FiNANCE FOR ENERGY MARKET RESEARCH CENTRE

Nesting Monte Carlo for high-dimensional Non Linear PDEs

Xavier Warin

Working Paper
RR-FiME-18-02

May 2018

Nesting Monte Carlo for high-dimensional Non Linear PDEs

Xavier Warin [*]

May 18, 2018

Abstract

A new method based on nesting Monte Carlo is developed to solve high-dimensional semi-linear PDEs. Convergence of the method is proved and its convergence rate studied. Results in high dimension for different kind of non-linearities show its efficiency.

1 Introduction

The resolution of Non Linear PDEs in high dimension is a challenging task due to the so-called "curse of dimensionality". Deterministic method can't cope with dimensions higher than 4 even using super-computers. In order to solve problems in higher dimension, effective resolution of Semi-Linear PDEs based on the BSDE approach first proposed by [21] were developed in [16] and [20]. A lot of literature on the subject has developed in recent years and the methodology has been extended to solve full non linear PDEs in [13], [24].
However because the methodology needs some basis functions to project conditional expectation, it faces the curse of dimensionality too by not being able to solve PDEs in dimension higher than six or seven.
Recently two new approaches have emerged in very high dimensions:

- The first is based on Deep Learning and uses deep neural networks [10], [9], [15]. The method seems to be effective in dimension over 100 but no proof of convergence is currently available and therefore we don't know its limitations.

[*]EDF R&D & FiME, Laboratoire de Finance des Marchés de l'Energie, xavier.warin@edf.fr

- The second is based on branching methods and is effective for non linearities polynomial in the solution u and its gradient Du. Convergence results are given in [18] and numerical results show that the PDEs can be solved in dimension 100. However the authors showed that the variance of the method explodes rapidly when the maturity grows or when the non linearity becomes important: numerical tests confirm that it is in fact the case. This methodology has being extended for other non linearities in [4] and [3] and the maturity problem is solved but at the price of the introduction of some grids meaning that the "curse of dimensionality" is back.

- The third is developed in [12], [11], [19] with an algorithm based on Picard iterations, multi-level techniques and automatic differentiation permitting to solve some high dimensional PDEs with non linearity in u and Du. The convergence of the algorithm is given and a lot of numerical examples show its efficiency in high dimension.

As for the branching method, the methodology proposed here is based on the Feynman-Kac representation of the PDEs coupled with the randomization of the time step proposed in [18]. This approach is combined with nesting Monte Carlo with a given depth. Then it is possible to get effective schemes to solve non-linear PDEs.
Because a truncation after a given number m of nesting is achieved, the method is biased.
For the demonstration of the convergence of the proposed schemes it is possible to follow classical approaches as the one used in [22]. It permits to understand how many particles to use at each nesting level and the number of nesting level m to take.
Classically the error is composed of a biased term and a variance term. It can be shown than the bias term goes to zero very quickly with m but to be effective we need to be able to take m below 5 or 6 such that the nesting Monte Carlo can be used. Therefore, as we will see, a limitation of the method will be that the maturity cannot be too large.

However the methodology proposed here has a lot of good properties :

- It is very simple to implement,

- It needs a very low memory to run on computers,

- Its convergence is independent on the dimension d of the problem,

- It is embarrassingly parallel so it can be run easily on super computers,

- If the different Lipschitz constants associated to the non-linearity are not too large, then the number of particles to take at each level to get a given accuracy is decreasing very fast giving a method very quickly converging.

Practically we will show that the method can be used on a wide set of cases and that we are able for example to solve all test cases proposed in [10], [9] for example.
The article has two parts :

- The first part is devoted to the resolution of the problem with linearity in u. The scheme is given, its convergence studied and numerical results in dimension 6 and 100 show the efficiency of the scheme.
- The second part is devoted to non linearities in Du. Based on automatic differentiation [14], we give a first scheme and show its convergence. We then introduce a second scheme using ideas in [26]. The second scheme permits to gain little theoretically but numerically we show that is is far more effective than the first proposed. We test the methods on problems with dimensions 10 to 100.

In the sequel we use the classical notation for $Y \in \mathbb{R}^d$, $||Y||_2 = \sqrt{\sum_{i=1}^d Y_i^2}$.
Given two matrix $A, B \in \mathbb{M}^d$, denote $A : B := \text{Trace}(AB^\top)$, $\mathbb{1}_d$ is the unit vector of $\mathbb{R}^d$ and $\mathbf{I}_d$ the identity matrix of $\mathbb{M}^d$.
All numerical experiments are achieved on a cluster using 8 nodes with a total of 224 cores and MPI is used for parallelization. The generation of random numbers is achieved using Tina's Random Number Generator Library [1]. All computational times are given for a configuration of Intel Xeon CPU E5-2680 v4 2.40GHz (Broadwell).

2 A first non linear case

In this section we study the case of a non linearity in u and we aim at solving the PDE for $t < T,\ x \in \mathbb{R}^d$:

$$
\begin{aligned}
(-\partial_t u - \mathcal{L}u)(t, x) &= f(t, x, u(t, x)), \\
u_T &= g,
\end{aligned}
\tag{1}
$$

where

$$
\mathcal{L}u(t, x) := \mu Du(t, x) + \frac{1}{2}\sigma\sigma^\top : D^2u(t, x), \tag{2}
$$

so that $\mathcal{L}$ is the generator associated to

$$\hat{X}_t = x + \mu t + \sigma dW_t, \tag{3}$$

with $\mu \in \mathbb{R}^d$, $\sigma \in \mathbb{M}^d$ is some constant matrix and W_t a d-dimensional Brownian motion.
We will use the following assumptions:

Assumption 2.1 *f is uniformly Lipschitz in u with constant K :*

$$|f(t,x,y) - f(t,x,w)| \leq K|y-w| \quad \forall t \in \mathbb{R}^+, x \in \mathbb{R}^d, (w,y) \in \mathbb{R}^2. \tag{4}$$

Assumption 2.2 *Equation* (1) *has a solution $u \in C^{1,2}([0,T] \times \mathbb{R}^d)$, such that*

- *u is θ-Hölder with $\theta \in (0,1]$ in time with constant $\hat{K}$:*

$$|u(t,x) - u(\tilde{t},x)| \leq \hat{K}|t-\tilde{t}|^\theta \qquad \forall (t,\tilde{t},x) \in [0,T] \times [0,T] \times \mathbb{R}^d,$$

- *$u(t,x)$ has a quadratic growth in x uniformly in t,*

In this section $\rho(x) = \lambda e^{-\lambda x}$ is the density of a random variable with exponential law.
Denote

$$\overline{F}(t) := \int_t^\infty \rho(s)ds = e^{-\lambda t} = 1 - F(t),$$

so that F is the cumulative distribution function of a random variable with density ρ.

2.1 Idea of the algorithm

We consider a sequence of i.i.d. random variables $(\tau_m)_{m\geq 1}$ of density ρ.
We consider the sequence defined by:

$$\begin{cases} T_0 &= 0, \\ T_{k+1} &= (T_k + \tau_k) \wedge T. \end{cases} \tag{5}$$

We further define $N_T = \inf\{n | T_{n+1} \geq T\}$.
We also consider a sequence of independent d-dimensional Brownian motion $(W_t^m)_{m\geq 1}$, which are independent of $(\tau_m)_{m\geq 1}$.

Define $W_t = W_t^1$ for all $t \in [0,T_1]$ and then for each k, define

$$W_t := W_{T_k} + W_{t-T_k}^{k+1}, \quad \text{for all } t \in [T_k, T_{k+1}]. \tag{6}$$

We define an associated diffusion process $(X_t)_{t\in[T_k,T_{k+1}]}$ by means of the following SDE:

$$X_t \ = \ X_{T_k} + \mu(t - T_k) + \sigma W^{k+1}_{t-T_k} \ \ t \in [T_k, T_{k+1}], \ \ \P\text{-a.s.}, \tag{7}$$

with $X_0 = x$.

Denoting by $\mathbb{E}_{t,x}$ the expectation operator conditional on $X_t = x$ at time t, from the Feynman-Kac formula the representation of the solution u valid under assumption 2.2 is given by:

$$\begin{aligned} u(0,x) =& \mathbb{E}_{0,x}\Big[\overline{F}(T)\frac{g(X_T)}{\overline{F}(T)} + \int_0^T \frac{f(t,X_t,u(t,X_t))}{\rho(t)}\rho(t)dt\Big] \\ =& \mathbb{E}_{0,x}[\phi(0,T_1,X_{T_1},u(T_1,X_{T_1}))], \end{aligned}$$

$$\phi(s,t,y,z) := \frac{\mathbf{1}_{\{t\geq T\}}}{\overline{F}(T-s)}g(y) + \frac{\mathbf{1}_{\{t<T\}}}{\rho(t-s)}f(t,y,z).$$

Recursively we have for $n < N_T$, noting $u_n = u(T_n, X_{T_n})$:

$$u_n = \mathbb{E}_{T_n,X_{T_n}}[\phi(T_n,T_{n+1},X_{T_{n+1}},u_{n+1})], \tag{8}$$

We further consider the truncated operator after p switches :

$$\begin{aligned} u^p_p =& g(X_{T_p}), \\ u^p_n =& \mathbb{E}_{T_n,X^n_{T_n}}[\phi(T_n,T_{n+1},X_{T_{n+1}},u^p_{n+1})], \quad n < p, \quad \text{defined if } T_n < T \end{aligned} \tag{9}$$

The goal of this section is to study the underlying algorithm when the resolution of equation (9) is achieved by nesting Monte Carlo. Starting from the ideas used in [18] we propose a nesting algorithm calculating all u^p_n by Monte Carlo. We have to show the bias associated to the algorithm goes to zero and that the global variance induced is controlled. In order to get a useful algorithm, we have to show that the bias goes to zero very quickly so that the number of switches to take is low: indeed it is well known that nesting Monte Carlo is subject to an explosion of the computer time. We will show that for many useful cases it is an effective approach.

2.2 Estimator and global error

Let set $p \in \mathbb{N}^+$. For $(N_0,..,N_{p-1}) \in \mathbb{N}^p$, we introduce the sets of i-tuple $Q_i = \{k = (k_1,...,k_i)\}$ for $i \in \{1,..,p\}$ where all components $k_j \in [1, N_{j-1}]$. Besides we define $Q^p = \cup_{i=1}^p Q_i$.
For $k = (k_1,...,k_i) \in Q_i$ we introduce the set $\tilde{Q}(k) = \{l = (k_1,..,k_i,m)/m \in$

$\{1,..,N_i\}\} \subset Q_{i+1}$. By convention $\tilde{Q}(\emptyset) = \{l = (m)/m \in \{1,..,N_0\}\} = Q_1$.
We define the following sequence τ_k of switching increments always i.i.d. random variables with density ρ for $k \in Q^p$, and a sequence of independent d-dimensional Brownian motion $(\bar{W}^k)$, which are independent of the $(\tau_k)_{k\in Q^p}$. Let us define the switching dates:

$$\begin{cases} T_{(j)} &= \tau_{(j)} \wedge T, j \in \{1,.,N_0\} \\ T_{\tilde{k}} &= (T_k + \tau_{\tilde{k}}) \wedge T, k = (k_1,..k_i) \in Q_i, \tilde{k} \in \tilde{Q}(k) \end{cases} \tag{10}$$

We define an associated diffusion process $(X_t^{\tilde{k}})_{t\geq 0}$ by means of the following SDE

$$\begin{aligned} X_t^{(i)} &= X_0^{\emptyset} + \mu t + \sigma \bar{W}_t^{(i)}, \quad t \in [0, T_{(i)}], i = 1, N_0 \\ X_t^{\tilde{k}} &= X_{T_k}^k + \mu(t - T_k) + \sigma \bar{W}_{t-T_k}^{\tilde{k}}, \text{ for } \tilde{k} \in \tilde{Q}(k), \quad t \in [T_k, T_{\tilde{k}}], \quad \P\text{-a.s.}, \end{aligned} \tag{11}$$

with $X_0^{\emptyset} = x$.

We consider the estimator defined by:

$$\begin{cases} \bar{u}_{\emptyset}^p = & \frac{1}{N_0} \sum_{j=1}^{N_0} \phi(0, T_{(j)}, X_{T_{(j)}}^{(j)}, \bar{u}_{(j)}^p), \\ \bar{u}_k^p = & \frac{1}{N_i} \sum_{\tilde{k}\in\tilde{Q}(k)} \phi(T_k, T_{\tilde{k}}, X_{T_{\tilde{k}}}^{\tilde{k}}, \bar{u}_{\tilde{k}}^p), \\ & \quad \text{for } k = (k_1, ..k_i) \in Q_i, 1 < i < p, T_k < T, \\ \bar{u}_{\tilde{k}}^p = & g(X_{T_{\tilde{k}}}^{\tilde{k}}) \quad \text{for } \tilde{k} \in Q_p. \end{cases} \tag{12}$$

Note that in the case where $T_{\tilde{k}} = T$ then $\phi(T_k, T_{\tilde{k}}, X_{T_{\tilde{k}}}^{\tilde{k}}, \bar{u}_{\tilde{k}}^p)$ is independent of $\bar{u}_{\tilde{k}}^p$ so that the recursion is stopped.

Proposition 2.3 *Under assumption 2.1, 2.2 , we have the following error given by the estimator* (12) *:*

$$\begin{aligned} \mathbb{E}((\bar{u}_{\emptyset}^p - u(0,x))^2) \leq \prod_{i=1}^{p} (1 + \frac{8}{N_{i-1}}) \frac{K^{2p} e^{\lambda T}}{\lambda^p} T^{2\theta} \hat{K}^2 \frac{T^p}{p\Gamma(p)} + \\ \sum_{i=0}^{p-1} \frac{K^{2i}}{N_i} \prod_{j=1}^{i} (1 + \frac{8}{N_{j-1}}) \frac{T^i e^{\lambda T}}{\lambda^i} \kappa_i \end{aligned}$$

with

$$\kappa_i = (\frac{4T}{\lambda(i+1)!} \sup_{t\in[0,T]} \mathbb{E}(f(t, X_t, u(t, X_t))^2) + 2\mathbb{E}(g(X_T)^2)(\frac{1_{i>0}}{i\Gamma(i)} + 1_{i=0}))$$

and $\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt$ *is the gamma function.*

Proof. First notice that due to the Lipschitz property in assumption 2.1 and the growth assumption on u, $\mathbb{E}(\int_0^T |f(t, X_t, u(t, X_t))| dt) < \infty$. Then notice that under assumption 2.2, the solution u of (1) satisfies a Feynman-Kac relation (see an adaptation of proposition 1.7 in [25]) so that for all $k \in Q_i$, and $\forall \tilde{k} \in \tilde{Q}(k)$,

$$u(T_k, X^k_{T_k}) = \mathbb{E}_{T_k, X^k_{T_k}}[\phi(T_k, T_{\tilde{k}}, X_{T^{\tilde{k}}_{\tilde{k}}}, u(T_{\tilde{k}}, X_{T^{\tilde{k}}_{\tilde{k}}}))], \tag{13}$$

Then for $k \in Q_i$, $i < p$, let us define:

$$\begin{aligned} E_k :=& \mathbb{E}_{T_k, X^k_{T_k}}((\bar{u}^p_k - u(T_k, X^k_{T_k}))^2 1_{T_k < T}) \\ =& V_k + B_k^2 \end{aligned} \tag{14}$$

where we note B_k the bias error for index k as:

$$B_k := (\mathbb{E}_{T_k, X^k_{T_k}}(\bar{u}^p_k) - u(T_k, X^k_{T_k})) 1_{T_k < T}, \tag{15}$$

and the variance term v_k of the estimator:

$$V_k := \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_k < T}(\bar{u}^p_k - \mathbb{E}_{T_k, X^k_{T_k}}(\bar{u}^p_k))^2). \tag{16}$$

