Numerical approximation of general Lipschitz BSDEs with branching processes

Bruno Bouchard*† Xiaolu Tan*‡ Xavier Warin§¶

October 30, 2017

Abstract

We extend the branching process based numerical algorithm of Bouchard et al. [3], that is dedicated to semilinear PDEs (or BSDEs) with Lipschitz nonlinearity, to the case where the nonlinearity involves the gradient of the solution. As in [3], this requires a localization procedure that uses a priori estimates on the true solution, so as to ensure the well-posedness of the involved Picard iteration scheme, and the global convergence of the algorithm. When, the nonlinearity depends on the gradient, the later needs to be controlled as well. This is done by using a face-lifting procedure. Convergence of our algorithm is proved without any limitation on the time horizon. We also provide numerical simulations 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

§EDF R&D & FiME, Laboratoire de Finance des Marchés de l'Energie

¶warin@edf.fr

1 Introduction

The aim of this paper is to extend the branching process based numerical algorithm proposed in Bouchard et al. [3] to general BSDEs in form:

$$Y_\cdot = g(X_T) + \int_\cdot^T f(X_s, Y_s, Z_s)\, ds - \int_\cdot^T Z_s^\top dW_s, \tag{1.1}$$

where W is a standard d-dimensional Brownian motion, $f : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is the driver function, $g : \mathbb{R}^d \to \mathbb{R}$ is the terminal condition, and X is the solution of

$$X = X_0 + \int_0^\cdot \mu(X_s)\, ds + \int_0^\cdot \sigma(X_s)\, dW_s, \tag{1.2}$$

with constant initial condition $X_0 \in \mathbb{R}^d$ and coefficients $(\mu, \sigma) : \mathbb{R}^d \mapsto \mathbb{R}^d \times \mathbb{R}^{d\times d}$, that are assumed to be Lipschitz[1]. From the PDE point of view, this amounts to solving the parabolic equation

$$\partial_t u + \mu \cdot Du + \frac{1}{2}\mathrm{Tr}[\sigma\sigma^\top D^2 u] + f\big(\cdot, u, Du\sigma\big) = 0,\ \ u(T, \cdot) = g.$$

The main idea of [3] was to approximate the driver function by local polynomials and use a Picard iteration argument so as to reduce the problem to solving BSDE's with (stochastic) global polynomial drivers, see Section 2, to which the branching process based pure forward Monte-Carlo algorithm of [12, 13, 14] can be applied. See for instance [16, 17, 18] for the related Feynman-Kac representation of the KPP (Kolmogorov-Petrovskii-Piskunov) equation.

This algorithm seems to be very adapted to situations where the original driver can be well approximated by polynomials with rather small coefficients on quite large domains. The reason is that, in such a situation, it is basically a pure forward Monte-Carlo method, see in particular [3, Remark 2.10(ii)], which can be expected to be less costly than the classical schemes, see e.g. [1, 4, 5, 11, 21] and the references therein. However, the numerical scheme of [3] only works when the driver function $(x, y, z) \mapsto f(x, y, z)$ is independent of z, i.e. the nonlinearity in the above equation does not depend on the gradient of the solution.

[1]As usual, we could add a time dependency in the coefficients f, μ and σ without any additional difficulty.

Importantly, the algorithm proposed in [3] requires the truncation of the approximation of the Y-component at some given time steps. The reason is that BSDEs with polynomial drivers may only be defined up to an explosion time. This truncation is based on a priori estimates of the true solution. It ensures the well-posedness of the algorithm on an arbitrary time horizon, its stability, and global convergence.
In the case where the driver also depends on the Z component of the BSDE, a similar truncation has to be performed on the gradient itself. It can however not be done by simply projecting Z on a suitable compact set at certain time steps, since Z only maters up to an equivalent class of $([0,T]\times\Omega, dt\times d\mathbb{P})$. Alternatively, we propose to use a face-lift procedure at certain time steps, see (2.10). Again this time steps depend on the explosion times of the corresponding BSDEs with polynomial drivers. Note that a similar face-lift procedure is used in Chassagneux, Elie and Kharroubi [6][2] in the context of the discrete time approximation of BSDEs with contraint on the Z-component.

We prove the convergence of the scheme. The very good performance of this approach is illustrated in Section 4 by a numerical test case.

Notations: All over this paper, we view elements of $\mathbb{R}^d$, $d\geq 1$, as column vectors. Transposition is denoted by the superscript $^\top$. We consider a complete filtered probability space $(\Omega,\mathcal{F},\mathbb{F}=(\mathcal{F}_t)_{t\leq T},\mathbb{P})$ supporting a d-dimensional Brownian motion W. We simply write $\mathbb{E}_t[\cdot]$ for $\mathbb{E}[\cdot|\mathcal{F}_t]$, $t\leq T$. We use the standard notations $\mathbf{S}_2$ (resp. $\mathbf{L}_2$) for the class of progressively measurable processes ξ such that $\|\xi\|_{\mathbf{S}_2}:=\mathbb{E}[\sup_{[0,T]}\|\xi\|^2]^{\frac{1}{2}}$ (resp. $\|\xi\|_{\mathbf{L}_2}:=\mathbb{E}[\int_0^T\|\xi_s\|^2ds]^{\frac{1}{2}}$) is finite. The dimension of the process ξ is given by the context. For a map $(t,x)\mapsto\psi(t,x)$, we denote by $\partial_t\psi$ is derivative with respect to its first variable and by $D\psi$ and $D^2\psi$ its Jacobian and Hessian matrix with respect to its second component.

2 Approximation of BSDE using local polynomial drivers and the Picard iteration

For the rest of this paper, let us consider the (decoupled) forward-backward system (1.1)-(1.2) in which f and g are both bounded and Lipschitz contin-

[2]We are grateful to the authors for several discussions on this subject.

uous, and σ is non-degenerate such that there is a constant $a_0 > 0$ satisfying

$$\sigma\sigma^\top(x) \geq |a_0|^2 \mathrm{I}_d,\ \forall x \in \mathbb{R}^d. \tag{2.1}$$

We also assume that μ, σ, $D\mu$ and $D\sigma$ are all bounded and continuous. In particular, (1.1)-(1.2) has a unique solution $(X, Y, Z) \in \mathbf{S}_2 \times \mathbf{S}_2 \times \mathbf{L}_2$. The above conditions indeed imply that $|Y| + \|\sigma(X)^{-1}Z\| \leq M$ on $[0, T]$, for some $M > 0$.

Remark 2.1. *The above assumptions can be relaxed by using standard localization or mollification arguments. For instance, one could simply assume that g has polynomial growth and is locally Lipschitz. In this case, it can be truncated outside a compacted set so as to reduce to the above. Then, standard estimates and stability results for SDEs and BSDEs can be used to estimate the additional error in a very standard way. See e.g. [7].*

2.1 Local polynomial approximation of the generator

As in [3], our first step is to approximate the driver f by a driver $f_\circ$ that has a local polynomial structure. The difference is that it now depends on both components of the solution of the BSDE. Namely, let

$$f_\circ(x, y, z, y', z') := \sum_{j=1}^{j_\circ} \sum_{\ell \in L} c_{j,\ell}(x) y^{\ell_0} \prod_{q=1}^{q_\circ} (b_q(x)^\top z)^{\ell_q} \varphi_j(y', z'), \tag{2.2}$$

in which $(x, y, z, y', z') \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d$, $L := \{0, \cdots, L_\circ\}^{q_\circ+1}$ (with $L_\circ, q_\circ \in \mathbb{N}$), the functions $(b_q)_{0\leq q\leq q_\circ}$ and $(c_{j,\ell}, \varphi_j)_{\ell\in L, 1\leq j\leq j_\circ}$ (with $j_\circ \in \mathbb{N}$) are continuous and satisfy

$$|c_{j,\ell}| \leq C_L\ ,\ \|b_q\| \leq 1,\ \ |\varphi_j| \leq 1,\ \text{and}\ \|D\varphi_j\| \leq L_\varphi, \tag{2.3}$$

for all $1 \leq j \leq j_\circ$, $0 \leq q \leq q_\circ$ and $\ell \in L$, for some constants $C_L, L_\varphi \geq 0$.
For a good choice of the local polynomial $f_\circ$, we can assume that

$$(x, y, z) \mapsto \bar{f}_\circ\big(x, y, z\big) := f_\circ\big(x, y, z, y, z\big)$$

is globally bounded and Lipschitz. Then, the BSDE

$$\bar{Y}_\cdot = g(X_T) + \int_\cdot^T \bar{f}_\circ(X_s, \bar{Y}_s, \bar{Z}_s)\, ds - \int_\cdot^T \bar{Z}_s^\top\, dW_s, \tag{2.4}$$

