}_{t_{i+1}-t}} \frac{1}{M} V^M_{t_{i+1}-t}\big|\hat{v}^m(t_{i+1}, X^{x,(k)}_{t_{i+1}}) - v^m(t_{i+1}, X^{x,(k)}_{t_{i+1}})\big|\Big] \\
= \;& \mathbb{E}\Big[\sum_{k\in\mathcal{K}_{t_{i+1}-t}} \frac{1}{M} V^M_{t_{i+1}-t}\mathbb{E}\Big[\big|\hat{v}^m(t_{i+1}, X^{x,(k)}_{t_{i+1}}) - v^m(t_{i+1}, X^{x,(k)}_{t_{i+1}})\big|\Big|\bar{\mathcal{F}}\Big]\Big] \\
\le \;& \frac{1}{M}\|\mathbb{E}[|\hat{v}^m - v^m|]\|_{t_{i+1}}\mathbb{E}\Big[q_{t_{i+1}-t}V^M_{t_{i+1}-t}\Big] \;\le\; \frac{M^1_h}{M}\|\mathbb{E}[|\hat{v}^m - v^m|]\|_{t_{i+1}}.
\end{aligned}$$

Similarly, for the second term, one has

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{k\in\bar{\mathcal{K}}_{t_{i+1}-t}\setminus\mathcal{K}_{t_{i+1}-t}} 2L_\varphi V^M_{t_{i+1}-t}\big|\hat{v}^{m-1}(T_k, X^{x,(k)}_{T_k}) - v^{m-1}(T_k, X^{x,(k)}_{T_k})\big|\Big] \\
\le \;& 2L_\varphi M_h^2\|\mathbb{E}[|\hat{v}^{m-1} - v^{m-1}|]\|_{t_i}.
\end{aligned}$$

Notice that $\|\mathbb{E}\left[|\tilde{v}^m - \hat{v}^m|\right]\|_{t_i} \le \mathcal{E}(\hat{\mathbb{E}})$ by Assumption 2.4. Hence,

$$\begin{aligned}
\|\mathbb{E}[|\hat{v}^m - v^m|]\|_{t_i} \;\le\; & \mathcal{E}(\hat{\mathbb{E}}) + 2L_\varphi M_h^2\|\mathbb{E}[|\hat{v}^{m-1} - v^{m-1}|]\|_{t_i} \\
& + \frac{M^1_h}{M}\|\mathbb{E}[|\hat{v}^m - v^m|]\|_{t_{i+1}}.
\end{aligned}$$

We now appeal to Proposition A.4 to obtain

$$\begin{aligned}
\|\mathbb{E}[|\hat{v}^m - v^m|]\|_{t_i} \;&\le\; \mathcal{E}(\hat{\mathbb{E}})\left(\sum_{i=1}^m C^i + \sum_{i'=2}^{N_h-i}\left(\sum_{j_1=1}^m \cdots \sum_{j_{i'}=1}^{j_{i'-1}} C^{m-j_{i'}}C^{i'-1}\right)\right) \\
&\le\; \mathcal{E}(\hat{\mathbb{E}})(1+N_h)\frac{(m+N_h)^{N_h}}{N_h!}C^{m+N_h},
\end{aligned}$$

with $C := (2L_\varphi M_h^2) \vee \frac{M_h^1}{M} \vee 1$. □

3 Example of application

In this section, we consider a toy example of application. Let us set $\mathbf{X} := [\underline{x}, \bar{x}]$ with $\underline{x} = \pi/8$ and $\bar{x} = 7\pi/8$, and consider the solution X of (2.1) with

$$\mu(x) = 0.1 \times (\frac{\pi}{2} - x) \;\text{ and }\; \sigma(x) := 0.2 \times (\bar{x} - x)(x - \underline{x}).$$

We then take

$$f(x,y) = \mu(x)\left(\sqrt{1-y^2}\mathbf{1}_{|y|\le\bar{y}} + \sqrt{1-\bar{y}^2}\mathbf{1}_{|y|>\bar{y}}\right) + \frac{1}{2}\sigma(x)^2 y$$

with $\bar{y} := \cos(\underline{x})$. As can be seen on Figure 1, the Lipschitz constant of the driver is rather large. However, a simple application of Itô's lemma shows that the solution of (2.2) with $g = \cos$ is given by $Y = \cos(X)$, which will be used to assess the precision of our estimator.

The driver f is approximated by polynomials of order two that are weighted by localizing functions. Namely, we fix $A_j^\upsilon := (y_j - \upsilon, y_{j+1} + \upsilon]$ for $j = 1, \ldots, 5$, with $\upsilon := 10^{-5}$ and

$$y_1 = -y_6 = \infty,\ y_2 = -y_5 = \bar{y},\ y_3 = -y_4 = \cos(x_{\frac{N_X+1}{4}-1}),$$

in which $\{x_1, \ldots, x_{N_X}\}$ are equidistant points with $x_1 = \underline{x}$ and $x_{N_X} = \bar{x}$. Then, $f_{\ell_\circ}$ is defined as

$$f_{\ell_\circ}(x,y,y') = \sum_{j=1}^{5}\left(\mu(x)(a_{j0} + a_{j1}y + a_{j2}y^2) + \frac{1}{2}\sigma(x)^2 y\right)\varphi_j(y')$$