Let us begin with the variance term.
Note that using the equation (13) and the $\bar{u}^p_k$ definition given by equation (12):

$$\begin{aligned} (\bar{u}^p_k - \mathbb{E}_{T_k, X^k_{T_k}}(\bar{u}^p_k))^2 1_{T_k < T} = [& \frac{1}{N_i} \sum_{\tilde{k} \in \tilde{Q}(k)} (1_{T_{\tilde{k}} < T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})} - \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}} < T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})})) + \\ & \frac{1}{N_i} \sum_{\tilde{k} \in \tilde{Q}(k)} (1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)}{\overline{F}(T - T_k)} - \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)}{\overline{F}(T - T_k)}))]^2 \end{aligned}$$

so that

$$\begin{aligned} V_k \leq & 2\mathbb{E}_{T_k, X^k_{T_k}}[(\frac{1}{N_i} \sum_{\tilde{k} \in \tilde{Q}(k)} (1_{T_{\tilde{k}} < T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})} - \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}} < T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})})))^2] + \\ & 2\mathbb{E}_{T_k, X^k_{T_k}}[(\frac{1}{N_i} \sum_{\tilde{k} \in \tilde{Q}(k)} (1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)}{\overline{F}(T - T_k)} - \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)}{\overline{F}(T - T_k)})))^2] \end{aligned}$$

and using that for independent random variables x_i

$$\mathbb{E}[(\frac{1}{N} \sum_{i=1}^N (x_i - E(x_i)))^2] = \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}[(x_i - \mathbb{E}(x_i))^2]$$

$$V_k \leq \frac{2}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}} [(1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})} - \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})}))^2] +$$
$$\frac{2}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}} [(1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)}{\overline{F}(T - T_k)} - \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)}{\overline{F}(T - T_k)}))^2]. \tag{17}$$

Developing

$$\begin{aligned} A =& 1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})} - \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})}) \\ =& (1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}) - f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})}) + \\ & (1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})} - \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})})) \end{aligned}$$

that we inject in (17) so that using the relation $\mathbb{E}(|x - \mathbb{E}(x)|^2) \leq \mathbb{E}(x^2)$, V_k is bounded following:

$$\begin{aligned} V_k \leq & 4 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}<T} (\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}) - f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2) + \\ & 4 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}} ((1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}}))} - \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})}))^2) + \\ & 2 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)^2}{\overline{F}(T - T_k)^2}). \end{aligned} \tag{18}$$

Using that for X, Y random variables $\mathbb{E}((X - E(Y))^2) = \mathbb{E}((X - E(X))^2) + (\mathbb{E}(X - Y))^2$, then for $\tilde{k} \in \tilde{Q}(k)$:

$$
\begin{aligned}
I =&\mathbb{E}_{T_k, X^k_{T_k}}((1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})} - \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})}))^2) \\
=&\mathbb{E}_{T_k, X^k_{T_k}}((1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})} - \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})}))^2)+ \\
&(\mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})}) - \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}))}{\rho(\tau_{\tilde{k}})}))^2 \\
\leq&\mathbb{E}_{T_k, X^k_{T_k}}((1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{(k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})} - \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T} \frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})}))^2)+ \\
&\mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T}(\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}) - f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2)
\end{aligned}
\tag{19}
$$

where the last inequality is obtained by Jensen.
Plugging (19) in (18) and using the relation $\mathbb{E}(|x - \mathbb{E}(x)|^2) \leq \mathbb{E}(x^2)$ we get :

$$
\begin{aligned}
V_k \leq&8 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T}(\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, \bar{u}^p_{\tilde{k}}) - f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2)+ \\
&4 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T}(\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2)+ \\
&2 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)^2}{\overline{F}(T - T_k)^2}).
\end{aligned}
\tag{20}
$$

Using the Lipschitz property of f and the tower property :

$$
\begin{aligned}
V_k \leq&8 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}}((\frac{K}{\rho(\tau_{\tilde{k}})})^2 E_{\tilde{k}})+ \\
&4 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}}<T}(\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, u(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2)+ \\
&2 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}} \geq T} \frac{g(X^{\tilde{k}}_T)^2}{\overline{F}(T - T_k)^2}).
\end{aligned}
\tag{21}
$$

Now we take care of the bias term.

$$B_k^2 = (\frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}(k)} E_{T_k,X^k_{T_k}}(1_{T_{\tilde{k}}<T}(\frac{f(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}},\bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})} - \frac{f(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}},u(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})))^2 \tag{22}$$

Using the fact that all expectations for the $\tilde{k}$ are the same, we get :

$$B_k^2 = \frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}(k)} E_{T_k,X^k_{T_k}}(1_{T_{\tilde{k}}<T}(\frac{f(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}},\bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})} - \frac{f(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}},u(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})}))^2.$$

Then using Jensen :

$$B_k^2 \leq \frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}(k)} E_{T_k,X^k_{T_k}}(1_{T_{\tilde{k}}<T}(\frac{f(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}},\bar{u}^p_{\tilde{k}})}{\rho(\tau_{\tilde{k}})} - \frac{f(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}},u(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2).$$

So that using the Lipschitz property of f and the tower property :

$$B_k^2 \leq \frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}(k)} E_{T_k,X^k_{T_k}}(\frac{K^2}{\rho(\tau_{\tilde{k}})^2}E_{\tilde{k}}). \tag{23}$$

Plugging (21) and (23) in (14) we get:

$$\begin{aligned} E_k \leq & \frac{1}{N_i}(1+\frac{8}{N_i})\sum_{\tilde{k}\in\tilde{Q}(k)} E_{T_k,X^k_{T_k}}(\frac{K^2}{\rho(\tau_{\tilde{k}})^2}E_{\tilde{k}})+ \\ & 4\frac{1}{N_i^2}\sum_{\tilde{k}\in\tilde{Q}(k)} \mathbb{E}_{T_k,X^k_{T_k}}(1_{T_{\tilde{k}}<T}(\frac{f(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}},u(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2)+ \\ & 2\frac{1}{N_i^2}\sum_{\tilde{k}\in\tilde{Q}(k)} \mathbb{E}_{T_k,X^k_{T_k}}(1_{T_{\tilde{k}}\geq T}\frac{g(X^{\tilde{k}}_T)^2}{\bar{F}(T-T_k)^2}). \end{aligned} \tag{24}$$

We can iterate to get $E_\emptyset$ using the tower property

$$\begin{aligned} E_\emptyset \leq & \prod_{i=1}^{p}\frac{1}{N_{i-1}}(1+\frac{8}{N_{i-1}})\sum_{\tilde{k}^1\in\tilde{Q}(\emptyset)}\cdots\sum_{\tilde{k}^p\in\tilde{Q}(\tilde{k}^{p-1})}\mathbb{E}[\frac{K^{2p}}{\prod_{j=1}^{p}\rho(\tau_{\tilde{k}^j})^2}E_{\tilde{k}^p}]+ \\ & \sum_{i=0}^{p-1}\frac{K^{2i}}{N_i^2}\prod_{j=1}^{i}\frac{1}{N_{j-1}}(1+\frac{8}{N_{j-1}})\sum_{\tilde{k}^1\in\tilde{Q}(\emptyset)}\cdots\sum_{\tilde{k}^{i+1}\in\tilde{Q}(\tilde{k}^i)}\mathbb{E}[1_{T_{\tilde{k}^{i+1}}<T}\frac{4f(T_{\tilde{k}^{i+1}},X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}},u(T_{\tilde{k}^{i+1}},X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}}))^2}{\prod_{j=1}^{i+1}\rho(\tau_{\tilde{k}^j})^2}+ \\ & 1_{T_{\tilde{k}^{i+1}}>T}1_{T_{\tilde{k}^i}<T}\frac{2g(X^{\tilde{k}^{i+1}}_T)^2}{\prod_{j=1}^{i}\rho(\tau_{\tilde{k}^j})^2\bar{F}(T-T_{\tilde{k}^i})^2}] \end{aligned} \tag{25}$$

where $E_{\tilde{k}^p} = 1_{T_{\tilde{k}^p}<T}(g(X^{\tilde{k}^p}_{T_{\tilde{k}^p}}) - u(T_{\tilde{k}^p}, X^{\tilde{k}^p}_{T_{\tilde{k}^p}}))^2$.
We now bound the two terms in the last summation. Using the fact that ρ corresponds to the density of an exponential law:

$$
\begin{aligned}
D :=& \mathbb{E}[1_{T_{\tilde{k}^{i+1}}<T} \frac{f(T_{\tilde{k}^{i+1}}, X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}}, u(T_{\tilde{k}^{i+1}}, X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}}))^2}{\prod_{j=1}^{i+1}\rho(\tau_{\tilde{k}^j})^2}] \\
\leq& \mathbb{E}[\frac{1_{T_{\tilde{k}^{i+1}}<T}}{\prod_{j=1}^{i+1}\rho(\tau_{\tilde{k}^j})} \frac{f(T_{\tilde{k}^{i+1}}, X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}}, u(T_{\tilde{k}^{i+1}}, X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}}))^2}{\prod_{j=1}^{i+1}\rho(\tau_{\tilde{k}^j})}] \\
\leq& \frac{e^{\lambda T}}{\lambda^{i+1}} \mathbb{E}[1_{T_{\tilde{k}^{i+1}}<T} \frac{f(T_{\tilde{k}^{i+1}}, X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}}, u(T_{\tilde{k}^{i+1}}, X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}}))^2}{\prod_{j=1}^{i+1}\rho(\tau_{\tilde{k}^j})}] \\
=& \frac{e^{\lambda T}}{\lambda^{i+1}} \int_0^T \int_0^{T-t^1} .. \int_0^{T-t^1-..-t^i} \mathbb{E}[f(\sum_{j=1}^{i+1} t_j, X^{\tilde{k}^{i+1}}_{\sum_{j=1}^{i+1} t^j}, u(\sum_{j=1}^{i+1} t^j, X^{\tilde{k}^{i+1}}_{\sum_{j=1}^{i+1} t^j}))^2] dt^1..dt^{i+1} \\
\leq& \frac{e^{\lambda T}}{\lambda^{i+1}} \frac{T^{i+1}}{(i+1)!} \sup_{t\in[0,T]} \mathbb{E}(f(t, X_t, u(t, X_t))^2),
\end{aligned}
\tag{26}
$$

where $\mathbb{E}(f(t, X_t, u(t, X_t))^2) < \infty$ due to the Lipschitz condition on f and the quadratic growth of u.
We know deal with the last term using the fact that distribution of X_T is independent

of the switching dates :

$$
\begin{aligned}
H_i :=& \mathbb{E}[1_{T_{\tilde{k}^{i+1}}>T} 1_{T_{\tilde{k}^i}<T} \frac{g(X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}})^2}{\prod_{j=1}^{i} \rho(\tau_{\tilde{k}^j})^2 \bar{F}(T-T_{\tilde{k}^i})^2}] \\
=& \mathbb{E}[1_{T_{\tilde{k}^{i+1}}>T} \frac{1_{T_{\tilde{k}^i}<T}}{\prod_{j=1}^{i} \rho(\tau_{\tilde{k}^j})^2 \bar{F}(T-T_{\tilde{k}^i})^2}]\mathbb{E}(g(X_T)^2) \\
=& \mathbb{E}[1_{T_{\tilde{k}^{i+1}}>T} 1_{T_{\tilde{k}^i}<T} \frac{e^{2\lambda T_{\tilde{k}^i}}}{\lambda^{2i}} e^{2\lambda(T-T_{\tilde{k}^i})}]\mathbb{E}(g(X_T)^2) \\
=& \frac{e^{2\lambda T}}{\lambda^{2i}} \mathbb{E}[1_{T_{\tilde{k}^{i+1}}>T} 1_{T_{\tilde{k}^i}<T}]\mathbb{E}(g(X_T)^2) \\
=& \frac{e^{2\lambda T}}{\lambda^{2i}} \mathbb{E}(g(X_T)^2) \int_0^T \lambda^i x^{i-1} \frac{e^{-\lambda x}}{\Gamma(i)} \int_{T-x}^{\infty} \lambda e^{-\lambda y} dy dx \\
=& \frac{e^{\lambda T}}{\lambda^i} \frac{T^i}{i\Gamma(i)} \mathbb{E}(g(X_T)^2) \quad \text{ for } i > 0, \\
H_0 =& \mathbb{E}[1_{T_{\tilde{k}^1}>T} \frac{g(X^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}}})^2}{\bar{F}(T)^2}] \\
=& e^{\lambda T}\mathbb{E}(g(X_T)^2).
\end{aligned} \tag{27}
$$

where we have used the fact that $T_{\tilde{k}^i}$ follows a gamma law with density $\lambda^i x^{i-1} \frac{e^{-\lambda x}}{\Gamma(i)}$. At last using the Hölder property of u with respect to t:

$$
\begin{aligned}
F :=& \mathbb{E}[1_{T_{\tilde{k}^p}<T} \frac{K^{2p}}{\prod_{j=1}^{p} \rho(\tau_{\tilde{k}^j})^2} E_{\tilde{k}^p}] \\
=& \mathbb{E}[1_{T_{\tilde{k}^p}<T} \frac{K^{2p}}{\prod_{j=1}^{p} \rho(\tau_{\tilde{k}^j})^2} (g(X^{\tilde{k}^p}_{T_{\tilde{k}^p}}) - u(T_{\tilde{k}^p}, X^{\tilde{k}^p}_{T_{\tilde{k}^p}}))^2] \\
\leq& K^{2p} T^{2\theta} \hat{K}^2 \mathbb{E}[1_{T_{\tilde{k}^p}<T} \frac{1}{\prod_{j=1}^{p} \rho(\tau_{\tilde{k}^j})^2}] \\
=& \frac{K^{2p}}{\lambda^{2p}} T^{2\theta} \hat{K}^2 \mathbb{E}[1_{T_{\tilde{k}^p}<T} e^{2\lambda T_{\tilde{k}^p}}] = \frac{K^{2p}}{\lambda^{2p}} T^{2\theta} \hat{K}^2 \int_0^T e^{\lambda x} \frac{\lambda^p x^{p-1}}{\Gamma(p)} dx \\
\leq& \frac{K^{2p} e^{\lambda T}}{\lambda^p} T^{2\theta} \hat{K}^2 \frac{T^p}{p\Gamma(p)}.
\end{aligned} \tag{28}
$$

Plugging (26),(27) and (28) in (25)

$$
\begin{aligned}
E_{\emptyset} \leq & \prod_{i=1}^{p} \frac{1}{N_{i-1}}(1+\frac{8}{N_{i-1}}) \sum_{\tilde{k}^1 \in \tilde{Q}(\emptyset)} \ldots \sum_{\tilde{k}^p \in \tilde{Q}(\tilde{k}^{p-1})} \frac{K^{2p} e^{\lambda T}}{\lambda^p} T^{2\theta} \hat{K}^2 \frac{T^p}{p\Gamma(p)} + \\
& \sum_{i=0}^{p-1} \frac{K^{2i}}{N_i} \prod_{j=1}^{i} (1+\frac{8}{N_{j-1}}) \frac{T^i e^{\lambda T}}{\lambda^i} (\frac{4T}{\lambda(i+1)!} \sup_{t \in [0,T]} \mathbb{E}(f(t, X_t, u(t, X_t))^2) + \\
& 2\mathbb{E}(g(X_T)^2)(\frac{1_{i>0}}{i\Gamma(i)} + 1_{i=0})) \\
= & \prod_{i=1}^{p} (1+\frac{8}{N_{i-1}}) \frac{K^{2p} e^{\lambda T}}{\lambda^p} T^{2\theta} \hat{K}^2 \frac{T^p}{p\Gamma(p)} + \\
& \sum_{i=0}^{p-1} \frac{K^{2i}}{N_i} \prod_{j=1}^{i} (1+\frac{8}{N_{j-1}}) \frac{T^i e^{\lambda T}}{\lambda^i} (\frac{4T}{\lambda(i+1)!} \sup_{t \in [0,T]} \mathbb{E}(f(t, X_t, u(t, X_t))^2) + \\
& 2\mathbb{E}(g(X_T)^2)(\frac{1_{i>0}}{i\Gamma(i)} + 1_{i=0}))
\end{aligned} \tag{29}
$$

which is the desired result. Observe that it is necessary to solve accurately each inner iteration with enough simulations in order to get convergence. This is due to the B_k estimation (23) for which we have to take enough simulations to avoid bias propagation. The result is classical in nested Monte Carlo : not enough convergence in inner iteration can lead to a bias on upper iterations.
The convergence result is quite obvious:

- the bias propagates multiplied at each switching dates by a square of the Lipschitz constant but decrease due to the fact that the probability that the branching dates doesn't reach T goes to zero. In fact using Stirling formula $\Gamma(p) \simeq \sqrt{2\pi(p-1)}(\frac{p-1}{e})^{p-1}$ we see that the bias term goes to 0 exponentially fast meaning that for not too long maturities only small values of p are needed to reach a very good accuracy.