has a unique solution $(\bar Y, \bar Z) \in \mathbf{S}_2 \times \mathbf{L}_2$, and standard estimates imply that $(\bar Y, \bar Z)$ provides a good approximation of (Y, Z) whenever $\bar f_\circ$ is a good approximation of f:

$$\|Y - \bar Y\|_{\mathbf{S}_2} + \|Z - \bar Z\|_{\mathbf{L}_2} \le C_\circ \, \|(f - \bar f_\circ)(X, Y, Z)\|_{\mathbf{L}_2}, \tag{2.5}$$

for some $C_\circ > 0$ that depends on the global Lipschitz constant of $\bar f_\circ$ (but not on the precise expression of $\bar f_\circ$), see e.g. [7].
One can think at the $(c_{j,\ell})_{\ell \in L}$ as the coefficients of a polynomial approximation of f in terms of $(y, (b_q(x)^\top z))_{q \le q_\circ}$ on a subset A_j, the A_j's forming a partition of $[-M, M]^{1+d}$. 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.
The choice of the basis functions $(b_q)_{q \le q_\circ}$ as well as $(\varphi_j)_{1 \le j \le j_\circ}$ will obviously depend on the application, but it should in practice typically be constructed such that the sets

$$A_j := \{(y, z) \in \mathbb{R} \times \mathbb{R}^d : \varphi_j(y, z) = 1\} \tag{2.6}$$

are large and the intersection between the supports of the φ_j's are small. See [3, Remark 2.10(ii)] and below. Finally, since the function $\bar f_\circ$ is chosen to be globally bounded and Lipschitz, by possibly adjusting the constant M, we can assume without loss of generality that

$$|\bar Y| + \|\bar Z^\top \sigma(X)^{-1}\| \;\le\; M. \tag{2.7}$$

For later use, let us recall that $\bar Y$ is related to the unique bounded and continuous viscosity solution $\bar u$ of

$$\partial_t \bar u + \mu \cdot D\bar u + \frac{1}{2} \mathrm{Tr}[\sigma\sigma^\top D^2 \bar u] + \bar f_\circ\big(\cdot, \bar u, D\bar u \sigma\big) = 0, \;\; \bar u(T, \cdot) = g,$$

though

$$\bar Y = \bar u(\cdot, X). \tag{2.8}$$

Moreover,

$$\bar u \text{ is bounded by } M \text{ and } M\text{-Lipschitz.}$$

2.2 Picard iteration with truncation and face-lifting

Our next step is to introduce a Picard iteration scheme to approximate the solution $\bar{Y}$ of (2.4) so as to be able to apply the branching process based forward Monte-Carlo approach of [12, 13, 14] to each iteration: given $(\bar{Y}^{m-1}, \bar{Z}^{m-1})$, use the representation of the BSDE with driver $f_\circ(X, \cdot, \cdot, \bar{Y}^{m-1}, \bar{Z}^{m-1})$. However, although the map $(y,z) \mapsto \bar{f}_\circ(x,y,z) = f_\circ(x,y,z,y,z)$ is globally Lipschitz, the map $(y,z) \mapsto f_\circ(x,y,z,y',z')$ is a polynomial, given fixed (x,y',z'), and hence only locally Lipschitz in general. In order to reduce to a Lipschitz driver, we need to truncate the solution at certain time steps, that are smaller than the natural explosion time of the corresponding BSDE with (stochastic) polynomial driver. As in [3], it can be performed by a simple truncation at the level of the first component of the solution. As for the second component, that should be interpreted as a gradient, a simple truncation does not make sense, the gradient needs to be modified by modifying the function itself. Moreover, from the BSDE viewpoint, Z is only defined up to an equivalent class on $([0,T] \times \Omega, dt \times d\mathbb{P})$, so that changing its value at a finite number of given times does not change the corresponding BSDE. We instead use a face-lifting procedure, as in [6].

More precisely, let us define the operator $\mathcal{R}$ on the space of bounded functions $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ by

$$\mathcal{R}[\psi] := (-M) \vee \left[\sup_{p \in \mathbb{R}^d} \big(\psi(\cdot + p) - \delta(p) \big) \right] \wedge M,$$

where

$$\delta(p) := \sup_{q \in [-M,M]^d} (p^\top q) = M \sum_{i=1}^{d} |p_i|.$$

The operation $\psi \mapsto \sup_{p \in \mathbb{R}^d} \big(\psi(\cdot + p) - \delta(p) \big)$ maps ψ into the smallest M-Lipschitz function above ψ. This is the so-called face-lifting procedure, which has to be understood as a (form of) projection on the family of M-Lipschitz functions, see e.g. [2, Exercise 5.2], see also Remark 2.2 below. The outer operations in the definition of $\mathcal{R}$ are just a natural projection on $[-M, M]$.

Let now $(h_\circ, M_{h_\circ})$ be such that (A.3) and (A.4) in the Appendix hold. The constant $h_\circ$ is a lower bound for the explosion time of the BSDE with driver $(y,z) \mapsto f_\circ(x,y,z,y',z')$ for any fixed (x,y',z'). Let us then fix $h \in (0, h_\circ)$

such that $N_h := T/h \in \mathbb{N}$, and define

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

Our algorithm consists in using a Picard iteration scheme to solve (2.4), which re-localize the solution at each time step of the grid $\mathbb{T}$ by applying operator $\mathcal{R}$.

Namely, using the notation $X^{t,x}$ to denote the solution of (1.2) on $[t, T]$ such that $X_t^{t,x} = x \in \mathbb{R}^d$, we initialize our Picard scheme by setting

$$\begin{aligned} &Y_T^{x,0} = \bar{Y}_T^{x,0} = g(x) \\ &(Y^{x,0}, Z^{x,0}) = (\bar{Y}^{x,0}, \bar{Z}^{x,0}) = (\mathrm{y}, D\mathrm{y})(\cdot, X^{t_i,x}) \ \text{ on } [t_i, t_{i+1}),\ i \le N_h - 1, \end{aligned}$$

in which y is a continuous function, M-Lipschitz in space, continuously differentiable in space on $[0, T) \times \mathbb{R}^d$ and such that $|\mathrm{y}| \le M$ and $\mathrm{y}(T, \cdot) = g$. Then, given $(\bar{Y}^{x,m-1}, \bar{Z}^{x,m-1})$, for $m \ge 1$, we define $(\bar{Y}^{x,m}, \bar{Z}^{x,m})$ as follows:

1. For $i = N_h$, set $\bar{\mathrm{u}}_{t_i}^m = \bar{\mathrm{u}}_T^m := g$
2. For $i < N_h$, given $(\bar{Y}^{x,m-1}, \bar{Z}^{x,m-1})$:
   (a) Let $(Y_\cdot^{x,m}, Z_\cdot^{x,m})$ be the unique solution on $[t_i, t_{i+1})$ of

   $$\begin{aligned} Y_\cdot^{x,m} &= \bar{\mathrm{u}}_{t_{i+1}}^m (X_{t_{i+1}}^{t_i,x}) \\ &\quad + \int_\cdot^{t_{i+1}} f_\circ\big(X_s^{t_i,x}, Y_s^{x,m}, Z_s^{x,m}, \bar{Y}_s^{x,m-1}, \bar{Z}_s^{x,m-1}\big) ds \\ &\quad - \int_\cdot^{t_{i+1}} (Z_s^{x,m})^\top dW_s. \end{aligned} \tag{2.9}$$

   (b) Let $\mathrm{u}_{t_i}^m : x \in \mathbb{R}^d \mapsto Y_{t_i}^{x,m}$, and set

   $$\bar{\mathrm{u}}_{t_i}^m := \mathcal{R}[\mathrm{u}_{t_i}^m]. \tag{2.10}$$

   (c) Set $\bar{Y}^{x,m} := Y^{x,m}$ on $(t_i, t_{i+1}]$, $\bar{Y}_{t_i}^{x,m} := \bar{\mathrm{u}}_{t_i}^m(x)$, and $\bar{Z}^{x,m} := Z^{x,m}$ on $[t_i, t_{i+1})$, for $x \in \mathbb{R}^d$.
3. We finally define $\bar{Y}_T^m = g(X_T)$ and

   $$(\bar{Y}^m, \bar{Z}^m) := (\bar{Y}^{X_{t_i},m}, \bar{Z}^{X_{t_i},m}) \ \text{ on } [t_i, t_{i+1}),\ i \le N_h. \tag{2.11}$$

In above, the existence and uniqueness of the solution $(Y^{x,m}, Z^{x,m})$ to (2.9) is ensured by Proposition A.1. The projection operation in (2.10) is consistent with the behavior of the solution of (2.4), recall (2.7), and it is crucial to control the explosion of $(\bar{Y}^m, \bar{Z}^m)$ and therefore to ensure both the stability and the convergence of the scheme. This procedure is non-expansive, as explained in the following Remark, and therefore can not alter the convergence of the scheme.