where

$$\varphi_j(y') = \begin{cases} \frac{y'-y_j+\upsilon}{2\upsilon} & \text{if} \quad y' \in A_j^\upsilon \cap [y_j - \upsilon, y_j + \upsilon) \\ 1 & \text{if} \quad y' \in A_j^\upsilon \cap [y_j + \upsilon, y_{j+1} - \upsilon] \\ 1 - \frac{y'-y_{j+1}+\upsilon}{2\upsilon} & \text{if} \quad y' \in A_j^\upsilon \cap [y_{j+1} - \upsilon, y_{j+1} + \upsilon) \\ 0 & \text{if} \quad y' \notin A_j^\upsilon \end{cases}$$

and

$$\begin{aligned}
(a_{10}, a_{11}, a_{12}) &= (a_{50}, a_{51}, a_{52}) = ((1-\bar{y}^2)^{\frac{1}{2}}, 0, 0) \\
(a_{20}, a_{21}, a_{22}) &= (a_{40}, a_{41}, a_{42}) \\
&\qquad\qquad = ((1-(y_3)^2)^{\frac{1}{2}} - a_{21}y_3, \frac{(1-(y_3)^2)^{\frac{1}{2}} - (1-\cos(x_2)^2)^{\frac{1}{2}}}{y_3 - \cos(x_2)}, 0) \\
(a_{30}, a_{31}, a_{32}) &= (1, 0, -\frac{1-(1-(y_3)^2)^{\frac{1}{2}}}{(y_3)^2}).
\end{aligned}$$

In Figure 2, we plot the approximation of $x \mapsto f(x, \cos(x), \cos(x))$ by $x \mapsto f_{\ell_\circ}(x, \cos(x), \cos(x))$, that drives the driver's approximation error, recall (2.8).

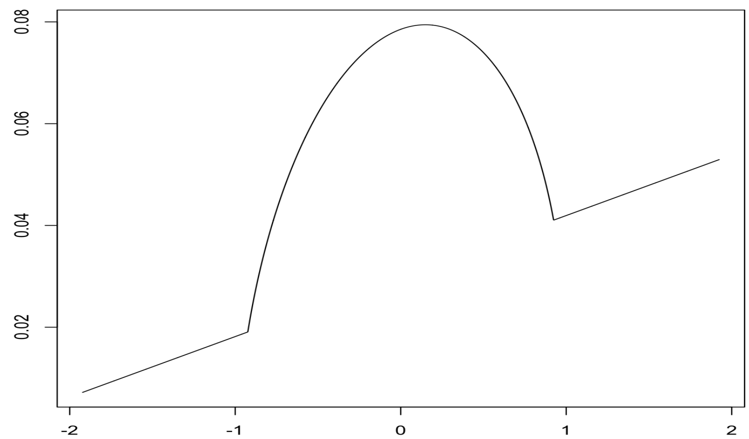

Figure 1: Driver $f(\pi/4, \cdot)$.

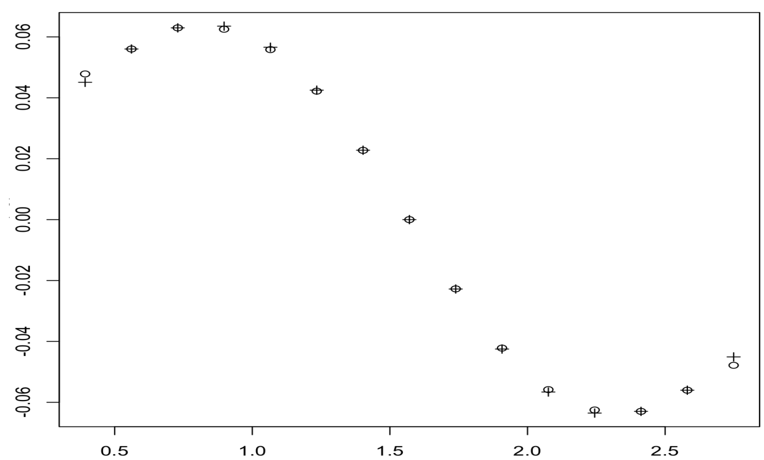

Figure 2: Approximation of the driver - Crosses : $f(\cdot, \cos)$. Circles: $f_{\ell_\circ}(\cdot, \cos, \cos)$.

It is very good except at the boundary points, which should not have a major impact given our mean-reverting dynamics for X.

To construct the approximation operator $\hat{\mathbb{E}}[\hat{V}]$. The time interval $[0, T)$ is divided into N_T intervals $[s_i, s_{i+1})$, $0 \leq i \leq N_T - 1$, of equal length, with $s_0 = 0$ and $s_{N_T} = T$. The branching density ρ is taken as the exponential law density of parameter $\lambda = 0.6$, but the branching times T_k are replaced by $\min(s_i \ : s_i \geq T_k, i \leq N_T)$, if $T_k \leq T$. We take $p_0 = p_1 = p_2 = 1/3$. We draw

N independent path of the Brownian particles system $(W^{(k),n}; k \in K)_{n \le N}$ (up to T) to which is associated the sequence of numbers of children, branching and birth times $(\zeta_k^n, \delta_k^n, T_k^n; k \in K)_{n \le N}$. The index sets $\mathcal{K}^n$ and $\bar{\mathcal{K}}^n$ are defined correspondingly. Let $\bar{\Phi}[x](\cdot, W)$ be the map that associates to x the Euler scheme of (2.1) starting from $X_0 = x$ on the grid $(s_i)_{i \le N_T}$. Then, we set $\bar{X}^{x_l,(k),n} := \bar{\Phi}[x](\cdot, W^{(k),n})$ for each $l \le N_X$. A typical path starting from $\pi/2$ is provided in Figure 3.

