FiNANCE FOR ENERGY
MARKET RESEARCH CENTRE

Computing Expectations for General SDE with Pure Monte Carlo Methods

Mahamadou Doumbia, Nadia Oudjane, Xavier Warin

Working Paper
RR-FiME-16-01

January 2016

DAUPHINE UNIVERSITÉ PARIS

CREST Centre de Recherche en Économie et Statistique

ÉCOLE POLYTECHNIQUE

EDF

Computing expectations for general SDE with pure Monte Carlo methods

Mahamadou Doumbia   Nadia Oudjane *   Xavier Warin *

January 11, 2016

Abstract

We develop a pure Monte Carlo method to compute $\mathbb{E}(g(X_T))$ where g is a bounded and Lipschitz function and X_t an Ito process. This approach extends the method proposed in [7] to the general multidimensional case with a SDE with varying coefficients. A variance reduction method relying on interacting particle systems is also developped.

Key words: Exact simulation of SDEs, linear parabolic PDEs
MSC2010: Primary 65C05, 60J60; secondary 60J85, 35K10

1 Introduction

Let $d \geq 1$ and W be a $d-$dimensional Brownian motion. We introduce the process X defined as the unique strong solution of the multi-dimensional SDE with coefficients satisfying the usual Lipschitz conditions :

$$\begin{cases} dX_t^{0,x_0} &= b(t, X_t^{0,x_0})dt + \sigma(t, X_t^{0,x_0})dW_t, \\ X_0^{0,x_0} &= x_0, \end{cases} \tag{1.1}$$

where $b : [0,T] \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is the drift and $\sigma : [0,T] \times \mathbb{R}^d \longrightarrow \mathcal{S}^d$ is the diffusion of the process, $\mathcal{S}^d$ being the set of $d \times d$ dimensional matrices.

In this paper, we are interested in a Monte Carlo approach to compute an expectation of the form

$$u(t,x) := \mathbb{E}[g(X_T^{t,x})] \ . \tag{1.2}$$

When no explicit solution is available, the classical method to solve equation (1.2) consists in using a discretization scheme of (1.1) (for example Euler scheme [8], Milstein scheme [9], Burrage schemes [2]) and the error can be decomposed as a sum of an error due to the discretization time step δt and a statistical error of order $1/\sqrt{N}$ due to the Monte Carlo method for a number N of simulations.

In this paper we propose to extend a method originally developed in [7]. The main idea developed in this seminal paper is to start by simulating exactly a SDE :

$$\begin{cases} d\hat{X}_t^{0,x_0} &= \hat{b}(t, \hat{X}_t^{0,x_0})dt + \hat{\sigma}(t, \hat{X}_t^{0,x_0})dW_t \\ \hat{X}_0^{0,x_0} &= x_0 \ , \end{cases}$$

*EDF R&D & FiME, Laboratoire de Finance des Marchés de l'Energie (www.fime-lab.org)

where the coefficients $\hat{b}$ and $\hat{\sigma}$ are updated at independent exponential switching times. Then the change in coefficients in SDE (1.1) is taken into account by a representation using the automatic differentiation technique developed in [6]. By carefully choosing the coefficients $\hat{b}$, $\hat{\sigma}$, the authors were able to provide a finite variance method in the case where the diffusion coefficient is constant or with a general diffusion term but without drift and in dimension one. However, the variance of the resulting estimator is proved to be infinite in the most general case.
To bypass this drawback, we extend the original framework to more general switching times and exploit the switching time distribution to control the estimator variance. We prove that under suitable assumptions on the switching times distribution, we can provide a finite variance estimate of the solution of (1.2) in the most general case with drift and diffusion coefficients both varying. For instance, the Gamma distribution is proved to verify those assumptions as soon as the shape parameter κ satisfies $\kappa \leq \frac{1}{2}$.
Another contribution consists in proposing an original interacting particle scheme that helps to stabilize even more the estimator. This approach results in a new estimator combining both branching and interacting particle techniques. The new estimator is proved to be unbiased with finite variance. Finally, numerical tests confirm the interest of our new algorithm showing significant variance reduction in various examples.

2 Notations

Let $C_b^{1,2}([0,T]\times\mathbb{R}^d,\mathbb{R})$ denote the set of continuously differentiable bounded functions with bounded derivatives of order 1 for the time variable and bounded derivatives up to order 2 for the space variable. Let $\mathcal{L}$ denote the infinitesimal generator associated with (1.1) such that for any sufficiently regular function $\varphi : [0,T]\times\mathbb{R}^d \mapsto \mathbb{R}$ in the domain of $\mathcal{L}$, $\mathcal{L}\varphi$ is given as the real valued function such that

$$(\mathcal{L}\varphi)(t,x) = b(t,x).D\varphi(t,x) + \frac{1}{2}a(t,x) : D^2\varphi(t,x) \ , \quad \text{for all } (t,x)\in[0,T]\times\mathbb{R}^d \ , \tag{2.1}$$

where $a(t,x) := \sigma(t,x)\sigma(t,x)^T$, $A : B := tr(AB^T)$ and D (resp. D^2) denotes the differential operator of order 1 (resp. of order 2) w.r.t. the space variable x. Let us consider a real value bounded continuous function g defined on $\mathbb{R}^d$. By the Feynman-Kac formula it is well-known that if there exists $v^* \in C_b^{1,2}([0,T]\times\mathbb{R}^d,\mathbb{R})$ solution of the linear Partial Differential Equation (PDE)

$$\begin{cases} \partial_t v + \mathcal{L}v = 0 \\ v(T,x) = g(x) \ , \end{cases} \tag{2.2}$$

then this PDE has a unique classical solution $v^*(t,x) = u(t,x) = \mathbb{E}[g(X_T^{t,x})]$. In the sequel $\|x\|$ stands for the L_∞ norm of a vector or a matrix x.
All along this paper, we will consider this specific situation assuming the following assumption.

Assumption 1. *1. The linear PDE (2.2) admits a unique classical solution $v^* \in C_b^{1,2}$.*

2. The diffusion $\sigma(t,x)$ is non-degenerated such that for some constant $\epsilon_0 > 0$:

$$a(t,x) \geq \epsilon_0 \mathbb{I}, \quad \forall (t,x)\in[0,T]\times\mathbb{R}^d.$$

3. b and a are uniformly Lipschitz w.r.t. the space variable i.e. there exists a finite constant L such that for any $(t,x,x') \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d$

$$\|b(t,x) - b(t,x')\| + \|a(t,x) - a(t,x')\| \leq L\|x - x'\| \ .$$

4. b and a are uniformly 1/2-Hölder continuous w.r.t. variable t i.e. there exists a finite constant H such that for any $(t,t',x) \in [0,T] \times [0,T] \times \mathbb{R}^d$

$$\|b(t,x) - b(t',x)\| + \|a(t,x) - a(t',x)\| \leq H|t - t'|^{1/2} \ .$$

For a fixed point $(\tilde{t},\tilde{x}) \in [0,T] \times \mathbb{R}^d$, we introduce some operators and processes that will be useful in the sequel

- $\mathcal{L}^{\tilde{t},\tilde{x}}$ the differential operator similar to $\mathcal{L}$ with the drift and diffusion frozen at $(\tilde{t},\tilde{x})$ such that for any regular function φ in the domain of $\mathcal{L}^{\tilde{t},\tilde{x}}$

$$\mathcal{L}^{\tilde{t},\tilde{x}}\varphi(t,x) = b(\tilde{t},\tilde{x}).D\varphi(t,x) + \frac{1}{2}a(\tilde{t},\tilde{x}) : D^2\varphi(t,x) \ , \quad \text{for all } (t,x) \in [0,T] \times \mathbb{R}^d \ , \tag{2.3}$$

- $(\tilde{X}_t^{\tilde{t},\tilde{x},t_0,x_0})_{t \geq t_0}$ the Gaussian process with infinitesimal operator $\mathcal{L}^{\tilde{t},\tilde{x}}$ defined by

$$\tilde{X}_t^{\tilde{t},\tilde{x},t_0,x_0} = x_0 + b(\tilde{t},\tilde{x})(t - t_0) + \sigma(\tilde{t},\tilde{x})(W_t - W_{t_0}) \ . \tag{2.4}$$

  for a given initial condition $(t_0,x_0) \in [0,T] \times \mathbb{R}^d$.

- $h^{*,\tilde{t},\tilde{x}} : [0,T] \times \mathbb{R}^d \mapsto \mathbb{R}$ involving the unique solution v^* of (2.2) is defined by

$$h^{*,\tilde{t},\tilde{x}}(t,x) := (b(t,x) - b(\tilde{t},\tilde{x})).Dv^*(t,x) + \frac{1}{2}(a(t,x) - a(\tilde{t},\tilde{x})) : D^2v^*(t,x) \ . \tag{2.5}$$

  Notice that $h^{*,\tilde{t},\tilde{x}}$ is a well defined continuous function since $v^* \in C_b^{1,2}$ and in particular

$$h^{*,\tilde{t},\tilde{x}}(t,x) = \mathcal{L}v^*(t,x) - \mathcal{L}^{\tilde{t},\tilde{x}}v^*(t,x) \quad \text{for all } (t,x) \in [0,T] \times \mathbb{R}^d \ . \tag{2.6}$$

3 Probabilistic representation using a branching process

Recalling [7], the following representation holds

Lemma 3.1. *Suppose that Assumption 1 holds and* $\tilde{X}^{\tilde{t},\tilde{x}}$ *is the Gaussian process defined in* (2.4)*, then u defined by* (1.2) *and its (bounded and continuous) derivatives Du and* D^2u *are solutions of the system*

$$\begin{array}{lcl} u(t,x) & = & \mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x}) + \int_t^T H^{\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}, Du(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}), D^2u(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}))\, ds] \\ Du(t,x) & = & \mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x})\mathcal{M}_{t,T}^{\tilde{t},\tilde{x}} + \int_t^T H^{\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}, Du(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}), D^2u(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}))\mathcal{M}_{t,s}^{\tilde{t},\tilde{x}}\, ds] \\ D^2u(t,x) & = & \mathbb{E}[g(\tilde{X}_T^{t,x,t,x})\mathcal{V}_{t,T}^{t,x} + \int_t^T H^{t,x}(s,\tilde{X}_s^{t,x,t,x}, Du(s,\tilde{X}_s^{t,x,t,x}), D^2u(s,\tilde{X}_s^{t,x,t,x}))\mathcal{V}_{t,s}^{t,x}\, ds] \ , \end{array} \tag{3.1}$$

where for any $(\tilde{t},\tilde{x}) \in [0,T] \times \mathbb{R}^d$ *the function* $H^{\tilde{t},\tilde{x}} : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{S}^d \mapsto \mathbb{R}$ *is such that*

$$H^{\tilde{t},\tilde{x}}(t,x,y,z) = (b(t,x) - b(\tilde{t},\tilde{x})).y + \frac{1}{2}(a(t,x) - a(\tilde{t},\tilde{x})) : z \ , \tag{3.2}$$

$\mathcal{M}_{t,s}^{\tilde{t},\tilde{x}}$ *and* $\mathcal{V}_{t,s}^{\tilde{t},\tilde{x}}$ *are respectively the first and second order Malliavin weights associated with the process* $\tilde{X}^{\tilde{t},\tilde{x}}$ *that is using* $\delta_{t,s}W = W_s - W_t$

$$\mathcal{M}_{t,s}^{\tilde{t},\tilde{x}} := (\sigma(\tilde{t},\tilde{x})^{-1})^T \frac{\delta_{t,s}W}{s-t} \ , \quad \textit{and} \quad \mathcal{V}_{t,s}^{\tilde{t},\tilde{x}} := (\sigma(\tilde{t},\tilde{x})^{-1})^T \frac{\delta_{t,s}W\delta_{t,s}W^T - (s-t)\mathbb{I}}{(s-t)^2}\sigma(\tilde{t},\tilde{x})^{-1} \ . \tag{3.3}$$

Proof. The proof relies on the uniqueness property of classical solutions of PDEs satisfying the Feynman-Kac representation. Notice that under Assumption 1, u is the unique classical solution of (2.2). Of course, thanks to equation (2.6), for any $(\tilde{t},\tilde{x}) \in [0,T] \times \mathbb{R}^d$ u is also a $C_b^{1,2}$ solution of the following linear PDE

$$\partial_t u + \mathcal{L}^{\tilde{t},\tilde{x}} u + h^{*,\tilde{t},\tilde{x}} = 0 \ .$$

Then one can use again Feynman-Kac formula to represent the unique solution u of the above PDE as

$$u(t,x) = \mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x}) + \int_t^T h^{*,\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x})\,ds] \ . \tag{3.4}$$

Finally observe that

$$h^{*,\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}) = H^{\tilde{t},\tilde{x}}(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x}, Du(s,\tilde{X}_s^{\tilde{t},\tilde{x}}), D^2u(s,\tilde{X}_s^{\tilde{t},\tilde{x},t,x})) \ . \tag{3.5}$$

The equations relative to Dv and D^2v are obtained by applying Elworthy's formula [5] (which simply results here in the Likelihood ratio of Broadie and Glasserman [1]) in (3.4) and by using some technical estimates placed in the Appendix 9 to be able to differentiate under the time integral.

$\square$

Let τ be a random time independent of the Brownian W following the density f supposed to be positive on $[0,\infty]$. One can rewrite representation (3.1) by using a change of measure to replace the time integral by an expectation according to the random time τ.

$$\begin{aligned}
u(t,x) &= \frac{\mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x})\mathbf{1}_{\tau \geq T-t}]}{1-F(T-t)} \\
&\quad + \mathbb{E}[\frac{H^{\tilde{t},\tilde{x}}(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}, Du(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}), D^2u(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}))}{f(\tau)}\,\mathbf{1}_{\tau<T-t}] \\
Du(t,x) &= \frac{\mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x})\mathcal{M}_{t,T}^{\tilde{t},\tilde{x}}\mathbf{1}_{\tau \geq T-t}]}{1-F(T-t)} \\
&\quad + \mathbb{E}[\frac{H^{\tilde{t},\tilde{x}}(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}, Du(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}), D^2u(t+\tau,\tilde{X}_{t+\tau}^{\tilde{t},\tilde{x},t,x}))}{f(\tau)}\mathcal{M}_{t,t+\tau}^{\tilde{t},\tilde{x}}\,\mathbf{1}_{\tau<T-t}] \\
D^2u(t,x) &= \frac{\mathbb{E}[g(\tilde{X}_T^{t,x,t,x})\mathcal{V}_{t,T}^{t,x}\mathbf{1}_{\tau \geq T-t}]}{1-F(T-t)} \\
&\quad + \mathbb{E}[\frac{H^{t,x}(t+\tau,\tilde{X}_{t+\tau}^{t,x,t,x}, Du(t+\tau,\tilde{X}_{t+\tau}^{t,x,t,x}), D^2u(t+\tau,\tilde{X}_{t+\tau}^{t,x,t,x}))}{f(\tau)}\mathcal{V}_{t,t+\tau}^{t,x}\,\mathbf{1}_{\tau<T-t}]
\end{aligned} \tag{3.6}$$

where F is the cumulative distribution of f. We will now apply recursively this representation (3.6) by considering a sequence of i.i.d. random times (τ_k).

