

Exact simulation of multi-dimensional stochastic differential equations *

Pierre Henry-Labordère[†] Xiaolu Tan[‡] Nizar Touzi[§]

April 23, 2015

Abstract

We develop a weak exact simulation technique for a process X defined by a multi-dimensional stochastic differential equation (SDE). Namely, for a Lipschitz function g , we propose a simulation based approximation of the expectation $\mathbb{E}[g(X_{t_1}, \dots, X_{t_n})]$, which by-passes the discretization error. The main idea is to start instead from a well-chosen simulatable SDE whose coefficients are up-dated at independent exponential times. Such a simulatable process can be viewed as a regime-switching SDE, or as a branching diffusion process with one single living particle at all times. In order to compensate for the change of the coefficients of the SDE, our main representation result relies on the automatic differentiation technique induced by Elworthy's formula from Malliavin calculus, as exploited by Fournié et al. [10] for the simulation of the Greeks in financial applications.

Unlike the exact simulation algorithm of Beskos and Roberts [3], our algorithm is suitable for the multi-dimensional case. Moreover, its implementation is a straightforward combination of the standard discretization techniques and the above mentioned automatic differentiation method.

Key words. Exact simulation of SDEs, regime switching diffusion, linear parabolic PDEs.

1 Introduction

Let $d \geq 1$, $T > 0$ and W be a d -dimensional Brownian motion, $\mu : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{S}^d$ be the drift and diffusion coefficients, where \mathbb{S}^d denotes the collection of all $d \times d$ dimensional matrices. Under standard assumptions on these

*X. Tan and N. Touzi gratefully acknowledge the financial support of the ERC 321111 Rofirm, the ANR Isotace, and the Chairs Financial Risks (Risk Foundation, sponsored by Société Générale) and Finance and Sustainable Development (IEF sponsored by EDF and CA).

[†]Société Générale, Global Market Quantitative Research, pierre.henry-labordere@sgcib.com

[‡]CEREMADE, University of Paris-Dauphine, tan@ceremade.dauphine.fr

[§]Ecole Polytechnique Paris, Centre de Mathématiques Appliquées, nizar.touzi@polytechnique.edu

coefficients, we introduce the process X defined as the unique strong solutions of the multi-dimensional SDE,

$$X_0 = x_0, \quad \text{and} \quad dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t,$$

Our main interest in this paper is on the Monte-Carlo approximation of the expectation

$$V_0 := \mathbb{E}[g(X_T)], \quad (1.1)$$

for some function $g : \mathbb{R}^d \rightarrow \mathbb{R}$. In the standard literature, see e.g. Kloeden and Platen [14], such approximations are based on the discretization of the SDE, thus inducing a discrete-time approximation error with magnitude depending on the order of the scheme. The error estimate of the Monte-Carlo approximation results from the combination of the discretization error and the statistical error. Consequently, optimizing the overall computational effort leads typically to an error with rate strictly smaller than the rate of the statistical error.

In order to restore the rate of the error to the rate of the statistical error, one needs to by-pass the discretization error. The first attempt was achieved by Beskos and Roberts [3], and Beskos, Papaspiliopoulos and Roberts [4] in the context one-dimensional homogeneous SDEs. Their method first reduces the SDE to the constant diffusion case by the so-called Lamperti's transformation. Next, they use the Girsanov measure change theorem to remove the drift term. In order to compensate for the change of drift, they propose a (time-consuming) rejection method to simulate the corresponding Radon-Nikodym derivative. We also refer to Jourdain and Sbail [13] for an extension to functionals depending on the arithmetic average of the diffusion process.

In this paper, we introduce an exact simulation method, of completely different nature than [3], which allows for possibly multi-dimensional SDEs with time dependent coefficients. More extensions to the path-dependent case are also explored in the last part of the paper.

The main idea is to start instead from a well-chosen simulatable SDE

$$\hat{X}_0 = x_0, \quad \text{and} \quad d\hat{X}_t = \hat{\mu}(t, \hat{X}_t) dt + \hat{\sigma}(t, \hat{X}_t) dW_t,$$

with coefficient functions $\hat{\mu}$ and $\hat{\sigma}$ which are updated at independent exponential times. Such a process can be viewed as a regime-switching SDE, or as a branching diffusion process with one single living particle at all times, and is chosen so as to be exactly simulatable, i.e. without discretization error. In order to compensate for the change of the coefficients of the SDE, our main representation result relies on the automatic differentiation technique induced by Elworthy's formula from Malliavin calculus, as exploited by Fournié et al. [10] for the simulation of the Greeks in financial applications. This leads to a representation formula in the spirit of that derived by Bally and Kohatsu-Higa [2, Section 6.1] as an application of their parametrix method for SDEs.