Remark 2.2. *Let ψ, ψ' be two measurable and bounded maps on $\mathbb{R}^d$. Then, $\sup_{p\in\mathbb{R}^d} |\psi(\cdot + p) - \delta(p) - \psi'(\cdot + p) + \delta(p)| = \sup_{x\in\mathbb{R}^d} |\psi(x) - \psi'(x)|$, and therefore $\|\mathcal{R}[\psi] - \mathcal{R}[\psi']\|_\infty \le \|\psi - \psi'\|_\infty$. In particular, since $\bar{u}$ defined through* (2.8) *is M-Lipschitz in its space variable and bounded by M, we have $\bar{u}(t,\cdot) = \mathcal{R}[\bar{u}(t,\cdot)]$ for $t \le T$ and therefore*

$$\|\mathcal{R}[\psi] - \bar{u}(t,\cdot)\|_\infty \le \|\psi - \bar{u}(t,\cdot)\|_\infty$$

for all $t \le T$ and all measurable and bounded map ψ.

Also note that, if we had $(\bar{Y}_t^{m-1}, \bar{Z}_t^{m-1}) \in A_j$ if and only if $(\bar{Y}_t, \bar{Z}_t) \in A_j$, for all $j \le j_\circ$, then we would have $(\bar{Y}^{m-1}, \bar{Z}^{m-1}) = (\bar{Y}, \bar{Z})$, recall (2.6) and the definition of $\bar{f}_\circ$ in terms of $f_\circ$. This means that we do not need to be very precise on the original prior, whenever the sets A_j can be chosen to be large.

From the theoretical viewpoint, the error due to the above Picard iteration scheme can be deduced from classical arguments. Recall that $(h_\circ, M_{h_\circ})$ is such that (A.3) and (A.4) in the Appendix hold.

Theorem 2.1. *For each $m \ge 0$, the algorithm defined in* 1.-2.-3. *above provides the unique solution $(\bar{Y}^m, \bar{Z}^m) \in \mathbf{S}_2 \times \mathbf{L}_2$. Moreover, it satisfies $|\bar{Y}^m| \vee \|\bar{Z}^{m\top}\sigma(X)^{-1}\| \le M_{h_\circ}$, and there exists a measurable map $(\bar{u}^m, \bar{v}^m) : [0,T] \times \mathbb{R}^d \mapsto \mathbb{R}^{1+d}$, that is continuous on $\cup_{i<N_h}(t_i, t_{i+1}) \times \mathbb{R}^d$, such that $\bar{u}^m(t_i,\cdot)$ is continuous on $\mathbb{R}^d$ for all $i \le N_h$, and*

$$\begin{aligned} \bar{Y}^m &= \bar{u}^m(\cdot, X) \text{ on } [0,T]\ \mathbb{P}\text{-a.s.} \\ \bar{Z}^m &= \bar{v}^m(\cdot, X)\ dt \times d\mathbb{P}\text{-a.e. on } [0,T] \times \Omega. \\ \bar{v}^{m\top} &= D\bar{u}^m \sigma \text{ on } \cup_{i<N_h}(t_i, t_{i+1}) \times \mathbb{R}^d. \end{aligned} \tag{2.12}$$

Moreover, for any constant $\rho \in (0,1)$, there is some constant $C_\rho > 0$ such that

$$|\bar{Y}_t^m - \bar{Y}_t|^2 + \mathbb{E}_t[\int_t^T \|\bar{Z}_s^m - \bar{Z}_s\|^2 ds] \ \le\ C_\rho\, \rho^m, \quad \text{for all } t \le T.$$

Proof. i) Recall from Remark 2.2 that $\mathcal{R}$ maps bounded functions into M-Lipschitz functions that are bounded by M. Then, by Proposition A.1 in the Appendix, the solutions $(Y^{x,m}, Z^{x,m})$ as well as $(\bar{Y}^{x,m}, \bar{Z}^{x,m})$ are uniquely defined in and below (2.9). Moreover, one has $|\bar{Y}^{x,m}_t| \vee \|(\bar{Z}^{x,m}_t)^\top \sigma(X^{t_i,x}_t)^{-1}\| \le M_{h_\circ}$, for all $x \in \mathbb{R}^d$, $i < N_h$ and $t \in [t_i, t_{i+1})$. As a consequence, $(\bar{Y}^m, \bar{Z}^m)$ is also uniquely defined and satisfies $|\bar{Y}^m| \vee \|\bar{Z}^{m\top}\sigma(X)^{-1}\| \le M_{h_\circ}$. Using again Proposition A.1, one has the existence of $(\bar{u}^m, \bar{v}^m)$ satisfying the condition in the statement.

ii) We next prove the convergence of the sequence $(\bar{Y}^m, \bar{Z}^m)_{m\ge 0}$ to $(\bar{Y}, \bar{Z})$. Since $\{(\bar{Y}^{x,m}, \bar{Z}^{x,m}), x \in \mathbb{R}^d\}$ is uniformly bounded, the generator $f_\circ(x, y, z, y', z')$ in (2.9) can be considered to be uniformly Lipschitz in (y, z) and (y', z'). Assume that the corresponding Lipschitz constants are L_1 and L_2.

Let us set $\Theta^{x,m} := (Y^{x,m}, Z^{x,m})$ and define $(\Delta Y^{x,m}, \Delta Z^{x,m}) := (Y^{x,m} - \bar{Y}^x, Z^{x,m} - \bar{Z}^x)$ where $\bar{\Theta}^x := (\bar{Y}^x, \bar{Z}^x)$ denotes the solution of

$$\bar{Y}^x_\cdot = \bar{u}(t_{i+1}, X^{t_i,x}_{t_{i+1}}) + \int_\cdot^{t_{i+1}} \bar{f}_\circ\big(X^{t_i,x}_s, \bar{Y}^x_s, \bar{Z}^x_s\big) ds - \int_\cdot^{t_{i+1}} (\bar{Z}^x_s)^\top dW_s,$$

on each $[t_i, t_{i+1}]$, recall (2.8). In the following, we fix $\beta > 0$ and use the notation

$$\|\xi\|_{\beta,t} := \mathbb{E}_t\Big[\int_t^{t_{i+1}} e^{\beta s}|\xi_s|^2 ds\Big]^{\frac{1}{2}} \ \text{ for } \xi \in \mathbf{L}_2,\ t \in [t_i, t_{i+1}).$$

Fix $t \in [t_i, t_{i+1})$. By applying Itô's formula to $(e^{\beta s}(\Delta Y^{x,m+1}_s)^2)_{s\in[t,t_{i+1}]}$ and then taking expectation, we obtain

$$\begin{aligned}
&e^{\beta t}|\Delta Y^{x,m+1}_t|^2 + \beta\|\Delta Y^{x,m+1}\|^2_{\beta,t} + 2\|\Delta Z^{x,m+1}\|^2_{\beta,t} \\
&\le \mathbb{E}_t\big[e^{\beta t_{i+1}}(\Delta Y^{x,m+1}_{t_{i+1}-})^2\big] \\
&+ 2\mathbb{E}_t\Big[\int_t^{t_{i+1}} e^{\beta s}\Delta Y^{x,m+1}_s\big(f_\circ(X^{t_i,x}_s, \Theta^{x,m+1}_s, \Theta^{x,m}_s) - f_\circ(X^{t_i,x}_s, \bar{\Theta}^x_s, \bar{\Theta}^x_s)\big)ds\Big].
\end{aligned}$$

Using the Lipschitz property of $f_\circ$ and the inequality $\lambda + \frac{1}{\lambda} \ge 2$ for all $\lambda > 0$, it follows that, for all $\lambda_1, \lambda_2 > 0$,

$$\begin{aligned}
&e^{\beta t}|\Delta Y^{x,m+1}_t|^2 + (\beta - (2L_1 + \lambda_1 L_1 + \lambda_2 L_2))\|\Delta Y^{x,m+1}\|^2_{\beta,t} \\
&+ (2 - \frac{L_1}{\lambda_1})\|\Delta Z^{x,m+1}\|^2_{\beta,t} \\
&\le \mathbb{E}_t\big[e^{\beta t_{i+1}}|\Delta Y^{x,m+1}_{t_{i+1}-}|^2\big] + \frac{L_2}{\lambda_2}\|\Delta Y^{x,m}\|^2_{\beta,t} + \frac{L_2}{\lambda_2}\|\Delta Z^{x,m}\|^2_{\beta,t}.
\end{aligned} \tag{2.13}$$

iii) Let us now choose $1 > \rho = \rho_0 > \rho_1 > \cdots > \rho_{N_h} > 0$ such that

$$(m+1)e^{\beta h} \le \big(\frac{\rho_k}{\rho_{k+1}}\big)^{m+1} \text{ for all } m \ge 0. \tag{2.14}$$