Let us introduce a non regular (stochastic) mesh of the interval $[t_0, T]$,

$$\pi := (T_0 := t_0 < T_1 < \cdots < T_k \cdots < T_{N_T} < T_{N_T+1} = T) \ , \tag{3.7}$$

characterized by the Markov chain (T_k) defined by

$$\begin{cases} T_0 & = & t_0 \\ T_{k+1} & = & T_k + \delta T_{k+1} \ , \text{for } k \in \mathbb{N} \text{ where} \\ \delta T_{k+1} & = & \tau_{k+1} \wedge (T - (T_k + \tau_{k+1}))^+ \ , \end{cases} \tag{3.8}$$

where (τ_k) is an i.i.d. sequence of random times distributed according the common probability density f. Notice that (T_k) defines a Markov chain with an absorbing state, T. (T_k) will define the so-called *switching time*.
The random integer N_T is defined as the following stopping time

$$N_T = \inf\{n \mid T_{n+1} \geq T\} \ . \tag{3.9}$$

From now on, the following assumption will be in force.

Assumption 2. *The density distribution f underlying the random variables τ_k is such that the stopping time N_T is almost surely finite.*

Remark 3.1. *In the sequel, we will consider an i.i.d. sequence (τ_k) of Gamma variables with parameters $(\kappa > 0, \theta > 0)$ recalling that the Gamma density with parameter $(\kappa > 0, \theta > 0)$ is given by*

$$f_\Gamma^{\kappa,\theta}(s) = \frac{s^{\kappa-1}e^{-s/\theta}}{\Gamma(\kappa)\theta^\kappa} \ , \quad \textit{for all } s > 0 \ , \tag{3.10}$$

where Γ is the Gamma Euler function. With that specific choice of density, one can verify that N_T defined in (3.9) *is a.s. finite. This can be proved easily, after remarking that by Borel Canteli Lemma it is enough to prove that there exists a sequence of integers (M_n) increasing to infinity such that $\sum_n \mathbb{P}(N_T > M_n) < \infty$ which can be proved by showing that $\sum_n \mathbb{P}(\sum_{k=1}^{M_n} \tau_k \leq T) < \infty$.*

For a given mesh π (defined as in (3.7) (3.8)), we consider the following sequence (defining a Markov chain conditionally to the mesh π)

$$\begin{cases} X_0 = X_{T_0} = x_0 \\ X_{k+1} = X_k + b(T_k, X_k)\delta T_{k+1} + \sigma(T_k, X_k)\delta W_{k+1} \ , \end{cases} \tag{3.11}$$

where $\delta W_{k+1} := W_{T_{k+1}} - W_{T_k}$. For the sake of simplicity, we will often note b_k or σ_k instead of $b(T_k, X_{T_k})$ or $\sigma(T_k, X_{T_k})$.

Using representation (3.6) with $(\tilde{t}, \tilde{x}) = (T_k, X_k)$ and $\tau = \tau_{k+1}$, conditioning with respect to $(\tilde{t}, \tilde{x})$ one gets for any integer $k \geq 0$

$$u(T_k, X_k) \quad = \quad \frac{\mathbb{E}[g(\tilde{X}_T^{T_k, X_k})\mathbf{1}_{T_{k+1}=T}]}{1 - F(T - T_k)} + \mathbb{E}[H_{k+1}\, \mathbf{1}_{T_{k+1}<T}] \tag{3.12}$$

with

$$H_{k+1} := \frac{H^{T_k,X_k}(T_{k+1},\tilde{X}^{T_k,X_k}_{T_{k+1}}, Du(T_{k+1},\tilde{X}^{T_k,X_k}_{T_{k+1}}), D^2u(T_{k+1},\tilde{X}^{T_k,X_k}_{T_{k+1}}))}{f(\delta T_{k+1})} \ .$$

and $\tilde{X}^{T_k,X_k}_s = \tilde{X}^{T_k,X_k,T_k,X_k}_s$ for $s \geq T_k$.
The derivatives Du and D^2u in H_{k+1} are given by applying the representation (3.6) with $(\tilde{t},\tilde{x}) = (T_{k+1}, X_{k+1})$ and $\tau = \tau_{k+2}$, conditioning with respect to $(\tilde{t},\tilde{x})$ one gets for any integer $k \geq 0$

$$\begin{aligned} Du(T_{k+1},X_{k+1}) &= \frac{\mathbb{E}[g(\tilde{X}^{T_{k+1},X_{k+1}}_T)\mathcal{M}^{T_{k+1},X_{k+1}}_{T_{k+1},T}\mathbf{1}_{T_{k+2}=T}]}{1-F(T-T_{k+1})} + \mathbb{E}[H_{k+2}\mathcal{M}^{T_{k+1},X_{k+1}}_{T_{k+1},T_{k+2}}\mathbf{1}_{T_{k+2}<T}] \\ Du(T_{k+1},X_{k+1}) &= \frac{\mathbb{E}[g(\tilde{X}^{T_{k+1},X_{k+1}}_T)\mathcal{V}^{T_{k+1},X_{k+1}}_{T_{k+1},T}\mathbf{1}_{T_{k+2}=T}]}{1-F(T-T_{k+1})} + \mathbb{E}[H_{k+2}\mathcal{V}^{T_{k+1},X_{k+1}}_{T_{k+1},T_{k+2}}\mathbf{1}_{T_{k+2}<T}] \end{aligned} \tag{3.13}$$

Let us introduce the sequence of weights $(P_k)_{k\geq 1}$ such that for $k = 1,\cdots,N_T$

$$\begin{cases} P_{k+1} &= \frac{M_{k+1}+\frac{1}{2}V_{k+1}}{f(\delta T_k)} \ , \\ M_{k+1} &= \delta b_k.(\sigma_k^{-1})^T\frac{\delta W_{k+1}}{\delta T_{k+1}}, \quad \text{with} \quad \delta b_k := b_k - b_{k-1} \\ V_{k+1} &= \delta a_k : (\sigma_k^{-1})^T\frac{\delta W_{k+1}\delta W_{k+1}^T - \delta T_{k+1}\mathbb{I}}{(\delta T_{k+1})^2}\sigma_k^{-1} \ , \quad \text{with} \quad \delta a_k := a_k - a_{k-1} \ . \end{cases} \tag{3.14}$$

Following the same lines as the proof of Theorem 2.3 in [7], one can derive by recurrence a representation formula for u as the expectation of an exactly simulatable variable. Before one has to introduce some new assumptions.

Assumption 3. *The coefficients b and a are uniformly bounded i.e. there exists a finite constant M such that for any $(t,x) \in [0,T]\times\mathbb{R}^d$*

$$\|b_t(x)\| \leq M \ , \quad \|a_t(x)\| \leq M \ .$$

Proposition 3.2. *Under assumptions 1, 2 and 3 , the following representation holds*

$$u(t_0,x_0) \ := \ \mathbb{E}[g(X^{t_0,x_0}_T)] = \mathbb{E}[\tfrac{g(X_{N_T+1})}{1-F(\delta T_{N_T+1})}\textstyle\prod_{k=2}^{N_T+1}P_k] \ , \tag{3.15}$$

with the convention $\prod_{k\in\varnothing} = 1$.

Remark 3.2. 1. *Using an Exponential distribution for f, one recovers the representation given in [7].*

2. *The proof relies on a recurrence argument and on representation* (3.6)*. It proves that any v satisfying the equation* (3.1) *is given by the above explicit equation* (3.15)*. Notice that this a posteriori proves the uniqueness of the solution of* (3.1)*.*

We next define a second representation that will be interesting in order to get some finite variance estimator for some given switching distribution, f. Following [7], one can introduce antithetic variables to control the variance induced by the last time step. Let $(\mathcal{F}_t)$ denote the filtration generated by the Brownian i.e. $\mathcal{F}_t := \sigma(W_s, s\in[0,t])$ and recall that π denotes the stochastic sequence $(T_k)_{k=0,\cdots,N_T+1}$ independent of the Brownian. Observe that

$$\mathbb{E}[M_{N_T+1}|\mathcal{F}_{T_{N_T}}\vee\pi] = \mathbb{E}[V_{N_T+1}|\mathcal{F}_{T_{N_T}}\vee\pi] = \mathbb{E}[P_{N_T+1}|\mathcal{F}_{T_{N_T}}\vee\pi] = 0 \ .$$

Hence replacing $g(X_{N_T+1})$ by $g(X_{N_T+1})-g(X_{N_T})$ in (3.15) does not change the expectation since due to tower property :

$$\mathbb{E}[\frac{g(X_{N_T})}{1-F(\delta T_{N_T+1})}\prod_{k=2}^{N_T+1} P_k] = \mathbb{E}[\frac{g(X_{N_T})}{1-F(\delta T_{N_T+1})}P_{N_T+1}\prod_{k=2}^{N_T} P_k] = 0 \ .$$

Notice that the following decomposition holds whenever $N_T \geq 1$

$$g(X_{N_T+1})\prod_{k=2}^{N_T+1} P_k = g(X_{N_T+1})\frac{M_{N_T+1}}{f(\delta T_{N_T})}\prod_{k=2}^{N_T} P_k + \frac{1}{2}g(X_{N_T+1})\frac{V_{N_T+1}}{f(\delta T_{N_T})}\prod_{k=2}^{N_T} P_k \ .$$

Then using antithetic variables for the second term in the r.h.s. of the above equality yields the following estimator.

Proposition 3.3. *Under assumptions 1, 2 and 3 , the following representation holds*

$$u(t_0,x_0) = \mathbb{E}[\beta\prod_{k=2}^{N_T} P_k \mathbf{1}_{N_T\geq 1}] + \mathbb{E}[\frac{g(X_1)}{1-F(\delta T_1)}\mathbf{1}_{N_T=0}] \ , \tag{3.16}$$

where $\beta := \frac{1}{2}(\beta_1+\beta_2)$ *with*

$$\left\{\begin{array}{lll} \beta_1 & := & \frac{g(X_{N_T+1})-g(X_{N_T})}{1-F(\delta T_{N_T+1})}\frac{M_{N_T+1}+\frac{1}{2}V_{N_T+1}}{f(\delta T_{N_T})}, \\ \beta_2 & := & \frac{g(\hat{X}_{N_T+1})-g(X_{N_T})}{1-F(\delta T_{N_T+1})}\frac{-M_{N_T+1}+\frac{1}{2}V_{N_T+1}}{f(\delta T_{N_T})} \end{array}\right. \tag{3.17}$$

and $\hat{X}_{N_T+1} = X_{N_T} + b_{N_T}\delta T_{N_T+1} - \sigma_{N_T}\delta W_{N_T+1}$.

4 Variance Analysis in the case of Gamma distribution

The previous representation given by Proposition 3.3 is general but the variance associated to the estimator is generally infinite as it is the case when f is an exponential density. From now on, we will suppose that the density $f = f_\Gamma^{\kappa,\theta}$ is the Gamma density (3.10) with parameters (κ,θ) with cumulative distribution $F = F_\Gamma^{\kappa,\theta}$.
First, we will introduce the following assumptions.

Assumption 4. *The following assertions hold*

1. $g \in C_b^{1,2}$.
2. $\kappa \leq \frac{1}{2}$.

Now, we can state the following proposition.

Proposition 4.1. *Under Assumption 1, 3 and 4, the estimator defined by* (3.16) *in Proposition 3.3 has finite variance.*

Proof. Let $\bar{\mathcal{F}}_k$ denote the sigma-field generated by the Brownian up to the random time T_k i.e. $\bar{\mathcal{F}}_k := \sigma(W_t, t \in [0, T_k])$ and $\mathcal{H}^n := \sigma(T_k\,,\, 0 \le k \le N_T + 1) \vee \{N_T = n\}$. Let us consider the first term in the sum (3.16).

$$\mathbb{E}[\Big(\beta \mathbf{1}_{N_T \ge 1} \prod_{k=2}^{N_T} P_k\Big)^2] \;=\; \sum_{n=1}^{\infty} \mathbb{E}[\Big(\beta \prod_{k=2}^{n} P_k\Big)^2 |N_T = n]\mathbb{P}(N_T = n) \tag{4.1}$$

The proof will be decomposed into several steps

1. Bounding $\mathbb{E}[\beta^2 | \bar{\mathcal{F}}_n \vee \mathcal{H}^n]$
   First considering M_{k+1} and V_{k+1} one easily obtains

$$\mathbb{E}[M_{k+1}^4 \,|\bar{\mathcal{F}}_k \vee \mathcal{H}^n] \le C \frac{(\delta b_k)^4}{(\delta T_{k+1})^2}\;, \quad \text{and} \quad \mathbb{E}[V_{k+1}^4 \,|\bar{\mathcal{F}}_k \vee \mathcal{H}^n] \le C \frac{(\delta a_k)^4}{(\delta T_{k+1})^4}\;. \tag{4.2}$$

   Notice that in the sequel, C will denote finite constants that may change from line to line that does not depend on k or n but only on the characteristics of the problem (T, the bounds or Lipschitz constants related to g, b, σ, a). Then consider the general term of the sum (4.1).

$$\mathbb{E}[\beta^2 \prod_{k=2}^{n} P_k^2 \,|N_T = n] \;=\; \mathbb{E}\Big[\mathbb{E}[\beta^2 | \bar{\mathcal{F}}_n \vee \mathcal{H}^n]\Big(\prod_{k=2}^{n} P_k\Big)^2 \,|N_T = n\Big]\;.$$

   We get

$$\begin{aligned} \mathbb{E}[\beta^2 | \bar{\mathcal{F}}_n \vee \mathcal{H}^n] \;\le\; & \frac{C}{(1 - F_\Gamma^{\kappa,\theta}(T))^2} \mathbb{E}[(g(X_{n+1}) - g(X_n))^2 \frac{M_{n+1}^2}{f(\delta t_n)^2} | \bar{\mathcal{F}}_n \vee \mathcal{H}^n] + \\ & \frac{C}{(1 - F_\Gamma^{\kappa,\theta}(T))^2} \mathbb{E}[(g(X_{n+1}) + g(\hat{X}_{n+1}) - 2g(X_n))^2 \frac{V_{n+1}^2}{f(\delta t_n)^2} | \bar{\mathcal{F}}_n \vee \mathcal{H}^n] \end{aligned} \tag{4.3}$$

   Consider the first term on the r.h.s. of inequality (4.3), by the Lipschitz property of g, the boundness of b,σ and using the fact that σ is uniformly bounded away from zero, we obtain

$$\begin{aligned} \mathbb{E}[|g(X_{n+1}) - g(X_n)|^2 \frac{M_{n+1}^2}{(f_\Gamma^{\kappa,\theta}(\delta t_n))^2} | \bar{\mathcal{F}}_n \vee \mathcal{H}^n] \;&\le\; C \frac{\|\delta b_n\|^2}{(f_\Gamma^{\kappa,\theta}(\delta T_n))^2} \\ &\le\; C \|\delta b_n\|^2 (\delta T_n)^{2(1-\kappa)} \end{aligned} \tag{4.4}$$

   Consider the second term of (4.3). By Assumption 4.3 ($g \in C_b^{1,2}$) one can apply Ito and obtain

$$|g(X_{n+1}) + g(X_n + b_n \delta T_{n+1} - \sigma_n \delta W_{n+1}) - 2g(X_n)| \;\le\; C \delta T_{n+1}\;.$$