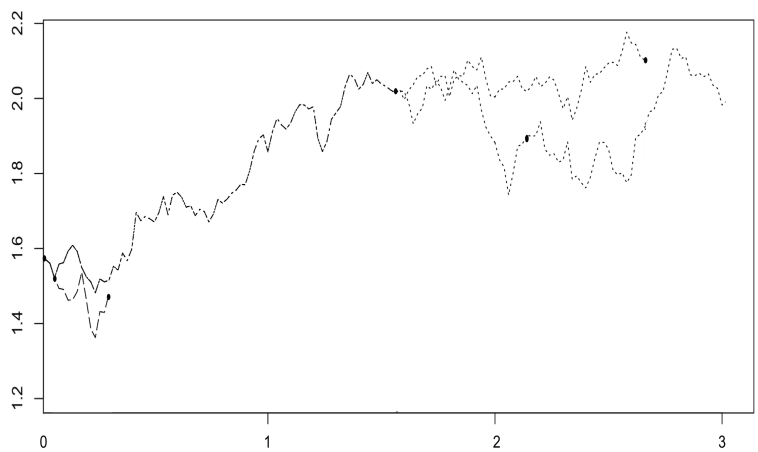

Figure 3: A typical simulated path of the branching diffusion starting from $\pi/2$ on $[0,3]$. Bullets denote branching or killing times.

The simplest algorithm reads as follows. We fix $\mathrm{y}(t,\cdot) = (t/T)\cos$, $\hat{v}^m(T,\cdot) = \cos$, and then set, for $\kappa \ge 1$, $m \ge 0$, $i < N_T/\kappa$ with $s_{i\kappa} \in [t_{i'}, t_{i'+1})$ and $l \le N_X$,

$$\hat{\mathbb{E}}\left[\hat{V}_{s_{i\kappa},x_l}(\hat{v}^m(t_{i'+1},\cdot), \hat{v}^{m-1})\right] := \frac{1}{N}\sum_{n=1}^{N} \hat{V}^n_{s_{i\kappa},x_l}(\hat{v}^m(t_{i'+1},\cdot), \hat{v}^{m-1})$$

where

$$\hat{V}^n_{s_{i\kappa},x_l}(\phi,\phi') := \Big(\prod_{k\in\mathcal{K}^n_{t_{i'+1}-s_{i\kappa}}} \hat{G}^n_{s_{i\kappa},x_l}(\phi,k)\Big)\Big(\prod_{k\in\bar{\mathcal{K}}^n_{t_{i'+1}-s_{i\kappa}}\setminus\mathcal{K}^n_{t_{i'+1}-s_{i\kappa}}} \hat{A}^n_{s_{i\kappa},x_l}(\phi',k)\Big),$$

$$\hat{G}^n_{s_{i\kappa},x_l}(\phi,k) := \frac{\phi(\bar{X}^{x_l,(k),n}_{t_{i'+1}-s_{i\kappa}})}{\bar{F}(t_{i'+1}-s_{i\kappa}-T^n_{k-})},$$

$$\hat{A}^n_{s_{i\kappa},x_l}(\phi',k) := \frac{\sum_{j=1}^{j_\circ} a_{j,\xi^n_k}(\bar{X}^{x_l,(k),n}_{T^n_k})\varphi_j(\phi'(s_{i\kappa}+T^n_k,\bar{X}^{x_l,(k),n}_{T_k}))}{p_{\xi^n_k}\,\rho(\delta^n_k)}.$$

For $m \ge 0$, $(\hat{v}^m(s_{i\kappa},\cdot))_{i<N_T/\kappa}$ is extended to $\mathbf{X}$ by a simple barycentric linearization, and $(\hat{v}^m(\cdot,x))_{x\in\mathbf{X}}$ is extended to $[0,T]$ by setting $\hat{v}^m(t,x) := \hat{v}^m(s_{(i+1)\kappa},x)$ if $t\in(s_{i\kappa},s_{(i+1)\kappa}]$. In particular, each function $\hat{v}^m$ is computed on a time grid that is κ times rougher than the one used to construct the Euler scheme of the branching system.
In practice, we proceed slightly differently than the Picard iteration in the form described in Section 2.4. We shall instead use a mixed Picard iteration and we drop the index m for more clarity. First, we set $\hat{v}(s_j,\cdot) = \hat{v}(T,\cdot)$ for $j > N_T - \kappa$. Then, one can compute $\hat{v}(s_{N_T-\kappa},\cdot)$ as above, based on y, and set $\hat{v}(s_j,\cdot) = \hat{v}(s_{N_T-\kappa},\cdot)$ for $N_T - \kappa \ge j > N_T - 2\kappa$. This allows to compute immediately, $\hat{v}(s_{N_T-2\kappa},\cdot)$, since it only requires the knowledge of $\hat{v}(s_j,\cdot)$ for $N_T \ge j > N_T - 2\kappa$. We then set $\hat{v}(s_j,\cdot) = \hat{v}(s_{N_T-2\kappa},\cdot)$ for $N_T - 2\kappa \ge j > N_T - 3\kappa$, from which we can compute $\hat{v}(s_{N_T-3\kappa},\cdot)$. We go on this way. The estimation $\hat{v}(s_j,\cdot)$ corresponds to a unique Picard iteration for $N_T \ge j > N_T - 2\kappa$. But, around T, we expect to be very precise with only one, as y is based on the terminal condition. The estimation $\hat{v}(s_j,\cdot)$ corresponds to a mix between a unique and two Picard iterations for $N_T - \kappa \ge j > N_T - 2\kappa$, and so on. We therefore increase automatically the number of Picard iterations when we go further from the terminal horizon.