- The variance term can be bounded by

$$\sum_{i=0}^{p-1} \frac{C_i}{N_i}$$

  with C_i going to zero very quickly meaning that we should take N_i with decreasing values.
  Besides if we consider series $\{(N_0^j, .., N_{p-1}^j)\}_{j>0}$ such that the corresponding $(\bar{u}_{\emptyset}^p)^j$ goes to the bias term, it is reasonable to take $(N_0^j, .., N_{p-1}^j) = M^j(N_0^0, .., N_{p-1}^0)$ where M^j goes to infinity.

Remark 2.4 *For f regular it could be tempting to try to use the ideas developed in [22]. The approximation in this article for f regular uses the fact the bias goes to* 0 *in order to declare that terms depending on the square of the bias are negligible compared to terms depending on the bias at each iteration of the nesting procedure. This is not true in our case.*

Remark 2.5 *The previous result in proposition 2.3 is also valid for more complex SDE as soon as the SDE can be simulated exactly.*

Remark 2.6 *We could have a tighter expression for the variance terms by keeping* $||g(X_T) - \mathbb{E}(g(X_T))||_2^2$ *and* $||f(t, X_t, u(t, X_t)) - \mathbb{E}(f(t, X_t, u(t, X_t)))||_2^2$ *instead of* $||g(X_T)||_2^2$ *and* $||f(t, X_t, u(t, X_t))||_2^2$ *respectively.*

Remark 2.7 *When the coefficients μ and σ are not constant the methodology is exactly the same except that the SDE has to be approximated by an Euler Scheme. Two ways to implement it can be used :*

- *the first consists in getting the coefficients of the SDE on a fixed grid with a given time step picking the values from the grid. Then the error added due to the discretization is classical [23].*
- *a second numerically more effective consists in using an Euler scheme between the switching dates essentially meaning that the Euler grid depends on the trajectory. This approach is suggested in [26].*

At last we see that the method converges with a speed independent of the dimension of the problem meaning that it is possible to solve non linear PDEs in very high dimension.

2.3 Numerical results for the first non linear case

We will first study a first toy case in high dimension then we will move to a realistic test case in finance.

2.3.1 A first toy example

In this first case we take 2 maturities $T = 1$, $T = 2$. The SDE coefficients are $\mu = \frac{\mu_0}{d}\mathbb{1}_d$, $\sigma = \frac{\sigma_0}{\sqrt{d}}\mathbf{I}_d$ with $\mu_0 = 0.2$, $\sigma_0 = 1$.

We take for $x \in \mathbb{R}^d$, $g(x) = \cos(\sum_{i=1}^d x_i)$ and the non linearity:

$$f(t,x,u) = \cos(\sum_{i=1}^d x_i)(a + \frac{\sigma_0^2}{2})e^{a(T-t)} + \sin(\sum_{i=1}^d x_i)\mu_0 e^{a(T-t)} -$$
$$r\cos(\sum_{i=1}^d x_i)^2 e^{2a(T-t)} + r(-e^{a(T-t)} \vee (u \wedge e^{a(T-t)}))^2$$

with $a = 0.1$, $r = 0.1$,$d = 100$.
Equation (1) admits the classical solution $u(t,x) = e^{a(T-t)}cos(\sum_{i=1}^d x_i)$. The Lipschitz constant associated to f is $K = 2re^{aT}$ and the solution is Lipschitz in time ($\theta = 1$) with a Lipschitz constant $\hat{K} = ae^{aT}$.
Notice that

$$\mathbb{E}(g(X_T)^2) \leq 1,$$

and

$$\sup_{t\in[0,T]} \mathbb{E}(f(t,X_t,u(t,X_t))^2) \leq (a + \frac{\sigma_0^2}{2} + \mu_0)^2 e^{2a(T-t)}.$$

In tables 1, 2 we give the different coefficients associated to error expression in proposition 2.3:

- Bias p corresponds to term

$$b(p) = \frac{K^{2p}e^{\lambda T}}{\lambda^p} T^{2\theta}\hat{K}^2 \frac{T^p}{p\Gamma(p)}, \tag{30}$$

- Var i corresponds to the variance term

$$v(i) = K^{2i}\frac{T^i e^{\lambda T}}{\lambda^i}\kappa_i. \tag{31}$$

We check that the bias and the variance terms decrease rapidly with p for small maturities.
In tables 3 and 4, for $T = 1$ and $T = 2$, $\lambda = 0.4$, we give the level p and the number of particles to take at each level to reach a given accuracy.

On figure 1, we plot for different values of λ the solution with $T = 1$ obtained with one or two switches with a number of particles $N_0 = 1000 \times 2^{\text{ipart}}$ and $N_1 = 50 \times 2^{\text{ipart}}$. With one switch the solution is clearly biased while the bias is indistinguishable from 0 with 2 switches whatever the λ taken.

On figure 2, we plot for different values of λ the solution with $T = 2$ obtained with one, two or three switches with a number of particles $N_0 = 1100 \times 2^{\text{ipart}}$, $N_1 = 110 \times 2^{\text{ipart}}$, $N_2 = 25 \times 2^{\text{ipart}}$. For all λ, we have to take three switches to have a good precision. The best solution seem to be reached with $\lambda = 0.2$.

λ	0.2	0.4	0.8
Bias p= 1	0.01458	0.008902	0.00664
Bias p= 2	0.00178	0.0005437	0.0002028
Bias p= 3	0.000145	2.213e-05	4.128e-06
Bias p= 4	8.854e-06	6.759e-07	6.302e-08
Bias p= 5	4.326e-07	1.651e-08	7.697e-10
Var i = 0	21.54	14.65	13.15
Var i = 1	2.929	1.077	0.5374
Var i = 2	0.2628	0.05125	0.01371
Var i = 3	0.01753	0.001791	0.0002515
Var i = 4	0.0009291	4.93e-05	3.588e-06

Table 1: Coefficient in the error analysis in proposition 2.3 for $T = 1$

λ	0.2	0.4	0.8
Bias p= 1	0.2125	0.1585	0.1764
Bias p= 2	0.0634	0.02364	0.01316
Bias p= 3	0.01261	0.002352	0.0006542
Bias p= 4	0.001881	0.0001754	2.44e-05
Bias p= 5	0.0002245	1.047e-05	7.28e-07
Var i = 0	59.96	46.95	57.2
Var i = 1	18.78	7.668	5.005
Var i = 2	3.912	0.8287	0.2856
Var i = 3	0.6101	0.06674	0.01202
Var i = 4	0.07596	0.004276	0.0003996

Table 2: Coefficient in the error analysis in proposition 2.3 for $T = 2$

i	0	1
N_i	129684	5299

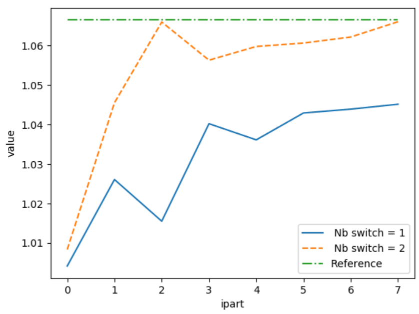
Table 3: Number of particles to take for $p = 2, \lambda = 0.4, T = 1$, an accuracy $B_p + \sum_{i=0}^{p-1} \frac{v(i)}{N_i} \prod_{j=0}^{i}(1 + \frac{8}{N_{j-1}}) = 1.088E - 03$.

2.3.2 A second test case

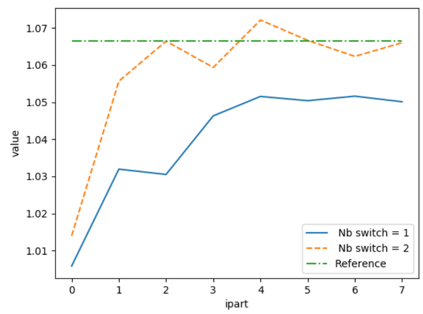
We use the test case presented in [17]. This is a test case in low dimension but the author gives some numerical bounds on the solution so that we can compare our methodology to some deep learning solution.

i	0	1	2
N_i	59885	9780	1057

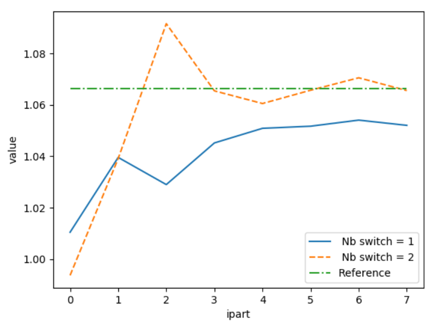
Table 4: Number of particles to take for $p = 3, \lambda = 0.4, T = 2$, an accuracy $B_p + \sum_{i=0}^{p-1} \frac{\nu(i)}{N_i} \prod_{j=0}^{i}(1 + \frac{8}{N_{j-1}}) = 4.720E-03$.



$\lambda = 0.2$.



$\lambda = 0.4$.



$\lambda = 0.8$.

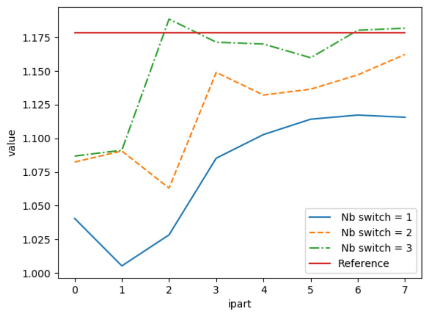
Figure 1: Convergence for different number of switches for case 1, $T = 1$.

The author considers the PDE obtained from a CVA valuation problem.

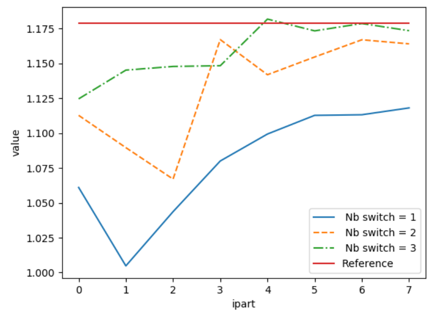
$$\mathcal{L}u(t,x) := \mu Du(t,x) + \frac{1}{2}\sigma\sigma^\top : D^2u(t,x), \tag{32}$$

taking $\mu = -\frac{\sigma_0^2}{2}\mathbb{I}_d$, $\sigma = \sigma_0 \mathbf{I}_d$, and a non-linearity

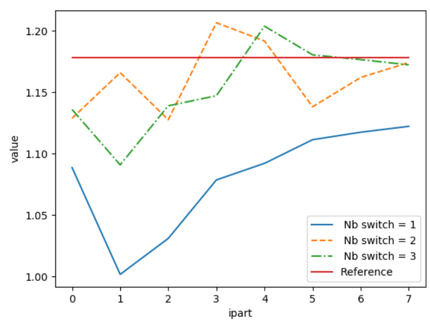
$$f(t,x,u) = \beta(u^+ - u),$$



$\lambda = 0.2.$



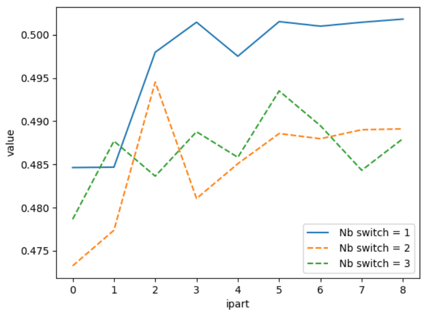
$\lambda = 0.4.$



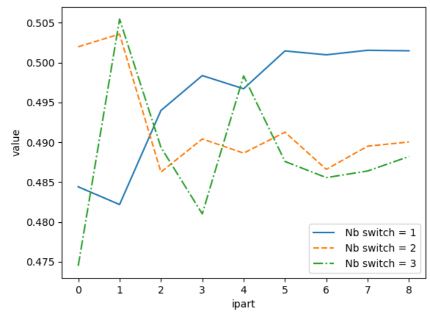
$\lambda = 0.8.$

Figure 2: Convergence for different number of switches for case 1, $T = 2$.

with $\beta = 0.03, \sigma_0 = 0.2$. The initial value for the SDE is $X_0 = \mathbb{1}_d$. The final function $g(x) = \sum_{i=1}^{d}(1 - 2 1_{e^{x_i}>1})$ and $T = 1$. Some bounds on the solution in dimension till 6 are given. For $d = 6$ a lower bound calculated is 48.80, an upper bound is 48.83 whereas with $\beta = 0$ the solution is 47.73.
On this simple problem we don't try to optimize the number of particles taken at each level nor the λ taken. On figure 3, we plot for $\lambda = 0.1$ and $\lambda = 0.2$, the solution obtained with one, two or three switches with a number of particles $N_0 = 36000 \times 2^{\text{ ipart}}$, $N_1 = 140 \times 2^{\text{ ipart}}$, $N_2 = 2^{\text{ ipart}}$. With $\lambda = 0.1$ with 3 switches and $ipart = 8$ we get 0.4880 while with $\lambda = 0.2$ we get 0.4882 so that both values are in the very tight bounds proposed in [17].



$\lambda = 0.1.$



$\lambda = 0.2.$

Figure 3: Convergence of the scheme on the CVA case.

2.3.3 A third test case

In this part we take a test case coming from [10] modeling the valuation of an European claim in dimension 100 using a Black-Scholes dynamic of the assets supposing the existence of a default risk. The default is modeled by the first jump time of a Poisson process with intensity Q. When a default occurs, the claim's holder receive only a fraction $\delta \in [0, 1]$ of the current value. We want to valuate the claim conditionally that the default hasn't occurred yet. The dynamic of an asset S_t with trend μ_0 and volatility σ_0 following the BS model satisfies

$$S_t = S_0 e^{(\mu_0 - \frac{\sigma_0^2}{2})t + \sigma_0 W_t}, \tag{33}$$

such that taking $X_t = \log(S_t)$, the X_t dynamic follows

$$dX_t = (\mu_0 - \frac{\sigma_0^2}{2})dt + \sigma_0 dW_t. \tag{34}$$

Supposing that all the assets are independent and follow the same equation (33), the value of the claim given by [10] can be equivalently given as the solution at date 0 and point $x = \log(100)\mathbb{1}_d$ of (1) where

$$\mathcal{L}u(t, x) := \mu Du(t, x) + \frac{1}{2}\sigma\sigma^\top : D^2u(t, x), \tag{35}$$

with $\mu = (\mu_0 - \frac{\sigma_0^2}{2})\mathbb{1}_d$, $\sigma = \sigma_0 \mathbf{I}_d$.
Following [10], the final function g satisfies for $x \in \mathbb{R}^d$,

$$g(x) = \min_{i=1}^{100}(e^{x_i}),$$

while the non-linearity is given by

$$f(t,x,u) = -\left((1-\delta)\min\{\gamma^h, \max\{\gamma^l, \frac{\gamma^h-\gamma^l}{v^h-v^l}(u-v^h)+\gamma^h\}\}+R\right)u$$

where R is the interest rate of the riskless asset (see [2]). We take the same parameters as in [10] so $T = 1$, $\delta = \frac{2}{3}$, $\mu_0 = 0.02$, $\sigma_0 = 0.2$, $v^h = 50$, $v^l = 70$, $\gamma^h = 0.2$, $\gamma^l = 0.02$. As pointed out by [10], the solution obtained by Monte Carlo ignoring the default risk is approximately 60.78. This reference can also be obtained by the algorithm taking $f = 0$ as shown on figure 4.