   This implies still using the boundness of b, σ and using the fact that σ is uniformly bounded away from zero :

$$\begin{aligned} \mathbb{E}[(g(X_{n+1}) + g(\hat{X}_{n+1}) - 2g(X_n))^2 \frac{V_{n+1}^2}{f(\delta t_n)^2} | \bar{\mathcal{F}}_n \vee \mathcal{H}^n] \;&\le\; C(\delta T_{n+1})^2 \frac{\|\delta a_n\|^2}{\delta T_{n+1}^2 f_\Gamma^{\kappa,\theta}(\delta T_n)^2} \\ &\le\; C(\delta T_n)^{2(1-\kappa)} \|\delta a_n\|^2\;. \end{aligned} \tag{4.5}$$

Injecting (4.4) and (4.5) into (4.3) finally yields

$$\mathbb{E}[\beta^2\,|\bar{\mathcal{F}}_n \vee \mathcal{H}^n] \le C(\delta T_n)^{2(1-\kappa)}\Big(\|\delta b_n\|^2 + \|\delta a_n\|^2\Big)\ . \tag{4.6}$$

2. Bounding $\mathbb{E}[\|\delta b_k\|^2 + \|\delta a_k\|^2\,|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n]$

$$\begin{aligned}
\mathbb{E}[\|\delta b_k\|^2\,|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n] &= \mathbb{E}[\|b(T_k, X_{T_k}) - b(T_{k-1}, X_{T_{k-1}})\|^2\,|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n] \\
&\le 2\mathbb{E}[\|b(T_k, X_{T_k}) - b(T_k, X_{T_{k-1}})\|^2 + \\
&\quad\ 2\|b(T_k, X_{T_{k-1}}) - b(T_{k-1}, X_{T_{k-1}})\|^2\,|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n] \\
&\le C(1+\delta T_k)\delta T_k + C\delta T_k \le C\delta T_k
\end{aligned}$$

using the fact that b is Lipschitz w.r.t. the space variable and 1/2-Hölder continuous w.r.t. the time variable. With the same development on δa_k one finally gets

$$\mathbb{E}[\|\delta b_k\|^2 + \|\delta a_k\|^2\,|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n] \quad\le\quad C\delta T_k\ , \tag{4.7}$$

Similarly we obtain

$$\mathbb{E}[\|\delta b_k\|^4 + \|\delta a_k\|^4\,|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n] \le C(\delta T_k)^2\ , \tag{4.8}$$

3. Bounding $\mathbb{E}[P_{k+1}^4\,|\bar{\mathcal{F}}_k \vee \mathcal{H}^n]$
Using (4.2), we obtain

$$\begin{aligned}
\mathbb{E}[P_{k+1}^4\,|\bar{\mathcal{F}}_k \vee \mathcal{H}^n] &= \mathbb{E}[(M_{k+1} + \frac{1}{2}V_{k+1})^4(\delta T_k)^{4(1-\kappa)}\theta^{4\kappa}\Gamma^4(\kappa)e^{4\delta T_k/\theta}\Big|\bar{\mathcal{F}}_k \vee \mathcal{H}^n] \\
&\le C\Big(\|\delta b_k\|^4 + \frac{\|\delta a_k\|^4}{(\delta T_{k+1})^2}\Big)\frac{(\delta T_k)^{4(1-\kappa)}}{(\delta T_{k+1})^2} \\
&\le CC_k^2\frac{(\delta T_k)^{4(1-\kappa)}}{(\delta T_{k+1})^4}\ ,
\end{aligned} \tag{4.9}$$

observing that $\delta T_{k+1} \le T \le C$ and with the C_k defined by

$$C_k := \|\delta b_k\|^2 + \|\delta a_k\|^2\ . \tag{4.10}$$

Observe in particular that by (4.8), $\mathbb{E}[C_k^2\,|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n] \le C(\delta T_k)^2$. Using the tower property of expectation and bound (4.6) yields

$$\begin{aligned}
\mathbb{E}[\beta^2\prod_{k=1}^{n-1}P_{k+1}^2\,|N_T = n] &= \mathbb{E}\Big[E[\beta^2\,|\bar{\mathcal{F}}_n \vee \mathcal{H}^n]\prod_{k=1}^{n-1}P_{k+1}^2\,|N_T = n\Big] \\
&\le C\mathbb{E}\Big[(\delta T_n)^{2(1-\kappa)}\Big(\|\delta a_n\|^2 + \|\delta b_n\|^2\Big)\prod_{k=1}^{n-1}P_{k+1}^2\,|N_T = n\Big] \\
&\le C\mathbb{E}\Big[\mathbb{E}[(\delta T_n)^{2(1-\kappa)}C_nP_n^2\,|\bar{\mathcal{F}}_{n-1} \vee \mathcal{H}^n]\prod_{k=1}^{n-2}P_{k+1}^2\,|N_T = n\Big]\ .
\end{aligned}$$

By Cauchy-Schwarz, for any k we have

$$\begin{aligned}
\mathbb{E}[C_k P_k^2|\bar{\mathcal{F}}_{k-1}\vee\mathcal{H}^n] &\leq \Big(\mathbb{E}[C_k^2|\bar{\mathcal{F}}_{k-1}\vee\mathcal{H}^n]\Big)^{1/2}\Big(\mathbb{E}[P_k^4|\bar{\mathcal{F}}_{k-1}\vee\mathcal{H}^n]\Big)^{1/2}\\
&\leq C\delta T_k C_{k-1}\frac{(\delta T_{k-1})^{2(1-\kappa)}}{(\delta T_k)^2}\\
&\leq CC_{k-1}\frac{(\delta T_k)^{2(1-\kappa)}}{\delta T_k}\ . \qquad (4.11)
\end{aligned}$$

Hence, we obtain by recursion

$$\mathbb{E}[\beta^2\prod_{k=1}^{n-1}P_{k+1}^2\,|N_T=n] \;\leq\; C^{n+1}\mathbb{E}[(\delta T_n)^{2(1-\kappa)}\prod_{k=1}^{n-1}\frac{(\delta T_k)^{2(1-\kappa)}}{\delta T_{k+1}}\,|N_T=n]\ ,$$

observing that $\mathbb{E}[C_1|\bar{\mathcal{F}}_0\vee\mathcal{H}^n]\leq C\delta T_1\leq CT$.
Then recalling that $\kappa\leq 1/2$ implies $(\delta T_k)^{1-2\kappa}\leq T^{1-2\kappa}$ and finally yields

$$\mathbb{E}[\beta^2\prod_{k=1}^{n-1}P_{k+1}^2\,|N_T=n]\leq CC^n\ . \qquad (4.12)$$

4. Convergence of the sum $\sum_{k=1}^{\infty}C^n\mathbb{P}(N_T=n)$. Let us introduce $S_n=\sum_{k=1}^{n}\tau_k$, notice that $S_n\sim\Gamma(n\kappa,\theta)$ with cumulative distribution

$$F_{S_n}(s)=\int_0^s\frac{r^{n\kappa-1}e^{-r/\theta}}{\Gamma(n\kappa)\theta^{n\kappa}}dr\ .$$

Hence one can bound $\mathbb{P}(N_T=n)$ as follows

$$\begin{aligned}
\mathbb{P}(N_T=n) &\leq \mathbb{P}(N_T=n)\\
&\leq \mathbb{P}(S_n\leq T)\\
&\leq \int_0^T\frac{r^{n\kappa-1}}{\Gamma(n\kappa)\theta^{n\kappa}}dr=\frac{T^{n\kappa}}{n\kappa\Gamma(n\kappa)\theta^{n\kappa}}
\end{aligned}$$

This implies that

$$\sum_{n=1}^{\infty}C^n\mathbb{P}(N_T=n)\leq\sum_{n=1}^{\infty}\frac{\hat{C}^n}{n\kappa\Gamma(n\kappa)}\ ,$$

with $\hat{C}=C\frac{T}{\theta}$. Using the generalization of the Stirling formula $\Gamma(z)\sim z^{z-1/2}e^{-z}\sqrt{2\pi}$ one proves that $\frac{\hat{C}^n}{n\kappa\Gamma(n\kappa)}\sim\frac{\hat{C}^{\frac{1}{2\kappa}}e^{-\frac{1}{2}}}{\sqrt{2\pi}}(\frac{\hat{C}^{\frac{1}{\kappa}}e}{n\kappa})^{n\kappa+\frac{1}{2}}$ which is the general term of a convergent sum.

$\square$

Remark 4.1. *The convergence of the previous series relies on two facts :*

- *Conditionally to the number of jumps, $N_T = n$, one has to be able to bound for all $k < n$ $\mathbb{E}[\frac{1}{(f(\delta T_k))^2\delta T_k}|N_T = n]$, where f is the underlying probability density of δT_k. In the case of an exponential density, it is well-known that the conditional distribution, $\mathcal{L}(\delta T_k|N_T = n)$, is the uniform distribution on $[0, T]$, hence this integral is infinite. In the case of a Gamma distribution, $\frac{1}{(f(\delta T_k))^2\delta T_k} \le C\delta T_k^{1-2\kappa} \le C(\sup(T, 1))^{1-2\kappa}$ with $\kappa \le \frac{1}{2}$ (the occurrence of small jumps is high) so the integral is bounded whatever the conditional distribution, $\mathcal{L}(\delta T_k|N_T = n)$. Notice that one could consider other densities f with a smaller intensity of small jumps as soon as the conditional law $\mathcal{L}(\delta T_k|N_T = n)$ ensures that $\mathbb{E}[\frac{1}{(f(\delta T_k))^2\delta T_k}|N_T = n]$ is bounded.*

- *The sum $\sum_{k=1}^{\infty} C^n \mathbb{P}(N_T = n)$ has to converge. By increasing the intensity of small jumps as explained at point 1., we expect that $\mathbb{P}(N_T = n)$ will decrease more slowly with n. This results in a tradeoff one has to achieve : increasing small jumps intensity to be able to bound $\mathbb{E}[\frac{1}{(f(\delta T_k))^2\delta T_k}|N_T = n]$ but not too strongly to ensure the convergence of $\sum_{k=1}^{\infty} C^n \mathbb{P}(N_T = n)$.*

Consequently, the representation (3.16) provides a Monte Carlo approach to compute $\mathbb{E}[g(X_T)]$, by simulating the regime switching process (3.11) instead of the SDE (1.1) which would potentially require to implement a stochastic Euler discretization scheme. However, even though our estimator is proved to have finite variance, one can observe in practice huge variances due to the product of a random number of terms P_k that could potentially take values greater that one. This expectation of products is *by nature* not a good candidate for Monte Carlo estimation. Hence, we propose to use a resampling procedure to change this expectation of products in a product of expectations which is known to be much more stable for estimation.

5 Resampling method for branching processes

In this section, we propose to introduce an interacting particle system (in the same vein as those thoroughly discussed in the reference books [3] and [4]) to approximate $u(t_0, x_0)$. We will prove that the resulting estimator has finite variance under the same assumptions required to bound the variance of estimator (3.16). However, in practice, the new estimator relying on interacting particle systems will show better performances providing smaller variances in many examples, as illustrated in Section 6.

5.1 A Feynman-Kac measure representation

First we have to express $u(t_0, x_0)$ as an integral according to a Feynman-Kac measure. Let us consider the Markov chain consisting of the sequence of random variables $\tilde{X}_k := (T_k, X_k)$, where (T_k) and (X_k) are given respectively by the dynamics (3.8) and (3.11). In the sequel, we note $\tilde{X}_{0:k} := (\tilde{X}_0, \cdots, \tilde{X}_k)$ the path valued Markov chain. Let us introduce, for any integer $k \ge 0$, the real valued function $\tilde{G}_k$ depending on the path $\tilde{x}_{0:k} \in E_k := (\mathbb{R}_+ \times \mathbb{R}^d)^{k+1}$ with the notations $\tilde{x}_{0:k} := (\tilde{x}_0, \cdots, \tilde{x}_k)$ and $\tilde{x}_p := (t_p, x_p) \in \mathbb{R}_+ \times \mathbb{R}^d$ such that

$$\tilde{G}_k(\tilde{x}_{0:k}) := \begin{cases} 1 & \text{if } k = 0 \text{ or } k = 1 \\ \frac{\tilde{M}_k(\tilde{x}_{0:k}) + \frac{1}{2}\tilde{V}_k(\tilde{x}_{0:k})}{f_\Gamma^{\kappa,\theta}(\delta t_{k-1})} & \text{if } k \ge 2 \text{ and } \delta t_{k-1}\delta t_k > 0 \\ 1 & \text{elsewhere .} \end{cases} \tag{5.1}$$

with $\delta t_{k+1} := t_{k+1} - t_k$ and where the real valued functions $\tilde{M}_{k+1}$, $\tilde{V}_{k+1}$ and $\delta\tilde{W}_{k+1}$ are such that for any $\tilde{x}_{0:k+1} \in E_{k+1}$

$$\begin{aligned} \tilde{M}_{k+1}(\tilde{x}_{0:k+1}) &:= \begin{cases} (b(t_k, x_k) - b(t_{k-1}, x_{k-1}).(\sigma(t_k, x_k)^{-1})^T \frac{\delta\tilde{W}_{k+1}(\tilde{x}_{0:k+1})}{\delta t_{k+1}} & \text{if } \delta t_{k+1} > 0 \\ 1 \quad \text{elsewhere} \end{cases} \\ \tilde{V}_{k+1}(\tilde{x}_{0:k+1}) &:= \begin{cases} (a(t_k, x_k) - a(t_{k-1}, x_{k-1})) : \frac{B_{k+1}(\tilde{x}_{0:k+1})}{(\delta t_{k+1})^2} & \text{if } \delta t_{k+1} > 0 \\ 1 \quad \text{elsewhere} , \end{cases} \end{aligned} \tag{5.2}$$

with

$$\begin{aligned} B_{k+1}(\tilde{x}_{0:k+1}) &:= (\sigma(t_k, x_k)^{-1})^T \Big(\delta\tilde{W}_{k+1}(\tilde{x}_{0:k+1})\delta\tilde{W}_{k+1}(\tilde{x}_{0:k+1})^T - \delta t_{k+1}\mathbb{I}\Big)\sigma(t_k, x_k)^{-1} \\ \delta\tilde{W}_{k+1}(\tilde{x}_{0:k+1}) &:= \sigma(t_k, x_k)^{-1}(x_{k+1} - x_k - b(t_k, x_k)\delta t_k) . \end{aligned}$$

Observe that $\tilde{G}_{k+1}$ does not really depend on the whole path $\tilde{x}_{0:k+1}$, but only on $(\tilde{x}_{k-1}, \tilde{x}_k, \tilde{x}_{k+1})$, for $k > 0$. Recalling (3.14), notice that the following identity holds

$$\tilde{G}_k(\tilde{X}_{0:k}) = P_k \ , \quad \mathbb{P} \ a.s. \quad \text{for all} \quad k = 2, \cdots, N_T \ .$$