In Figure 4, we plot the solution $x \mapsto \cos(x)$ and the confidence interval obtained by computing the mean estimated value over 100 independent estimations $\pm$ twice the standard deviation computed over these 100 estimations, for $N = 10^3$, $N_X = 31$, $N_T = 50$, $\kappa = 10$ and $N_h = 1$. As can be seen, the algorithm is already quite efficient with only a rather small number of simulations. Figure 5 provides the same curves in the case $N = 10.10^3$.
In Figure 6 and Figure 7, we consider the case $T = 2$ with $N_h = 2$, $N_T = 140$ and $N_X = 47$. We use $N = 10^3$ and $N = 10.10^3$ simulations, respectively.

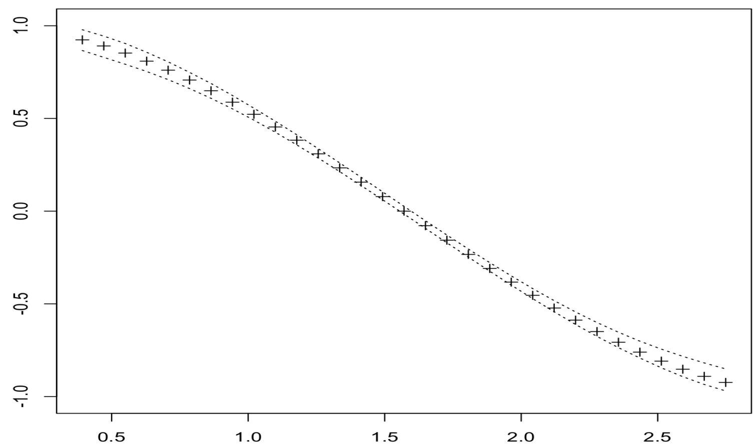

Figure 4: $T = N_h = 1$ with $N = 10^3$ - Crosses: cos function. Dotted lines: mean of estimations $\pm$ 2 standard deviation computed over the estimated values.

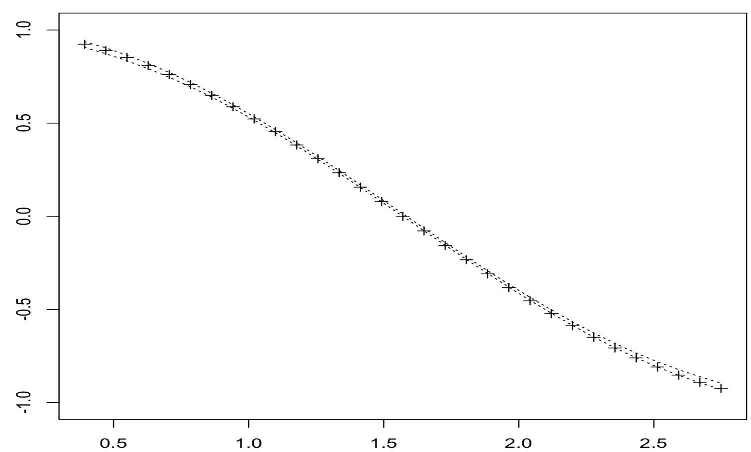

Figure 5: $T = N_h = 1$ with $N = 10.10^3$ - Crosses: cos function. Dotted lines: mean of estimations $\pm$ 2 standard deviation computed over the estimated values.

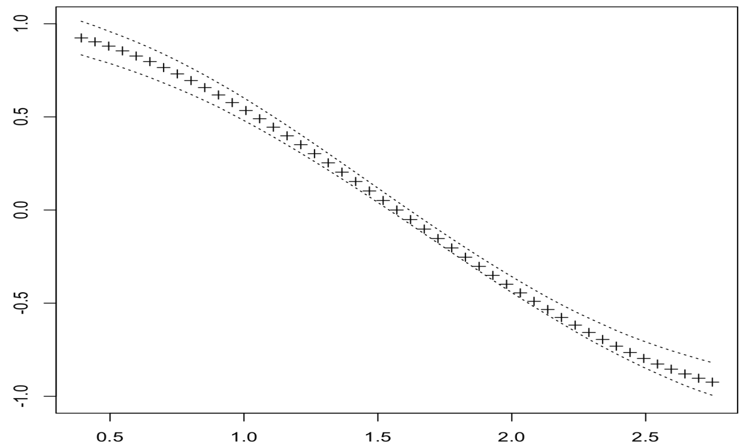

Figure 6: $T = N_h = 2$ with $N = 10^3$ - Crosses: cos function. Dotted lines: mean of estimations ± 2 standard deviation computed over the estimated values.