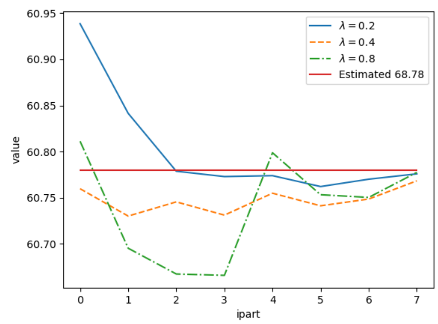


Figure 4: Convergence of the scheme on the Black Scholes case without default taking $f = 0$ in the algorithm with a number of particle $N = 100000 \times 2^{\text{ipart}}$.

Let us study the value of the different terms :

- $K = (1-\delta)\gamma^h + R = 0.086$,
- $\mathbb{E}(g(X_T)^2) \leq \mathbb{E}((X_T^1)^2) = S_0^2 e^{(2\mu+\sigma_0^2)T} \simeq 10000$ where X^1 stands for the first asset log value,
- Noting $\hat{u}$ the solution without default

$$\begin{aligned}\sup_{t\in[0,1]} \mathbb{E}(f(t,X_t,u(t,X_t))^2) \leq & K^2 \sup_{t\in[0,1]} \mathbb{E}(u(t,X_t)^2) \\ \leq & K^2 \sup_{t\in[0,1]} \mathbb{E}(\hat{u}(t,X_t)^2) \leq K^2 \sup_{t\in[0,1]} \mathbb{E}((X_t^1)^2) \\ = & K^2 S_0^2 \sup_{t\in[0,1]} e^{(2\mu+\sigma_0^2)t} \simeq 50.\end{aligned}$$

- At last the solution may not be uniformly Lipschitz in time but remember that $\hat{K}^2 T^{2\theta}$ is a bound from an expression $\psi = \sup_{t\in[0,T]} \mathbb{E}((\hat{u}(t,X_t) - g(X_t))^2)$

so that using the previous estimations

$$\psi \leq 2 \sup_{t\in[0,T]} \mathbb{E}(u(t,X_t)^2 + g(X_t)^2)$$
$$\leq 2S_0^2 \sup_{t\in[0,1]} (e^{(2\mu+\sigma_0^2)t} + e^{(2\mu+\sigma_0^2)t}) \simeq 40000$$

so that $\hat{K}^2T^{2\theta}$ can be replace 40000.

In table 5, we give the coefficients of equations (30) and (31) involved in proposition 2.3. By taking $\lambda = 0.8$, $p = 3$, on table 6 we give the number of particles to

λ	0.2	0.4	0.8
Bias p= 1	1988.0	1214.0	905.4
Bias p= 2	37.32	11.4	4.25
Bias p= 3	0.4672	0.07134	0.0133
Bias p= 4	0.004387	0.0003349	3.122e-05
Bias p= 5	3.295e-05	1.258e-06	5.863e-08
Var i = 0	28450.0	33540.0	49120.0
Var i = 1	1031.0	618.3	457.0
Var i = 2	19.13	5.77	2.139
Var i = 3	0.238	0.036	0.006682
Var i = 4	0.002226	0.0001687	1.567e-05

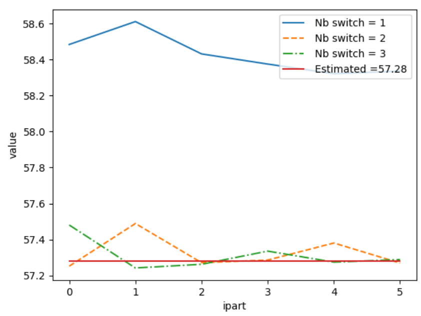
Table 5: Coefficient in the error analysis in proposition 2.3 for the Black Scholes case with default risk.

take to have an accuracy of 0.01.

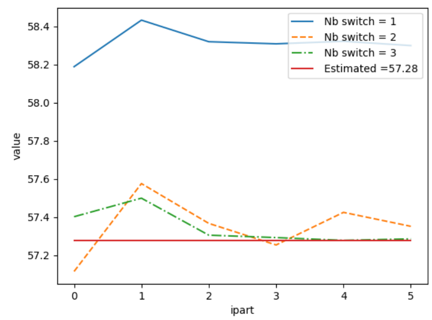
i	0	1	2
N_i	110796	1030	4

Table 6: Number of particles to take for $p = 3, \lambda = 0.8$, an accuracy $B_p + \sum_{i=0}^{p-1} \frac{v(i)}{N_i} \prod_{j=0}^{i} (1 + \frac{8}{N_{j-1}}) = 4.1E-2$.

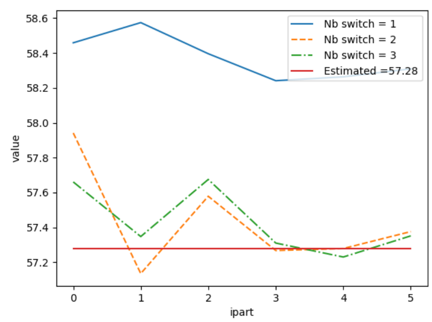
On figure 5, we plot for different values of λ the solution obtained with one, two or three switches with a number of particles $N_0 = 36000 \times 2^{\text{ ipart}}$, $N_1 = 40 \times 2^{\text{ ipart}}$, $N_2 = 2^{\text{ ipart}}$. The solution seems to be 57.28 (value obtained for 3 switches with both $\lambda = 0.1$ and $\lambda = 0.2$ and also obtained with deep learning techniques [5]) and close to the value obtained in [10] who game 57.30.
Two switches are enough to get a good accuracy. For example 57.27 is reached with two switches taking $N_0 = 1152000$, $N_1 = 4480$ in 90 seconds with $\lambda = 0.2$.



$\lambda = 0.2.$



$\lambda = 0.4.$



$\lambda = 0.8.$

Figure 5: Convergence of the scheme on the Black Scholes case with default

3 The semi linear case

In this section we extend the previous scheme obtained to the semi-linear case. To simplify the setting, without restriction, we just take a function depending on Du:

$$\begin{aligned}(-\partial_t u - \mathcal{L}u)(t,x) &= f(t,x,Du(t,x)),\\ u_T &= g, \quad t < T,\ x \in \mathbb{R}^d,\end{aligned} \tag{36}$$

where $\mathcal{L}$ is always given by equation (2) and the dynamic of the underlying SDE is still given by (3) with $\mu \in \mathbb{R}^d$, and $\sigma \in \mathbb{M}^d$ is here some constant non-degenerated matrix.
We will take the same kind of assumption as in the previous section:

Assumption 3.1 *f is uniformly Lipschitz in Du with constant K :*

$$|f(t,x,y) - f(t,x,w)| \le K\|y-w\|_2 \qquad \forall t \in [0,T], x \in \mathbb{R}^d, (w,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{37}$$

Assumption 3.2 *Equation* (36) *has a solution $u \in C^{1,2}([0,T]\times\mathbb{R}^d)$, such that*

- *Du is θ-Hölder with $\theta \in (0,1]$ in time with constant $\hat{K}$:*

$$\|Du(t,x) - Du(\tilde{t},x)\| \le \hat{K}|t-\tilde{t}|^\theta \qquad \forall (t,\tilde{t},x) \in [0,T]\times[0,T]\times\mathbb{R}^d,$$

- *$u(t,x)$ and $Du(t,x)$ have a quadratic growth in x uniformly in t.*

Assumption 3.3 *g is uniformly Lipschitz such that for $\tilde{K} > 0$:*

$$|g(x) - g(y)| \le \tilde{K}\|x-y\|_2 \quad \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d.$$

3.1 General idea of the algorithm

We will propose two algorithms that are some extensions of the algorithm previously given.
As in the previous section, the sequence $(T_i)_{i\ge 0}$ is defined by equation (5) but the $(\tau_m)_{m\ge 1}$ are i.i.d. random variables of density ρ which follow a general gamma distribution so that

$$\rho(x) = \lambda^u x^{u-1}\frac{e^{-\lambda x}}{\Gamma(u)}, u > 0 \tag{38}$$

and the associated cumulated distribution function is

$$F(x) = \frac{\gamma(u,\lambda x)}{\Gamma(u)},$$

where $\gamma(s,x) = \int_0^x t^{s-1}e^{-t}dt$ is the incomplete gamma function.
In order to have a converging method we will see that we will have to take $u < 1$ in ρ expression (38) excluding the exponential distribution. This a weaker constraint than in [18] where, using branching for some polynomial non-linearities, converging results were only obtained for $u < 0.5$.

Under the regularity assumption on u, from the Feynman-Kac formula, the representation of the solution u is

$$\begin{aligned} u(0,x) =& \mathbb{E}_{0,x}\Big[\overline{F}(T)\frac{g(W_T)}{\overline{F}(T)} + \int_0^T \frac{f(t,X_t,Du(t,X_t))}{\rho(t)}\rho(t)dt\Big] \\ =& \mathbb{E}_{0,x}[\hat{\phi}(0,T_1,X_{T_1},Du(T_1,X_{T_1}))], \end{aligned} \tag{39}$$

with

$$\hat{\phi}(s,t,x,z) := \frac{\mathbf{1}_{\{t\geq T\}}}{\overline{F}(T-s)} g(x) + \frac{\mathbf{1}_{\{t<T\}}}{\rho(t-s)} f(t,x,z). \tag{40}$$

then we define $Du(T_1, X_{T_1})$ using the automatic differentiation rule :

$$Du(T_1, X_{T_1}) = \mathbb{E}_{T_1, X_{T_1}}\Big[\sigma^{-\top}\frac{W_{T_2} - W_{T_1}}{T_2 - T_1}\phi(T_1, T_2, X_{T_1}, X_{T_2}, Du(T_2, X_{T_2}))\Big], \tag{41}$$

with

$$\phi(s,t,x,y,z) := \frac{\mathbf{1}_{\{t\geq T\}}}{\overline{F}(T-s)}(g(y) - g(x)) + \frac{\mathbf{1}_{\{t<T\}}}{\rho(t-s)} f(t,y,z). \tag{42}$$

where the $g(x)$ acts as a control variate term.
The automatic differentiation used here is based on the Malliavin integration by parts formula (see [14] for its use in the context of Monte Carlo approximation and the extension to other sensitivities) and has been used in a similar context as the one presented here in [8], [18].
Recursively we define for $n < N_T$:

$$Du_n = \mathbb{E}_{T_n, X_{T_n}}[\sigma^{-\top}\frac{W_{T_{n+1}} - W_{T_n}}{T_{n+1} - T_n}\phi(T_n, T_{n+1}, X_{T_n}, X_{T_{n+1}}, Du_{n+1})], \tag{43}$$

As in the previous section we besides consider the truncated operator after p switches:

$$\begin{aligned}
u_0^p =& \mathbb{E}(\hat{\phi}(0, T_1, X_{T_1}, Du_1^p))\\
Du_n^p =& \mathbb{E}_{T_n, X_{T_n}^n}[\sigma^{-\top}\frac{W_{T_{n+1}} - W_{T_n}}{T_{n+1} - T_n}\phi(T_n, T_{n+1}, X_{T_n}, X_{T_{n+1}}, Du_{n+1}^p)], \quad 1 \leq n < p\\
Du_p^p =& Dg(X_{T_p}).
\end{aligned}$$

The goal of the following section is to present two algorithms based on the previously defined recursion and to show their convergence.

3.2 A first estimator

We take the same notations as in the section 2.2 for the set Q_i, $i < p$, the set $\tilde{Q}(k)$ for $k \in Q_i$. The τ_k are as before some switching increments. They are always i.i.d. random variables with density ρ and for $k \in Q^p$ the $\bar{W}^k$ are some independent d-dimensional Brownian motions, independent of the $(\tau_k)_{k\in Q^p}$ too. The switching

dates are defined by equation (5) and $(X_t^{\tilde{k}})_{t\geq 0}$ defined by (11).
We propose the following estimator defined by:

$$
\begin{cases}
\bar{u}_\emptyset^p = & \frac{1}{N_0}\sum_{j=1}^{N_0}\hat{\phi}(0, T_{(j)}, X_{T_{(j)}}^{(j)}, D\bar{u}_{(j)}^p),\\
D\bar{u}_k^p = & \frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}(k)}\phi(T_k, T_{\tilde{k}}, X_{T_k}^k, X_{T_{\tilde{k}}}^{\tilde{k}}, D\bar{u}_{\tilde{k}}^p)\sigma^{-\top}\frac{\bar{W}_{T_{\tilde{k}}-T_k}^{\tilde{k}}}{T_{\tilde{k}}-T_k}\\
& \quad \text{for } k=(k_1,k_2,..k_i)\in Q_i, i<p,\\
D\bar{u}_{\tilde{k}}^p = & Dg(X_{T_{\tilde{k}}}^{\tilde{k}}) \quad \text{ for } \tilde{k}\in Q_p
\end{cases}
\tag{44}
$$

Remark 3.4 *An estimator of the gradient at the initial date is off course available too as*

$$
D\bar{u}_\emptyset^p = \frac{1}{N_0}\sum_{j=1}^{N_0}\phi(0, T_{(j)}, x, X_{T_{(j)}}^{(j)}, D\bar{u}_{(j)}^p)\sigma^{-\top}\frac{\bar{W}_{T_{(j)}}}{T_{(j)}}
$$

Proposition 3.5 *Suppose that ρ is the density of a gamma law so that $\rho(x) = \lambda^u x^{u-1}\frac{e^{-\lambda x}}{\Gamma(u)}$, suppose that $u<1$, and suppose that assumptions 3.1, 3.2, 3.3 are satisfied then there are two functions C and $\tilde{C}$ depending on σ and one $\bar{C}$ depending on μ, σ and T such that using estimator* (44) *to solve equation* (36)*, we have the error estimate*

$$
\begin{aligned}
\mathbb{E}((\bar{u}_\emptyset^p - u(0,x))^2) \leq & \prod_{i=1}^{p}(1+\frac{8}{N_{i-1}})\frac{\Gamma(u)^p e^{\lambda T}}{\lambda^p}\frac{T^{(1-u)p+1+2\theta}}{(1-u)^{p-1}(2-u)}C(\sigma)^{p-1}\hat{K}^2K^{2p} + \\
& 4\sum_{i=0}^{p-1}\frac{K^{2i}}{N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})\frac{\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^{i+1}}\frac{T^{(1-u)(i+1)+1}}{(1-u)^i(2-u)}\tilde{C}(\sigma)^i\hat{F} + \\
& 2\sum_{i=1}^{p-1}\frac{K^{2i}}{N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})\frac{\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^i}\frac{T^{(1-u)i+1}}{(1-u)^{i-1}(2-u)}\frac{\bar{C}(\mu,\sigma,T)C(\sigma)^{i-1}\tilde{K}^2}{\Gamma(u)-\gamma(u,\lambda T)} + \\
& \frac{2}{N_0}\frac{\Gamma(u)}{\Gamma(u)-\gamma(u,\lambda T)}\mathbb{E}(g(X_T)^2)
\end{aligned}
\tag{45}
$$

where

$$
\hat{F} = \sup_{t\in[0,T]}\mathbb{E}[f(t,X_t,Du(t,X_t))^4]^{\frac{1}{2}}, \tag{46}
$$

Proof. Under assumption 3.2, the solution u of (36) satisfies a Feynman-Kac relation so that for all $\tilde{k}\in\tilde{Q}(\emptyset)$

$$
u(0,x) = \mathbb{E}_{0,x}[\hat{\phi}(0, T_{\tilde{k}}, X_{T_{\tilde{k}}}^{\tilde{k}}, Du(T_{\tilde{k}}, X_{T_{\tilde{k}}}^{\tilde{k}}))], \tag{47}
$$

and that for all $k \in Q_i$ with $i < p$, and $\forall \tilde{k} \in \tilde{Q}(k)$, the gradient is given by:

$$Du(T_k, X^k_{T_k}) = \mathbb{E}_{T_k, X^k_{T_k}}[\sigma^{-\top} \frac{\bar{W}^{\tilde{k}}_{T_{\tilde{k}}-T_k}}{T_{\tilde{k}} - T_k} \phi(T_k, T_{\tilde{k}}, X^k_{T_k}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))], \tag{48}$$