In the sequel, it will appear to be crucial to consider positive *potential functions* with uniformly bounded conditional variances, more specifically such that $\sup_{\tilde{x}_{0:k} \in E_k} \mathbb{E}[G^2_{k+1}(\tilde{X}_{0:k+1}) \,|\tilde{X}_{0:k} = \tilde{x}_{0:k}] < \infty$, thus we define the potential functions $(G_k)_{k\geq 0}$ (depending implicitly on T) such that for any $k \geq 0$ and for any $\tilde{x}_{0:k} \in E_k$,

$$G_k(\tilde{x}_{0:k}) := \begin{cases} 1 & \text{if } k = 0 \\ |\tilde{G}_1(\tilde{x}_{0:1})|(\delta t_1)^{1-\kappa}\sqrt{c_1(\tilde{x}_{0:1})} & \text{if } k = 1, \ t_2 < T, \ \text{and } \delta t_2 > 0 \\ |\tilde{G}_k(\tilde{x}_{0:k})|\sqrt{\frac{c_k(\tilde{x}_{0:k})}{c_{k-1}(\tilde{x}_{0:k-1})}}\Big(\frac{\delta t_k}{\delta t_{k-1}}\Big)^{1-\kappa} & \text{if } k \geq 2 \ , \ \delta t_{k-1}\delta t_k > 0 \ , \\ 1 & \text{elsewhere} \ , \end{cases} \tag{5.3}$$

where the real valued function c_k is defined on E_k by

$$c_k(\tilde{x}_{0:k}) = |\delta t_k| + \|b(t_k, x_k) - b(t_{k-1}, x_{k-1})\|^2 + \|a(t_k, x_k) - a(t_{k-1}, x_{k-1})\|^2 \tag{5.4}$$

Notice that this definition of c_k is such that $c_k(\tilde{X}_{0:k}) = C_k + \delta T_k$ where C_k was defined in (4.10), hence

$$G^2_k(\tilde{X}_{0:k}) = \frac{C_k + \delta T_k}{C_{k-1} + \delta T_{k-1}}\Big(\frac{\delta t_k}{\delta t_{k-1}}\Big)^{2(1-\kappa)} P^2_k \ , \quad \mathbb{P} \ a.s. \quad \text{for all} \quad k = 2, \cdots, N_T \ .$$

Then observe that one can prove an inequality similar as (4.11) with C_k replaced by $c_k(\tilde{X}_{0:k})$

$$\begin{aligned} \mathbb{E}[c_k(\tilde{X}_{0:k})P^2_k|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n] &\leq \Big(\mathbb{E}[c^2_k(\tilde{X}_{0:k})|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n]\Big)^{1/2}\Big(\mathbb{E}[P^4_k|\bar{\mathcal{F}}_{k-1} \vee \mathcal{H}^n]\Big)^{1/2} \\ &\leq C c_{k-1}(\tilde{X}_{0:k-1})\frac{(\delta T_{k-1})^{2(1-\kappa)}}{\delta T_k} \ , \end{aligned} \tag{5.5}$$

which yields as announced, that for all $\tilde{x}_{0:k-1} \in E_{k-1}$

$$\mathbb{E}[G^2_k(\tilde{X}_{0:k}) \,|\tilde{X}_{0:k-1} = \tilde{x}_{0:k-1}] \leq C < \infty \ . \tag{5.6}$$

Notice that $\prod_{k=2}^{N_T} P_k = H_{N_T+1}(\tilde{X}_{0:N_T+1}) \prod_{k=0}^{N_T} G_k(\tilde{X}_{0:k}) S_k(\tilde{X}_{0:k})$, $\quad \mathbb{P}$ $a.s.$ where for any $k \geq 0$ and for any $\tilde{x}_{0:k} \in E_k$,

$$S_k(\tilde{x}_{0:k}) := Sign(\tilde{G}_k(\tilde{x}_{0:k})) \ , \tag{5.7}$$

and

$$H_{k+1}(\tilde{x}_{0:k+1}) := \begin{cases} \frac{1}{(\delta t_k)^{1-\kappa}\sqrt{c_k(\tilde{x}_{0:k})}} & \text{if } k \geq 1 \text{ and } \delta t_k > 0 \\ 1 & \text{elsewhere .} \end{cases} \tag{5.8}$$

Let us introduce $\beta_{n+1} := \frac{1}{2}\beta_{1,n+1} + \frac{1}{2}\beta_{2,n+1}$ defined on E_{n+1} such that $\beta_{1,1}(\tilde{x}_{0:1}) = \beta_{2,1}(x_{0:n+1}) = \frac{1}{(1-F_\Gamma^{\kappa,\theta}(\delta t_1))} g(x_1)$ and for any $n \geq 1$

$$\begin{cases} \beta_{1,n+1}(\tilde{x}_{0:n+1}) & := & \frac{g(x_{n+1})-g(x_n)}{1-F_\Gamma^{\kappa,\theta}(\delta t_{n+1})} \frac{M_{n+1}(\tilde{x}_{0:n+1})+\frac{1}{2}\tilde{V}_{n+1}(\tilde{x}_{0:n+1})}{f_\Gamma^{\kappa,\theta}(\delta t_n)} \\ \beta_{2,n+1}(\tilde{x}_{0:n+1}) & := & \frac{g(\hat{x}_{n+1})-g(x_n)}{1-F_\Gamma^{\kappa,\theta}(\delta t_{n+1})} \frac{-M_{n+1}(\tilde{x}_{0:n+1})+\frac{1}{2}\tilde{V}_{n+1}(\tilde{x}_{0:n+1})}{f_\Gamma^{\kappa,\theta}(\delta t_n)} \end{cases} , \tag{5.9}$$

with $\hat{x}_{n+1} = x_n + b(t_n, x_n)\delta t_{n+1} - \sigma(t_n, x_n)\delta\tilde{W}_{n+1}(\tilde{x}_{0:n+1})$.
Recalling (3.16), observe that

$$\begin{aligned} u(t_0, x_0) &= \mathbb{E}[\beta \prod_{k=2}^{N_T} P_k \mathbf{1}_{N_T \geq 1}] + \mathbb{E}[\frac{g(X_1)}{1-F_\Gamma^{\kappa,\theta}(\delta T_1)} \mathbf{1}_{N_T=0}] \\ &= \mathbb{E}[(\beta_{N_T+1} H_{N_T+1})(\tilde{X}_{0:N_T+1})(S_{0:N_T} G_{0:N_T})(\tilde{X}_{0:N_T})] \ , \end{aligned} \tag{5.10}$$

where to simplify the notation $G_{p:q}$ (resp. $S_{p:q}$) denotes the product $\prod_{k=p}^q G_k$ (resp. $\prod_{k=p}^q S_k$), with in particular $G_{p,q} = \mathbf{1}$ when $p > q$, where $\mathbf{1}$ denotes the function which takes the unique value 1. Now, we can define the sequence of non negative measures $(\gamma_k)_{k \geq 0}$ such that for any real valued bounded test function φ defined on $E_n := (\mathbb{R}_+ \times \mathbb{R}^d)^{n+1}$, we have

$$\gamma_k(\varphi) := \mathbb{E}[\varphi(\tilde{X}_{0:k}) \prod_{p=0}^{k-1} G_p(\tilde{X}_{0:p})] = \mathbb{E}[\varphi(\tilde{X}_{0:k}) G_{0:k-1}(\tilde{X}_{0:k-1})] \ , \quad \text{for } k \geq 1 \ . \tag{5.11}$$

We set by convention $\gamma_0 := \mu_0$ where μ_0 denotes the probability distribution, $\mathcal{L}(\tilde{X}_0)$, of the initial condition $\tilde{X}_0 = (t_0, x_0)$ i.e. $\mu_0 := \mathcal{L}(\tilde{X}_0) = \delta_{(t_0,x_0)}$. Gathering (5.10) together with the above definition one readily obtains the following proposition expressing $u(t_0, x_0)$ as an integral w.r.t. the non-negative measures γ_n.

Remark 5.1. *The weights used in equation 5.3 can be generalized with* $\alpha \in [\frac{1}{2}, 1-\kappa]$ *as*

$$G_k(\tilde{x}_{0:k}) := \begin{cases} 1 & \textit{if } k = 0 \\ |\tilde{G}_1(\tilde{x}_{0:1})| (\delta t_1)^\alpha \sqrt{c_1(\tilde{x}_{0:1})} & \textit{if } k = 1, \ t_2 < T, \textit{ and } \delta t_2 > 0 \\ |\tilde{G}_k(\tilde{x}_{0:k})| \sqrt{\frac{c_k(\tilde{x}_{0:k})}{c_{k-1}(\tilde{x}_{0:k-1})}} \left(\frac{\delta t_k}{\delta t_{k-1}}\right)^\alpha & \textit{if } k \geq 2 \ , \ \delta t_{k-1}\delta t_k > 0 \ , \\ 1 & \textit{elsewhere} \ , \end{cases} \tag{5.12}$$

Proposition 5.1. *Under Assumptions 1 and 3, the following identity holds for any* $n \geq 1$

$$u(t_0, x_0) = \gamma_n(\varphi_n) \ , \tag{5.13}$$

where $(\varphi_n)_{n \geq 1}$ *is a sequence of real valued functions such that for any* $n \geq 1$ *and* $\tilde{x}_{0:n} \in E_n := (\mathbb{R}_+ \times \mathbb{R}^d)^{n+1}$

$$\varphi_n(\tilde{x}_{0:n}) := \mathbb{E}[(\beta_{N_T+1} H_{N_T+1})(\tilde{X}_{0:N_T+1})(S_{1:N_T} G_{n:N_T})(\tilde{X}_{0:N_T}) \,|\tilde{X}_{0:n} = \tilde{x}_{0:n}] \ . \tag{5.14}$$

Remark 5.2. *Observe that for a given $n \geq 1$, φ_n is defined by (5.14) as a conditional expectation of a terminal payoff delivered at a future random time $N_T + 1$, knowing the state of the Markov chain from time 0 to n. Hence, evaluating $\varphi_n(\tilde{x}_{0:n})$ is not trivial, for a given $\tilde{x}_{0:n}$, since it requires to compute a conditional expectation. However whenever $\tilde{x}_n = (t_n, x_n)$ is such that $t_n \geq T$, then the knowledge of $\tilde{X}_{0:n} = \tilde{x}_{0:n}$ determines completely both $N_T = q < n$ and $\tilde{X}_{0:N_T+1}$, which implies*

$$\begin{aligned}\varphi_n(\tilde{x}_{0:n}) &:= \mathbb{E}[(\beta_{N_T+1}H_{N_T+1})(\tilde{X}_{0:N_T+1})(S_{1:N_T}G_{n:N_T})(\tilde{X}_{0:N_T})\,|\tilde{X}_{0:n} = \tilde{x}_{0:n}] \\ &= (\beta_{q+1}H_{q+1})(\tilde{x}_{0:q+1})S_{1:q}(\tilde{x}_{0:q})\ .\end{aligned}$$

Now, let us introduce the sequence of probability measures (η_k) defined by normalization of $(\gamma_k)_{k\geq 1}$

$$\eta_k(\varphi) := \frac{\gamma_k(\varphi)}{\gamma_k(\mathbf{1})} = \frac{\mathbb{E}[\varphi(\tilde{X}_{0:k})G_{0:k-1}(\tilde{X}_{0:k-1})]}{\mathbb{E}[G_{0:k-1}(\tilde{X}_{0:k-1})]}\ , \quad \text{for any } k \geq 0\ , \tag{5.15}$$

where $\mathbf{1}$ denotes the function which takes the unique value 1. Observing that for $k \geq 1$, $\gamma_k(\mathbf{1}) = \gamma_{k-1}(G_{k-1})$, we obtain by recurrence

$$\begin{aligned}\gamma_k(\varphi) &= \eta_k(\varphi)\gamma_k(\mathbf{1}) \\ &= \eta_k(\varphi)\gamma_{k-1}(G_{k-1}) \\ &= \eta_k(\varphi)\eta_{k-1}(G_{k-1})\cdots\eta_0(G_0)\ .\end{aligned} \tag{5.16}$$

As announced, we have replaced the expectation of a product of functions by the product of expectations of functions, since for any $n \geq 1$

$$u(t_0, x_0) = \gamma_n(\varphi_n) = \mathbb{E}[\varphi_n(\tilde{X}_{0:n})] = \eta_n(\varphi_n)\eta_{n-1}(G_{n-1})\cdots\eta_0(G_0)\ .$$

Our objective is now to approximate the sequence of probability measures $(\eta_k)_{k\geq 0}$ by a sequence of empirical measures $(\eta_k^N)_{k\geq 0}$ based on a system of N particles to finally end up with an approximation of the type

$$u(t_0, x_0) \approx \eta_n^N(\varphi_n)\eta_{n-1}^N(G_{n-1})\cdots\eta_0^N(G_0)\ .$$

5.2 The particle approximation scheme

The sequence of approximating measures $(\eta_k^N)_{k\geq 0}$ will be defined by mimicking the dynamics of $(\eta_k)_{k\geq 0}$. Hence, we begin by describing this recursive dynamics.
First let K_k denote the transition kernel of the path valued Markov chain $(X'_k := \tilde{X}_{0:k})$ from $k-1$ to k for any integer $k \geq 1$. Recall that K_k can be considered both as an integral operator on the space of measurable functions defined on E_k and on the space of finite measures, $\mathcal{M}(E_{k-1})$, such that

- for any measurable test function f_k defined on E_k, $K_k(f_k)$ is a measurable function defined on E_{k-1} such that for any $x'_{k-1} \in E_{k-1}$

$$K_k(f_k)(x'_{k-1}) = \mathbb{E}[f_k(X'_k)\,|X'_{k-1} = x'_{k-1}] = \int_{y'_k\in E_k} K_k(x'_{k-1}, dy'_k)f_k(y'_k)\ ,$$

- for any finite measure m_{k-1} on E_{k-1}, $m_{k-1}K_k$ is a finite measure on E_k such that for any $x'_k \in E_k$

$$(m_{k-1}K_k)(dx'_k) = \int_{y'_{k-1} \in E_{k-1}} m_{k-1}(dy'_{k-1})K_k(y'_{k-1}, dx'_k) \ .$$

In particular, let μ_k denote the probability law underlying the random variable $X'_k := \tilde{X}_{0:k}$ (we will often write $\mu_k = \mathcal{L}(X'_k)$), for any $k \geq 0$. Then observe that $\mu_k K_{k+1} = \mu_{k+1}$ the probability law of $X'_{k+1} := \tilde{X}_{0:k+1}$. Besides, notice that if $\tilde{K}_k$ denotes the transition kernel of the Markov chain $(\tilde{X}_k)$ from $k-1$ to k, then the transition kernel K_k is obtained as the following cartesian product, for any $(y'_{k-1}, x'_k) := (y_{0:k-1}, dx_{0:k}) \in E_{k-1} \times E_k$

$$K_k(y'_{k-1}, dx'_k) = K_k(y_{0:k-1}, dx_{0:k}) = \delta_{y_{0:k-1}}(dx_{0:k-1}) \times \tilde{K}_k(y_{k-1}, dx_k) \ .$$