For $i = N_h - 1$, we have $\Delta Y^{x,m}_{t_{i+1}-} = 0$ for all $m \ge 1$. Choosing λ_1, λ_2 and $\beta > 0$ in (2.13) such that

$$\frac{L_2}{\lambda_2}\frac{1}{\beta - (2L_1 + \lambda_1 L_1 + \lambda_2 L_2)} \le \rho_{N_h} \text{ and } \frac{L_2}{\lambda_2}\frac{1}{2 - L_1/\lambda_1} \le \rho_{N_h},$$

it follows from (2.13) that, for $t \in [t_{N_h-1}, t_{N_h})$, $\|\Delta Y^{x,m+1}\|^2_{\beta,t} + \|\Delta Z^{x,m+1}\|^2_{\beta,t} \le C(\rho_i)^{m+1}$, for $m \ge 0$, where

$$C := \operatorname{esssup} \sup_{(s,x')\in[0,T]\times\mathbb{R}^d} e^{\beta T}\big(|\Delta Y^{x',0}_s|^2 + \|\Delta Z^{x',0}_s\|^2\big) < \infty,$$

and then, by (2.13) again,

$$|\Delta Y^{x,m+1}_t|^2 \le C(\rho_i)^{m+1}, \text{ for } t \in [t_i, t_{i+1}),\ i = N_h - 1,\ m \ge 0.$$

Recalling Remark 2.2, this shows that

$$|\bar{Y}^{x,m+1}_t - \bar{Y}^x_t|^2 \le C_i(\rho_i)^{m+1}, \text{ for } t \in [t_i, t_{i+1}),\ i = N_h - 1,\ m \ge 0, \tag{2.15}$$

in which

$$C_{N_h-1} := C.$$

Assume now that (2.15) holds true for $i + 1 \le N_h$ and some given $C_{i+1} > 0$. Recall that $\rho_i \ge \rho_{N_h}$. Applying (2.13) with the above choice of λ_1, λ_2 and β, we obtain

$$\begin{aligned}\|\Delta Y^{x,m+1}\|^2_{\beta,t} + \|\Delta Z^{x,m+1}\|^2_{\beta,t} &\le e^{\beta h} C_{i+1}(\rho_{i+1})^{m+1} \\ &\quad + \rho_i\big(\|\Delta Y^{x,m}\|^2_{\beta,t} + \|\Delta Z^{x,m}\|^2_{\beta,t}\big),\end{aligned}$$

which, by (2.14) and the fact that $\rho_i < 1$, induces that

$$\begin{aligned}\|\Delta Y^{x,m+1}\|^2_{\beta,t} + \|\Delta Z^{x,m+1}\|^2_{\beta,t} &\le (m+1)e^{\beta h} C_{i+1}(\rho_{i+1})^{m+1} \\ &\quad + (\rho_i)^{m+1}\big(\|\Delta Y^{x,0}\|^2_{\beta,t} + \|\Delta Z^{x,0}\|^2_{\beta,t}\big) \\ &\le C'_i\ (\rho_i)^{m+1}\end{aligned} \tag{2.16}$$

where

$$C'_i := C_{i+1} + C.$$

Let us further choose $\lambda_2 > 0$ such that $L_2/\lambda_2 \le \rho_{N_h}$, and recall that $\rho_i \ge \rho_{N_h}$. Then, using again (2.13), (2.14), (2.15) applied to $i+1$, we obtain, for $t \in [t_i, t_{i+1})$,

$$|\Delta Y_t^{x,m+1}|^2 \le e^{\beta h} C_{i+1} (\rho_{i+1})^{m+1} + C'_i (\rho_i)^{m+1} \le C_i (\rho_i)^{m+1},$$

so that it follows from Remark 2.2 that

$$|\bar{Y}_t^{x,m+1} - \bar{Y}_t^x|^2 \le C_i (\rho_i)^{m+1}, \tag{2.17}$$

where

$$C_i := e^{\beta h} C_{i+1} + C'_i.$$

Since $(\bar{Y}, \bar{Z}) = (\bar{Y}^{X_{t_i}}, \bar{Z}^{X_{t_i}})$ and $(\bar{Y}^m, \bar{Z}^m) = (\bar{Y}^{X_{t_i},m}, \bar{Z}^{X_{t_i},m})$ on each $[t_i, t_{i+1})$, this concludes the proof. □

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

We now explain how the solution of (2.9) on $[t_i, t_{i+1})$ can be represented by means of a branching diffusion system. We slightly adapt the arguments of [13].

Let us consider an element $(p_\ell)_{\ell \in L} \subset \mathbb{R}_+$ such that $\sum_{\ell \in L} p_\ell = 1$, set $K_n := \{(1, k_2, \ldots, k_n) : (k_2, \ldots, k_n) \in \{1, \ldots, (q_\circ + 1)L_\circ\}^{n-1}\}$ for $n \ge 1$, and $K := \cup_{n \ge 1} K_n$. Let $(W^k)_{k \in K}$ be a sequence of independent d-dimensional Brownian motions, $(\xi_k = (\xi_{k,q})_{0 \le q \le q_\circ})_{k \in K}$ and $(\delta_k)_{k \in K}$ be two sequences of independent random variables, such that

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

and

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

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

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

Given the above, we construct particles $X^{(k)}$ that have the dynamics (1.2) up to a killing time T_k at which they split in $\|\xi_k\|_1 := \xi_{k,0} + \cdots \xi_{k,q_\circ}$ different (conditionally) independent particles with dynamics (1.2) 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

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

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

$$W^{(k \oplus j)} := \left(W^{(k)}_{\cdot \wedge T_k} + W^{k \oplus j}_{\cdot \vee T_k} - W^{k \oplus j}_{T_k} \right) \mathbf{1}_{[0,T_{k \oplus j}]}, \ 1 \leq j \leq \|\xi_k\|_1,$$

in which we use the notation

$$(1, k_1, \ldots, k_{n-1}) \oplus j := (1, k_1, \ldots, k_{n-1}, j),$$

and

$$\bar{\mathcal{K}}_t^n := \{k \oplus j : k \in \bar{\mathcal{K}}_T^{n-1}, 1 \leq j \leq \|\xi_k\|_1 \text{ s.t. } t \in (0, T_{k \oplus j}]\}, \ \ \bar{\mathcal{K}}_t := \cup_{n \geq 1} \bar{\mathcal{K}}_t^n,$$

$$\mathcal{K}_t^n := \{k \oplus j : k \in \bar{\mathcal{K}}_T^{n-1}, 1 \leq j \leq \|\xi_k\|_1 \text{ s.t. } t \in (T_k, T_{k \oplus j}]\}, \ \ \mathcal{K}_t := \cup_{n \geq 1} \mathcal{K}_t^n.$$

Now observe that the solution X^x of (1.2) on $[0, T]$ with initial condition $X_0^x = x \in \mathbb{R}^d$ 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 $X^{x,(k)} := \Phi[x](\cdot, W^{(k)})$. Moreover, we define the $d \times d$-dimensional matrix valued tangent process $\nabla X^{x,(k)}$ defined on $[\![T_{k-}, T_k]\!]$ by

$$\nabla X^{x,(k)}_{T_{k-}} = I_d \tag{3.1}$$

$$d\nabla X_t^{x,(k)} = D\mu(X_t^{x,(k)}) \nabla X_t^{x,(k)} dt + \sum_{i=1}^{d} D\sigma_i(X_t^{x,(k)}) \nabla X_t^{x,(k)} dW_t^{(k),i},$$

where I_d denotes the $d \times d$-dimensional identity matrix, and σ_i denotes the i-th column of matrix σ.

Finally, we give a mark 0 to the initial particle (1), and, for every particle k, knowing $\xi_k = (\xi_{k,0}, \cdots, \xi_{k,q_\circ})$, we consider its offspring particles $(k \oplus j)_{j=1,\cdots,\|\xi_k\|_1}$ and give the first $\xi_{k,0}$ particles the mark 0, the next $\xi_{k,1}$ particles the mark 1, the next $\xi_{k,2}$ particles the mark 2, etc. Thus, every particles k carries a mark θ_k taking values in 0 to $q_\circ$.