We then introduce for $k \in Q_i$, $1 \leq i < p$:

$$E_k := \mathbb{E}_{T_k, X^k_{T_k}}(\|D\bar{u}^p_k - Du(T_k, X^k_{T_k})\|^2_2 1_{T_k < T}).$$

In exactly the same way as in the demonstration of proposition 2.3, we have the following result similar to the one given by equation (24):

$$\begin{aligned} E_k \leq & \frac{1}{N_i}(1 + \frac{8}{N_i}) \sum_{\tilde{k} \in \tilde{Q}(k)} E_{T_k, X^k_{T_k}}(\frac{(\bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}})^\top \sigma^{-1} \sigma^{-\top} \bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}}}{\tau^2_{\tilde{k}}} \frac{K^2}{\rho(\tau_{\tilde{k}})^2} E_{\tilde{k}}) + \\ & 4 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}<T}} \frac{(\bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}})^\top \sigma^{-1} \sigma^{-\top} \bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}}}{\tau^2_{\tilde{k}}} (\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2) + \\ & 2 \frac{1}{N_i^2} \sum_{\tilde{k} \in \tilde{Q}(k)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k} \geq T}} \frac{(\bar{W}^{\tilde{k}}_{T-T_k})^\top \sigma^{-1} \sigma^{-\top} \bar{W}^{\tilde{k}}_{T-T_k}}{(T - T_k)^2} \frac{(g(X^{\tilde{k}}_T) - g(X^k_{T_k}))^2}{\overline{F}(T - T_k)^2}) \end{aligned} \tag{49}$$

Using equation (24) with $k = \emptyset$ obtained in demonstration of proposition 2.3:

$$\begin{aligned} \mathbb{E}((\bar{u}^p_\emptyset - u(0, x))^2) \leq & \frac{1}{N_0}(1 + \frac{8}{N_0}) \sum_{\tilde{k} \in \tilde{Q}(\emptyset)} \mathbb{E}(\frac{K^2}{\rho(\tau_{\tilde{k}})^2} E_{\tilde{k}}) + \\ & 4 \frac{1}{N_0^2} \sum_{\tilde{k} \in \tilde{Q}(\emptyset)} \mathbb{E}(1_{T_{\tilde{k}<T}} (\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2) + \\ & 2 \frac{1}{N_0^2} \sum_{\tilde{k} \in \tilde{Q}(\emptyset)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k} \geq T}} \frac{g(X^{\tilde{k}}_T)^2}{\overline{F}(T - T_k)^2}). \end{aligned} \tag{50}$$

Iterating from 1 to $p - 1$ we get:

$$\mathbb{E}((\bar{u}^p_\emptyset - u(0, x))^2) \leq A_1 + 4A_2 + 2A_3, \tag{51}$$

where

$$A_1 = \prod_{i=1}^p \frac{1}{N_{i-1}}(1 + \frac{8}{N_{i-1}}) \sum_{\tilde{k}^1 \in \tilde{Q}(\emptyset)} \ldots \sum_{\tilde{k}^p \in \tilde{Q}(\tilde{k}^{p-1})} B_1(\tilde{k}^1, .., \tilde{k}^P) \tag{52}$$

$$A_2 = \sum_{i=0}^{p-1} \frac{K^{2i}}{N_i^2} \prod_{j=1}^{i} \frac{1}{N_{j-1}}(1+\frac{8}{N_{j-1}}), \sum_{\tilde{k}^1 \in \tilde{Q}(\emptyset)} ... \sum_{\tilde{k}^{i+1} \in \tilde{Q}(\tilde{k}^i)} B_2(\tilde{k}^1, .., \tilde{k}^{i+1}), \tag{53}$$

and

$$A_3 = \sum_{i=0}^{p-1} \frac{K^{2i}}{N_i^2} \prod_{j=1}^{i} \frac{1}{N_{j-1}}(1+\frac{8}{N_{j-1}}) \sum_{\tilde{k}^1 \in \tilde{Q}(\emptyset)} ... \sum_{\tilde{k}^{i+1} \in \tilde{Q}(\tilde{k}^i)} B_3(\tilde{k}^1, .., \tilde{k}^{i+1}), \tag{54}$$

where noting

$$\begin{aligned}
\psi_k &= \frac{(\bar{W}^k_{\tau_k})^\top \sigma^{-1} \sigma^{-\top} \bar{W}^k_{\tau_k}}{\tau_k^2}, \\
\hat{f}_k &= f(T_k, X^k_{T_k}, Du(T_k, X^k_{T_k})),
\end{aligned} \tag{55}$$

we have:

$$\begin{aligned}
B_1(\tilde{k}^1, .., \tilde{k}^P) =& \mathbb{E}[\prod_{j=1}^{p} \frac{K^2}{\rho(\tau_{\tilde{k}^j})^2} \prod_{j=2}^{p} \psi_{\tilde{k}^j} E_{\tilde{k}^p}], \\
B_2(\tilde{k}^1, .., \tilde{k}^{i+1}) =& \mathbb{E}\big[1_{T_{\tilde{k}^{i+1}} < T} (\hat{f}_{\tilde{k}^{i+1}})^2 \prod_{j=1}^{i+1} \frac{1}{\rho(\tau_{\tilde{k}^j})^2} \prod_{j=2}^{i+1} \psi_{\tilde{k}^j}\big], \\
B_3(\tilde{k}^1, .., \tilde{k}^{i+1}) =& \mathbb{E}\big[1_{T_{\tilde{k}^{i+1}} \geq T} 1_{T_{\tilde{k}^i} < T} \frac{(g(X_T^{\tilde{k}^{i+1}}) - 1_{i \geq 1} g(X^{\tilde{k}^i}_{T_{\tilde{k}^i}}))^2}{\bar{F}(T - T_{\tilde{k}^i})^2} \prod_{j=1}^{i} \frac{1}{\rho(\tau_{\tilde{k}^j})^2} \prod_{j=2}^{i} \psi_{\tilde{k}^j} \\
& \frac{(\bar{W}^{\tilde{k}^{i+1}}_{T-T_{\tilde{k}^i}})^\top \sigma^{-1} \sigma^{-\top} \bar{W}^{\tilde{k}^{i+1}}_{T-T_{\tilde{k}^i}}}{(T - T_{\tilde{k}^i})^2}\big],
\end{aligned} \tag{56}$$

where $E_{\tilde{k}^p} = 1_{T_{\tilde{k}^p} < T}(Dg(X^{\tilde{k}^p}_{T_{\tilde{k}^p}}) - Du(T_{\tilde{k}^p}, X^{\tilde{k}^p}_{T_{\tilde{k}^p}}))^2$.
We first bound B_1.
We introduce $\psi_k = \frac{\phi_k}{\tau_k}$, where

$$\phi_k = G_k^\top \sigma^{-1} \sigma^{-\top} G_k, \tag{57}$$

and $G_k \in \mathbb{R}^d$ is composed of centered unitary independent Gaussian random variables.
We introduce $C(\sigma) = \mathbb{E}(\phi_k)$ which is independent of k. Using the Hölder property

of Du, the tower property, the independence of the τ_k an the ϕ_k, we get:

$$\begin{aligned}B_1(\tilde{k}^1,..,\tilde{k}^P) =&\mathbb{E}[\prod_{j=1}^{p}\frac{K^2}{\rho(\tau_{\tilde{k}^j})^2}(\prod_{j=2}^{p}\frac{1}{\tau_{\tilde{k}^j}})\mathbb{E}(E_{\tilde{k}^p}\prod_{j=2}^{p}\phi_{\tilde{k}^j}|\tau_{\tilde{k}^1},..,\tau_{\tilde{k}^p})]\\ \leq&\hat{K}^2T^{2\theta}K^{2p}C(\sigma)^{p-1}E[1_{T_{\tilde{k}^p<T}}\prod_{j=1}^{p}\frac{1}{\rho(\tau_{\tilde{k}^j})^2}\prod_{j=2}^{p}\frac{1}{\tau_{\tilde{k}^j}}].\end{aligned}$$

Using the expression for the density of ρ given by (38), we have the bound:

$$\begin{aligned}H :=&E[1_{T_{\tilde{k}^p<T}}\prod_{j=1}^{p}\frac{1}{\rho(\tau_{\tilde{k}^j})^2}\prod_{j=2}^{p}\frac{1}{\tau_{\tilde{k}^j}}]\\ \leq&\frac{\Gamma(u)^pe^{\lambda T}}{\lambda^p}\int_0^T\frac{1}{x^{u-1}}dx\quad(\int_0^T\frac{1}{x^u}dx)^{p-1}\\ =&\frac{\Gamma(u)^pe^{\lambda T}}{\lambda^p}\frac{T^{(1-u)p+1}}{(1-u)^{p-1}(2-u)},\end{aligned}\tag{58}$$

so that

$$B_1(\tilde{k}^1,..,\tilde{k}^P)\leq\frac{\Gamma(u)^pe^{\lambda T}}{\lambda^p}\frac{T^{(1-u)p+1+2\theta}}{(1-u)^{p-1}(2-u)}C(\sigma)^{p-1}\hat{K}^2K^{2p}.\tag{59}$$

In a similar way introducing $\tilde{C}(\sigma)=\mathbb{E}(\phi_k^2)^{\frac{1}{2}}$ and using the tower rule, the independence of the different random variables and Cauchy Schwartz,

$$\begin{aligned}B_2(\tilde{k}^1,..,\tilde{k}^{i+1}) =&\ \mathbb{E}(1_{T_{\tilde{k}^{i+1}}<T}\mathbb{E}(\hat{f}^2_{\tilde{k}^{i+1}}\prod_{j=2}^{i+1}\phi_{\tilde{k}^j}/\tau_{\tilde{k}^1},..,\tau_{\tilde{k}^{i+1}})\frac{1}{\rho(\tau_{\tilde{k}^1})^2}\prod_{j=2}^{i+1}\frac{1}{\rho(\tau_{\tilde{k}^j})^2\tau_{\tilde{k}^j}})\\ \leq&\ \sup_{t\in[0,T]}\mathbb{E}[f(t,X_t,Du(t,X_t))^4]^{\frac{1}{2}}\mathbb{E}(\prod_{j=2}^{i+1}\phi^2_{\tilde{k}^j})^{\frac{1}{2}}E(1_{T_{\tilde{k}^{i+1}}<T}\frac{1}{\rho(\tau_{\tilde{k}^1})^2}\prod_{j=2}^{i+1}\frac{1}{\rho(\tau_{\tilde{k}^j})^2\tau_{\tilde{k}^j}})\\ \leq&\ \frac{\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^{i+1}}\frac{T^{(1-u)(i+1)+1}}{(1-u)^i(2-u)}\tilde{C}(\sigma)^i\sup_{t\in[0,T]}\mathbb{E}[f(t,X_t,Du(t,X_t))^4]^{\frac{1}{2}}.\end{aligned}\tag{60}$$

For B_3 we divide the calculation into two cases :

$$\begin{aligned}B_3(\tilde{k}^1) =&\mathbb{E}(1_{\tau_{\tilde{k}^1}>T}\frac{1}{\bar{F}(T)^2}\mathbb{E}(g(X_T)^2))\\ =&\frac{1}{\bar{F}(T)}\mathbb{E}(g(X_T)^2)\\ =&\frac{\Gamma(u)}{\Gamma(u)-\gamma(u,\lambda T)}\mathbb{E}(g(X_T)^2).\end{aligned}\tag{61}$$

For $i > 0$, using the Lipschitz property of g, the tower property, the independence of the different random variables, and noting

$$\bar{C}(\mu, \sigma, T) = \mathbb{E}[(2\|\mu\|_2^2 T + 2\|\sigma G_{\tilde{k}^{i+1}}\|_2^2)\phi_{\tilde{k}^{i+1}}],$$

we have the following bound:

$$\begin{aligned}
B_3(\tilde{k}^1, ..., \tilde{k}^{i+1}) \leq & \mathbb{E}[1_{T_{\tilde{k}^{i+1}} > T} 1_{T_{\tilde{k}^i} < T} \tilde{K}^2 \frac{(2\|\mu\|_2^2 T + 2\|\sigma G_{\tilde{k}^{i+1}}\|_2^2)\phi_{\tilde{k}^{i+1}}}{\bar{F}(T)^2} \frac{C(\sigma)^{i-1}}{\rho(\tau_{\tilde{k}^1})^2} \prod_{j=2}^{i} \frac{1}{\rho(\tau_{\tilde{k}^j})^2 \tau_{\tilde{k}^j}}] \\
\leq & \frac{\bar{C}(\mu, \sigma, T)\tilde{K}^2}{\bar{F}(T)^2} C(\sigma)^{i-1} \frac{\Gamma(u)^i e^{\lambda T}}{\lambda^i} \int_0^T \frac{1}{x^{u-1}} dx \quad (\int_0^T \frac{1}{x^u} dx)^{i-1} \int_T^\infty \rho(x) dx \\
= & C(\sigma)^{i-1} \frac{\Gamma(u)^{i+1}}{\lambda^i} \frac{T^{(1-u)i+1}}{(1-u)^{i-1}(2-u)} \frac{\bar{C}(\mu, \sigma, T)\tilde{K}^2 e^{\lambda T}}{\Gamma(u) - \gamma(u, \lambda T)}.
\end{aligned} \tag{62}$$

Plugging (59), (60), (61), (62) into (52), (53) and (54) that we insert into (51) gives the desired estimation.