However, an arbitrary choice of $\hat{\mu}$ and $\hat{\sigma}$ leads in general to a problem of simulation of a random variable with infinite variance and even non-integrable. This was also observed in [2]. As a second main contribution, our choice of the coefficients $\hat{\mu}$ and $\hat{\sigma}$

is designed so that the induced representation involves a random variable with finite variance. Consequently, the error of approximation of the corresponding Monte Carlo approximation results from the classical central limit theorem.

The idea of using a branching diffusion representation for a class of semilinear PDEs for the purpose of numerical approximation was introduced in [11, 12]. On one hand, the present setting is simpler as it involves one single living particle at each point in time. However, the correction for the replaced coefficients involves the gradient and the Hessian of the value function, a feature which was avoided in [11, 12] by restricting the class of semilinear PDEs. This major difference is solved in the present paper by the Monte Carlo automatic differentiation technique.

The rest of the paper is organized as follows. In Section 2, we provide a general result on the regime switching diffusion representation of the value V_0 defined in (1.1), by considering a SDE with replaced coefficients. Then in Section 3, we restrict to the constant diffusion coefficient case. With a good choice of $\hat{\mu}$ and $\hat{\sigma}$ as well as the Malliavin weight, we obtain an exact simulation estimator of finite variance. In Section 4, we consider a one-dimensional SDE with a general diffusion coefficient but zero drift, and also obtain an exact simulation estimator of finite variance. Some numerical examples are contained in Section 5. Finally, Section 6 provides further discussions to more general multi-dimensional SDEs, and extensions to the path-dependent case.

2 Regime switching diffusion representation

Let $d \geq 1$, $T > 0$, W be a standard d -dimensional Brownian motion, μ, σ be bounded continuous functions from $[0, T] \times \mathbb{R}^d$ to \mathbb{R}^d and \mathbb{S}^d respectively, satisfying

$$|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq L|x - y|; \quad (t, x, y) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d, \quad (2.1)$$

for some constant $L > 0$. For $(t, x) \in [0, T] \times \mathbb{R}^d$, we denote by $(X_s^{t,x})_{t \leq s \leq T}$ the unique strong solution of the SDE

$$X_t = x, \quad \text{and} \quad dX_s = \mu(s, X_s) ds + \sigma(s, X_s) dW_s, \quad s \geq t. \quad (2.2)$$

Let $x_0 \in \mathbb{R}^d$ and $g : \mathbb{R}^d \rightarrow \mathbb{R}$ be some bounded continuous function, we define

$$u(t, x) := \mathbb{E}[g(X_T^{t,x})] \quad \text{and} \quad V_0 := u(0, x_0). \quad (2.3)$$

By the Itô formula, the function $u(t, x)$ can be characterized by means of the linear parabolic PDE

$$\partial_t u + \mu \cdot Du + a : D^2 u = 0, \quad \text{on} \quad [0, T] \times \mathbb{R}^d, \quad (2.4)$$

with terminal condition $u(T, x) = g(x)$, where $a(\cdot) := \frac{1}{2}\sigma\sigma^T(\cdot)$, and $A : B := \text{Tr}(AB^T)$ for any two $d \times d$ dimensional matrices $A, B \in \mathbb{S}^d$, and D, D^2 denote the gradient and Hessian operators with respect to the space variable x .

2.1 Regime switching diffusion representation

For general coefficient functions μ and σ , an exact simulation of X is a difficult task. Our main idea is to simulate another SDE with some other coefficient functions $\hat{\mu}$ and $\hat{\sigma}$ along some exponential times, and then to correct the induced error using some weight functions.

Let $\beta > 0$ be a fixed positive constant, $(\tau_i)_{i>0}$ be a sequence of i.i.d. $\mathcal{E}(\beta)$ -exponential random variables, which is independent of the Brownian motion W . We define

$$T_k := \left(\sum_{i=1}^k \tau_i \right) \wedge T, \quad k \geq 0, \quad \text{and} \quad N_t := \max \{k : T_k < t\}. \quad (2.5)$$

Then $(N_t)_{0 \leq t \leq T}$ is a Poisson process with intensity β and arrival times $(T_k)_{k>0}$, and $T_0 = 0$.