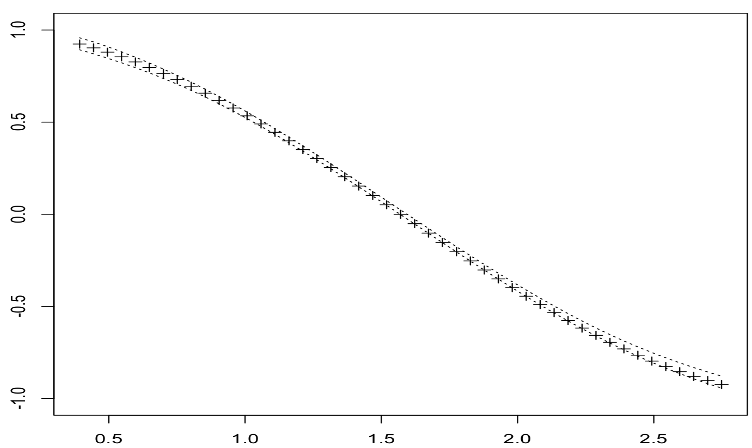

Figure 7: $T = N_h = 2$ with $N = 10.10^3$ - Crosses: cos function. Dotted lines: mean of estimations ± 2 standard deviation computed over the estimated values.

A Appendix

A.1 Technical lemmas

Lemma A.1. *The ordinary differential equation* $\eta'(t) = \sum_{\ell=0}^{\ell_\circ} 2C_{\ell_\circ}\eta(t)^\ell$ *with initial condition* $\eta(0) = M > 0$ *has a unique solution on* $[0, h_\circ]$ *for*

$$h_\circ := \frac{(\ell_\circ - 1)(1 - M)_+ + (1 \vee M)^{-(\ell_\circ - 1)}}{(\ell_\circ + 1)(\ell_\circ - 1)2C_{\ell_\circ}}. \tag{A.1}$$

Moreover, it is bounded on $[0, h_\circ]$ *by*

$$M_{h_\circ} := \max\Big(1, \big((1\vee M)^{1-\ell_\circ} + (\ell_\circ - 1)(1-M)^+ - h_\circ \ell_\circ(\ell_\circ - 1)2C_{\ell_\circ}\big)^{(1-\ell_\circ)^{-1}}\Big). \quad \text{(A.2)}$$

Consequently, one has, for all $t \in [0, h_\circ]$,

$$\mathbb{E}\Big[\Big(\prod_{k\in\mathcal{K}_t} \frac{M}{\bar F(t - T_{k-})}\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_t\setminus\mathcal{K}_t} \frac{2C_{\ell_\circ}}{p_{\xi_k}\rho(\delta_k)}\Big)\Big] \;\le\; \eta(t) \;\le\; M_{h_\circ}. \qquad \text{(A.3)}$$

Proof. i) We first claim that

$$\int_M^{M_{h_\circ}} \frac{dy}{2C_{\ell_\circ}(1 + y + \cdots + y^{\ell_\circ})} \ge h_\circ. \qquad \text{(A.4)}$$

Then, for every $t \in [0, h_\circ]$, there is some constant $M(t) \le M_{h_\circ} < \infty$ such that

$$\int_M^{M(t)} \frac{dy}{2C_{\ell_\circ}(1 + y + \cdots + y^{\ell_\circ})} \;=\; t = \int_0^t ds.$$

This means that $(M(t))_{t\in[0,h_\circ]}$ is a bounded solution (and hence the unique solution) of $\eta'(t) = \sum_{\ell=0}^{\ell_\circ} 2C_{\ell_\circ}\eta(t)^\ell$ with initial condition $\eta(0) = M > 0$. In particular, it is bounded by $M_{h_\circ}$.

ii) Let us now prove (A.4). Notice that $y^k \le 1 \vee y^{\ell_\circ}$ for any $y \ge 0$ and $k = 0, \cdots, \ell_\circ$. Then, it is enough to prove that

$$\int_M^{M_{h_\circ}} \Big(1 \wedge \frac{1}{y^{\ell_\circ}}\Big) dy \ge h_\circ(\ell_\circ + 1)2C_{\ell_\circ}. \qquad \text{(A.5)}$$

By direct computation, the l.h.s. of (A.5) equals

$$(M_{h_\circ} - M)\mathbf{1}_{\{M_{h_\circ}\le 1\}} + \Big((1-M)^+ + \frac{1}{\ell_\circ - 1}\big((1\vee M)^{1-\ell_\circ} - M_{h_\circ}^{1-\ell_\circ}\big)\Big)\mathbf{1}_{\{M_{h_\circ}>1\}}.$$

When $h_\circ$ satisfies (A.1), it is easy to check that (A.5) holds true.

iii) We now prove (A.3). Recall that $\bar{\mathcal{K}}_t^n$ denotes the collection of all particles in $\bar{\mathcal{K}}_t$ of generation n. Set

$$\chi_t^n := \Big(\prod_{k\in\cup_{j=1}^n \mathcal{K}_t^j} \frac{M}{\bar F(t - T_{k-})}\Big)\Big(\prod_{k\in\cup_{j=1}^n(\bar{\mathcal{K}}_t^j\setminus\mathcal{K}_t^j)} \frac{2C_{\ell_\circ}}{p_{\xi_k}\rho(\delta_k)}\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_t^{n+1}} \eta(t - T_{k-})\Big).$$

Since $\bar{\mathcal{K}}_t^n$ has only finite number of particles, the random variable χ_t^n is uniformly bounded. Then by exactly the same arguments as in (A.6) and (A.7) below, and by repeating this argument over n, one has

$$\eta(t) = M + \int_0^t \sum_{\ell=0}^{\ell_\circ} 2C_{\ell_\circ}\eta(s)^\ell ds = \mathbb{E}\big[\chi_t^1\big] = \mathbb{E}\big[\chi_t^n\big], \ \forall n \geq 1.$$

It follows by Fatou Lemma that

$$\mathbb{E}\Big[\Big(\prod_{k\in\mathcal{K}_t} \frac{M}{\bar{F}(t-T_{k-})}\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_t\setminus\mathcal{K}_t} \frac{2C_{\ell_\circ}}{p_{\xi_k}\rho(\delta_k)}\Big)\Big] = \mathbb{E}\big[\lim_{n\to\infty}\chi_t^n\big] \ \leq \ \lim_{n\to\infty}\mathbb{E}[\chi_t^n] = \eta(t).$$

□

For completeness, we provide here the proof the representation formula of Proposition 2.1 and of the technical lemma that was used in the proof of Proposition 2.6.