3.3 A second estimator

In this section we present a scheme derived from [26].
We extend the notations from the previous section 3.2. Let set $p \in \mathbb{N}^*$. We construct the sets Q_i^o for $i = 1, .., p$, such that $Q_1^o = \{(k_1), (k_1^-)\}$ where $k_1 \in \{1, .., N_1\}$, so that to a particle noted $(k_1) \in Q_1$, we associate a antithetic particle noted k_1^-. Then the set Q_i^o are defined by recurrence :

$$Q_{i+1}^o = \{(k_1, .., k_i, k_{i+1})/(k_1, .., k_i) \in Q_i^o, k_{i+1} \in \{1, .., N_{i+1}, 1^-, .., N_{i+1}^-\}\}$$

To a particle $k = (k_1, .., k_i) \in Q_i^o$ we associate its original particle $o(k) \in Q_i$ such that $o(k) = (\hat{k}_1, ..\hat{k}_i)$ where $\hat{k}_j = l$ if $k_j = l$ or l^-. Further, when $k = (k_1, \cdots, k_i)$ is such that $k_i \in \mathbb{N}$, we denote $k^- := (k_1, \cdots, k_{i-1}, k_i^-)$.
By convention $T_k = T_{o(k)}$, $\tau_k = \tau_{o(k)}$ and $\bar{W}_t^k = \bar{W}_t^{o(k)}$. For $k = (k_1, ..., k_i) \in Q_i^o$ we introduce the set

- $\tilde{Q}^o(k) = \{l = (k_1, .., k_i, m)/m \in \{1, .., N_i\}\} \subset Q_{i+1}^o$
- and $\hat{Q}^o(k) = \{l = (k_1, .., k_i, m)/m \in \{1, .., N_i, 1^-, .., N_i^-\}\} \subset Q_{i+1}^o$

For $k = (k_1, .., k_i) \in Q_i^o$ and $\tilde{k} = (k_1, .., k_i, k_{i+1}) \in \hat{Q}^o(k)$ we define the following trajectories :

$$W_s^{\tilde{k}} := W_{T_k}^k + \mathbf{1}_{k_{i+1} \in \mathbb{N}} \bar{W}_{s-T_k}^{o(\tilde{k})} - \mathbf{1}_{k_{i+1} \notin \mathbb{N}} \bar{W}_{s-T_k}^{o(\tilde{k})}, \quad \text{and} \tag{63}$$

$$X_s^{\tilde{k}} := x + \mu s + \sigma W_s^{\tilde{k}}, \quad \forall s \in [T_k, T_{\tilde{k}}]. \tag{64}$$

Using the previous definitions, we consider the estimator defined by:

$$\begin{cases} \bar{u}_0^p = & \frac{1}{N_0}\sum_{j=1}^{N_0} \frac{\left(\hat{\phi}\left(0,T_{(j)},X_{T_{(j)}}^{(j)},D\bar{u}_{(j)}^p\right)+\hat{\phi}\left(0,T_{(j)},X_{T_{(j)}}^{(j)^-},D\bar{u}_{(j)^-}^p\right)\right)}{2}), \\ D\bar{u}_k^p = & \frac{1}{N_i}\sum_{\tilde{k}\in\tilde{Q}^o(k)} \tilde{W}^{\tilde{k}}\frac{1}{2}(\hat{\phi}(T_k,T_{\tilde{k}},X_{T_{\tilde{k}}}^{\tilde{k}},D\bar{u}_{\tilde{k}}^p) - \hat{\phi}(T_k,T_{\tilde{k}},X_{T_{\tilde{k}}}^{\tilde{k}^-},D\bar{u}_{\tilde{k}^-}^p)), \\ & \quad \text{for } k=(k_1,..k_i)\in Q_i^o, i<p, \\ D\bar{u}_{\tilde{k}}^p = & Dg(X_{T_{\tilde{k}}}^{\tilde{k}}) \quad \text{for } \tilde{k}\in Q_p^o \end{cases} \tag{65}$$

where $\hat{\phi}$ is defined by equation (40) and $\tilde{W}^{\tilde{k}} = \sigma^{-\top}\frac{\bar{W}^{o(\tilde{k})}_{T_{\tilde{k}}-T_k}}{T_{\tilde{k}}-T_k}$.
The idea is that, for a given $k=(k_1,..,k_i)\in Q_i^o$ and a given $\tilde{k}\in\tilde{Q}^o(k)$, if we have $\bar{W}_{T_{\tilde{k}}-T_k}$ very small then

$$\begin{aligned} \frac{1}{2}(\hat{\phi}(T_k,T_{\tilde{k}},X_{T_{\tilde{k}}}^{\tilde{k}},D\bar{u}_{\tilde{k}}^p) - \hat{\phi}(T_k,T_{\tilde{k}},X_{T_{\tilde{k}}}^{\tilde{k}^-},D\bar{u}_{\tilde{k}^-}^p)) \simeq & D_u\hat{\phi}(T_k,T_{\tilde{k}},X_{T_{\tilde{k}}}^{\tilde{k}},D\bar{u}_{\tilde{k}}^p)(D\bar{u}_{\tilde{k}}^p - D\bar{u}_{\tilde{k}^-}^p)+ \\ & D_x\hat{\phi}(T_k,T_{\tilde{k}},X_{T_{\tilde{k}}}^{\tilde{k}},D\bar{u}_{\tilde{k}}^p)(X_{T_{\tilde{k}}}^{\tilde{k}} - X_{T_{\tilde{k}}}^{\tilde{k}^-}) \\ \simeq & C\sqrt{\tau^{\tilde{k}}}. \end{aligned}$$

So $\tilde{W}^{\tilde{k}}\frac{1}{2}(\hat{\phi}(T_k,T_{\tilde{k}},X_{T_{\tilde{k}}}^{\tilde{k}},D\bar{u}_{\tilde{k}}^p) - \hat{\phi}(T_k,T_{\tilde{k}},X_{T_{\tilde{k}}}^{\tilde{k}^-},D\bar{u}_{\tilde{k}^-}^p))$, as function of $\tau^{\tilde{k}}$, should be bounded, and we hope that estimator (65) has a much smaller variance than estimator (44).

Remark 3.6 *As in a previous scheme an estimation of the gradient is obtained as:*

$$D\bar{u}_0^p = \frac{1}{N_0}\sum_{j=1}^{N_0}\frac{1}{2}\sigma^{-\top}\frac{W_{T_{(j)}}}{T_{(j)}}(\hat{\phi}(0,T_{(j)},X_{T_{(j)}}^{(j)},D\bar{u}_{(j)}^p) - \hat{\phi}(0,T_{(j)},X_{T_{(j)}}^{(j)^-},D\bar{u}_{(j)^-}^p))$$

We need other assumptions on the solution and the driver to fully exploit this scheme:

Assumption 3.7 *Du is uniformly Lipschitz in x such that for $\bar{K}>0$:*

$$\|Du(t,x)-Du(t,y)\|_2 \le \bar{K}\|x-y\|_2 \quad \forall (t,x,y)\in[0,T]\times\mathbb{R}^d\times\mathbb{R}^d.$$

Assumption 3.8 *f is uniformly Lipschitz in x such that there exists $\underline{K}>0$*

$$|f(t,x,z)-f(t,y,z)| \le \underline{K}\|x-y\|_2, \quad \forall (t,x,y,z)\in[0,T]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d.$$

We now give the error estimate with the second scheme

Proposition 3.9 *Suppose that ρ is the density of a gamma law so that $\rho(x) = \lambda^u x^{u-1} \frac{e^{-\lambda x}}{\Gamma(u)}$, suppose that $u < 1$, and suppose that assumptions 3.1, 3.2, 3.3,3.7 and 3.8 are satisfied then there are two functions C and $\tilde{C}$ depending on σ and one $\bar{C}$ depending on K, $\bar{K}$, $\underline{K}$ and σ such that using estimator* (65) *to solve equation* (36)*, we have the error estimate*

$$\mathbb{E}((\bar{u}^p_\emptyset - u(0,x))^2) \leq \prod_{i=1}^{p}(1+\frac{8}{N_{i-1}})\frac{\Gamma(u)^p e^{\lambda T}}{\lambda^p}\frac{T^{(1-u)p+1+2\theta}}{(1-u)^{p-1}(2-u)}C(\sigma)^{p-1}\hat{K}^2 K^{2p} +$$
$$\sum_{i=1}^{p-1}\frac{K^{2i}}{N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})\bar{C}(\sigma,K,\bar{K},\underline{K})C(\sigma)^{i-1}\frac{\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^{i+1}}\frac{T^{(1-u)i+3-u}}{(2-u)^2(1-u)^{i-1}} +$$
$$\sum_{i=1}^{p-1}\frac{K^{2i}}{2N_i}\prod_{j=1}^{i}(1+\frac{8}{N_{j-1}})C(\sigma)^{i-1}\frac{\tilde{K}^2\Gamma(u)^{i+1}e^{\lambda T}}{\lambda^i(\Gamma(u)-\gamma(u,\lambda T))}\tilde{C}(\sigma)\frac{T^{(1-u)i+1}}{(2-u)(1-u)^{i-1}} +$$
$$\frac{4}{N_0}\frac{\Gamma(u)}{\lambda}e^{\lambda T}\frac{T^{2-u}}{2-u}\hat{F} + \frac{2}{N_0}\frac{\Gamma(u)}{\Gamma(u)-\gamma(u,\lambda T)}\mathbb{E}(g(X_T)^2) \tag{66}$$

with $\hat{F}$ given by equation (46)*.*

Proof. First notice that under assumption 3.2, u satisfies (39) and then, for $j \in [0, N_0]$

$$u(0,x) = \mathbb{E}_{0,x}[\frac{\hat{\phi}(0,T_{(j)},X^{(j)}_{T_{(j)}},Du(T_{(j)},X^{(j)}_{T_{(j)}})) + \hat{\phi}(0,T_{(j)},X^{(j)^-}_{T_{(j)}},Du(T_{(j)},X^{(j)^-}_{T_{(j)}})}{2}],$$

and because u satisfies equation (41), we have that for $k \in Q^o_i$, $\tilde{k} \in \tilde{Q}^o(k)$,

$$Du(T_k, X^k_{t_k}) = \mathbb{E}_{T_k,X^k_{t_k}}(\sigma^{-\top}\frac{\bar{W}^{o(\tilde{k})}_{T_{\tilde{k}}-T_k}}{T_{\tilde{k}}-T_k}\frac{1}{2}(\hat{\phi}(T_k,T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}},Du(T_{\tilde{k}},X^{\tilde{k}}_{T_{\tilde{k}}})) - \hat{\phi}(T_k,T_{\tilde{k}},X^{\tilde{k}^-}_{T_{\tilde{k}}},Du(T_{\tilde{k}},X^{\tilde{k}^-}_{T_{\tilde{k}}})))$$

Then we can introduce for $k \in Q_i$, $1 \leq i < p$:

$$E_k := \mathbb{E}_{T_k,X^k_{T_k}}(||D\bar{u}^p_k - Du(T_k,X^k_{T_k})||^2_2 1_{T_k<T}).$$

In exactly the same way as in the demonstration of proposition 3.5, we have the following result similar to the one given by equation (49):

$$
\begin{aligned}
E_k \leq & \frac{1}{N_i}(1+\frac{8}{N_i}) \sum_{\tilde{k}\in\tilde{Q}^o(k)} E_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}<T} \frac{(\bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}})^\top \sigma^{-1}\sigma^{-\top} \bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}}}{\tau^2_{\tilde{k}}} \frac{1}{4\rho(\tau_{\tilde{k}})^2} (f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, D\bar{u}^p_{\tilde{k}}) - \\
& f(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}, D\bar{u}^p_{\tilde{k}^-}) - f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}})) + f(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}})))^2) + \\
& 4\frac{1}{N_i^2} \sum_{\tilde{k}\in\tilde{Q}^o(k)} \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}<T} \frac{(\bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}})^\top \sigma^{-1}\sigma^{-\top} \bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}}}{\tau^2_{\tilde{k}}} \\
& (\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}})) - f(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}))}{2\rho(\tau_{\tilde{k}})})^2) + \\
& 2\frac{1}{N_i^2} \sum_{\tilde{k}\in\tilde{Q}^o(k)} \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}\geq T} \frac{(\bar{W}^{\tilde{k}}_{T-T_k})^\top \sigma^{-1}\sigma^{-\top} \bar{W}^{\tilde{k}}_{T-T_k}}{(T-T_k)^2} \frac{(g(X^{\tilde{k}}_T) - g(X^{\tilde{k}^-}_T))^2}{4\overline{F}(T-T_k)^2}).
\end{aligned}
\tag{67}
$$

Using $(a+b)^2 \leq 2a^2 + 2b^2$, the Lipschitz property of f, we get

$$
\begin{aligned}
E_k \leq & \frac{1}{2N_i}(1+\frac{8}{N_i}) \sum_{\tilde{k}\in\hat{Q}^o(k)} E_{T_k, X^k_{T_k}} (\frac{(\bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}})^\top \sigma^{-1}\sigma^{-\top} \bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}}}{\tau^2_{\tilde{k}}} \frac{K^2}{\rho(\tau_{\tilde{k}})^2} E_{\tilde{k}}) + \\
& \frac{1}{N_i^2} \sum_{\tilde{k}\in\tilde{Q}^o(k)} \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}<T} \frac{(\bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}})^\top \sigma^{-1}\sigma^{-\top} \bar{W}^{\tilde{k}}_{\tau_{\tilde{k}}}}{\tau^2_{\tilde{k}}} \\
& (\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}})) - f(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2) + \\
& \frac{1}{2N_i^2} \sum_{\tilde{k}\in\tilde{Q}^o(k)} \mathbb{E}_{T_k, X^k_{T_k}} (1_{T_{\tilde{k}}\geq T} \frac{(\bar{W}^{\tilde{k}}_{T-T_k})^\top \sigma^{-1}\sigma^{-\top} \bar{W}^{\tilde{k}}_{T-T_k}}{(T-T_k)^2} \frac{(g(X^{\tilde{k}}_T) - g(X^{\tilde{k}^-}_T))^2}{\overline{F}(T-T_k)^2}).
\end{aligned}
\tag{68}
$$

Similarly to (50)

$$
\begin{aligned}
D =&\mathbb{E}((\bar{u}^p_\emptyset - u(0,x))^2) \\
\leq& \frac{1}{N_0}(1+\frac{8}{N_0}) \sum_{\tilde{k}\in\tilde{Q}^o(\emptyset)} \mathbb{E}(1_{T_{\tilde{k}<T}} \frac{K^2}{4\rho(\tau_{\tilde{k}})^2}(f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, D\bar{u}^p_{\tilde{k}})+ \quad (69)\\
& f(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}, D\bar{u}^p_{\tilde{k}^-}) - f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}})) - f(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}})))^2))+\\
& 4\frac{1}{N_0^2}\sum_{\tilde{k}\in\tilde{Q}^o(\emptyset)} \mathbb{E}(1_{T_{\tilde{k}<T}}(\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}})) + f(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}^-}_{T_{\tilde{k}}}))}{2\rho(\tau_{\tilde{k}})})^2)+\\
& 2\frac{1}{N_0^2}\sum_{\tilde{k}\in\tilde{Q}^o(\emptyset)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}\geq T}} \frac{(g(X^{\tilde{k}}_T) + g(X^{\tilde{k}^-}_T))^2}{4\overline{F}(T-T_k)^2}) \quad (70)
\end{aligned}
$$

so that using the Lipschitz property of f

$$
\begin{aligned}
\mathbb{E}((\bar{u}^p_\emptyset - u(0,x))^2) \leq& \frac{1}{2N_0}(1+\frac{8}{N_0}) \sum_{\tilde{k}\in\hat{Q}^o(\emptyset)} \mathbb{E}(1_{T_{\tilde{k}<T}} \frac{K^2}{\rho(\tau_{\tilde{k}})^2} E_{\tilde{k}})+\\
& 2\frac{1}{N_0^2}\sum_{\tilde{k}\in\hat{Q}^o(\emptyset)} \mathbb{E}(1_{T_{\tilde{k}<T}}(\frac{f(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}, Du(T_{\tilde{k}}, X^{\tilde{k}}_{T_{\tilde{k}}}))}{\rho(\tau_{\tilde{k}})})^2)+\\
& \frac{1}{N_0^2}\sum_{\tilde{k}\in\hat{Q}^o(\emptyset)} \mathbb{E}_{T_k, X^k_{T_k}}(1_{T_{\tilde{k}\geq T}} \frac{g(X^{\tilde{k}}_T)^2}{\overline{F}(T-T_k)^2}). \quad (71)
\end{aligned}
$$

Then we get that

$$
\mathbb{E}((\bar{u}^p_\emptyset - u(0,x))^2) \leq \hat{A}_1 + \hat{A}_2 + \hat{A}_3, \quad (72)
$$

where the terms $\hat{A}_1$, $\hat{A}_2$ and $\hat{A}_3$ are given by:

- $$
\hat{A}_1 = \prod_{i=1}^{p} \frac{1}{2N_{i-1}^2}(1+\frac{8}{N_{i-1}}) \sum_{\tilde{k}^1\in\hat{Q}^o(\emptyset)} \dots \sum_{\tilde{k}^p\in\hat{Q}^o(\tilde{k}^{p-1})} B_1(\tilde{k}^1, .., \tilde{k}^P) \quad (73)
$$

  where B_1 is given by (56) and bounded by (59),

- $$\hat{A}_2 = \frac{2}{N_0^2} \sum_{\tilde{k}^1 \in \hat{Q}^o(\emptyset)} B_2(\tilde{k}^1) + \sum_{i=1}^{p-1} \frac{K^{2i}}{N_i^2} \prod_{j=1}^{i} \frac{1}{2N_{j-1}}(1 + \frac{8}{N_{j-1}}) \sum_{\tilde{k}^1 \in \hat{Q}^o(\emptyset)} ... \sum_{\tilde{k}^i \in \hat{Q}^o(\tilde{k}^{i-1})} \sum_{\tilde{k}^{i+1} \in \tilde{Q}^o(\tilde{k}^i)} \hat{B}_2(\tilde{k}^1, .., \tilde{k}^{i+1}), \tag{74}$$

  where B_2 is given by equation (56) and bounded by equation (60) so that

  $$B_2(\tilde{k}^1) \leq \frac{\Gamma(u)}{\lambda} e^{\lambda T} \frac{T^{2-u}}{2-u} \hat{F}, \tag{75}$$

  with $\hat{F}$ given by equation (46), and where using notation given by (55)

  $$\hat{B}_2(\tilde{k}^1, .., \tilde{k}^{i+1}) = \mathbb{E}[1_{T_{\tilde{k}^{i+1}} < T} (\hat{f}_{\tilde{k}^{i+1}} - \hat{f}_{(\tilde{k}^{i+1})^-})^2 \prod_{j=1}^{i+1} \frac{1}{\rho(\tau_{\tilde{k}^j})^2} \prod_{j=2}^{i+1} \psi_{\tilde{k}^j}], \tag{76}$$