Next, let $(\hat{\mu}, \hat{\sigma}) : (s, y, t, x) \in [0, T] \times \mathbb{R}^d \times [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{S}^d$ be continuous in t and Lipschitz in x . The starting point for our exact discretization method is the process \hat{X} defined by

$$\hat{X}_0 := x_0 \quad \text{and} \quad d\hat{X}_t = \hat{\mu}(\Theta_t, t, \hat{X}_t)dt + \hat{\sigma}(\Theta_t, t, \hat{X}_t)dW_t, \quad (2.6)$$

with $\Theta_t := (T_{N_t}, \hat{X}_{T_{N_t}})$. In other words, the process \hat{X} is defined recursively by, $\hat{X}_0 = x_0$ and for all $k \geq 0$,

$$\hat{X}_{T_{k+1}} = \hat{X}_{T_k} + \int_{T_k}^{T_{k+1}} \hat{\mu}(T_k, X_{T_k}, s, \hat{X}_s)ds + \int_{T_k}^{T_{k+1}} \hat{\sigma}(T_k, X_{T_k}, s, \hat{X}_s)dW_s.$$

We also introduce, for all $k > 0$, $\Delta W_t^k := W_{(T_{k-1}+t) \wedge T_k} - W_{T_{k-1}}$. It is clear that the sequence of processes $(\Delta W_t^k)_{k>0}$ are mutually independent.

Notice that the above system is defined with initial data $(0, x_0)$ and $\theta_0 = (0, x_0)$ on the time horizon $[0, T]$. Similarly, we can also define the system with other initial data. For $t \in [0, T]$, we denote $T_k^t := (t + \sum_{i=1}^k \tau_i) \wedge T$, $k \geq 0$, and $(N_s^t)_{t \leq s \leq T}$ the corresponding shifted Poisson process. We also introduce the increments of the Brownian motion $\Delta W_t^{k,t}$ along $(T_k^t)_{k \geq 0}$ defined as above.

For $(t, x, \theta) \in [0, T] \times \mathbb{R}^d \times [0, T] \times \mathbb{R}^d$, we define the process $(\hat{X}_s^{t,x,\theta})_{t \leq s \leq T}$ as the unique strong solution of

$$\hat{X}_t^{t,x,\theta} := x, \quad d\hat{X}_s^{t,x,\theta} = \hat{\mu}(\Theta_s^{t,x,\theta}, s, \hat{X}_s^{t,x,\theta})ds + \hat{\sigma}(\Theta_s^{t,x,\theta}, s, \hat{X}_s^{t,x,\theta})dW_s. \quad (2.7)$$

with $\Theta_s^{t,x,\theta} := \theta$ for $s \in [t, T_1^t]$ and $\Theta_s^{t,x,\theta} := (T_{N_s^t}^t, \hat{X}_{T_{N_s^t}^t}^{t,x,\theta})$ for $s \in (T_1^t, T]$.

For the sake of simplicity, we also denote

$$\hat{X}^{t,x} := \hat{X}^{t,x,(t,x)}, \quad \text{and} \quad \hat{X}^{t,x,y} := \hat{X}^{t,x,(t,y)}, \quad \text{for all } y \in \mathbb{R}^d.$$

We first formulate an assumption on the existence of Malliavin weights associated to SDE (2.6).

Assumption 2.1. For all $\theta \in [0, T) \times \mathbb{R}^d$, and $(t, x) \in [0, T) \times \mathbb{R}^d$, there is a pair of random functions $(\widehat{\mathcal{W}}_\theta^1(\cdot), \widehat{\mathcal{W}}_\theta^2(\cdot))$, called Malliavin weights, depending only on $(t, x, T_1^t, \Delta W^{1,t})$ and taking value in $\mathbb{R}^d \times \mathbb{S}^d$, such that

$$D^i \mathbb{E}[\phi(T_1^t, \widehat{X}_{T_1^t}^{t,x,\theta})] = \mathbb{E}[\phi(T_1^t, \widehat{X}_{T_1^t}^{t,x,\theta}) \widehat{\mathcal{W}}_\theta^i(t, x, T_1^t, \Delta W^{1,t})], \quad i = 1, 2,$$

where D, D^2 denote the gradient and Hessian operators with respect to the variable x , and $\phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is an arbitrary bounded measurable function stratifying $x \mapsto \phi(t, x)$ is continuous for each $t \in [0, T]$.

Let $a(\cdot) := \frac{1}{2} \sigma \sigma^T(\cdot)$ and $\hat{a}(\cdot) := \frac{1}{2} \hat{\sigma} \hat{\sigma}^T(\cdot)$. For $(t, x, y) \in [0, T) \times \mathbb{R}^d \times \mathbb{R}^d$, we denote

$$\widehat{\Theta}_0^{t,x,y} = (t, y), \quad \text{and then} \quad \widehat{\Theta}_k^{t,x,y} = (T_k^t, \widehat{X}_{T_k^t}^{t,x,y}), \quad \text{for } k > 0.$$