Proposition A.2. *The representation formula of Proposition 2.1 holds.*

Proof. We only provide the proof on $[t_{N_h-1}, T]$, the general result is obtained by induction. It is true by construction when m is equal to 0. Let us now fix $m \geq 1$.

First, Lemma A.1 shows that the random variable $V_{t,x}^m$ is integrable.

Next, Set $(1)+ := \{(1,j), j \leq \ell_\circ\} \cap \bar{\mathcal{K}}_T$ and define $\mathcal{K}_t(1) := \mathcal{K}_t \cap (1)+$ and $\bar{\mathcal{K}}_t(1) := \bar{\mathcal{K}}_t \cap (1)+$. For ease of notations, we write $X^x := X^{x,((1))}$. Then, for all $(t,x) \in [t_{N_h-1}, T] \times \mathbb{R}^d$,

$$\mathbb{E}[V_{t,x}^m] = \mathbb{E}\left[\frac{g(X_{T-t}^x)}{\bar{F}(T-t)}\mathbf{1}_{\{T_{(1)}\geq T-t\}}\right] + \mathbb{E}\left[\mathbf{1}_{\{T_{(1)}<T-t\}} \frac{\sum_{j=1}^{j_\circ} a_{j,\xi_{(1)}}(X_{T_{(1)}}^x)\varphi_j(v^{m-1}(t+T_{(1)}, X_{T_{(1)}}^x))}{p_{\xi_{(1)}}\,\rho(\delta_{(1)})} R_{t,x}^m\right]$$

where

$$R_{t,x}^m := \Big(\prod_{k\in\mathcal{K}_{T-t}(1)} G_{t,x}(k)\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_{T-t}(1)\setminus\mathcal{K}_{T-t}(1)} A_{t,x}^m(k)\Big)$$

satisfies

$$\mathbb{E}[R_{t,x}^m|\mathcal{F}_{T_{(1)}}] = \prod_{k\in(1)+} v^m\big(T_{(1)}, X_{T_{(1)}}^{t,x}\big) = \big[v^m(T_{(1)}, X_{T_{(1)}}^x)\big]^{\xi_{(1)}},$$

by (2.16). On the other hand, (2.15) and (2.16) imply

$$\mathbb{E}\left[\frac{g(X^x_{T-t})}{\bar{F}(T-t)}\mathbf{1}_{\{T_{(1)}\geq T-t\}}\right] = \mathbb{E}[g(X^x_{T-t})] \tag{A.6}$$

and

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{\{T_{(1)}<T-t\}}\frac{\sum_{j=1}^{j_\circ} a_{j,\xi_{(1)}}(X^x_{T_{(1)}})\varphi_j(v^{m-1}(t+T_{(1)},X^x_{T_{(1)}}))}{p_{\xi_{(1)}}\,\rho(\delta_{(1)})}[v^m(T_{(1)},X^x_{T_{(1)}})]^{\xi_{(1)}}\right]\\
&=\mathbb{E}\left[\int_0^{T-t}\frac{\sum_{j=1}^{j_\circ} a_{j,\xi_{(1)}}(X^x_s)\varphi_j(v^{m-1}(t+s,X^x_s))}{p_{\xi_{(1)}}}[v^m(s,X^x_s)]^{\xi_{(1)}}ds\right]\\
&=\mathbb{E}\left[\int_0^{T-t}\sum_{j=1}^{j_\circ}\sum_{\ell\leq\ell_\circ} a_{j,\ell}(X^x_s)\varphi_j(v^{m-1}(t+s,X^x_s))[v^m(s,X^x_s)]^{\ell}ds\right]\\
&=\mathbb{E}\left[\int_0^{T-t} f_{\ell_\circ}(X^x_s,v^m(t+s,X^x_s),v^{m-1}(t+s,X^x_s))ds\right].
\end{aligned} \tag{A.7}$$

Combining the above implies that

$$v^m(t,X_t) = \mathbb{E}\left[g(X_T)+\int_t^T f_{\ell_\circ}(X_s,v^m(s,X_s),v^{m-1}(s,X_s))ds\Big|\mathcal{F}_t\right],$$

and the required result follows by induction. □

Lemma A.3. *Let $(x^i,y^i)_{i\leq I}$ be a sequence of real numbers. Then,*

$$\left|\prod_{i=1}^I x^i - \prod_{i=1}^I y^i\right| \leq \sum_{i\in I}\Big(|x^i-y^i|\prod_{j\neq i}\max(|x^j|,|y^j|)\Big).$$