- $$\hat{A}_3 = \frac{1}{N_0^2} \sum_{\tilde{k}^1 \in \hat{Q}^o(\emptyset)} \hat{B}_3(\tilde{k}^1) + \sum_{i=1}^{p-1} \frac{K^{2i}}{N_i^2} \prod_{j=1}^{i} \frac{1}{2N_{j-1}}(1 + \frac{8}{N_{j-1}}) \sum_{\tilde{k}^1 \in \hat{Q}^o(\emptyset)} ... \sum_{\tilde{k}^i \in \hat{Q}^o(\tilde{k}^{i-1})} \sum_{\tilde{k}^{i+1} \in \tilde{Q}^o(\tilde{k}^i)} \hat{B}_3(\tilde{k}^1, .., \tilde{k}^{i+1}), \tag{77}$$

  where using notation given by (55)

  $$\hat{B}_3(\tilde{k}^1, .., \tilde{k}^{i+1}) = \mathbb{E}[1_{T_{\tilde{k}^{i+1}} > T} 1_{T_{\tilde{k}^i} < T} \frac{(g(X_T^{\tilde{k}^{i+1}}) - 1_{i>1} g(X_{T_{\tilde{k}^i}}^{(\tilde{k}^{i+1})^-}))^2}{\bar{F}(T - T_{\tilde{k}^i})^2} \prod_{j=1}^{i} \frac{1}{\rho(\tau_{\tilde{k}^j})^2} \prod_{j=2}^{i} \psi_{\tilde{k}^j} \frac{(\bar{W}_{T-T_{\tilde{k}^i}}^{\tilde{k}^{i+1}})^\top \sigma^{-1} \sigma^{-\top} \bar{W}_{T-T_{\tilde{k}^i}}^{\tilde{k}^{i+1}}}{(T - T_{\tilde{k}^i})^2}] \tag{78}$$

We can bound $\hat{B}_2$ using definitions in equations (55) and(57)

$$\hat{B}_2(\tilde{k}^1, .., \tilde{k}^{i+1}) = \mathbb{E}[1_{T_{\tilde{k}^{i+1}} < T} \mathbb{E}((\hat{f}_{\tilde{k}^{i+1}} - \hat{f}_{(\tilde{k}^{i+1})^-})^2 \prod_{j=2}^{i+1} \phi_{\tilde{k}^j} | \tau_{\tilde{k}^1}, ..., \tau_{\tilde{k}^{i+1}}) \frac{1}{\tau_{\tilde{k}^{i+1}} \rho(\tau_{\tilde{k}^{i+1}})^2} \prod_{j=1}^{i} \frac{1}{\rho(\tau_{\tilde{k}^j})^2} \prod_{j=2}^{i} \frac{1}{\tau_{\tilde{k}^j}}].$$

Using assumptions 3.1, 3.7 and 3.8,

$$
\begin{aligned}
|\hat{f}_{\tilde{k}^{i+1}} - \hat{f}_{(\tilde{k}^{i+1})^-}| \leq & 2\bar{K} \|\sigma W^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}} - T_{\tilde{k}^i}}\|_2 + K\|Du(T_{\tilde{K}^{i+1}}, X^{\tilde{k}^{i+1}}_{T_{\tilde{K}^{i+1}}}) - Du(T_{\tilde{K}^{i+1}}, X^{(\tilde{k}^{i+1})^-}_{T_{\tilde{K}^{i+1}}})\|_2 \\
\leq & 2(\underline{\mathrm{K}} + K\bar{K})\|\sigma W^{\tilde{k}^{i+1}}_{T_{\tilde{k}^{i+1}} - T_{\tilde{k}^i}}\|_2
\end{aligned}
$$

so that :

$$
\begin{aligned}
R = & \mathbb{E}((\hat{f}_{\tilde{k}^{i+1}} - \hat{f}_{(\tilde{k}^{i+1})^-})^2 \prod_{j=2}^{i+1} \phi_{\tilde{k}^j} | \tau_{\tilde{k}^1}, ..., \tau_{\tilde{k}^{i+1}}) \\
\leq & \bar{C}(\sigma, K, \bar{K}, \underline{\mathrm{K}}) C(\sigma)^{i-1} \tau_{\tilde{k}^{i+1}}.
\end{aligned}
$$

Then

$$
\begin{aligned}
\hat{B}_2(\tilde{k}^1, .., \tilde{k}^{i+1}) \leq & \bar{C}(\sigma, K, \bar{K}, \underline{\mathrm{K}}) C(\sigma)^{i-1} \mathbb{E}[1_{T_{\tilde{k}^{i+1}} < T} \frac{1}{\rho(\tau_{\tilde{k}^{i+1}})^2} \prod_{j=1}^{i} \frac{1}{\rho(\tau_{\tilde{k}^j})^2} \prod_{j=2}^{i} \frac{1}{\tau_{\tilde{k}^j}}] \\
\leq & \bar{C}(\sigma, K, \bar{K}, \underline{\mathrm{K}}) C(\sigma)^{i-1} \frac{\Gamma(u)^{i+1} e^{\lambda T}}{\lambda^{i+1}} (\int_0^T \frac{1}{x^{u-1}} dx)^2 (\int_0^T \frac{1}{x^u} dx)^{i-1}, \\
= & \bar{C}(\sigma, K, \bar{K}, \underline{\mathrm{K}}) C(\sigma)^{i-1} \frac{\Gamma(u)^{i+1} e^{\lambda T}}{\lambda^{i+1}} \frac{T^{(1-u)i+3-u}}{(2-u)^2 (1-u)^{i-1}}.
\end{aligned} \tag{79}
$$

Similarly

$$
\hat{B}_3(\tilde{k}^1) = B_3(\tilde{k}^1) \tag{80}
$$

with B_3 given by (61).
For the general B_3 term for $i > 0$, using assumption 3.3 and using the notation $\bar{C}(\sigma) = \mathbb{E}[4\phi_{\tilde{k}^{i+1}} \|\sigma G_{\tilde{k}^{i+1}}\|_2^2]$:

$$
\begin{aligned}
\hat{B}_3(\tilde{k}^1, .., \tilde{k}^{i+1}) = & \mathbb{E}[1_{T_{\tilde{k}^{i+1}} > T} 1_{T_{\tilde{k}^i} < T} \frac{(g(X_T^{\tilde{k}^{i+1}}) - g(X^{(\tilde{k}^{i+1})^-}_{T_{\tilde{k}^i}}))^2}{(T - T_{\tilde{k}^i}) \bar{F}(T - T_{\tilde{k}^i})^2} \frac{1}{\rho(\tau_{\tilde{k}^1})^2} \prod_{j=2}^{i} \frac{1}{\tau_{\tilde{k}^j} \rho(\tau_{\tilde{k}^j})^2} \prod_{j=2}^{i+1} \phi_{\tilde{k}^j}] \\
\leq & C(\sigma)^{i-1} \frac{\tilde{K}^2}{\bar{F}(T)^2} \mathbb{E}[4\phi_{\tilde{k}^{i+1}} \|\sigma G_{\tilde{k}^{i+1}}\|_2^2] \mathbb{E}(1_{T_{\tilde{k}^{i+1}} > T} 1_{T_{\tilde{k}^i} < T} \frac{1}{\rho(\tau_{\tilde{k}^1})^2} \prod_{j=2}^{i} \frac{1}{\tau_{\tilde{k}^j} \rho(\tau_{\tilde{k}^j})^2}) \\
\leq & C(\sigma)^{i-1} \frac{\tilde{K}^2 e^{\lambda T} \Gamma(u)^i}{\lambda^i \bar{F}(T)^2} \bar{C}(\sigma) \int_0^T \frac{1}{x^{u-1}} dx (\int_0^T \frac{1}{x^u} dx)^{i-1} \int_T^\infty \rho(x) dx \\
\leq & C(\sigma)^{i-1} \frac{\tilde{K}^2 \Gamma(u)^i e^{\lambda T}}{\bar{F}(T) \lambda^i} \bar{C}(\sigma) \frac{T^{(1-u)i+1}}{(2-u)(1-u)^{i-1}}
\end{aligned} \tag{81}
$$

Plugging (59), (75), (79),(80),(81) in (73), (74), (77) and (73), (74), (77) in (72) give the result.

Remark 3.10 *The result obtained is however a little bit disappointing: in the case of the linear driver the result can be improved and it can be shown that the error goes to zero even using an exponential law. In the general case, proposition 3.9 gives us that the variance is finite using an exponential law for ρ only if $p \leq 2$.*

Remark 3.11 *In the case of non constant coefficients, most of the time it is necessary to use an Euler scheme. As ready pointing out in [26], the first estimator has an exploding variance because the integration by part has to be achieved on the first time step of the Euler scheme using a mesh of size Δt. It gives a Malliavin weight in $O(\frac{1}{\sqrt{\Delta t}})$ leading to an explosion in variance as the step size goes to zero. This second estimator doesn't suffer from this problem.*

3.4 Numerical tests for the semi-linear case

In this section we give some numerical results illustrating the previous results obtained. In the whole section the number of particles taken at each level will be a sequence $(N_i^{ipart})_{i\geq 0}$ indexed by *ipart* such that:

$$N_i^{ipart} = N_i^0 \times 2^{ipart}. \tag{82}$$

3.5 A Bürgers test case.

We take the test case proposed in [9] which is derived from a test in [6]. We take the same parameters as in [9]: $\mu = 0$, $\sigma = d\mathbf{I}_d$, $T = 1$, the driver is given for $x \in \mathbb{R}^d, y \in \mathbb{R}, z \in \mathbb{R}^d$ by

$$f(t,x,y,z) = (y - \frac{2+d}{2d})(d\sum_{i=1}^d z_i),$$

and the final function is :

$$g(x) = \frac{e^{T+\frac{1}{d}\sum_{i=1}^d x_i}}{1+e^{T+\frac{1}{d}\sum_{i=1}^d x_i}}.$$

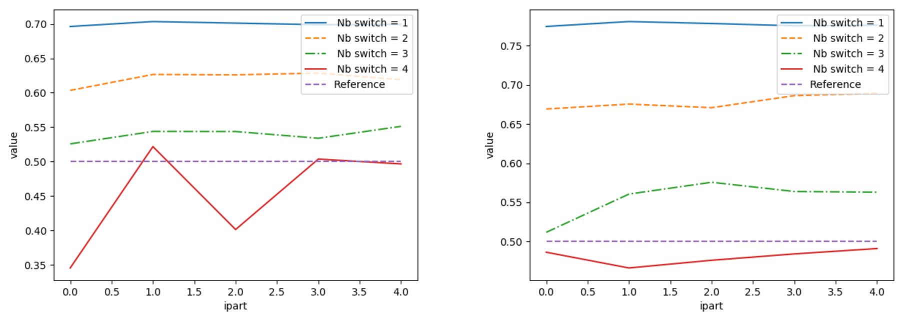
The explicit solution given by [9] is

$$u(t,x) = \frac{e^{t+\frac{1}{d}\sum_{i=1}^d x_i}}{1+e^{t+\frac{1}{d}\sum_{i=1}^d x_i}}.$$

We solve the problem in dimension $d = 10$ and $d = 20$ at date $t = 0$ and for $x = 0\,\mathbb{1}_d$ such that the reference is equal to 0.5.
We test the different schemes for a gamma law given by (38) with $\lambda = 0.1$ and

$\lambda = 0.2$.
For the first scheme we take $u = 0.8$ in (38) and give the results obtained on figure 6 and 7 for estimator (44) with $(N_0^0, N_1^0, N_2^0, N_3^0) = (1000, 40, 40, 30)$ in (82).
In dimension 10 and 20, we obtain very good results with 4 switches and $ipart = 4$,

- getting in dimension 10 a value 0.496 for $\lambda = 0.1$ in 80 seconds and 0.4910 for $\lambda = 0.2$ in 1000 seconds,
- getting in dimension 20 a value 0.501 for $\lambda = 0.1$ in 350 seconds and 0.5006 for $\lambda = 0.2$ in 1400 seconds.



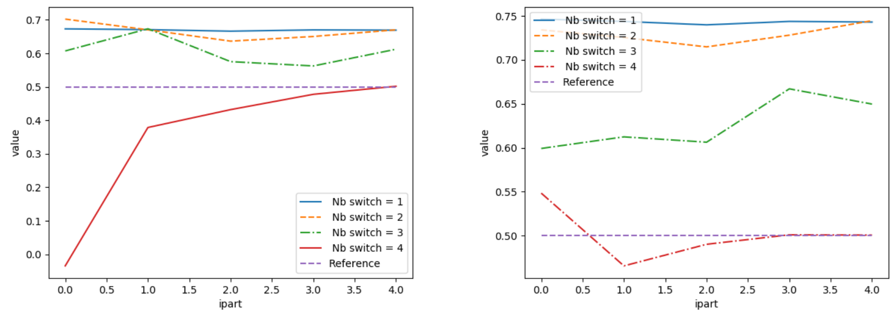
$\lambda = 0.1$. $\lambda = 0.2$.

Figure 6: Bürgers case: convergence with estimator (44) in dimension 10 using a gamma law with $u = 0.8$.

For the second scheme (65), taking $u = 0.9$ in (38), we give the results obtained on figure 8 and 9 for estimator 44 with $(N_0^0, N_1^0, N_2^0, N_3^0) = (1000, 40, 40, 30)$ in (82). In dimension 10 and 20 , we obtain very good results with 4 switches and $ipart = 3$,

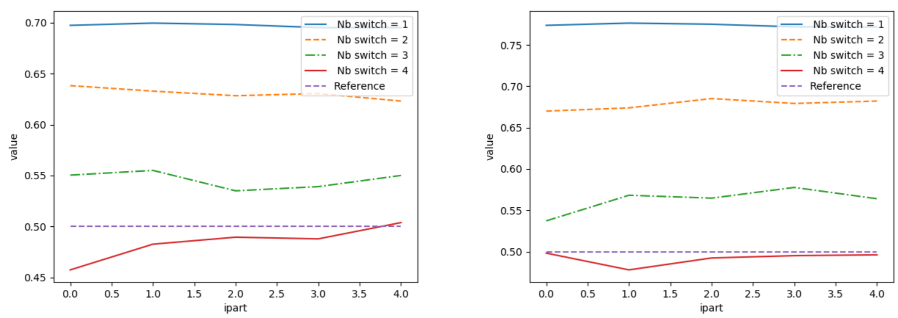
- getting in dimension 10 a value 0.487 for $\lambda = 0.1$ in 27 seconds and 0.49509 for $\lambda = 0.2$ in 130 seconds,
- getting in dimension 20 a value 0.5040 for $\lambda = 0.1$ in 40 seconds and 0.4979 for $\lambda = 0.2$ in 186 seconds.

At last we use an exponential law for ρ, leading to $u = 1$ in (38) and we use the estimator (65). We take $(N_0^0, N_1^0, N_2^0, N_3^0) = (1000, 40, 40, 4)$ in (82) and only give the results on figure 10 for the most difficult case $d = 20$.
With 4 switches, $ipart = 3$ we get very good results :



$\lambda = 0.1$ $\lambda = 0.2$

Figure 7: Bürgers case: convergence with estimator (44) in dimension 20 using a gamma law with $u = 0.8$.