and for $k > 0$,

$$\begin{aligned} \Delta f_k^{t,x,y} &:= (\mu, a)(T_k^t, \widehat{X}_{T_k^t}^{t,x,y}) - (\hat{\mu}, \hat{a})(\widehat{\Theta}_{k-1}^{t,x,y}, T_k^t, \widehat{X}_{T_k^t}^{t,x,y}) \in \mathbb{R}^d \times \mathbb{S}^d, \\ \widehat{\mathcal{W}}_k^{t,x,y} &:= (\widehat{\mathcal{W}}_{\widehat{\Theta}_k^{t,x,y}}^1, \widehat{\mathcal{W}}_{\widehat{\Theta}_k^{t,x,y}}^2)(T_k^t, \widehat{X}_{T_k^t}^{t,x,y}, T_{k+1}^t, \Delta W^{k+1,t}) \in \mathbb{R}^d \times \mathbb{S}^d, \end{aligned}$$

with the weight functions $(\widehat{\mathcal{W}}_\theta^1(\cdot), \widehat{\mathcal{W}}_\theta^2(\cdot))$ given in Assumption 2.1. We then define

$$\widehat{\psi}^{t,x,y} := e^{\beta(T-t)} \left(g(\widehat{X}_T^{t,x,y}) - g(\widehat{X}_{T_{N_T^t}}^{t,x,y}) \mathbf{1}_{\{N_T^t > 0\}} \right) \beta^{-N_T^t} \prod_{k=1}^{N_T^t} (\Delta f_k^{t,x,y} \bullet \widehat{\mathcal{W}}_k^{t,x,y}), \quad (2.8)$$

where $(p, P) \bullet (q, Q) := p \cdot q + P : Q$ for all $p, q \in \mathbb{R}^d, P, Q \in \mathbb{S}^d$. Here we use the convention $\prod_{k=1}^0 = 1$.

Assumption 2.2. For all initial data $(t, x) \in [0, T) \times \mathbb{R}^d$, there is a neighborhood A_x of x such that for all $y \in A_x$, the sequence

$$\left(|g(\widehat{X}_T^{t,x,y}) - g(\widehat{X}_{T_{N_T^t}}^{t,x,y}) \mathbf{1}_{\{N_T^t > 0\}}| \mathbf{1}_{\{N_T^t \leq n\}} + |\Delta f_{n+1}^{t,x,y}| \mathbf{1}_{\{N_T^t > n\}} \right) \prod_{k=1}^{n \wedge N_T^t} |\beta^{-1} \Delta f_k^{t,x,y} \bullet \widehat{\mathcal{W}}_k^{t,x,y}|,$$

$n > 0$, is uniformly integrable.

The following result provides our main alternative representation of the function u introduced in (2.3).

Theorem 2.3. Let Assumptions 2.1 and 2.2 hold true, and suppose that $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$. Then for all $(t, x) \in [0, T] \times \mathbb{R}^d$, $\widehat{\psi}^{t,x,y}$ is integrable and $u(t, x) = \mathbb{E}[\widehat{\psi}^{t,x,y}]$ for all $y \in A_x$, where A_x is a neighborhood of x in Assumption 2.2.

Our interest in this representation is that it derives a weak exact simulation scheme for the solution of stochastic differential equations, whenever the regime switching diffusion $\widehat{X}^{t,x,y}$ and the corresponding Malliavin weights $\widehat{\mathcal{W}}_k^{t,x,y}$ can be exactly simulated. This will be developed in the subsequent sections.

Remark 2.4. (i) By the Feynmann-Kac formula, the condition $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ implies that u is the unique classical solution of PDE (2.4).

(ii) The condition that $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ may be relaxed in the concrete applications of Theorem 2.3. This will be indeed performed in Sections 3 and 4 by exploiting the integrability of the Malliavin weights $(\widehat{\mathcal{W}}_\theta^1, \widehat{\mathcal{W}}_\theta^2)$ of Assumption 2.1.

(iii) In the following sections, we will discuss how to choose $\hat{\mu}_k$ and $\hat{\sigma}_k$ and then compute the weight functions $(\widehat{\mathcal{W}}_\theta^1, \widehat{\mathcal{W}}_\theta^2)$ in different cases, so as to ensure that Assumptions 2.1 and 2.2 are satisfied.