Given the above construction, we can now provide the branching process based representation of $(\bar{Y}^m)_{m\geq 0}$. We assume here that $(\bar{u}^{m-1}, \bar{v}^{m-1})$ defined in (2.12) are given for some $m \geq 1$, recall that $(\bar{u}^0, \bar{v}^0) = (\mathrm{y}, D\mathrm{y})$ by construction. We set $\tilde{u}^m(T, \cdot) = g$. Then, for $i = N_h - 1, \cdots, 0$, we define $(\tilde{u}^m, \tilde{v}^m)$ on each interval $[t_i, t_{i+1})$ recursively by

$$
\begin{aligned}
\tilde{u}^m(t,x) &:= \mathbb{E}\left[U^m_{t,x}\right]\mathbf{1}_{t\neq t_i} + \mathcal{R}[\mathbb{E}\left[U^m_{t,\cdot}\right]](x)\mathbf{1}_{t=t_i} \\
\tilde{v}^m(t,x) &:= \mathbb{E}\big[V^m_{t,x}\big]\sigma(x),
\end{aligned}
\tag{3.2}
$$

for $(t,x) \in [t_i, t_{i+1}) \times \mathbb{R}^d$, in which

$$
\begin{aligned}
U^m_{t,x} &:= \Big[\prod_{k\in\mathcal{K}_{t_{i+1}-t}} G^m_{t,x}(k)\mathcal{W}_{t,x}(k)\Big]\Big[\prod_{k\in\bar{\mathcal{K}}_{t_{i+1}-t}\setminus\mathcal{K}_{t_{i+1}-t}} A^m_{t,x}(k)\mathcal{W}_{t,x}(k)\Big] \\
V^m_{t,x} &:= U^m_{t,x}\,\frac{1}{T_{(1)}}\int_0^{T_{(1)}} \left[\sigma^{-1}(X^{x,(1)}_s)\nabla X^{x,(1)}_s\right]^\top dW^{(1)}_r
\end{aligned}
$$

where

$$
G^m_{t,x}(k) := \frac{\tilde{u}^m\big(t_{i+1}, X^{x,(k)}_{t_{i+1}-t}\big) - \tilde{u}^m\big(t_{i+1}, X^{x,(k-)}_{T_{k-}}\big)\mathbf{1}_{\{\theta(k)\neq 0\}}}{\bar{F}(t_{i+1}-t-T_{k-})},
$$

$$
A^m_{t,x}(k) := \frac{\sum_{j=1}^{j_\circ} c_{j,\xi_k}\big(X^{x,(k)}_{T_k}\big)\varphi_j\big((\bar{u}^{m-1}, \bar{v}^{m-1})(t+T_k, X^{x,(k)}_{T_k})\big)}{p_{\xi_k}\,\rho(\delta_k)},
$$

$$
\mathcal{W}_{t,x}(k) := \mathbf{1}_{\{\theta_k=0\}} + \frac{\mathbf{1}_{\{\theta_k\neq 0\}}}{T_k - T_{k-}} b_{\theta_k}(X^{x,(k)}_{T_k}) \cdot \int_{T_{k-}}^{T_k} \left[\sigma^{-1}(X^{x,(k)}_s)\nabla X^{x,(k)}_s\right]^\top dW^{(k)}_r.
$$

Compare with [13, (3.4) and (3.10)].

The next proposition shows that $(\tilde{u}^m, \tilde{v}^m)$ actually coincides with $(\bar{u}^m, \bar{v}^m)$ in (2.12), a result that follows essentially from [13]. Nevertheless, to be more precise on the square integrability of $U^m_{t,x}$ and $V^m_{t,x}$, one will fix a special density function ρ as well as probability weights $(p_\ell)_{\ell\in L}$. Recall again that $(\bar{u}^m, \bar{v}^m)$ are defined in Theorem 2.1 and satisfy (2.12), and that $(h_\circ, M_\circ)$ are chosen such that (A.3)-(A.4) in the Appendix hold.

Proposition 3.1. *Let us choose* $\rho(t) = \frac{1}{3}t^{-2/3}\mathbf{1}_{\{t\in[0,1]\}}$ *and* $p_\ell = \frac{\|c_\ell\|_\infty}{\|c\|_{1,\infty}}$ *with* $\|c\|_{1,\infty} := \sum_{\ell\in L}\|c_\ell\|_\infty$*. Let* $h'_\circ$ *and* $M_{h'_\circ}$ *be given by* (A.8) *and* (A.9)*. Assume that* $h \in (0, h_\circ \wedge h'_\circ)$*. Then,*

$$\mathbb{E}[|U^m_{t,x}|^2] \vee \mathbb{E}[\|V^m_{t,x}\|^2] \le (M_{h'_\circ})^2, \quad \textit{for all } m \ge 1 \textit{ and } (t,x) \in [0,T]\times\mathbb{R}^d.$$

Moreover, $\tilde{u}^m = \bar{u}^m$ *on* $[0,T]\times\mathbb{R}^d$ *and* $\tilde{v}^m = \bar{v}^m$ *on* $\cup_{i\le N_h-1}(t_i,t_{i+1})\times\mathbb{R}^d$*.*

The proof of the above mimics the arguments of [13, Theorem 3.12] and is postponed to the Appendix A.2.

Remark 3.1. *The integrability and representation results in Proposition 3.1 hold true for a large class of parameters* ρ*,* $(p_\ell)_{\ell\in L}$ *and* $(c_\ell)_{\ell\in L}$ *(see e.g. [13, Section 3.2] for more details). We restrict to a special choice of parameters in Proposition 3.1 in order to compute explicitly the lower bound* $h'_\circ$ *for the explosion time as well as the upper bound* $M_{h'_\circ}$ *for the variance of the estimators.*

Remark 3.2. *a. The above scheme requires the computation of conditional expectations that involve the all path* $(\tilde{u}^m,\tilde{v}^m)(\cdot,X^{t_i,x})$ *on each* (t_i,t_{i+1})*, for all* $x\in\mathbb{R}^d$*, and therefore the use of an additional time space grid on which* $\tilde{u}^m$ *and* $\tilde{v}^m$ *are estimated by Monte-Carlo. For* $t_i < t < t_{i+1}$*, the precision does not need to be important because the corresponding values are only used for the localization procedure, see the discussion just after Remark 2.2, and the grid does not need to be fine. The space grid should be finer at the times in* $\mathbb{T}$ *because each* $\tilde{u}^m(t_i,\cdot)$ *is used per se as a terminal condition on* $(t_i,t_{i+1}]$*, and not only for the localization of the polynomial. This corresponds to the Method B in [3, Section 3]. One can also consider a very fine time grid* $\mathbb{T}$ *and avoid the use of a sub-grid. This is Method A in [3, Section 3]. The numerical tests performed in [3] suggest that Method A is more efficient.*
b. From a numerical viewpoint, $\tilde{v}^m$ *can also be estimated by using a finite difference scheme based on the estimation of* $\tilde{u}^m$*. It seems to be indeed more stable.*
c. Obviously, one can also adapt the algorithm to make profit of the ghost particle or of the normalization techniques described in [20], which seem to reduce efficiently the variance of the estimations, even when ρ *is the exponential distribution.*

4 Numerical example

This section is dedicated to a simple example in dimension one showing the efficiency of the proposed methodology. We use the Method A of [3, Section 3] together with the normalization technique described in [20] and a finite difference scheme to compute $\tilde{v}^m$. In particular, this normalization technique allows us to take ρ as an exponential density rather than that in Proposition 3.1. See Remark 3.2.

We consider the SDE with coefficients

$$\mu(x) = -0.5(x+0.2) \quad \text{and} \quad \sigma(x) = 0.1\left(1 + 0.5((x \vee 0) \wedge 1)\right).$$

The maturity is $T = 1$ and the non linearity in (1.1) is taken as

$$\begin{aligned} f(t,x,y,z) =& \hat{f}(y,z) + \frac{1}{2}e^{\frac{t-T}{2}}\left(\cos(x)\frac{\sigma(x)^2}{2} - \frac{1}{2}(1+\cos(x)) + \mu(x)\sin(x)\right) \\ & - \frac{2}{4 + |\sin(x)(1+\cos(x))|e^{t-T}} \end{aligned}$$

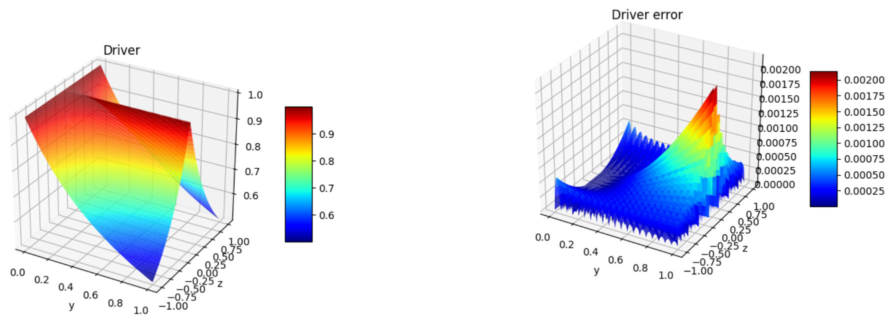
where

$$\hat{f}(y,z) = \frac{1}{2(1+|yz|)}. \tag{4.1}$$

It is complemented by the choice of the terminal condition $g(x) = \frac{1+\cos(x)}{2}$, so that an analytic solution is available :

$$u(t,x) = \frac{1+\cos(x)}{2}e^{\frac{t-T}{2}}.$$

We use the algorithm to compute an estimation of $u(0,\cdot)$ on $\mathbf{X} := [-1., 1.]$. To construct our local polynomials approximation of $\hat{f}$, we use a linear spline interpolation in each direction, obtained by tensorization, and leading to a local approximation on the basis 1, y, z, yz on each mesh of a domain $[0,1] \times [-1,1]$. Figure 1 displays the function $\hat{f}$ and the error obtained by a discretization of 20×20 meshes.