Now we can describe the dynamics of $(\eta_k)_{k\geq 0}$ with k. For any real valued test function f_k defined on E_k, the following identities holds

$$\begin{aligned}
\eta_k(f_k) &:= \frac{\gamma_k(f_k)}{\gamma_k(\mathbf{1})} \\
&= \frac{\mu_k(f_k G_{1:k-1})}{\mu_k(G_{1:k-1})} \ , \quad \text{where } \mu_k := \mathcal{L}(X'_k) = \mathcal{L}(\tilde{X}_{0:k}) \ \text{, and } G_{1:k} := \prod_{p=1}^{k} G_p \\
&= \frac{\mu_{k-1}(K_k(f_k)G_{1:k-1})}{\mu_{k-1}(G_{1:k-1})} \quad \text{by the tower property of conditional expectation} \\
&= \frac{\gamma_{k-1}(K_k(f_k)G_{k-1})}{\gamma_{k-1}(G_{k-1})} \quad \text{by definition (5.11) of } \gamma_{k-1} \\
&= \frac{\eta_{k-1}(K_k(f_k)G_{k-1})}{\eta_{k-1}(G_{k-1})} \quad \text{by dividing the numerator and denominator by } \gamma_{k-1}(\mathbf{1}) \\
&= ((G_{k-1} \cdot \eta_{k-1})K_k)(f_k) \ ,
\end{aligned}$$

where the $\cdot$ sign denotes the projective product between a non-negative function G defined on E and a non-negative measure $\mu \in \mathcal{M}^+(E)$ returning the probability measure $G \cdot \mu$ such that

$$(G \cdot \mu)(dx) := G(x)\mu(dx)/\mu(G) \ . \tag{5.17}$$

Hence, one can describe the evolution from η_{k-1} to η_k into two steps

$$\eta_{k-1} \xrightarrow{\text{Correction}} \hat{\eta}_{k-1} := G_{k-1} \cdot \eta_{k-1} \xrightarrow{\text{Evolution}} \eta_k := \hat{\eta}_{k-1}K_k \ , \tag{5.18}$$

In other words, the sequence of probability measures (η_k) satisfies the following recursion

$$\begin{cases}
\eta_0 = \mu_0 \ , \quad \text{where } \mu_0 := \mathcal{L}(X'_0) = \mathcal{L}(\tilde{X}_0) \\
\hat{\eta}_k := G_k \cdot \eta_k \ , \quad \text{for all} \quad 1 \leq k \leq n \ , \\
\eta_{k+1} = \hat{\eta}_k K_k \ , \quad \text{for all} \quad 1 \leq k \leq n \ .
\end{cases} \tag{5.19}$$

An Interacting Particle System will be used to approximate the sequence of probability measures $(\eta_k)_{0\le k\le n}$ by a sequence of empirical probability measures $(\eta_k^N)_{0\le k\le n}$, such that for all $1 \le k \le n$, η_k^N is associated with an N-samples $(\xi_k^{1,N}, \cdots, \xi_k^{N,N})$ approximately distributed according to η_k. To simplify the notation, we will often drop the exponent N and write $(\xi_k^i)_{i=1,\cdots N}$ instead of $(\xi_k^{i,N})_{i=1,\cdots N}$. The recursive evolution described by (5.19) is approximated by the following dynamics:

$$\begin{cases} \eta_0^N = \mu_0 \\ \hat{\eta}_k^N = G_k \cdot \eta_k^N \ , \quad \text{for all} \quad 1 \le k \le n \\ \eta_{k+1}^N = S^N(\hat{\eta}_k^N K_k) \ , \quad \text{for all} \quad 1 \le k \le n \ , \end{cases} \tag{5.20}$$

where $S^N(\mu)$ denotes the empirical measure associated to an N-sample $(\xi^1, \cdots, \xi^N)$ i.i.d. according to μ, that is

$$S^N(\mu) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi^i} \ , \quad \text{where } (\xi^1, \cdots, \xi^N) \text{ i.i.d. } \sim \mu \ .$$

Hence, the algorithm proceeds as follows. Recalling that $G_0 = \mathbf{1}$, we initiate the algorithm by generating N i.i.d. random variables $(\xi_1^1, \cdots, \xi_1^N)$ according to μ_0, then we set

$$\eta_1^N = S^N(G_0 \cdot \mu_0) = S^N(\mu_0) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_1^i} \ . \tag{5.21}$$

The evolution of the discrete measures, $(\eta_k^N)_{0\le k\le n}$, (where N denotes the size of the particle system) between two iterations k and $k+1$, consists into three steps:

1. **Weighting**: each particle is weighted according to the value of the current potential function G_k. For all $i \in \{1, \cdots, N\}$, we compute $\omega_k^i = \dfrac{G_k(\xi_k^i)}{\sum_{j=1}^N G_k(\xi_k^j)}$ and we set $\hat{\eta}_k^N = \displaystyle\sum_{i=1}^{N} \omega_{k+1}^i \, \delta_{\xi_{k+1}^i}$.

2. **Selection**: N i.i.d. random variables $(\hat{\xi}_k^1, \cdots, \hat{\xi}_k^N)$ are generated according to the weighted discrete probability distribution $\hat{\eta}_k^N = \displaystyle\sum_{i=1}^{N} \omega_k^i \, \delta_{\xi_k^i}$. More specifically, for all $i \in \{1, \cdots, N\}$, an index $I \in \{1, \cdots, N\}$ is generated independently with probability $\mathbb{P}(I = j) = \omega_k^j$ and we set $\hat{\xi}_k^i = \xi_k^I$.

3. **Mutation**: Each selected particle evolves independently according to the dynamics K_{k+1}. This produces a new particle system $(\xi_{k+1}^1, \cdots, \xi_{k+1}^N)$. More specifically, for all $i \in \{1, \cdots, N\}$, we generate independently ξ_{k+1}^i according to the conditional distribution $\mathcal{L}(X'_{k+1} | X'_k = \hat{\xi}_k^i)$, then we set

$$\eta_{k+1}^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_{k+1}^i} \ . \tag{5.22}$$

For all $k \geq 1$, let us introduce γ_k^N, the particle approximation of γ_k based on η_k^N defined by recursion (5.20) and such that for any real valued measurable test function f_k defined on E_k,

$$\gamma_k^N(f_k) = \eta_k^N(f_k) \prod_{0 \leq p \leq k-1} \eta_p^N(G_p) \ . \tag{5.23}$$

We begin by stating a Lemma that will be crucial to prove the convergence of our new estimator.

Lemma 5.2. *Let $(X'_n)_{n\geq 0}$ be a Markov chain (with initial distribution μ_0 and transition kernel K_k) defined on a sequence of measurable spaces $(E_n, \mathcal{E}_n)_{n\geq 0}$ and $(G_n)_{n\geq 0}$ be a sequence of positive measurable functions defined on $(E_n, \mathcal{E}_n)_{n\geq 0}$ such that there exists a finite constant $A \geq 2$ such that*

$$\sup_{x'_0 \in E_0} G_0(x'_0) \leq A \ , \quad \text{and} \quad \sup_{x'_{p-1} \in E_{p-1}} \mathbb{E}[G_p^2(X'_p) | X'_{p-1} = x'_{p-1}] \leq A \ , \quad \text{for any } p \geq 1 \ . \tag{5.24}$$

We consider the sequence of Feynman-Kac measures (γ_n) such that for any measurable real valued function f_n defined on E_n,

$$\gamma_n(f_n) := \mathbb{E}[f_n(X'_n) \prod_{k=0}^{n-1} G_k(X'_k)] \ . \tag{5.25}$$

Let (γ_n^N) be a sequence of particle approximation measures of (γ_n) defined similarly as in (5.23), with $(\eta_p^N)_{0\leq p}$ defined by (5.20). For a given $n \geq 1$, let us consider a real valued measurable function f_n defined on E_n such that there exists a finite positive constant B such that

$$\sup_{x'_{p-1} \in E_{p-1}} |\mathbb{E}[f_n^2(X'_n) G_{p:n-1}^2(X'_{p:n}) | X'_{p-1} = x'_{p-1}] \leq B \quad \text{for any } p = 1, \cdots n \ , \tag{5.26}$$

for a given positive constant C. Then the particle approximation $\gamma_n^N(f_n)$ is unbiased with finite variance, more precisely

$$\mathbb{E}[\gamma_n^N(f_n)] = \gamma_n(f_n) \ , \ \text{and} \quad \mathbb{E}[\big(\gamma_n^N(f_n) - \gamma_n(f_n)\big)^2] \leq 2B \frac{A^{n+2}}{N} \quad \text{for } N \geq A^{n+1} \ . \tag{5.27}$$

The proof of this Lemma relies on the formalism developed in the reference books [3, 4]. However, we had to carry out an original proof to take into account our specific framework where the potential functions G_k are unbounded which is not considered to our knowledge in the existing literature. The proof is placed in the Appendix 8.

We are now in a position to state the main result of this section.

Theorem 5.3. *Suppose that Assumptions 1, 3 and 4 are satisfied. For any $n \geq 1$, the resampling estimator $\gamma_n^N(\varphi_n)$ defined by (5.23) is unbiased with finite variance. More precisely,*

$$\mathbb{E}[\gamma_n^N(\varphi_n)] = u(t_0, x_0) \ , \quad \text{and} \quad \mathbb{E}[\big(\gamma_n^N(\varphi_n) - u(t_0, x_0)\big)^2] \leq \frac{C^{n+2}}{N} \quad \text{for } N \geq C^{n+1} \ , \tag{5.28}$$

where $(\varphi_n)_{n\geq 1}$ is a sequence of real valued functions defined on E_n by (5.14) and C is a constant depending only on the characteristics of the problem (T, the bounds or Lipschitz constants related to g, b, σ, a).

Remark 5.3. *Computing $\gamma_n^N(\varphi_n)$ reduces to compute the following product of empirical means*

$$\begin{aligned}\gamma_n^N(\varphi_n) &= \eta_n^N(\varphi_n)\eta_{n-1}^N(G_{n-1})\cdots\eta_0^N(G_0)\\ &= \left(\frac{1}{N}\sum_{i=1}^N \varphi_n(\xi_n^i)\right)\left(\frac{1}{N}\sum_{i=1}^N G_{n-1}(\xi_{n-1}^i)\right)\cdots\left(\frac{1}{N}\sum_{i=1}^N G_0(\xi_0^i)\right) ,\end{aligned}$$

where $(\xi_k^i)_{1\leq i\leq N}$ is the particle system at the kth iteration of the algorithm as stated by (5.22). This in particular requires to compute $\varphi_n(\xi_n^i)$ for each particle of the final particle system $(\xi_n^i)_{i=1,\cdots N}$. Recalling Remark 5.2, this may require to compute a conditional expectation. In practice, one chooses n large enough such that most of particles have already reached time T after n iterations implying that for most particles $\varphi_n(\xi_n^i)$ can be computed explicitly. In the rare cases of particles ξ_n^i that have not reached yet time T, the computation of $\varphi_n(\xi_n^i)$ that should normally require to compute a conditional expectation is approximated by one simulation according to

$$\mathcal{L}((\beta_{N_T+1}H_{N_T+1}S_{0:N_T+1})(\tilde{X}_{0:N_T+1})G_{n:N_T}(\tilde{X}_{0:N_T})\,|\tilde{X}_{0:n}=\xi_n^i) .$$

Notice that it would be interesting to consider the estimator

$$\gamma_{n_N}^N(\varphi_{n_N}) , \quad \textit{with} \quad n_N = \inf\{n\,|\,\xi_n^i \textit{ has reached } T \textit{ for all } i=1,\cdots N\} .$$

This will be left for future work.

Proof. Theorem 5.3 is a direct consequence of Proposition 5.1 stating that $\gamma_n(\varphi_n)=u(t_0,x_0)$ and of Lemma 5.2 after having verified that there exists a finite positive constant C for which the bounds (5.24) and (5.26) are verified. Observe that (5.24) is automatically implied by (5.6). Let us consider (5.26), similarly to the proof of Proposition 4.1 one obtains

$$\begin{aligned}&\mathbb{E}[\varphi_n^2(\tilde{X}_{0:n})G_{p:n-1}^2(\tilde{X}_{p:n})|\tilde{X}_{0:p-1}=\tilde{x}_{0:p-1}]\\ &=\textstyle\sum_{q=0}^{\infty}\mathbb{E}[\varphi_n^2(\tilde{X}_{0:n})G_{p:n-1}^2(\tilde{X}_{p:n})|\tilde{X}_{0:p-1}=\tilde{x}_{0:p-1},N_T=q]\mathbb{P}(N_T=q) .\end{aligned} \tag{5.29}$$

Now considering the general term of this sum for $q\geq p\geq 2$

$$\begin{aligned}&\mathbb{E}[(\beta_{N_T+1}^2H_{N_T+1}^2)(\tilde{X}_{0:N_T+1})\textstyle\prod_{k=p}^{N_T}G_k^2(\tilde{X}_{0:k})\,|\tilde{X}_{0:p-1}=\tilde{x}_{0:p-1},N_T=q]\\ &=\mathbb{E}[\beta^2\frac{1}{c_{p-1}(\tilde{X}_{0:p-1})}\frac{1}{(\delta T_{p-1})^{2(1-\kappa)}}\textstyle\prod_{k=p-1}^{q-1}P_{k+1}^2\,|\tilde{X}_{0:p-1}=\tilde{x}_{0:p-1},N_T=q]\\ &\leq C\mathbb{E}[(\delta T_q)^{2(1-\kappa)}\frac{1}{c_{p-1}(\tilde{X}_{0:p-1})}\frac{1}{(\delta T_{p-1})^{2(1-\kappa)}}c_q(\tilde{X}_{0:q})P_q^2\textstyle\prod_{k=p-1}^{q-2}P_{k+1}^2\,|\tilde{X}_{0:p-1}=\tilde{x}_{0:p-1},N_T=q] ,\end{aligned}$$

where C is a constant that may change from line to line. Recalling (5.5) finally gives

$$\begin{aligned}&\mathbb{E}[(\beta_{N_T+1}^2H_{N_T+1}^2)(\tilde{X}_{0:N_T+1})\textstyle\prod_{k=p}^{N_T}G_k^2(\tilde{X}_{0:k})\,|\tilde{X}_{0:p-1}=\tilde{x}_{0:p-1},N_T=q]\\ &\leq C^{q-p+1}\mathbb{E}[(\delta T_q)^{2(1-\kappa)}\frac{c_{p-1}(\tilde{X}_{0:p-1})}{c_{p-1}(\tilde{X}_{0:p-1})}\frac{1}{(\delta T_{p-1})^{2(1-\kappa)}}\textstyle\prod_{k=p-1}^{q-1}\frac{(\delta T_k)^{2(1-\kappa)}}{\delta T_{k+1}}\,|N_T=q]\\ &\leq C^{q-p+1} .\end{aligned}$$

We proceed similarly when $p=1$. We conclude by observing that the sum (5.29) is finite by the same argument as in the proof of Proposition 4.1. $\square$

6 Numerical simulations

In this section we analyse and compare the performances of the three approaches described previously

1. Monte Carlo simulation with exponential switching times;
2. Monte Carlo simulation with gamma switching times (with parameter $\kappa \leq 1/2$);
3. Interacting Particle Systems (IPS) with gamma switching times (with parameter $\kappa \leq 1/2$).