(iv) By definition, the Malliavin weight satisfies $\mathbb{E}[\widehat{\mathcal{W}}_k^{t,x,y}] = 0$, then the estimator $\widehat{\psi}^{t,x,y}$ in (2.8) is equivalent to the estimator

$$e^{\beta(T-t)} g(\widehat{X}_T^{t,x,y}) \beta^{-N_T^t} \prod_{k=1}^{N_T^t} (\Delta f_k^{t,x,y} \bullet \widehat{\mathcal{W}}_k^{t,x,y}).$$

However, in practice, the weight function $\widehat{\mathcal{W}}_k^{t,x,y}$ is typically of infinity variance, or even not integrable, in general. Indeed, as we will see in the following sections, $\widehat{\mathcal{W}}_k^{t,x,y}$ is generally of order $\frac{1}{\Delta T_{k+1}^t} = \frac{1}{T_{k+1}^t - T_k^t}$, where conditioning on $N_T^t = n$, $(T_1^t, \dots, T_{N_T^t}^t)$ follows the law of statistic order of uniform distribution on $[t, T]$. Then by direct computation, one knows $\mathbb{E}[1/\Delta T_{N_T^t+1}^t] = \infty$. In the definition of $\widehat{\psi}^{t,x,y}$ in (2.8), the additional term $-g(\widehat{X}_{T_{N_T^t}^t}^{t,x,y}) \mathbf{1}_{\{N_T^t > 0\}}$ can be seen as a control variate so as to guarantee the integrability of $\widehat{\psi}^{t,x,y}$.

2.2 A general error analysis

To solve problem V_0 in (2.3) by Monte-Carlo methods, there are generally two kinds of errors. The first is the discretization error when one uses a discretization method to simulate SDE (2.2), which depends on the time discretization size. The second is the statistical error when one estimates the expectation value by the empirical mean value of a large number of samples. By the central limit theorem, the statistical error depends on the variance of the Monte-Carlo estimator.

Let us first consider the discretization scheme. Suppose that the discretization error, with time step $\Delta t := \frac{T}{n}$ (i.e. n steps on $[0, T]$), is given by $C_0 \Delta t^\rho$ for some $\rho > 0$. Suppose in addition that the variance of the Monte-Carlo estimator is given by C_1 . Notice also that, given S sample paths, the computation effort is $M = CSn$ for some constant C . Then, it follows from the central limit theorem that the global error is given by $C_0 \Delta t^\rho + \frac{\sqrt{C_1 C}}{\sqrt{M/n}}$. We can now optimize the choice of the number of steps n in terms of the effort M :

$$E^D = \min_{n>0} \left(C_0 \left(\frac{T}{n} \right)^\rho + \frac{\sqrt{C_1 C}}{\sqrt{M/n}} \right) = C_2 M^{-\frac{\rho}{1+2\rho}}, \quad (2.9)$$

for some constant $C_2 > 0$.

For our exact simulation algorithm, there is no time discretization, then the computation effort M is proportional to the number of samples. Suppose that the above

exact simulation estimator (2.8) admits a variance C_3 , we obtain a global error

$$\mathbb{E}^{ES} = \sqrt{C_3} M^{-\frac{1}{2}}. \quad (2.10)$$

Generally speaking, the above estimator (2.8) uses additional randomness of exponential time $(\tau_i)_{i>0}$, which leads to a bigger variance of the estimator, i.e. $C_3 > C_1$. However, the exact simulation method will always be more interesting by comparing the order of the global error in (2.9) and (2.10).

2.3 Proof of Theorem 2.3

In preparation of the proof of Theorem 2.3, let us provide a technical lemma.

Lemma 2.5. *Assume $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$. Let $\beta > 0$, $\theta \in [0, T) \times \mathbb{R}^d$, $(t, x) \in [0, T) \times \mathbb{R}^d$, and $(\hat{X}^{t,x,\theta})$ be defined by SDE (2.7). Then*

$$\begin{aligned} u(t, x) = & \mathbb{E} \left[e^{\beta(T_1^t - t)} \left(g(\hat{X}_T^{t,x,\theta}) \mathbf{1}_{\{N_T^t = 0\}} \right. \right. \\ & \left. \left. + \beta^{-1} \Delta f_1^{t,x,\theta} \bullet (Du, D^2u)(T_1^t, \hat{X}_{T_1^t}^{t,x,\theta}) \mathbf{1}_{\{N_T^t > 0\}} \right) \right], \end{aligned} \quad (2.11)$$

where

$$\Delta f_1^{t,x,\theta} := (\mu, a)(T_1^t, \hat{X}_{T_1^t}^{t,x,\theta}) - (\hat{\mu}, \hat{a})(\theta, T_1^t, \hat{X}_{T_1^t}^{t,x,\theta}).$$

Proof. Let us denote the r.h.s. of (2.11) by $v(t, x)$, then $v(t, x)$ is a bounded continuous function since g , $\Delta f_1^{t,x,\theta}$ and (Du, D^2u) are all uniformly bounded and continuous.