The driver $\hat{f}$

Error on the driver due to the linear spline representation.

Figure 1: The driver $\hat{f}$ and its linear spline representation error for 20×20 splines.

The parameters affecting the convergence of the algorithm are:

- The couple of meshes (n_y, n_z) used in the spline representation of (4.1), where n_y (resp. n_z) is the number of linear spline meshes for the y (resp. z) discretization.

- The number of time steps N_h.

- The grid and the interpolation used on $\mathbf{X}$ at $t_0 = 0$ and for all dates t_i, $i = 1, \ldots, N_h$. Note that the size of the grid has to be adapted to the value of T, because of the diffusive feature of (1.2). All interpolations are achieved with the StOpt library (see [9, 10]) using a modified quadratic interpolator as in [19]. In the following, Δx denotes the mesh of the space discretization.

- The time step is set to 0.002 and we use an Euler scheme to approximate (1.2).

- The accuracy of the estimation of the expectations appearing in our algorithm. We compute the empirical standard deviation θ associated to each Monte Carlo estimation of the expectation in (3.2). We try to fix the number $\hat{M}$ of samples such that $\theta/\sqrt{\hat{M}}$ does not exceed a certain level, fixed at 0.000125, at each point of our grid. We cap this number of simulations at 510^5.

- The intensity, set to 0.4, of the exponential distribution used to define the random variables $(\delta_k)_{k \in K}$.

Finally, we take $M = 1$ in the definition of $\mathcal{R}$.

We only perform one Picard iteration with initial prior $(\tilde{u}^0, \tilde{v}^0) = (g, Dg\sigma)$.

On the different figures below, we plot the errors obtained on $\mathbf{X}$ for different values of N_h, (n_y, n_z) and Δx. We first use 20 time steps and an interpolation step of 0.1 In figure 2, we display the error as a function of the number of spline meshes. We provide two plots:

- On the left-hand side, n_y varies above 5 and n_z varies above 10,
- On the right-hand side, we add $(n_y, n_z) = (5, 5)$. It leads to a maximal error of 0.11, showing that a quite good accuracy in the spline representation in z is necessary.

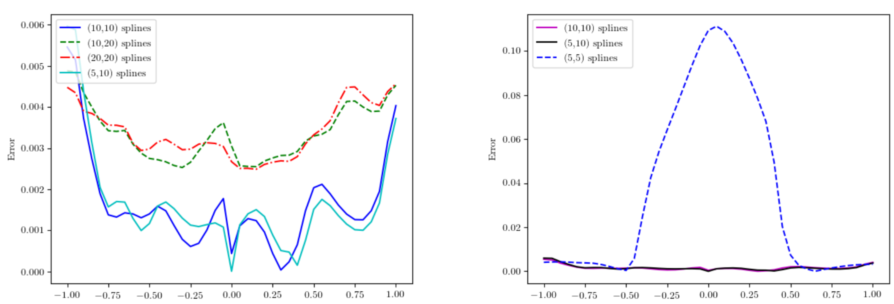


Figure 2: Error plot depending on the couple (n_y, n_z) for 20 time steps, a space discretization $\Delta x = 0.1$.

In figure 3, we plot the error obtained with $(n_y, n_z) = (20, 10)$ and a number of time steps equal to $N_h = 20$, for different values of Δx: the results are remarkably stable with the interpolation space discretization.

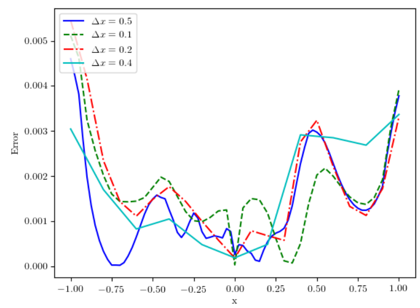


Figure 3: Error plot depending on Δx for $(n_y, n_z) = (20, 10)$, $N_h = 20$.

In Figure 4, we finally let the number of time steps N_h vary. Once again we give two plots:

- one with N_h above or equal to 20,
- one with small values of N_h.

The results clearly show that the algorithm produces bad results when N_h is too small: the time steps are too large for the branching method. In this case, it exhibits a large variance. When N_h is too large, then interpolation errors propagate leading also to a deterioration of our estimations.

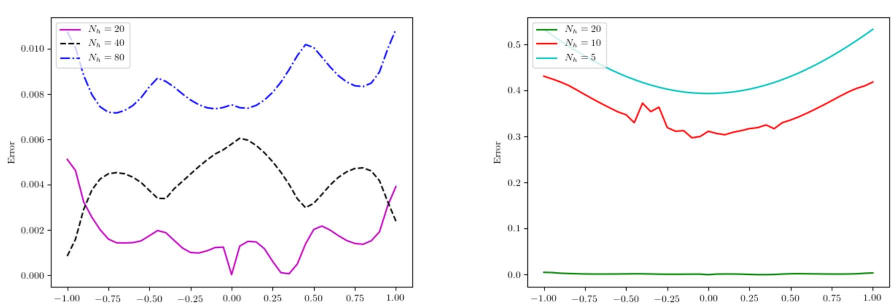


Figure 4: Error plot depending on N_h for $(n_y, n_z) = (20, 10)$, a space discretization $\Delta x = 0.1$

Numerically, it can be checked that the face-lifting procedure is in most of the cases useless when only one Picard iteration is used:

- When the variance of the branching scheme is small, the face-lifting and truncation procedure has no effect,

- When the variance becomes too large, the face-lifting procedure is regularizing the solution and this permits to reduce the error due to our interpolations.

In figure 5, we provide the estimation with and without face-lifting, obtained with $N_h = 10$, $(n_y, n_z) = (20, 10)$ and a space discretization $\Delta x = 0.1$.

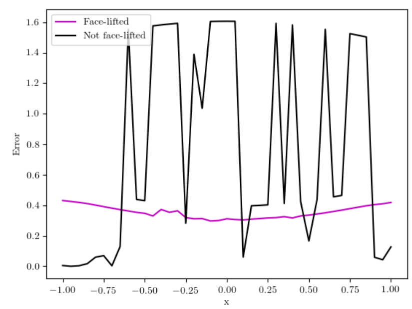


Figure 5: Error plot for $N_h = 10$ for $(n_y, n_z) = (20, 10)$, a space discretization $\Delta x = 0.1$ with and without face-lifting.

For $N_h = 20$ the computational time is equal to 320 seconds on a 3 year old laptop.

A Appendix

A.1 A priori estimates for the Picard iteration scheme

In this section, we let ∇X be the tangent process associated to X on $[0, T]$ by

$$\nabla X_0 = I_d \ , \ d\nabla X_t = D\mu(X_t)\nabla X_t dt + \sum_{i=1}^{d} D\sigma_i(X_t)\nabla X_t dW_t^i,$$

and we define

$$N_s^t := \left(\int_t^s \frac{1}{s-t} (\sigma(X_r)^{-1} \nabla X_r)^\top dW_r \right)^\top \nabla X_t^{-1}$$

for $t \le s \le T$. Standard estimates lead to

$$\mathbb{E}_t[\|N_s^t\|] \le C_{\mu,\sigma}(s-t)^{-\frac{1}{2}} \text{ for } t \le s \le T, \tag{A.1}$$

for some $C_{\mu,\sigma} > 0$ that only depends on $\|\mu\|_\infty$, $\|\sigma\|_\infty$, $\|D\mu\|_\infty$, $\|D\sigma\|_\infty$, a_0 in (2.1) and T. In particular, it does not depend on $\|X_0\|$. Up to changing this constant, we may assume that

$$\mathbb{E}_t[\|\nabla X_s \nabla X_t^{-1}\|] \le e^{C_{\mu,\sigma}(s-t)} \text{ for } t \le s \le T. \tag{A.2}$$

Set

$$\mathcal{D}_{M_{h_\circ}} := \mathbb{R}^d \times \{(y,z) \in \mathbb{R} \times \mathbb{R}^d : |y| \vee \|z^\top \sigma^{-1}\| \le M_{h_\circ}\}^2.$$

and let $M_{h_\circ} \ge M$ and $h_\circ \in (0,T]$ be such that

$$M_{h_\circ} \ge M + \sup_{\mathcal{D}_{M_{h_\circ}}} |f_\circ| \, h_\circ. \tag{A.3}$$

and

$$M_{h_\circ} \ge M e^{C_{\mu,\sigma} h_\circ} + C_{\mu,\sigma} \sup_{\mathcal{D}_{M_{h_\circ}}} |f_\circ| \, (h_\circ)^{\frac{1}{2}} \tag{A.4}$$

The existence of $h_\circ$ and $M_{h_\circ}$ follows from (2.2) and (2.3). Note that they do not depend on X_0.