We consider a one dimensional case ($d = 1$) and a four dimensional case ($d = 4$). Because the variance of the results is closely related to the diffusion coefficient variation, we will consider various examples with σ more and more space dependent.

In all cases, we consider

- a drift coefficient $b(t, x) = 1 - x$,
- an initial condition $x_0 = 1$,
- a terminal condition or payoff function $g(x) = (\frac{1}{d}\sum_{i=1}^{d} x_i - 1)^+$,
- a terminal time $T = 1$.

The parameters of the switching time distributions is $\lambda = 0.4$ for the exponential distribution. Even if the exponential distribution gives a theoretical infinite variance (in cases we consider here), the numerical variance observed is finite so it is interesting to compare the results obtained by the gamma distribution and the exponential distributions.
To exploit the parallelism, for each number of particles $n_{\text{part}} = 96N$, where 96 is the numbers of computational cores, we allocate N particles to each core. When Interacting Particle Systems (IPS) are used, a resampling estimator $\gamma_p^{N,j}(\varphi_p)$ is simulated independently on each core j and we return the average estimator : $\frac{1}{96}\sum_{j=1}^{96}\gamma_p^{N,j}(\varphi_p)$. Then the procedure is repeated independently for 1000 estimations, so as to approximate empirically the expectation and the variance of each estimator by the empirical average and variance computed on the 1000 estimates.
The whole procedure is then repeated for different values of $n_{\text{part}} = 4^q n_0$ from $q = 0$ to $q = 5$, with $n_0 = 10^5$. We reported on the graphs the evolution of the estimator expectation as a function of $\log(n_{\text{part}})$ and the related standard deviation is represented on log-log graphs. On each figure devoted to the standard deviation, the theoritical decrease at a rate $1/(n_{\text{part}})^{1/2}$ is represented by the plot of a line with slope -0.5.

6.1 One dimensional tests

6.1.1 $\sigma(t, x) = 0.5 + 0.2(x^2 \wedge 1)$

This first case shows some quite small variations of σ. The parameters of the gamma distribution are $\theta = 2.5$ and $\kappa = 0.5$. The reference value is 0.17466. Evolution of the global estimate and the standard deviation are given on figure 1

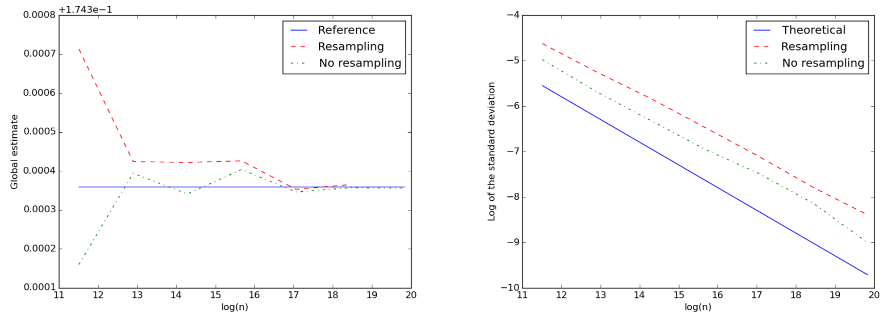

Figure 1: Estimation and standard deviation observed for case 1 with gamma distribution $\kappa = 0.5$, $\theta = 2.5$.

In this simple case easily converging, importance sampling doesn't improve the standard deviation. Besides the sampling has a cost. Taking 1600000 just add 15% in computational time, but while taking 410 millions particles, importance sampling doubles the computational cost.

6.1.2 $\sigma(t, x) = 0.5 + 0.4(x^2 \wedge 1)$

With this more interesting case, we plot results obtained with $\theta = 2.5$, $\kappa = 0.3$, $\kappa = 0.5$ with and without resampling. Besides we add the case of the exponential distribution. The reference value is 0.21408. Evolution of the global estimate and the standard deviation are given on figure 2.

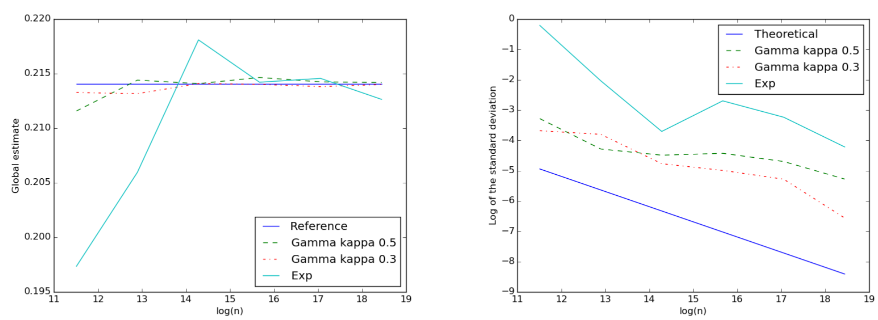

Figure 2: Estimation and standard deviation observed for case 2.

With or without importance sampling, the standard deviation is decreasing steadily while using gamma distributions. As we double the number of simulations, the standard deviation is roughly divided by two which is coherent with the theory. Using the exponential distribution with resampling, the numerical standard deviation goes to zero but at a pace impossible to quantify : this is the consequence of an infinite theoretical variance, whereas the convergence without importance sampling seems to be slow and erratic.
In the case of gamma distribution, the standard deviation with importance sampling is nearly half of the one without importance sampling clearly showing the interest in this method. Results with $\kappa = 0.5$ and $\kappa = 0.3$ are very similar especially with importance

sampling, but the number of jumps increases as κ decreases and the computational time is nearly doubled with $\kappa = 0.3$ indicating that the optimal choice is to take $\kappa = 0.5$.

6.1.3 $\sigma(t, x) = 0.5 \vee x^2 \wedge 1$

This case is more difficult than the two firsts. We drop the exponential distribution that still shows some erratic decrease in the standard deviation. The reference value is 0.2100. We keep $\theta = 2.5$ and we use $\kappa = 0.3$ and $\kappa = 0.5$. On figures 3 and 4, we show that without importance sampling the method is not clearly converging, while it is converging always with the same pace when importance sampling is used.

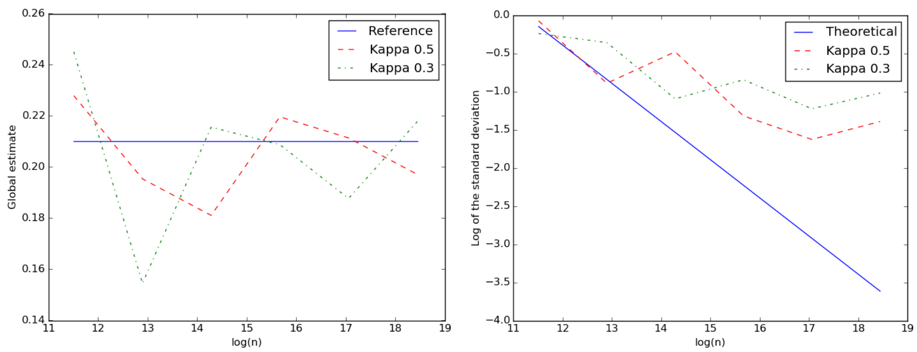

Figure 3: Estimation and standard deviation observed for case 3 without resampling.

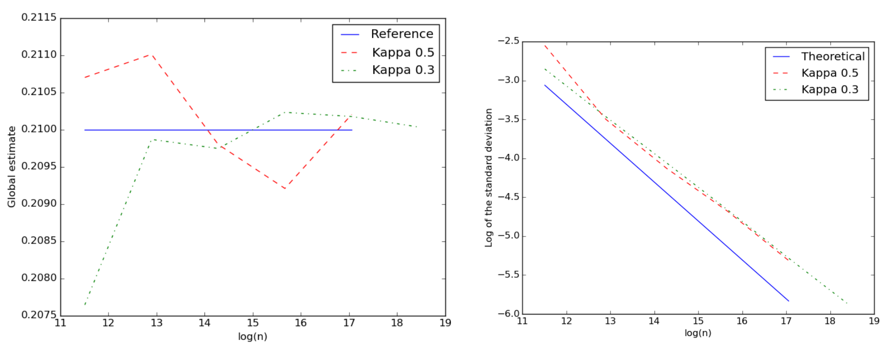

Figure 4: Estimation and standard deviation observed for case 3 with resampling.

6.2 Some four dimensional cases

The parameters of the SDE (1.1) are $x \in \mathbb{R}^4$: $b(t, x) = 1 - x$, $x_0 = 1$ and $\sigma(t, x) = (0.5 + a \min((\sum_{i=1}^4 x_i)^2, 1))\mathbb{I}$ for a a positive real number. Besides we take $g(x) = (\frac{1}{4} \sum_{i=1}^4 x_i - 1)^+$, $T = 1$.

6.2.1 $a = 0.4$

The reference solution is 0.11806. We keep $\theta = 2.5$ for the gamma distribution. For this first case, we plot on figures 5 6 the result obtained by the different distributions with and without importance sampling.

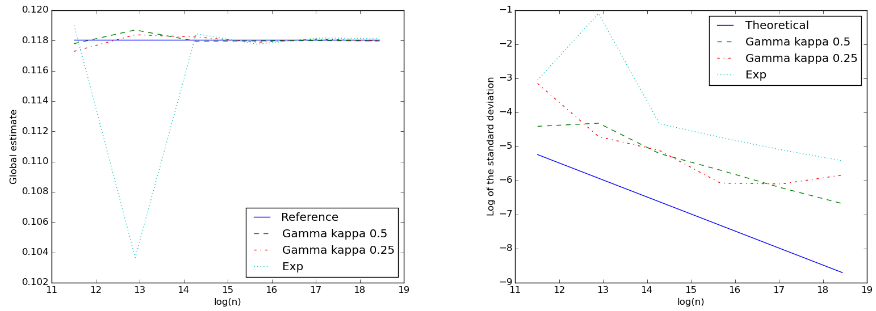

Figure 5: Estimation and standard deviation without resampling observed for the first 4 dimensional test case.

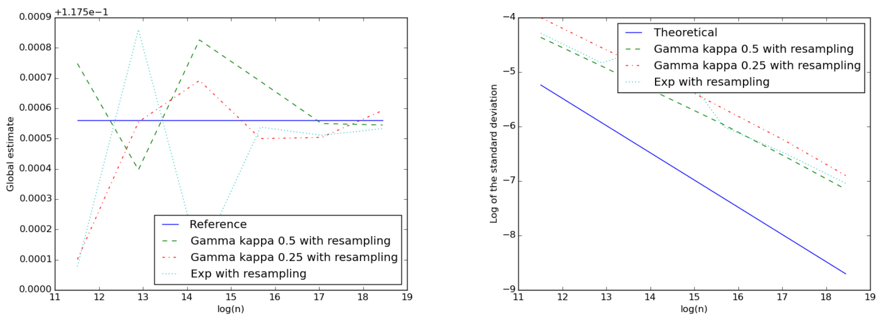

Figure 6: Estimation and standard deviation with resampling observed for the first 4 dimensional test case.

In this case with importance sampling with gamma functions, the log of the standard deviation decreases linearly. This is not the case with the exponential distribution. Without importance sampling, the decrease in standard deviation is not regular. On this test case importance sampling is effective by reducing the standard deviation by a factor rougly equal to 2 and by stabilizing the results.

6.2.2 $a = 0.6$

On this test case it was impossible to get convergence without importance sampling and with importance sampling with the exponential distribution. On figure 7, we only give the results obtained with importance sampling and the gamma distribution. We notice that the standard deviation calculated are far higher than in the previous case. We have difficulties to get a linear reduction in the standard deviation. The influence of θ parameter is not obvious on the curves, but because higher θ give higher jumps, it gives smaller computational times.

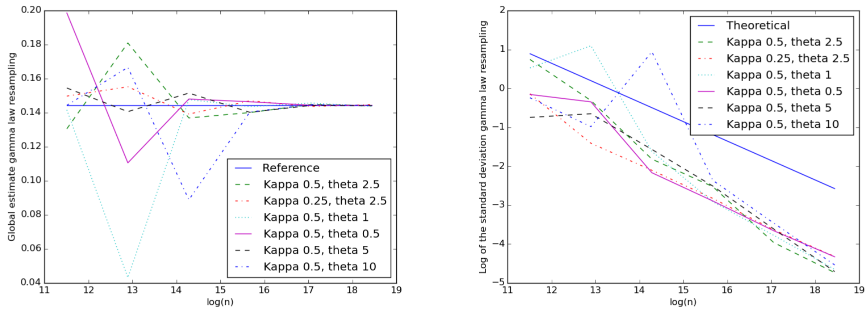

Figure 7: Estimation and standard deviation observed for the second 4 dimensional test case for gamma distributions with resampling.

7 Appendix: Algorithms

We give a description of the algorithm using interacting particle sytems with gamma switching times Note that here the importance sampling is applied during the whole algorithm and not only for a fixed number of steps.
In the sequel, g denotes the sample of a d dimensional Gaussian distribution and ω a sample of the one dimensional distribution with density *pdf* and cumulative distribution *cdf* used for time steps generation. This algorithm is decomposed in one initialization algorithm (1) for the first time step, and a resampling algorithm (3) and a final step calculating the solution (2).