(i) Let $0 < h < T - t$ be small enough, denote $T_h^t := T_1^t \wedge (t + h) = t + (\tau_1 \wedge h)$ and define a σ -field $\mathcal{F}_{T_h^t} := \sigma(T_h^t, W_r, r \leq T_h^t)$. Taking conditional expectation of the r.h.s. of (2.11) w.r.t. $\mathcal{F}_{T_h^t}$, and using Bayes formula, it follows that

$$\begin{aligned} v(t, x) = & \mathbb{E} \left[e^{\beta h} v(t + h, \hat{X}_{t+h}^{t,x,\theta}) \middle| \tau_1 \geq h \right] \mathbb{P}[\tau_1 \geq h] \\ & + \mathbb{E} \left[e^{\beta \tau_1} \beta^{-1} \Delta f_1^{t,x,\theta} \bullet (Du, D^2u)(t + \tau_1, \hat{X}_{t+\tau_1}^{t,x,\theta}) \middle| \tau_1 < h \right] \mathbb{P}[\tau_1 < h]. \end{aligned} \quad (2.12)$$

Notice that $\mathbb{P}[\tau_1 \geq h] = e^{-\beta h}$, $\mathbb{P}[\tau_1 < h] = 1 - e^{-\beta h} \approx \beta h$ when h is small, and the exponential law is memoryless. Taking $v(t, x)$ to the r.h.s. of (2.12), then dividing it by h , and then sending $h \rightarrow 0$, it follows by standard arguments that v is a bounded viscosity solution to PDE on $[0, T] \times \mathbb{R}^d$:

$$-\partial_t v - \hat{\mu}_\theta \cdot Dv - \hat{a}_\theta : D^2v - ((\mu - \hat{\mu}_\theta) \cdot Du + (a - \hat{a}_\theta) : D^2u) = 0, \quad (2.13)$$

with terminal condition $v(T, x) = g(x)$, where $(\hat{\mu}_\theta, \hat{a}_\theta)(\cdot) = (\hat{\mu}, \hat{a})(\theta, \cdot)$.

Clearly, u is a classical solution of PDE (2.13) with the same terminal condition. In the next step, we show that $v = u$ by a standard uniqueness argument, which concludes the proof.

(ii) The following partial comparison principle of PDE (2.13) is reported for completeness. By a variable change argument, it is equivalent to consider PDE, with $\beta > 0$,

$$\beta v - \partial_t v - \hat{\mu}_\theta \cdot Dv - \hat{a}_\theta : D^2v - ((\mu - \hat{\mu}_\theta) \cdot Du + (a - \hat{a}_\theta) : D^2u) = 0, \quad (2.14)$$

on $[0, T] \times \mathbb{R}^d$. Let v be a bounded viscosity super-solution of (2.14) and \bar{v} be a bounded classical sub-solution of (2.14) such that $v(T, \cdot) \geq \bar{v}(T, \cdot)$. We will prove that $v \geq \bar{v}$ on $[0, T] \times \mathbb{R}^d$ by contradiction.

Suppose that $\delta := (\bar{v} - v)(\bar{t}, \bar{x}) > 0$, for some points $(\bar{t}, \bar{x}) \in [0, T) \times \mathbb{R}^d$, and let $\varepsilon > 0$. By the boundedness of \bar{v} and v , we may find $(t_\varepsilon, x_\varepsilon) \in [0, T] \times \mathbb{R}^d$ such that

$$0 < \delta < \max_{(t,x) \in [0,T] \times \mathbb{R}^d} \left((\bar{v} - v)(t, x) - \frac{\varepsilon}{2} |x - \bar{x}|^2 \right) = (\bar{v} - v)(t_\varepsilon, x_\varepsilon) - \frac{\varepsilon}{2} |x_\varepsilon - \bar{x}|^2.$$

Notice that $\varepsilon |x_\varepsilon - \bar{x}|^2$ is uniformly bounded (in fact we may prove that $\varepsilon |x_\varepsilon - \bar{x}|^2 \rightarrow 0$), and therefore $\varepsilon (x_\varepsilon - \bar{x}) \rightarrow 0$ as $\varepsilon \rightarrow 0$. Further, since v is viscosity super-solution and \bar{v} is a smooth function which can serve as a test function at $(t_\varepsilon, x_\varepsilon)$, it follows that

$$\begin{aligned} 0 \leq & \left(\beta v - \partial_t \bar{v} - \hat{\mu}_\theta \cdot (D\bar{v} - \varepsilon(x_\varepsilon - y)) - \hat{a}_\theta : (D^2 \bar{v} - \varepsilon I_d) \right. \\ & \left. - ((\mu - \hat{\mu}_\theta) \cdot D\bar{v} + (a - \hat{a}_\theta) : D^2 \bar{v}) \right)(t_\varepsilon, x_\varepsilon). \end{aligned}$$