Proposition A.1. *Let $\tilde{g} : \mathbb{R}^d \mapsto \mathbb{R}$ be bounded by M and M-Lipschitz. Fix $T' \in [0, h_\circ]$. Let $(\tilde{U}, \tilde{V}) : [0,T'] \times \mathbb{R}^d \mapsto \mathbb{R} \times \mathbb{R}^d$ be measurable such that $|\tilde{U}| \vee \|\tilde{V}^\top \sigma^{-1}\| \le M_{h_\circ}$. Then, there exists a unique bounded solution on $[0,T']$ to*

$$\tilde{Y}_\cdot = \tilde{g}(X_{T'}) + \int_\cdot^{T'} f_\circ(X_s, \tilde{Y}_s, \tilde{Z}_s, (\tilde{U}, \tilde{V})(s, X_s)) \, ds - \int_\cdot^{T'} \tilde{Z}_s^\top \, dW_s. \tag{A.5}$$

It satisfies

$$|\tilde{Y}| \vee \|\tilde{Z}^\top \sigma(X)^{-1}\| \le M_{h_\circ} \text{ on } [0,T']. \tag{A.6}$$

Moreover, there exists a bounded continuous map $(\tilde{u}, \tilde{v}) : [0,T'] \times \mathbb{R}^d \mapsto \mathbb{R} \times \mathbb{R}^d$ such that $\tilde{Y} = \tilde{u}(\cdot, X)$ on $[0,T']$ $\mathbb{P}$-a.s. and $\tilde{Z} = \tilde{v}(\cdot, X)$ $dt \times d\mathbb{P}$-a.e. on $[0,T'] \times \Omega$. It satisfies $\tilde{v}^\top = D\tilde{u}\sigma$ on $[0,T')$.

Proof. We construct the required solution by using Picard iterations. We set $(\tilde{Y}^0, \tilde{Z}^0) = (\mathrm{y}, D\mathrm{y})(\cdot, X)$, and define recursively on $[0, T']$ the couple $(\tilde{Y}^{n+1}, \tilde{Z}^{n+1})$ as the unique solution of

$$\tilde{Y}^{n+1}_{\cdot} = \tilde{g}(X_{T'}) + \int_{\cdot}^{T'} f_\circ(X_s, \tilde{Y}^n_s, \tilde{Z}^n_s, (\tilde{U}, \tilde{V})(s, X_s))\, ds - \int_{\cdot}^{T'} (\tilde{Z}^{n+1}_s)^\top\, dW_s,$$

whenever it is well-defined. It is the case for $n = 0$. We now assume that $(\tilde{Y}^n, \tilde{Z}^n)$ is well-defined and such that $|\tilde{Y}^n| \vee \|\sigma(X)^{-1}\tilde{Z}^n\| \leq M_{h_\circ}$ on $[0, T']$ for some $n \geq 0$. Then,

$$|\tilde{Y}^{n+1}_{\cdot}| \leq \|\tilde{g}\|_\infty + \sup_{\mathcal{D}_{M_{h_\circ}}} |f_\circ|\, h_\circ \leq M_{h_\circ},$$

in which we used (A.3) for the second inequality. On the other hand, up to using a mollifying argument, one can assume that $\tilde{g}$ is C^1_b and that (U, V) is Lipschitz. Then, it follows from the same arguments as in [15, Theorem 3.1, Theorem 4.2] that $\tilde{Z}^{n+1}$ admits the representation

$$\begin{aligned}(\tilde{Z}^{n+1}_t)^\top =& \mathbb{E}_t \left[D\tilde{g}(X_{T'})\nabla X_{T'} \nabla X_t^{-1}\right] \sigma(X_t) \\ &+ \mathbb{E}_t \left[\int_t^{T'} f_\circ(X_s, \tilde{Y}^n_s, \tilde{Z}^n_s, (U, V)(s, X_s)) N^t_s ds\right] \sigma(X_t).\end{aligned}$$

By combining the above together with (A.1) and (A.2), we obtain that

$$\|(\tilde{Z}^{n+1}_t)^\top \sigma(X_t)^{-1}\| \leq M e^{C_{\mu,\sigma}(T'-t)} + C_{\mu,\sigma} \sup_{\mathcal{D}_{M_{h_\circ}}} |f_\circ|\, (h_\circ)^{\frac{1}{2}} \leq M_{h_\circ},$$

in which we used (A.4) for the second inequality. The above proves that the sequence $(\tilde{Y}^n, \tilde{Z}^n)_{n\geq 0}$ is uniformly bounded on $[0, T']$. Therefore, we can consider $f_\circ$ as a Lipschitz generator, and hence $(\tilde{Y}^n, \tilde{Z}^n)_{n\geq 0}$ is in fact a Picard iteration that converges to a solution of (A.5) with the same bound.
The existence of the maps $\tilde{u}$ and $\tilde{v}$ such that $\tilde{v}^\top = D\tilde{u}\sigma$ follows from [15, Theorem 3.1] applied to (A.5) when $(\tilde{g}, f_\circ(\cdot, y, z, (U, V)(t, \cdot)))$ is C^1_b, uniformly in $(t, y, z) \in [0, T] \times \mathbb{R}^{d+1}$. The representation result of [15, Theorem 4.2] combined with a simple approximation argument, see e.g. [8, (ii) of the proof of Proposition 3.2], then shows that the same holds on $[0, T')$ under our conditions. □

A.2 Proof of the representation formula

We adapt the proof of [13, Theorem 3.12] to our context. We proceed by induction. In all this section, we fix

$$(t, x) \in [t_i, t_{i+1}) \times \mathbb{R}^d,$$

and assume that the result of Proposition 3.1 holds up to rank $m-1 \geq 0$ on $[0,T]$ (with the convention $U_\cdot^0 = y$, $V_\cdot^0 := Dy$), and up to rank m on $[t_{i+1}, T]$. In particular, we assume that $\tilde{u}^m(t_{i+1}, \cdot) = \bar{u}^m(t_{i+1}, \cdot)$.
We fix $\varrho = 4$ and define

$$C_{1,\varrho} := M^\varrho \vee \sup_{t\leq s,\ x\in\mathbb{R}^d,\ q=1,\cdots,q_\circ,\ \|\xi\|\leq M} \mathbb{E}\Big[\Big|\big(\xi \cdot (X_s^{t,x} - x)\big)\big(b_q(x) \cdot \overline{\mathcal{W}}_{t,x,s}\big)\Big|^\varrho\Big],$$

$$C_{2,\varrho} := \sup_{t\leq s\leq t_{i+1},\ x\in\mathbb{R}^d,\ q=1,\cdots,q_\circ} \mathbb{E}\Big[\big|\sqrt{s-t}\ b_q(x) \cdot \overline{\mathcal{W}}_{t,x,s}\big|^\varrho\Big]$$

where

$$\overline{\mathcal{W}}_{t,x,s} := \frac{1}{s-t}\int_t^s \big[\sigma^{-1}(X_r^{t,x})\nabla X_r^{t,x}\big]^\top dW_r$$

in which $\nabla X^{t,x}$ is the tangent process of $X^{t,x}$ with initial condition I_d at t. We then set

$$\hat{C}_{1,\varrho} := \frac{C_{1,\varrho}}{\bar{F}(T)^{\varrho-1}}, \qquad \hat{C}_{2,\varrho} := C_{2,\varrho}\ j_\circ \sup_{j\leq j_\circ, \ell\in L, t\in(0,h_\circ]} \Big(\frac{\|c_{j,\ell}\|_\infty}{p_\ell}\frac{t^{-\varrho/(2(\varrho-1))}}{\rho(t)}\Big)^{\varrho-1}.$$

Since $\bar{F}$ is non-increasing and $\tilde{u}^m(t_{i+1}, \cdot) = \bar{u}^m(t_{i+1}, \cdot)$ is bounded by M and M-Lipschitz, direct computations imply that

$$\mathbb{E}[|U_{t,x}^m|^\varrho] \vee \mathbb{E}\|V_{t,x}^m\|^\varrho] \leq \mathbb{E}\Big[\Big(\prod_{k\in\mathcal{K}_t} \frac{\hat{C}_{1,\varrho}}{\bar{F}(t-T_{k-})}\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_t\setminus\mathcal{K}_t} \frac{\hat{C}_{2,\varrho}}{p_{\xi_k}\rho(\delta_k)}\Big)\Big]. \tag{A.7}$$