Proof. It suffices to observe that

$$\prod_{i=1}^I x^i - \prod_{i=1}^I y^i \;=\; (x^1-y^1)\prod_{i=2}^I x^i + y^1\Big(\prod_{i=2}^I x^i - \prod_{i=2}^I y^i\Big),$$

and to proceed by induction. □

Proposition A.4. *Let $c_1,c_2,c_3\geq 0$, and let $(u^i_m)_{m\geq 0,i\geq 0}$ be a sequence such that*

$$u^i_m \leq c_1 u^i_{m-1} + c_2 u^{i+1}_m + c_3 \;\; \textit{for } m\geq 1, i<N_h.$$

Then

$$u_m^i \leq c_1^m u_0^i + \sum_{i'=1}^{N_h-i} \left(\sum_{j_1=1}^{m} \sum_{j_2=1}^{j_1} \cdots \sum_{j_{i'}=1}^{j_{i'-1}} c_1^m c_2^{i'} u_0^{i+i'} \right)$$
$$+ c_3 \left(\sum_{i=1}^{m} c_1^i + \sum_{i'=2}^{N_h-i} \Big(\sum_{j_1=1}^{m} \sum_{j_2=1}^{j_1} \cdots \sum_{j_{i'}=1}^{j_{i'-1}} c_1^{m-j_{i'}} c_2^{i'-1} \Big) \right).$$

Proof. We have

$$u_m^i \leq (c_1)^m u_0^i + \sum_{j=1}^{m} (c_1)^{m-j} (c_2 u_m^{i+1} + c_3).$$

The required result then follows from a simple induction. □

A.2 More on the error analysis for the abstract numerical approximation

The regression error will depend on the regularity of v^m. Here we prove that $v^m(t,x)$ is Hölder in t and Lipschitz in x under additional conditions, and provide some estimates on the corresponding coefficients. Given $\phi : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, denote

$$[\phi]_{t_i} := \sup_{(t,x) \neq (t',x') \in [t_i, t_{i+1}] \times \mathbf{X}} \frac{|\phi(t,x) - \phi(t',x')|}{|t-t'|^{\frac{1}{2}} + |x - x'|}.$$

Since (μ, σ) is assumed to be Lipschitz, it is clear that there exists $L_X > 0$ such that for all $(t,x), (t',x') \in [0,T] \times \mathbf{X}$,

$$\|X_t^x - X_{t'}^{x'}\|_{\mathbf{L}^2} \leq L_X \Big(\sqrt{|t'-t|} + |x'-x| \Big). \tag{A.8}$$

Proposition A.5. *Suppose that $x \mapsto g(x)$ and $x \mapsto f_{\ell_\circ}(x,y,y')$ are uniformly Lipschitz with Lipschitz constants L_g and L_f. Let β and $\lambda_1, \lambda_2 > 0$ such that $\frac{L_2}{\lambda_2^2} T < 1$ and $\beta \geq 2L_1 + L_f \lambda_1^2 + L_2 \lambda_2^2$, then for all $m \geq 1$ and $i \leq N_h$,*

$$[v^m]_{t_i} \leq L_v := (1 + L_X) L_X \sqrt{\big(L_g^2 + \frac{L_f}{\beta \lambda_1^2}\big) e^{\beta T} / \big(1 - \frac{L_2}{\lambda_2^2} T\big)}$$
$$+ 2(1 + \ell_\circ) C_\ell (1 \vee (M_{h_\circ})^{\ell_\circ}) \sqrt{h_\circ}.$$

Proof. For ease of notations, we provide the proof for $t = 0$ only.
i) Let $x_1, x_2 \in \mathbb{R}^d$ and $Y^{m,1} := v^m(\cdot, X^{x_1})$, $Y^{m,2} := v^m(\cdot, X^{x_2})$, and denote $\Delta Y^m := Y^{m,1} - Y^{m,2}$, $\Delta X := X^{x_1} - X^{x_2}$, where X^{x_1} (resp. X^{x_2}) denotes the solution of SDE (2.1) with initial condition $X_0 = x_1$ (resp. $X_0 = x_2$). Using the same arguments as in the proof of Theorem 2.3, it follows that, for any $\beta \geq 2L_1 + L_f\lambda_1^2 + L_2\lambda_2^2$, one has

$$\mathbb{E}[e^{\beta t}(\Delta Y_t^{m+1})^2] \;\leq\; \mathbb{E}[e^{\beta T}(\Delta Y_T^{m+1})^2] \;+\; \frac{L_f}{\lambda_1^2}\mathbb{E}\Big[\int_t^T e^{\beta s}|\Delta X_s|^2 ds\Big] + \frac{L_2}{\lambda_2^2}\mathbb{E}\Big[\int_t^T e^{\beta s}(\Delta Y_s^m)^2 ds\Big] \tag{A.9}$$

and then

$$\begin{aligned}
\mathbb{E}\Big[\int_0^T e^{\beta t}(\Delta Y_t^{m+1})^2 dt\Big] \;\leq\;& T\mathbb{E}[e^{\beta T}(\Delta Y_T^{m+1})^2] + T\frac{L_f}{\lambda_1^2}\mathbb{E}\Big[\int_0^T e^{\beta s}|\Delta X_s|^2 ds\Big] \\
&+ T\frac{L_2}{\lambda_2^2}\mathbb{E}\Big[\int_0^T e^{\beta t}(\Delta Y_t^m)^2 dt\Big] \\
\leq\;& Te^{\beta T}\Big(L_g^2 + \frac{L_f}{\beta\lambda_1^2}\Big)L_X^2|x_1 - x_2|^2 \\
&+ T\frac{L_2}{\lambda_2^2}\mathbb{E}\Big[\int_0^T e^{\beta t}(\Delta Y_t^m)^2 dt\Big].
\end{aligned}$$