Since \bar{v} is a classical sub-solution of (2.14), this provides

$$0 \leq \beta(v - \bar{v})(s_\varepsilon, x_\varepsilon) + \varepsilon(\hat{\mu} \cdot (x_\varepsilon - \bar{x}) + \hat{a} : I_d) < -\beta\delta + \varepsilon(\hat{\mu} \cdot (x_\varepsilon - \bar{x}) + \hat{a} : I_d),$$

which is a contradiction to the fact that $\beta\delta > 0$ and $\varepsilon(x_\varepsilon - \bar{x}) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

Similarly, let v be a bounded viscosity sub-solution of (2.14) and \bar{v} be a bounded classical super-solution of (2.14) such that $v(T, \cdot) \leq \bar{v}(T, \cdot)$. It follows by the same arguments that $v \leq \bar{v}$ on $[0, T] \times \mathbb{R}^d$. \square

Proof of Theorem 2.3. (i) Let $(t, x) \in [0, T) \times \mathbb{R}^d$, $y \in A_x$, for all $n \geq 0$, we define

$$\begin{aligned} \hat{\psi}_n^{t,x,y} := & e^{\beta(T_n^t - t)} \left[\left(g(\hat{X}_T^{t,x,y}) - g(\hat{X}_{T_{N_T^t}^t}^{t,x,y}) \mathbf{1}_{\{N_T^t > 0\}} \right) \mathbf{1}_{\{N_T^t \leq n\}} \right. \\ & \left. + \beta^{-1} \left(\Delta f_{n+1}^{t,x,y} \bullet (Du, D^2 u)(T_{n+1}^t, \hat{X}_{T_{n+1}^t}^{t,x,y}) \right) \mathbf{1}_{\{N_T^t > n\}} \right] \prod_{k=1}^{n \wedge N_T^t} (\beta^{-1} \Delta f_k^{t,x,y} \bullet \widehat{\mathcal{W}}_k^{t,x,y}) \end{aligned}$$

We shall prove in the next step that $u(t, x) = \mathbb{E}[\hat{\psi}_n^{t,x,y}]$ for all $n \geq 0$. Since $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$, it follows from Assumption 2.2 that the sequence $(\hat{\psi}_n^{t,x,y})_{n \geq 0}$ is uniformly integrable. Then letting $n \rightarrow \infty$, we obtain

$$u(t, x) = \lim_{n \rightarrow \infty} \mathbb{E}[\hat{\psi}_n^{t,x,y}] = \mathbb{E}[\lim_{n \rightarrow \infty} \hat{\psi}_n^{t,x,y}] = \mathbb{E}[\hat{\psi}^{t,x,y}].$$

(ii) To conclude the proof, we now prove by induction that $u(t, x) = \mathbb{E}[\hat{\psi}_n^{t,x,y}]$ for all $n \geq 0$. First, the equality is true for $n = 0$ by Lemma 2.5.

Next, admitting that the claim holds true for some $n \geq 0$, we consider the case $n + 1$. Notice that by Assumption 2.1, we have

$$\mathbb{E} \left[e^{\beta(T_1^t - t)} \mathbf{1}_{\{N_T^t = 0\}} \widehat{\mathcal{W}}_{(t,y)}^i(t, x, T_1^t, \Delta W^{1,t}) \right] = 0.$$

Then by considering the conditional expectation $\mathbb{E}[\widehat{\psi}_n^{t,x,y}|T_1, \widehat{X}_{T_1^t}^{t,x,y}]$, and using again Assumption 2.1, it follows that for all c in a neighborhood of x , and $i = 1, 2$,

$$\begin{aligned} D^i u(t, x) &= \mathbb{E} \left[\mathbb{E} \left[\widehat{\psi}_n^{t,x,y} \middle| T_1, \widehat{X}_{T_1^t}^{t,x,y} \right] \widehat{\mathcal{W}}_{(t,y)}^i(t, x, T_1^t, \Delta W^{1,t}) \right] \\ &= \mathbb{E} \left[\widehat{\psi}_n^{t,x,y} \widehat{\mathcal{W}}_{(t,y)}^i(t, x, T_1^t, \Delta W^{1,t}) \right] \\ &= \mathbb{E} \left[\left(\widehat{\psi}_n^{t,x,y} - e^{\beta(T_1^t - t)} g(x) \mathbf{1}_{\{N_T^t = 0\}} \right) \widehat{\mathcal{W}}_{(t,y)}^i(t, x, T_1^t, \Delta W^{1,t}) \right]. \end{aligned}$$

Setting $y = x$ and inserting the above representation of $D^i u(t, x)$ in Lemma 2.5, it follows by straightforward application of the tower property that $u(t, x) = \mathbb{E}[\widehat{\psi}_{n+1}^{t,x,y}]$, which concludes the proof. \square