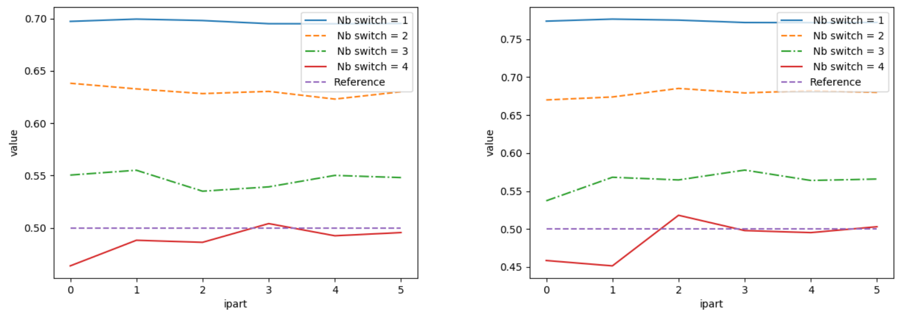


$\lambda = 0.1$. $\lambda = 0.2$.

Figure 8: Bürgers case: convergence with estimator (65) in dimension 10 using a gamma law with $u = 0.9$.

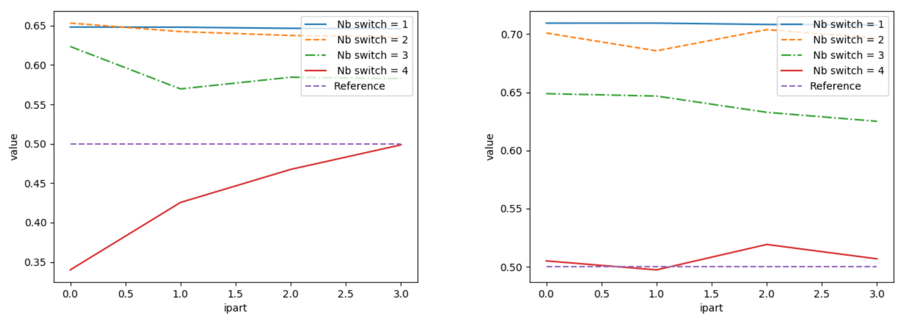
- A solution equal to 0.499 with a computational time equal to 14 seconds with $\lambda = 0.1$,
- A solution equal to 0.506 with a computational time equal to 55 seconds with $\lambda = 0.2$.

The variance using exponential laws seems to be lower and the generation of an exponential law takes far less time than with general gamma laws.



$\lambda = 0.1$. $\lambda = 0.2$.

Figure 9: Bürgers case: convergence with estimator (65) in dimension 20 with a gamma law with $u = 0.9$.



$\lambda = 0.1$. $\lambda = 0.2$.

Figure 10: Bürgers case: convergence with estimator (65) in dimension 20 with an exponential law.

3.5.1 A second case

We then take the HJB equation test case taken from [9],[10], [7]. As in [10] we solve the problem in dimension 100 to show the efficiency with the same characteristics as in [9],[10]:

$$\begin{aligned} \mu =& 0, \quad \sigma = \sqrt{2}\mathbf{I}_d, \quad T = 1, \\ f(t,x,z) =& -\theta \|z\|_2^2, \end{aligned}$$

such that a semi-explicit solution is available :

$$u(t,x) = -\frac{1}{\theta}\log\left(\mathbb{E}[e^{-\theta g(x+\sqrt{2}W_{T-t})}]\right). \tag{83}$$

In the example, we take the g function as in [9]:

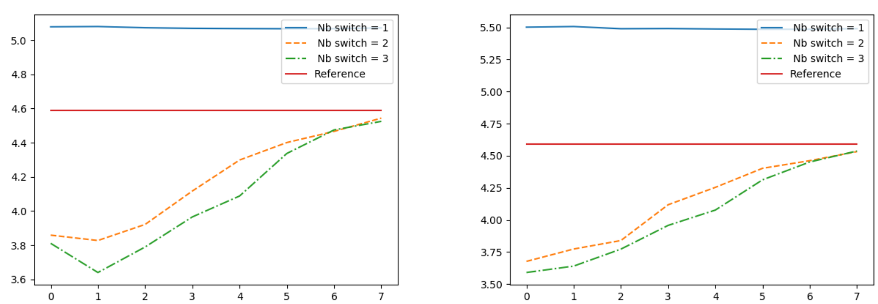
$$g(x) = \log\left(\frac{1+\|x\|_2^2}{2}\right),$$

and we want to estimate the solution at date $t = 0$ and for $x = 0\mathbb{1}_d$ using our algorithm.
We treat three cases with increasing difficulty by taking $\theta = 1$, then $\theta = 10$ then at last $\theta = 20$. The difficulty comes from an increasing value of the non linearity. Using a Monte Carlo approximation of equation (83) we get some references and a good approximation of the solution is 4.59 with $\theta = 1$, 4.49 with $\theta = 10$, and 4.36 with $\theta = 20$. In order to fit the framework we modify the non linearity to

$$f(t,x,z) = -\theta\min(\|z\|_2^2, 1),$$

but the truncation has in fact no effect on the method.

First for $\theta = 1$ we plot on figure 11 the results obtained by estimator (44) taking in equation (82): $(N_0^0, N_1^0, N_2^0) = (1000, 20, 20)$. The convergence with this scheme



$\lambda = 0.1$ $\lambda = 0.2$

Figure 11: HJB convergence case for $\theta = 1$ and estimator 44 using a gamma law with $u = 0.8$.

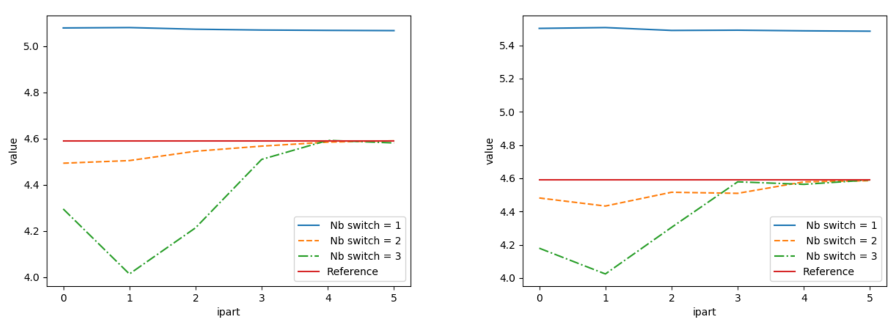
is quite slow for both $\lambda = 0.1$ and $\lambda = 0.2$. The number of switches to take to have a very good accuracy seems to be equal to 3 but the variance observed with this estimator is quite high. Experiments with $\theta = 10$ or $\theta = 20$ show that this kind of singularity is hard to cope with so high Lipschitz constants of the driver.

For the three values of θ on figures 12, 13 and 14 we plot the results for gamma law given by (38) with $u = 0.9$ and estimator (65) taking in equation (82):

- $(N_0^0, N_1^0, N_2^0) = (1000, 10, 1)$ for $\theta = 1$,
- $(N_0^0, N_1^0, N_2^0) = (1000, 10, 5)$ for $\theta = 10$,
- $(N_0^0, N_1^0, N_2^0) = (1000, 40, 20)$ for $\theta = 20$,

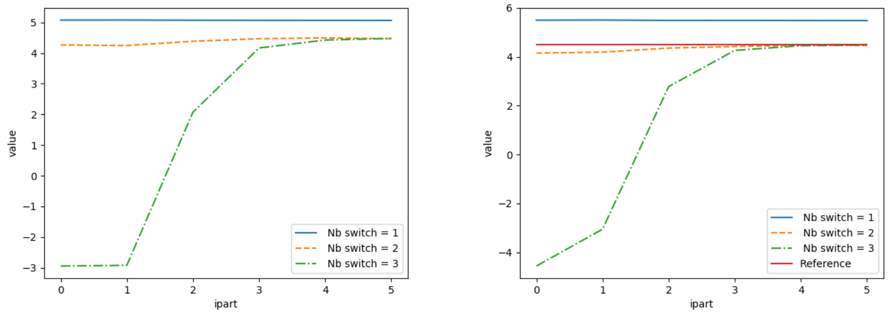
such that as the different Lipschitz constants increase we increase the numbers of samples taken in inner nesting.
As for the results obtained using estimator (65), a good accuracy is obtained using 2 switches:

- with $\theta = 1$, $\lambda = 0.1$, $ipart = 3$, we get 4.57 whereas taking $\lambda = 0.2$, $ipart = 4$ gives 4.58.
- with $\theta = 10$, $\lambda = 0.1$, $ipart = 3$, we get 4.47 whereas taking $\lambda = 0.2$, $ipart = 4$ gives 4.48.
- with $\theta = 20$, $\lambda = 0.1$, $ipart = 4$, we get 4.37 whereas taking $\lambda = 0.2$, $ipart = 4$ gives 4.33.



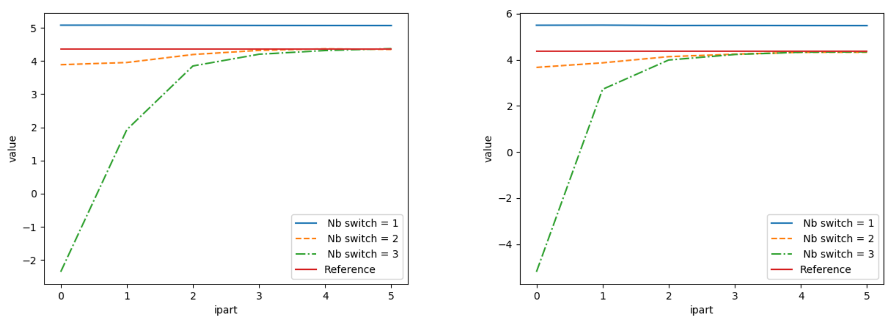
$\lambda = 0.1$ $\qquad\qquad$ $\lambda = 0.2$

Figure 12: HJB convergence case for $\theta = 1$ and estimator (65) using a gamma law with $u = 0.9$.



$\lambda = 0.1$ $\lambda = 0.2$

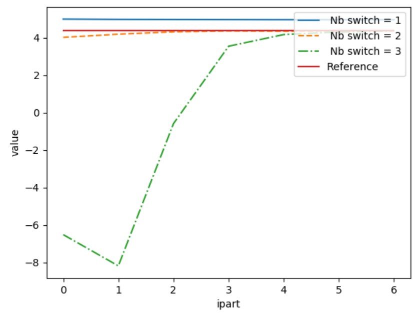
Figure 13: HJB convergence case for $\theta = 10$ and estimator (65) using a gamma law with $u = 0.9$.



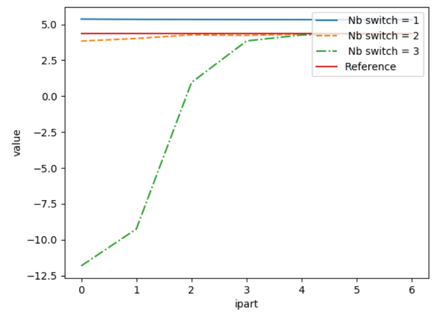
$\lambda = 0.1$ $\lambda = 0.2$

Figure 14: HJB convergence case for $\theta = 20$ and estimator (65) using a gamma law with $u = 0.9$.

At last on figure 15, we plot the solution obtained for the most difficult case ($\theta = 20$) using an exponential law for ρ and $(N_0^0, N_1^0, N_2^0) = (1000, 40, 10)$. A good solution is obtained in 10 seconds for $\lambda = 0.1$ with 2 switches with a precision of less than 0.5% using $(N_0, N_1) = (8000, 320)$. To get a very accurate solution with a precision of 0.1% with both λ it is then necessary to use 3 switches with $(N_0, N_1, N_2) = (64000, 2560, 640)$ and the computational time explodes to nearly 30000 seconds with $\lambda = 0.1$ and 80000 seconds with $\lambda = 0.2$.



$\lambda = 0.1$



$\lambda = 0.2$

Figure 15: HJB convergence case for $\theta = 20$ and estimator (65) using an exponential law.

4 Conclusion

An effective method to solve semi-linear equations has been developed and tested. The most effective way to solve these equations consists in taking the second scheme proposed to treat the gradient term. The limitation due to Lipschitz constant and the maturity of the problem can be easily postponed using cluster of CPU or perhaps GPU.
Besides even if cannot prove that the second scheme can be used with an exponential law, it seems to be the most effective. A better understanding of its efficiency could pave the way to solve full non linear PDEs.

References

[1] H. Bauke. Tinas random number generator library, 2011.

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

[3] B. Bouchard, X. Tan, and X. Warin. Numerical approximation of general lipschitz bsdes with branching processes. *arXiv preprint arXiv:1710.10933*, 2017.

[4] 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*, 23(4):241–263, 2017.

[5] Q. Chan and J. Mikael. Personal communication.

[6] J.-F. Chassagneux. Linear multistep schemes for bsdes. *SIAM Journal on Numerical Analysis*, 52(6):2815–2836, 2014.

[7] J.-F. Chassagneux, A. Richou, et al. Numerical simulation of quadratic bsdes. *The Annals of Applied Probability*, 26(1):262–304, 2016.

[8] M. Doumbia, N. Oudjane, and X. Warin. Unbiased monte carlo estimate of stochastic differential equations expectations. *ESAIM: Probability and Statistics*, 21:56–87, 2017.

[9] W. E, J. Han, and A. Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, 5(4):349–380, 2017.

[10] W. E, J. Han, and A. Jentzen. Overcoming the curse of dimensionality: Solving high-dimensional partial differential equations using deep learning. *arXiv preprint arXiv:1707.02568*, 2017.

[11] W. E, M. Hutzenthaler, A. Jentzen, and T. Kruse. On multilevel picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *arXiv preprint arXiv:1607.03295*, 46, 2016.

[12] W. E, M. Hutzenthaler, A. Jentzen, and T. Kruse. Linear scaling algorithms for solving high-dimensional nonlinear parabolic differential equations. *SAM Research Report*, 2017, 2017.

[13] A. Fahim, N. Touzi, and X. Warin. A probabilistic numerical method for fully nonlinear parabolic pdes. *The Annals of Applied Probability*, pages 1322–1364, 2011.

[14] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of malliavin calculus to monte carlo methods in finance. *Finance and Stochastics*, 3(4):391–412, 1999.

[15] M. Fujii, A. Takahashi, and M. Takahashi. Asymptotic expansion as prior knowledge in deep learning method for high dimensional bsdes. 2017.

[16] E. Gobet, J.-P. Lemor, X. Warin, et al. A regression-based monte carlo method to solve backward stochastic differential equations. *The Annals of Applied Probability*, 15(3):2172–2202, 2005.

[17] P. Henry-Labordere. Deep primal-dual algorithm for bsdes: Applications of machine learning to cva and im. 2017.

[18] P. Henry-Labordere, N. Oudjane, X. Tan, N. Touzi, and X. Warin. Branching diffusion representation of semilinear pdes and monte carlo approximation. *arXiv preprint arXiv:1603.01727*, 2016.

[19] M. Hutzenthaler and T. Kruse. Multi-level picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *arXiv preprint arXiv:1711.01080*, 2017.

[20] J.-P. Lemor, E. Gobet, X. Warin, et al. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5):889–916, 2006.

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

[22] T. Rainforth, R. Cornish, H. Yang, A. Warrington, and F. Wood. On the opportunities and pitfalls of nesting monte carlo estimators. *arXiv preprint arXiv:1709.06181*, 2017.

[23] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic analysis and applications*, 8(4):483–509, 1990.

[24] X. Tan. A splitting method for fully nonlinear degenerate parabolic pdes. *Electronic Journal of Probability*, 18, 2013.

[25] N. Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29. Springer Science & Business Media, 2012.

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

Finance for Energy Market Research Centre

Institut de Finance de Dauphine, Université Paris-Dauphine
1 place du Maréchal de Lattre de Tassigny
75775 PARIS Cedex 16
www.fime-lab.org