Since $\frac{L_2}{\lambda_2^2}T < 1$, this induces that

$$\mathbb{E}\Big[\int_0^T e^{\beta t}(\Delta Y_t^{m+1})^2 dt\Big] \;\leq\; \frac{Te^{\beta T}\big(L_g^2 + \frac{L_f}{\beta\lambda_1^2}\big)L_X^2|x_1 - x_2|^2}{1 - \frac{L_2}{\lambda_2^2}T}.$$

Plugging the above estimates into (A.9), it follows that

$$(\Delta Y_0^m)^2 \;\leq\; \hat{L}_v^2|x_1 - x_2|^2, \text{ with } \hat{L}_v^2 := \frac{\big(L_g^2 + \frac{L_f}{\beta\lambda_1^2}\big)L_X^2 e^{\beta T}}{1 - \frac{L_2}{\lambda_2^2}T}.$$

ii) For the Hölder property of v^m, it is enough to notice that for $t \leq h_\circ$,

$$\begin{aligned}
|v^m(0,x) - v^m(t,x)| \;\leq\;& \mathbb{E}\Big[|v^m(t, X_t^x) - v^m(t,x)| + \int_0^t |f(X_s^x, Y_s^m, Y_s^{m-1})|ds\Big] \\
\leq\;& \hat{L}_v L_X\sqrt{t} + 2(1+\ell_\circ)C_\ell(1 \vee (M_{h_\circ})^{\ell_\circ})t,
\end{aligned}$$

where the last inequality follows from the Lipschitz property of v^m in x and the fact that Y^m is uniformly bounded by $M_{h_\circ}$. We hence conclude the proof. □

Ackowledgements

This work has benefited from the financial support of the Initiative de Recherche "Méthodes non-linéaires pour la gestion des risques financiers" sponsored by AXA Research Fund.

Bruno Bouchard acknowledges the financial support of ANR Caesar.

Xiaolu Tan acknowledges the financial support of the ERC 321111 Rofirm, the ANR Isotace, and the Chairs Financial Risks (Risk Foundation, sponsored by Société Générale) and Finance and Sustainable Development (IEF sponsored by EDF and CA).

References

[1] K. B. Athreya and P. E. Ney, *Branching processes*, Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[2] V. Bally and P. Pages, *Error analysis of the quantization algorithm for obstacle problems*, Stochastic Processes & Their Applications, 106(1):1-40, 2003.

[3] C. Bender and R. Denk. *A forward scheme for backward SDEs*, Stochastic Processes & Their Applications 117(12): 1793-1812, 2007.

[4] M. Bossy, N. Champagnat, H. Leman, S. Maire, L. Violeau and M. Yvinec, *Monte Carlo methods for linear and non-linear Poisson-Boltzmann equation*, ESAIM:Proceedings and Surveys, 48:420-446, 2015.

[5] B. Bouchard and N. Touzi, *Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations*, Stochastic Process. Appl., 111(2):175-206, 2004.

[6] B. Bouchard, X. Warin, *Monte-Carlo valuation of American options: facts and new algorithms to improve existing methods*. In Numerical methods in finance (pp. 215-255). Springer Berlin Heidelberg, 2012.

[7] J. Cvitanic, I. Karatzas. *Backward stochastic differential equations with reflection and Dynkin games*, The Annals of Probability, 2024-2056, 1996.

[8] N. El Karoui, S. Peng, M.C. Quenez, *Backward stochastic differential equations in finance*, Mathematical finance 7(1), 1-71, 1997.

[9] E. Gobet, J.P. Lemor, X. Warin, *A regression-based Monte Carlo method to solve backward stochastic differential equations*. The Annals of Applied Probability, 15(3):2172-202, 2005.

[10] P. Henry-Labordère, *Cutting CVA's Complexity*, Risk magazine (Jul 2012).

[11] P. Henry-Labordere, N. Oudjane, X. Tan, N. Touzi, X. Warin, *Branching diffusion representation of semilinear PDEs and Monte Carlo approximation*. arXiv preprint, 2016.

[12] P. Henry-Labordere, X. Tan, N. Touzi *A numerical algorithm for a class of BSDEs via the branching process*. Stochastic Processes and their Applications. 28;124(2):1112-1140, 2014.

[13] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability, Vol. 23, Springer, 1992.

[14] H. P. McKean, *Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov*, Comm. Pure Appl. Math., Vol 28, 323-331, 1975.

[15] E. Pardoux and S. Peng, *Adapted solutions of backward stochastic differential equations*, System and Control Letters, 14, 55-61, 1990.

[16] A. Rasulov, G. Raimova, M. Mascagni, *Monte Carlo solution of Cauchy problem for a nonlinear parabolic equation*, Mathematics and Computers in Simulation, 80(6), 1118-1123, 2010.

[17] A.V. Skorokhod *Branching diffusion processes*, Theory of Probability & Its Applications, 9(3):445-449, 1964.

[18] D. W. Stroock, S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Springer, 1979.

[19] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*, American Mathematical Society, Graduate Studies in Mathematics, Volume 140, 2012.

[20] S. Watanabe, *On the branching process for Brownian particles with an absorbing boundary*, J. Math. Kyoto Univ. 4(2), 385-398, 1964.

[21] J. Zhang, *A numerical scheme for backward stochastic differential equations*, Annals of Applied Probability, 14(1), 459-488, 2004.