We first estimate the right-hand side, see (A.10) below.
Let us denote by $C_{\rm bdg}$ the constant in the Burkholder-Davis-Gundy inequality such that $\mathbb{E}\big[\sup_{0\leq t\leq T}|M_t|^\varrho\big] \leq C_{\rm bdg}\mathbb{E}[(\langle M\rangle_T)^{\frac{\varrho}{2}}]$ for any continuous martingale M with $M_0 = 0$. Denote further

$$C_0 := (3\times 3)^{\varrho-1}\big(1 + (\varrho/(\varrho-1))^\varrho\big)\big(1 + (1 + |\bar{\lambda}_{D\mu}T|^\varrho e^{C_Q T})\big),$$

where $\bar{\lambda}_{D\mu}$ the largest eigenvalue of the matrix $D\mu$, $\bar{\lambda}_{D\sigma}$ the largest eigenvalue of the matrix $D\sigma_i$, $i \leq d$, and $C_Q := \varrho\bar{\lambda}_{D\mu} + d\varrho(\varrho-1)\bar{\lambda}_{D\sigma}/2$. Define also $\bar{\lambda}_{(\sigma\sigma^\top)^{-1}}$ as the largest eigenvalue of matrix $(\sigma\sigma^\top)^{-1}$.

Lemma A.1. *Under the Assumptions of Proposition 3.1,*

$$\hat{C}_{1,\varrho} \le \hat{C}_1 := 2\Big(1 \vee M \vee 2^{\varrho-1}(M\sqrt{d})^{\varrho}\Big(C_0 + \|\mu\|^{\varrho} T^{\frac{\varrho}{2}} C_{BDG} C_0 \big(\bar{\lambda}_{(\sigma\sigma^\top)^{-1}}\big)^{\frac{\varrho}{2}}\Big)\Big),$$

and

$$\hat{C}_{2,\varrho} \le \hat{C}_2 := C_{\mathrm{bdg}}\ C_0\ \big(\bar{\lambda}_{(\sigma\sigma^\top)^{-1}}\big)^{\frac{\varrho}{2}}.$$

Proof. Let $\tilde{b} \in \mathbb{R}^d$ be a fixed vector. Set $Q^{t,x} := \nabla X^{t,x}\tilde{b}$. Then, it follows from direct computations that

$$\mathbb{E}\Big[\max_{[t,t_{i+1}]} \|Q^{t,x}\|^{\varrho}\Big] \le C_0 \|\tilde{b}\|^{\varrho}.$$

Further, remember that each b_q is assumed to be bounded by 1, so that $\|b_q\sigma^{-1}\|^2$ is uniformly bounded by $\bar{\lambda}_{(\sigma\sigma^\top)^{-1}}$. Then, direct computations lead to

$$C_{1,\varrho} \le 1 \vee M^{\varrho} \vee 2^{\varrho-1}\Big(C_0 + \|\mu\|^{q} T^{\frac{\varrho}{2}} C_{\mathrm{bdg}} C_0 \big(\bar{\lambda}_{(\sigma\sigma^\top)^{-1}}\big)^{\frac{\varrho}{2}}\Big),$$

and

$$C_{2,\varrho} \le C_{\mathrm{bdg}}\ C_0\ \big(\bar{\lambda}_{(\sigma\sigma^\top)^{-1}}\big)^{\frac{\varrho}{2}}.$$

It remains to use our specific choice of ρ and $(p_\ell)_{\ell\in L}$ in Proposition 3.1 to conclude. □

Let us now choose $h'_\circ$ and $M_{h'_\circ}$ such that

$$h'_\circ \ < \ 1 \wedge \frac{\hat{C}_1^{-(|L|-1)}}{(|L|+1)(|L|-1)\hat{C}_2}, \tag{A.8}$$

and

$$(M_{h'_\circ})^4 \ := \ \big(\hat{C}_1^{1-|L|} - h'_\circ(|L|+1)(|L|-1)\hat{C}_2\big)^{(1-|L|)^{-1}}. \tag{A.9}$$

Lemma A.2. *Let the conditions of Proposition 3.1 hold. Then, the ordinary differential equation $\eta'(t) = \sum_{\ell\in L} \hat{C}_2 \eta(t)^{\|\ell\|_1}$ with initial condition $\eta(0) = \hat{C}_1 \ge 1$ has a unique solution on $[0, h'_\circ]$, and it is bounded by $(M_{h'_\circ})^2$. Moreover,*

$$\mathbb{E}\Big[\Big(\prod_{k\in\mathcal{K}_t} \frac{\hat{C}_1}{\bar{F}(t-T_{k-})}\Big)\Big(\prod_{k\in\bar{\mathcal{K}}_t\setminus\mathcal{K}_t} \frac{\hat{C}_2}{p_{\xi_k}\rho(\delta_k)}\Big)\Big] \ \le \ \eta(t) \ \le \ (M_{h'_\circ})^4, \tag{A.10}$$

for all $t \in [0, h'_\circ]$.

Proof. The result follows from exactly the same arguments as in [3, Lemma A.1]. □

We can now conclude the proof of Proposition 3.1.

Proof of Proposition 3.1. In view of (A.7), Lemma A.2 implies that $\{|U^m_{t,x}|^2 + \|V^m_{t,x}\|^2, (t,x) \in [t_i, t_{i+1}) \times \mathbb{R}^d\}$ is uniformly integrable (with a bound that does not depend on $i < N_h$ for $0 < h \leq h'_\circ$). Then, arguing exactly as in [3, Proposition A.2] leads to $\tilde{u}^m = \bar{u}^m$ on $[t_i, t_{i+1})$. Combined with [13, Proposition 3.7], the uniform integrability also implies that $D\tilde{u}^m = \tilde{v}^m \sigma$ on $(t_i, t_{i+1}) \times \mathbb{R}^d$, and one can conclude from Theorem 2.1 that $\tilde{v}^m = \tilde{u}^m$ on $(t_i, t_{i+1}) \times \mathbb{R}^d$. By the induction hypothesis of the beginning of this section, this proves that the statements of Proposition 3.1 hold. □

Remark A.1. *The constants $\hat{C}_1$ and $\hat{C}_2$ (and hence $h'_\circ$ and $M_{h'_\circ}$) are clearly not optimal for applications. For instance, if $\sigma \equiv \sigma_\circ$, for some non-degenerate constant matrix, the constants $C_{1,\varrho}$ and $C_{2,\varrho}$ can be significantly simplified as shown in [13, Remark 3.9].*

Acknowledgements

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 and Xavier Warin acknowledge the financial support of ANR project CAESARS (ANR-15-CE05-0024)..

Xiaolu Tan acknowledges the financial support of the ERC 321111 Rofirm, the ANR Isotace (ANR-12-MONU-0013), and the Chairs Financial Risks (Risk Foundation, sponsored by Société Générale) and Finance and Sustainable Development (IEF sponsored by EDF and CA).

References

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

[2] B. Bouchard, J.F. Chassagneux. Fundamentals and advanced techniques in derivatives hedging. Springer Universitext, 2016.

[3] B. Bouchard, X. Tan , X. Warin and Y. Zou, *Numerical approximation of BSDEs using local polynomial drivers and branching processes*, Monte Carlo Methods and Applications, to appear.

[4] 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.

[5] 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.

[6] J.F. Chassagneux, R. Elie and I. Kharroubi, *A numerical probabilistic scheme for super-replication with Delta constraints*, arXiv, forthcoming.

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

[8] E. Fournié, J.M. Lasry, , J. Lebuchoux, P.L. Lions, N. Touzi, *Applications of Malliavin calculus to Monte Carlo methods in finance*. Finance and Stochastics, 3(4), 391–412, 1999.

[9] H. Gevret, J. Lelong, X. Warin , *The StOpt library*, `https://gitlab.com/stochastic-control/StOpt`

[10] H. Gevret, J. Lelong, X. Warin , *STochastic OPTimization library in C++*, EDF Lab, 2016.

[11] 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.

[12] P. Henry-Labordère, *Cutting CVA's Complexity*, Risk, 25(7), 67, 2012.

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

[14] 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.

[15] J. Ma and J. Zhang, *Representation theorems for backward stochastic differential equations*. The annals of applied probability, 12(4), 1390–1418, 2002.

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

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

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

[19] X. Warin, *Some Non-monotone Schemes for Time Dependent Hamilton-Jacobi-Bellman Equations in Stochastic Control*, Journal of Scientific Computing, 66(3), 1122-1147, 2016.

[20] X. Warin, *Variations on branching methods for non linear PDEs*, arXiv preprint arXiv:1701.07660, 2017

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