Algorithm 1 Initialization step for algorithm (4)

procedure INITIALIZATION($X, W, Kjump, \Delta T, T, \hat{\mu}, \hat{a}, R, SgnW$)
  $Kjump = 0$ ▷ First jump
  $R = 1$ ▷ Renormalization weight for importance sampling
  for $i = 1, M$ **do**
    $X(i) = x_0$, $W(i) = 1$, $SgnW(i) = 1$ ▷ Position,weight, and sign of the weights
    Sample g, ω ▷ for time step and Brownian
    $\Delta T(i) = (T - T_0) \wedge w$, $T(i) = T_0 + \Delta T(i)$ ▷ Time increment and time update
    $\hat{\mu}(i) = \mu(0, X(i))$, $\hat{\sigma}(i) = \sigma(0, X(i))$, $\hat{a}(i) = \frac{1}{2}\sigma(0, X(i))\sigma(0, X(i))^T$ ▷ Update SDE coefficient
    $X(i) = X(i) + \hat{\mu}(i)\Delta T(i) + \hat{\sigma}(i) g\sqrt{\Delta T(i)}$ ▷ One step for position
  end for
end procedure

Algorithm 2 Final step for algorithm (4)

procedure FINALIZE($X, \Delta T, T, \sigma, \hat{g}, W, \hat{W}, R, SgnW$)
    $S = 0$
    for $i = 1, M$ **do**
        $X_p = X(i) - \sqrt{\Delta T(i)}\sigma(i)\hat{g}(i)$
        $\hat{X} = X(i) - 2\sqrt{\Delta T(i)}\sigma(i)\hat{g}(i)$ ▷ Antithetic position
        if $T(i) = \Delta T(i)$ **then** ▷ Only on step
            $G = 0$
        else
            $G = g(T, X_p)$
        end if
        $E = (g(T, X(i)) - G)W(i)$
        $\hat{E} = (g(T, \hat{X}) - G)\hat{W}(i)$ ▷ Antithetic
        $S = S + \frac{1}{1-cfd(\Delta T(i))}\frac{E+\hat{E}}{2}RSgnW(i)$ ▷ take into account renormalization factor and sign of the weights
    end for
    $S = \frac{S}{M}$ ▷ Final estimation of the solution
end procedure

Algorithm 3 Importance sampling for algorithm (4)

procedure RESAMPLE($X, N, SgnW$,...)
    for $i = 1, M$ **do**
        if $T(i) < T$ **then**
            $\underline{W}(i) = |W(i)|$
            if $W(i) < 0$ **then**
                $SgnW(i) = -SgnW(i)$ ▷ Update of the product of signs for particle
            end if
        else
            $\underline{W}(i) = 1$ ▷ Particle arrived in T so weight equal to 1
        end if
    end for
    $R = R\frac{1}{M}\sum_{i=1}^{M}\underline{W}(i)$ ▷ Renormalization update according to average of weights
    Resample X according to weights $\underline{W}$
    update W, $\hat{W}$,$SgnW$, ΔT ... according resampling
end procedure

Algorithm 4 Algorithm for Feynman Kac resolution

Initialization($X, W, Kjump, \Delta T, T, \hat{\mu}, \hat{a}, R, SgnW$)
while $\min_i T(i) < T$ **do**
    for $i = 1, M$ **do**
        if $T(i) < T$ **then**
            $d = pdf(\Delta T(i))$, $\hat{\Delta T} = \Delta T(i)$ ▷ Store density, time step
            $\mu(i) = \mu(T(i), X(i))$, $\sigma(i) = \sigma(T(i), X(i))$, $a(i) = \frac{1}{2}\sigma(i)\sigma(i)^T$
            Sample g, w
            $\Delta T(i) = (T - T(i)) \wedge \omega$, $T(i) = T(i) + \Delta T(i)$ ▷ Time step update
            $X(i) = X(i) + \mu(i)\Delta T(i) + \sigma(i)\sqrt{\Delta T(i)}$ ▷ Position update
            $W_\mu = \frac{1}{d}(\mu(i) - \hat{\mu}(i)).((\sigma(i)^{-1})^T \frac{\omega}{\sqrt{\Delta T(i)}})$ ▷ Malliavin weight 1
            $W_a = \frac{1}{d} tr[(a(i) - \hat{a}(i)) \left((\sigma(i)^{-1})^T \frac{gg^T - \mathbb{I}}{\Delta T(i)} \sigma(i)^{-1}\right)^T]$ ▷ Malliavin weight 2
            $\hat{\mu}(i) = \mu(i)$, $\hat{a}(i) = a(i)$ ▷ for next step
            if $KJump = 0$ **then**
                if $T(i) < T$ **then**
                    $W(i) = (W_\mu + \frac{1}{2}W_a)\Delta T(i)$
                else
                    $W(i) = W_\mu + \frac{1}{2}W_a$, $\hat{W}(i) = -W_\mu + \frac{1}{2}W_a$, $\hat{g}(i) = g$ ▷ Antithetic weight
                end if
            else
                if $T(i) < T$ **then**
                    $W(i) = (W_\mu + \frac{1}{2}W_a)\frac{\Delta T(i)}{\hat{\Delta T}}$
                else
                    $W(i) = \frac{W_\mu + \frac{1}{2}W_a}{\hat{\Delta T}}$, $\hat{W}(i) = \frac{-W_\mu + \frac{1}{2}W_a}{\hat{\Delta T}}$, $\hat{g}(i) = g$ ▷ Antithetic weight
                end if
            end if
        end if
    end for
    $kJump = 1$
    if $\min_i T(i) < T$ **then**
        Resample($X, N, SgnW$,...)
    end if
end while
Finalize($X, \Delta T, T, \sigma, \hat{g}, W, \hat{W}, R, SgnW$)

8 Appendix: Usefull notations and proof of Lemma 5.2

8.1 Classical notations and results

Let us recall somme classical notations and results stated in [3, 4]. We consider a Markov chain (X'_n) (with initial probability μ_0 and transition kernel K_k) taking values on a sequence of measurable spaces $(E_n, \mathcal{E}_n)$ and a sequence of positive potential functions (G_n) defined on $(E_n, \mathcal{E}_n)$.

- For all $0 \leq p \leq n$, we define the sigma-finite measure $\gamma_p \in \mathcal{M}(E_p)$ such that for any real valued bounded test function f_p defined on E_p,

$$\gamma_p(f_p) = \mathbb{E}[\, f_p(X'_p) \prod_{0 \leq k \leq p-1} G_k(X'_k)\,] \ .$$

- For all $0 \leq p \leq n$, let us define the probability measure η_p obtained by normalization of γ_p and such that for any real valued bounded test function f_p defined on E_p,

$$\eta_p(f_p) = \frac{\gamma_p(f_p)}{\gamma_p(1)} \ .$$

  Notice that γ_p can be written as the following product involving the probability measures $\eta_1, \cdots \eta_p$,

$$\gamma_p(f_p) = \eta_p(f_p) \prod_{0 \leq k \leq p-1} \eta_k(G_k) \ .$$

- For all $1 \leq p \leq n$, let us introduce the nonlinear operator, Φ_p defined on the space of sigma-finite and non-negative measures $\mathcal{M}^+(E_{p-1})$ and taking values in $\mathcal{M}^+(E_p)$ such that

$$\Phi_p(m_{p-1}) := (G_{p-1} \cdot m_{p-1}) K_p \ , \quad \text{for any} \quad m_{p-1} \in \mathcal{M}^+(E_{p-1}) \ . \tag{8.1}$$

  Then the nonlinear evolution of (η_p) can be summarized by

$$\eta_{p+1} = \Phi_{p+1}(\eta_p) \ , \tag{8.2}$$

- For all $0 \leq p \leq n$, let us define the Feynman-Kac semi-group $Q_{p,n}$ associated to the distribution flow $(\gamma_p)_{1 \leq p \leq n}$ such that for all $x'_p \in E_p$ and any real valued bounded test function f_n defined on E_n,

$$Q_{p,n}(f_n)(x'_p) = \mathbb{E}[\, f_n(X'_n) \prod_{p \leq k \leq n-1} G_k(X'_k) \,|\, X'_p = x'_p\,] \ , \quad \text{with } Q_{n,n} = Id \ , \tag{8.3}$$

  Notice that for all $0 \leq p \leq n$, γ_n can be written as the transformation of γ_p via $Q_{p,n}$,

$$\gamma_n = \gamma_p Q_{p,n} \ . \tag{8.4}$$

  Moreover, for any sigma-finite non-negative measure $\mu \in \mathcal{M}^+(E_p)$,

$$\Phi_p(\mu) = \frac{\mu Q_{p-1,p}}{(\mu Q_{p-1,p})(1)} \ . \tag{8.5}$$

- We introduce (γ_n^N) the particle approximation sequence of measures as defined on Lemma 5.2. Notice that by definition (5.23) of γ_p^N the following relation holds

$$\eta_p^N(f_p) = \frac{\gamma_p^N(f_p)}{\gamma_p^N(1)} \ . \tag{8.6}$$

- Let f_n be a real valued test function defined on E_n, then the error between $\gamma_n^N(f_n)$ and $\gamma_n(f_n)$ can be decomposed as the sum of n "local errors" as follows, using relation (8.4)

$$\begin{aligned}(\gamma_n^N - \gamma_n)(f_n) &= \sum_{p=1}^n [\gamma_p^N Q_{p,n} - \gamma_{p-1}^N Q_{p-1,n}](f_n) \quad (\text{recalling that } \gamma_0^N := \gamma_0) \\ &= \sum_{p=1}^n [\gamma_p^N - \gamma_{p-1}^N Q_{p-1,p}]\big(Q_{p,n}(f_n)\big) \ ,\end{aligned}$$

Then using relations (8.5) and (8.6) and recalling that $\gamma_p^N(\mathbf{1}) = \gamma_{p-1}^N(G_{p-1})$ yields

$$(\gamma_n^N - \gamma_n)(f_n) = \sum_{p=1}^n \gamma_{p-1}^N(G_{p-1})[\eta_p^N - \Phi_p(\eta_{p-1}^N)]\big(Q_{p,n}(f_n)\big) \ . \tag{8.7}$$

- Let us introduce the following notations

$$\begin{aligned}\Gamma_p^N(f_p) &:= (\gamma_p^N - \gamma_p)(f_p) \ , \\ \Delta_p^N(f_p) &:= [\eta_p^N - \Phi_p(\eta_{p-1}^N)]\big(f_p\big) = \tfrac{1}{N}\textstyle\sum_{i=1}^N f_p(\xi_p^i) - \Phi_p(\eta_{p-1}^N)(f_p) \ .\end{aligned} \tag{8.8}$$

- For any $p \geq 0$, let us introduce the σ-algebra $\mathcal{G}_p$ generated by the particle system until the p-th generation, observe that since

$$\eta_p^N = S^N\big(\Phi_p(\eta_{p-1}^N)\big) = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_p^i} \ ,$$

where $(\xi_p^1, \cdots, \xi_p^N)$ are i.i.d. according to $\Phi_p(\eta_{p-1}^N)$ conditionally to $\mathcal{G}_{p-1}$. Thus

$$\mathbb{E}[\Delta_p^N(f_p)\,|\,\mathcal{G}_{p-1}] = \frac{1}{N}\sum_{i=1}^N \mathbb{E}[f_p(\xi_p^i)|\mathcal{G}_{p-1}] - \Phi_p(\eta_{p-1}^N)(f_p) = 0 \ , \quad \text{and} \quad \mathbb{E}[\Gamma_p^N(f_p)] = 0 \ . \tag{8.9}$$

- Moreover one can bound the conditional variance of $\Delta_p^N(f_p)$ as follows

$$\begin{aligned}\mathbb{E}[(\Delta_p^N(f_p))^2\,|\,\mathcal{G}_{p-1}] &= \mathbb{E}[\big((\eta_p^N - \Phi_p(\eta_{p-1}^N))(f_p)\big)^2\,|\,\mathcal{G}_{p-1}] \\ &= \mathbb{E}[\big(\frac{1}{N}\sum_{i=1}^N f_p(\xi_p^i) - \Phi_p(\eta_{p-1}^N)\big)^2\,|\,\mathcal{G}_{p-1}] \\ &= \frac{1}{N}[\Phi_p(\eta_{p-1}^N))(f_p^2) - \big(\Phi_p(\eta_{p-1}^N))(f_p)\big)^2] \\ &\leq \frac{1}{N}\Phi_p(\eta_{p-1}^N))(f_p^2) \\ &= \frac{1}{N}(G_{p-1}\cdot\eta_{p-1}^N)(K_p(f_p^2)) \ .\end{aligned} \tag{8.10}$$

We are now in a position to prove Lemma 5.2.

8.2 Proof of Lemma 5.2

Let us introduce the notation $G_{k,p} := Q_{k,p}(G_p)$ for any $p \geq k \geq 0$. Recalling (8.7) and notations (8.8) gives

$$
\begin{aligned}
\Gamma_k^N(G_{k,p}) &= \sum_{q=1}^{k} \gamma_{q-1}^N(G_{q-1})[\eta_q^N - \Phi_q(\eta_{q-1}^N)]\big(Q_{q,k}(G_{k,p})\big) \\
&= \gamma_{k-1}^N(G_{k-1})\Delta_k^N\big(Q_{k,k}(G_{k,p})\big) + \sum_{q=1}^{k-1} \gamma_{q-1}^N(G_{q-1})\Delta_q^N\big(Q_{q,k-1}(Q_{k-1,k}(G_{k,p}))\big) \\
&= \gamma_{k-1}^N(G_{k-1})\Delta_k^N(G_{k,p}) + \Gamma_{k-1}^N(G_{k-1,p}) \ .
\end{aligned}
$$

Using (8.9) stating that $\mathbb{E}[\Delta_k^N(G_{k,p}) \,|\, \mathcal{G}_{k-1}] = 0$ gives

$$
\mathbb{E}[\big(\Gamma_k^N(G_{k,p})\big)^2 \,|\, \mathcal{G}_{k-1}] = \big(\gamma_{k-1}^N(G_{k-1})\big)^2 \mathbb{E}[\big(\Delta_k^N(G_{k,p})\big)^2 \,|\, \mathcal{G}_{k-1}] + \big(\Gamma_{k-1}^N(G_{k-1,p})\big)^2 \ .
$$

Recalling assumption (5.24), observe that for any $x'_{k-1} \in E_{k-1}$

$$
\begin{aligned}
K_k(G_{k,p}^2)(x'_{k-1}) &= \mathbb{E}\big[(\mathbb{E}[G_{k:p}(X'_p)|X'_k])^2\big|X'_{k-1} = x'_{k-1}\big] \\
&\leq \mathbb{E}\big[G_{k:p}^2(X'_p)|X'_{k-1} = x'_{k-1}\big] \\
&\leq A^{p-k+1} < \infty \ .
\end{aligned}
$$

Then using the bound (8.10), with $p = k$ and $G_{k,p}$ as a test function implies

$$
\mathbb{E}[\big(\Delta_k^N(G_{k,p})\big)^2 \,|\, \mathcal{G}_{k-1}] \leq \frac{A^{p-k+1}}{N} \ .
$$

Using the above inequality and recalling that $\gamma_{k-1}^N(G_{k-1}) = \gamma_{k-1}(G_{k-1}) + \Gamma_{k-1}^N(G_{k-1})$ yields

$$
\mathbb{E}[\big(\Gamma_k^N(G_{k,p})\big)^2 \,|\, \mathcal{G}_{k-1}] \leq \frac{A^{p-k+1}}{N}[\big(\gamma_{k-1}(G_{k-1})\big)^2 + \big(\Gamma_{k-1}^N(G_{k-1})\big)^2] + \big(\Gamma_{k-1}^N(G_{k-1,p})\big)^2 \ .
$$

Again recall that by assumption (5.24)

$$
\gamma_{k-1}(G_{k-1}) := \mathbb{E}[G_{1:k-1}(X'_{k-1})] \leq (\mathbb{E}[G_{1:k-1}^2(X'_{k-1})])^{1/2} \leq A^k < \infty \ ,
$$

which finally yields

$$
\mathbb{E}[\big(\Gamma_k^N(G_{k,p})\big)^2] \leq \frac{A^{p+1}}{N}(1 + \mathbb{E}[\big(\Gamma_{k-1}^N(G_{k-1})\big)^2]) + \mathbb{E}[\big(\Gamma_{k-1}^N(G_{k-1,p})\big)^2] \ .
$$

Adding the above inequality from $k = 1$ to $k = p$ gives for any $p \leq n$

$$
\mathbb{E}[\big(\Gamma_p^N(G_p)\big)^2] \leq \frac{A^{p+1}}{N}\sum_{k=1}^{p}(1 + \mathbb{E}[\big(\Gamma_{k-1}^N(G_{k-1})\big)^2]) \ . \leq \frac{A^{n+1}}{N}\sum_{k=1}^{p}(1 + \mathbb{E}[\big(\Gamma_{k-1}^N(G_{k-1})\big)^2]) \ .
$$

We obtain by recursion

$$
\mathbb{E}[\big(\Gamma_p^N(G_p)\big)^2] \leq (1 + \frac{A^{n+1}}{N})^p - 1 \ . \tag{8.11}
$$

Now let us consider a test function f_n verifying assumption (5.26). Using again (8.9) stating that $\mathbb{E}[\Delta_k^N(G_{k,}) \,|\, \mathcal{G}_{k-1}] = 0$ gives

$$\mathbb{E}[\big(\Gamma_n^N(f_n)\big)^2] = \sum_{p=1}^n \mathbb{E}[\big(\gamma_{p-1}^N(G_{p-1})\big)^2 \big(\Delta_p^N(Q_{p,n}(f_n))\big)^2] \ .$$

By (8.10) and Assumption (5.26), we obtain $\mathbb{E}[\big(\Delta_p^N(Q_{p,n}(f_n))\big)^2 \,|\mathcal{G}_{p-1}] \leq B/N$ which yields

$$\begin{aligned}
\mathbb{E}[\big(\Gamma_n^N(f_n)\big)^2] &\leq \frac{B}{N}\sum_{p=1}^n \mathbb{E}[\big(\gamma_{p-1}^N(G_{p-1})\big)^2] \\
&\leq 2\frac{B}{N}\sum_{p=1}^n \Big(\big(\gamma_{p-1}(G_{p-1})\big)^2 + \mathbb{E}[\big(\Gamma_{p-1}^N(G_{p-1})\big)^2]\Big) \\
&\leq 2\frac{B}{N}\sum_{p=1}^n \Big(A^p + \mathbb{E}[\big(\Gamma_{p-1}^N(G_{p-1})\big)^2]\Big) \quad \text{since } (\gamma_{p-1}(G_{p-1}))^2 \leq A^p \ .
\end{aligned}$$

By (8.11) we finally get

$$\mathbb{E}[\big(\Gamma_n^N(f_n)\big)^2] \leq 2\frac{B}{N}\sum_{p=1}^n \Big(A^p + (1+\frac{A^{n+1}}{N})^{p-1} - 1\Big) \leq 2\frac{B}{N}\sum_{p=1}^n A^{p+1} \leq 2\frac{B}{N}A^{p+2} \ .$$

as soon as $N \geq A^{n+1}$ and $A \geq 2$.

9 Appendix: Technicalities related to the proof of Lemma 3.1

This section provides technical arguments allowing for differentiating under the integral sign that are necessary to prove the second and third identity of (3.1).

9.1 Concerning the second identity of (3.1)

Assume the first identity of (3.1) is verified. Let us introduce the real valued function such that for any $(s, \tilde{t}, \tilde{x}, t', x') \in [t', T] \times [0, T] \times \mathbb{R}^d \times [0, T] \times \mathbb{R}^d$

$$\phi^{\tilde{t},\tilde{x}}(s, t, x) := \mathbb{E}[h^{*,\tilde{t},\tilde{x}}(s, \tilde{X}_s^{\tilde{t},\tilde{x},t',x'})] \ , \tag{9.1}$$

where $(\tilde{X}_s^{\tilde{t},\tilde{x},t',x'})$ is the Gaussian process defined by (2.4) and $h^{*,\tilde{t},\tilde{x}}$ is the real valued function defined on $[0, T] \times \mathbb{R}^d$ by (2.5). Recalling identity (3.4) and using Fubini's lemma gives

$$u(t, x) = \mathbb{E}[g(\tilde{X}_T^{\tilde{t},\tilde{x},t,x})] + \int_t^T \phi^{\tilde{t},\tilde{x}}(s, t, x)\, ds \ . \tag{9.2}$$

Notice that by a simple application of Elworthy's formula [5] (which simply results here in the Likelihood ratio of Broadie and Glasserman [1]), we get

$$\begin{aligned}
D\phi^{\tilde{t},\tilde{x}}(s, t', x') &= \mathbb{E}[h^{*,\tilde{t},\tilde{x}}(s, \tilde{X}_s^{\tilde{t},\tilde{x},t',x'})\mathcal{M}_{t',s}^{\tilde{t},\tilde{x}}] \\
&= \int_{\mathbb{R}^d} \Big[h^{*,\tilde{t},\tilde{x}}(s, x' + b(\tilde{t},\tilde{x})(s-t') + \sqrt{s-t'}\sigma(\tilde{t},\tilde{x})u) \\
&\qquad \times \frac{(\sigma(\tilde{t},\tilde{x})^{-1})^T}{\sqrt{s-t'}} u p(u)\Big]\, du \ ,
\end{aligned} \tag{9.3}$$

where p denotes the centered and standard Gaussian density on $\mathbb{R}^d$ and $\mathcal{M}^{\tilde{t},\tilde{x}}_{t',s}$ is the Malliavin weight defined at (3.3). Recall that b and a are Lipschitz w.r.t. the space variable and 1/2-Hölder continuous w.r.t. the time variable as stated at item 3. and 4. of Assumption 1 and that Dv^* and D^2v^* are bounded as stated at item 1. of Assumption 1. Thus there exists a finite constant C depending on T and that may change from line to line such that

$$\|h^{*,\tilde{t},\tilde{x}}(s, x' + b(\tilde{t},\tilde{x})(s-t') + \sqrt{s-t'}\sigma(\tilde{t},\tilde{x})u)\| \quad \leq \quad C(1 + \|\tilde{x} - x'\| + \|b(\tilde{t},\tilde{x})\| + \|\sigma(\tilde{t},\tilde{x})\|\|u\|) \ .$$

Thus for any x' such that $\|x - x'\| \leq C$.

$$\|D\phi^{\tilde{t},\tilde{x}}(s,t',x')\| \quad \leq \quad C\frac{\|\sigma(\tilde{t},\tilde{x})^{-1}\|}{\sqrt{s-t'}}\Big(1 + \|b(\tilde{t},\tilde{x})\| + \|\sigma(\tilde{t},\tilde{x})\|\Big), \tag{9.4}$$

Since the term on the r.h.s of the above inequality is integrable w.r.t s on $[t',T]$ one can differentiate under the integral sign in (9.2) which ends the proof of the second identity of (3.1).

9.2 Concerning the third identity of (3.1)

Let us introduce the $\mathbb{R}^d$ valued function Ψ such that for any $(\tilde{t},\tilde{x},t',x') \in [0,T] \times \mathbb{R}^d \times [0,T] \times \mathbb{R}^d$

$$\Psi(\tilde{t},\tilde{x},t',x') := \int_t^T \psi(s,\tilde{t},\tilde{x},t',x')ds \tag{9.5}$$

where $\psi(s,\tilde{t},\tilde{x},t',x') := \psi_1(s,\tilde{t},\tilde{x},t',x') + \psi_2(s,\tilde{t},\tilde{x},t',x')$ with

$$\begin{aligned} \psi_1(s,\tilde{t},\tilde{x},t',x') &:= \tfrac{1}{T-t}\mathbb{E}[g(\tilde{X}^{\tilde{t},\tilde{x},t',x'}_T)\mathcal{M}^{\tilde{t},\tilde{x}}_{t',T}] \\ \psi_2(s,\tilde{t},\tilde{x},t',x') &:= \mathbb{E}[h^{\tilde{t},\tilde{x}}(s,\tilde{X}^{\tilde{t},\tilde{x},t',x'}_s)\mathcal{M}^{\tilde{t},\tilde{x}}_{t',s}] \ . \end{aligned} \tag{9.6}$$

Observe that $\psi = (\psi^1, \cdots, \psi^d)$ is differentiable w.r.t. the variable x' and for any $j = 1, \cdots ,, d$ and $i = 1, \cdots, d$

$$\frac{\partial \psi^j}{\partial x'_i}(s,\tilde{t},\tilde{x},t',x') = \frac{\partial \psi^j_1}{\partial x'_i}(s,\tilde{t},\tilde{x},t',x') + \frac{\partial \psi^j_2}{\partial x'_i}(s,\tilde{t},\tilde{x},t',x') \ ,$$

where

$$\begin{aligned} \tfrac{\partial \psi^j_1}{\partial x'_i}(s,\tilde{t},\tilde{x},t',x') &= \Big(\tfrac{1}{T-t}\mathbb{E}[g(\tilde{X}^{\tilde{t},\tilde{x},t',x'}_T)\mathcal{V}^{\tilde{t},\tilde{x}}_{t',T}]\Big)_{i,j} \\ \tfrac{\partial \psi^j_2}{\partial x'_i}(s,\tilde{t},\tilde{x},t',x') &= \Big(\mathbb{E}[h^{*,\tilde{t},\tilde{x}}(s,X^{\tilde{t},\tilde{x},t',x'}_s)\mathcal{V}^{\tilde{t},\tilde{x}}_{t',s}]\Big)_{i,j} \end{aligned} \tag{9.7}$$

Notice that Ψ does not depend on the pair $(\tilde{t},\tilde{x})$, hence one can fix $\tilde{t} = t' = t \in [0,T]$. Let us consider a fixed point $x \in \mathbb{R}^d$, two indexes $i,j \in \{1, \cdots d\}$, and a point $\tilde{x} \in \mathbb{R}^d$ such that each coordinate $\tilde{x}_\ell = x_\ell$ for any $\ell \neq i$ and $\tilde{x}_i$ is fixed at a given value. We want to prove that $\frac{\partial \Psi^j}{\partial x'_i}(t,\tilde{x},t,x)$ exists and is continuous (which implies the differentiability of Ψ) and to give an explicit expression for it. By the mean value theorem, there exists a real

$\theta^j(t, \tilde{x}_i, x, h) \in [-1, 1]$ (to simplify the notations we will forget the dependence on t) such that for any $i = 1, \cdots d$

$$\begin{aligned} \frac{1}{2h}[\Psi^j(t, \tilde{x}, t, x + he_i) - \Psi^j(t, \tilde{x}, t, x - he_i)] &= \int_t^T \frac{1}{2h}[\psi^j(s, t, \tilde{x}, t, x + he_i) - \psi^j(s, t, \tilde{x}, t, x - he_i)]\, ds \\ &= \int_t^T \frac{\partial \psi^j}{\partial x'}(t, \tilde{x}, t, x + \theta^j(\tilde{x}_i, x, h)he_i)\, ds \ , \end{aligned} \tag{9.8}$$

where e_i denotes the vector of $\mathbb{R}^d$ with zeros coordinates except for the i^{th} coordinate that equals 1. Consider the following equation w.r.t. the variable $\tilde{x}_i \in \mathbb{R}$

$$\lambda^{i,j}_{x,h}(\tilde{x}_i) = 0 \ , \quad \text{where} \quad \lambda^{i,j}_{x,h}(\tilde{x}_i) := x_i + \theta^j(\tilde{x}_i, x, h)h - \tilde{x}_i \ .$$

One can check that there exists a solution $\hat{x}_i(x, h)$ to this equation. Indeed, taking $\tilde{x}_i = x_i + h$ and $\tilde{x}_i = x_i - h$ and recalling that $\theta^j(\tilde{x}_i, x, h) \in [-1, 1]$, we obtain

$$\lambda^{i,j}_{x,h}(x_i + h) = h(\theta^j(x_i + h, x, h) - 1) \leq 0 \ , \quad \text{and} \quad \lambda^{i,j}_{x,h}(x_i - h) = h(\theta^j(x_i - h, x, h) + 1) \geq 0$$

which, by continuity of $\lambda^{i,j}_{x,h}$, implies the existence of a solution. Now we choose to take in equation (9.8), $\tilde{x}$ as the vector having the same coordinates as x except that $\tilde{x}_i = \hat{x}_i(x, h)$, this vector will be denoted by $\hat{x}(x, h)$. Observe that by construction choosing $\tilde{x} = \hat{x}(x, h)$ implies

$$\tilde{x} = x + \theta^j(\tilde{x}_i, x, h)he_i \ .$$

We are now interested in the limit of (9.8) as $h \to 0$. The technical point will consist in applying Lebesgue theorem to permute the limit with the integral sign. First, by Lebesgue theorem, and observing that $\hat{x}_i(x, h) \to x_i$ when $h \to 0$, we have

$$\begin{aligned} \lim_{h \to 0} \int_t^T \frac{\partial \psi^j_1}{\partial x'_i}(s, t, \hat{x}(x, h), t, \hat{x}(x, h))\, ds &= \int_t^T \lim_{h \to 0} \frac{\partial \psi^j_1}{\partial x'_i}(s, t, \hat{x}(x, h), t, \hat{x}(x, h))\, ds \\ &= \int_t^T \frac{\partial \psi^j_1}{\partial x'_i}(s, t, x, t, x)\, ds \ . \end{aligned}$$

Considering the integral term involving ψ^j_2, using the Lipschitz and Hölder properties of b and σ, we get

$$|h^{*,t,\hat{x}(x,h)}(s, \tilde{X}^{t,\hat{x}(x,h),t,\hat{x}(x,h)}_s)| \leq C(\hat{x}(x, h), t)\sqrt{s - t}$$

so

$$|\frac{\partial \psi^j_2}{\partial x'}(s, t, \hat{x}(x, h), t, \hat{x}(x, h))| \leq \frac{C(\hat{x}(x, h), t)}{\sqrt{s - t}}$$

where $C(\hat{x}(x, h), t)$ is locally bounded due to the non degeneracy hypothesis and the Lipschitz properties in assumption 1. The rhs of the previous equation is integrable so that we can use the Lebesgues Theorem,

$$\begin{aligned} \lim_{h \to 0} \int_t^T \frac{\partial \psi^j_2}{\partial x'_i}(s, t, \hat{x}(x, h), t, \hat{x}(x, h))\, ds &= \int_t^T \lim_{h \to 0} \frac{\partial \psi^j_2}{\partial x'_i}(s, t, \hat{x}(x, h), t, \hat{x}(x, h))\, ds \\ &= \int_t^T \frac{\partial \psi^j_2}{\partial x'_i}(s, t, x, t, x)\, ds \ . \end{aligned}$$

We finally obtain

$$\frac{\partial \Psi^j}{\partial x'_i}(\tilde{t}, \tilde{x}, t, x) = \left(\mathbb{E}[g(\tilde{X}_T^{t,x,t,x})\mathcal{V}_{t,T}^{t,x}]\right)_{i,j} + \int_t^T \left(\mathbb{E}[h^{*,t,x}(s, X_s^{t,x,t,x})\mathcal{V}_{t,s}^{t,x}]\right)_{i,j} ds$$

which ends the proof.

References

[1] M. Broadie and P. Glasserman. Estimating security prices using simulation. *Manag. Sc.*, 42:269–285, 1996.

[2] K. Burrage, P. M. Burrage, and Tian P. Numerical methods for strong solutions of stochastic differential equations: an overview. *Proceedings of the Royal Society of London*, Series 1 460:373–402, 2004.

[3] P. Del Moral. *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York, 2004. Genealogical and interacting particle systems with applications.

[4] P. Del Moral. *Mean field simulation for Monte Carlo integration*. Monographs on Statistics and applied Probability. Chapman and Hall, 2013.

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

[6] E. Fournié, J.M. Lasry, J. Lebuchoux, and Touzi N. Some applications of malliavin calculus to monte carlo methods in finance. *Finance and Stochastics*, 3:391–412, 1999.

[7] P. Henry-Labordère, X. Tan, and N. Touzi. Exact simulation of multi-dimensional stochastic differential equations. 2015.

[8] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[9] G. N. Milstein. Approximate integration of stochastic differential equations. *Theory of Probability & Its Applications*, 19(3):557–600, 1975.

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