3 The constant diffusion coefficient case

In this section, we restrict to the constant diffusion coefficient case,

$$X_0 = x_0, \quad dX_t = \mu(t, X_t) dt + \sigma_0 dW_t, \quad (3.1)$$

for some non-degenerate matrix $\sigma_0 \in \mathbb{S}^d$, the objective is to compute

$$V_0 = \mathbb{E}[g(X_T)].$$

for some Lipschitz function $g : \mathbb{R}^d \rightarrow \mathbb{R}$. We will discuss how to choose $\hat{\mu}(\cdot)$ and $\hat{\sigma}(\cdot)$, and then how to compute the associated Malliavin weight function $(\widehat{\mathcal{W}}_\theta^1, \widehat{\mathcal{W}}_\theta^2)$, to ensure the conditions in Theorem 2.3.

Assumption 3.1. *The drift function $\mu(t, x)$ is bounded continuous in (t, x) , uniformly $\frac{1}{2}$ -Hölder in t and uniformly Lipschitz in x , i.e. for some constant $L > 0$,*

$$\left| \mu(t, x) - \mu(s, y) \right| \leq L \left(\sqrt{|t - s|} + |x - y| \right), \quad \forall (s, x), (t, y) \in [0, T] \times \mathbb{R}^d. \quad (3.2)$$

3.1 The algorithm

Recall that the random variable N_T and the sequence $(T_k)_{k=1, \dots, N_T+1}$ are defined by (2.5) from a sequence of i.i.d. $\mathcal{E}(\beta)$ -exponential random variables $(\tau_i)_{i>0}$, which is independent of the Brownian motion W . For simplicity, denote

$$\Delta W_{T_k} := \Delta W_{\Delta T_k}^k = W_{T_k} - W_{T_{k-1}}, \quad k > 0.$$

In this simplified context, we propose to choose

$$\hat{\mu}(s, y, t, x) = \mu(s, y) \quad \text{and} \quad \hat{\sigma}(\cdot) \equiv \sigma_0, \quad (3.3)$$

so that the process \widehat{X} in (2.6) can be given by $\widehat{X}_0 = x_0$ and

$$\widehat{X}_{T_{k+1}} := \widehat{X}_{T_k} + \mu(T_k, \widehat{X}_{T_k}) \Delta T_{k+1} + \sigma_0 \Delta W_{T_{k+1}}, \quad k = 0, 1, \dots, N_T.$$

In the present case, the increment $\widehat{X}_{T_{k+1}} - \widehat{X}_{T_k}$, conditional on (T_k, \widehat{X}_{T_k}) , is Gaussian. Then, we may provide Malliavin weights by direct integration by parts using the explicit gaussian density. This is the so-called likelihood ratio method in Broadie and Glasserman [5]. In the multi-dimensional case a possible choice of the Malliavin weights is:

$$\widehat{\mathcal{W}}_\theta^1(\cdot, \delta t, \delta w) := (\sigma_0^T)^{-1} \frac{\delta w}{\delta t} \quad \text{and} \quad \widehat{\mathcal{W}}_\theta^2(\cdot, \delta t, \delta w) := (\sigma_0^T)^{-1} \frac{\delta w \delta w^T - \delta t I_d}{\delta t^2} \sigma_0^{-1}. \quad (3.4)$$

Notice that the last Malliavin weights satisfy Assumption 2.1. Then our estimator is given by

$$\widehat{\psi} := e^{\beta T} \left[g(\widehat{X}_T) - g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right] \beta^{-N_T} \prod_{k=1}^{N_T} \overline{\mathcal{W}}_k^1, \quad (3.5)$$

with

$$\overline{\mathcal{W}}_k^1 := \frac{(\mu(T_k, \widehat{X}_{T_k}) - \mu(T_{k-1}, \widehat{X}_{T_{k-1}})) \cdot (\sigma_0^T)^{-1} \Delta W_{T_{k+1}}}{\Delta T_{k+1}}. \quad (3.6)$$

Theorem 3.2. *Suppose that Assumption 3.1 holds true, and g is Lipschitz. Then with the choice (3.3) of $(\hat{\mu}, \hat{\sigma})$, for all intensity constant $\beta > 0$,*

$$\mathbb{E}[(\widehat{\psi})^2] < \infty; \quad \text{and moreover,} \quad V_0 = u(0, x_0) = \mathbb{E}[\widehat{\psi}].$$

Proof. (i) We first show that $\mathbb{E}[(\widehat{\psi})^2] < \infty$. For simplicity, we denote $\Delta \widehat{X}_k := \widehat{X}_{T_k} - \widehat{X}_{T_{k-1}}$ for $k > 0$. Let L_g be the Lipschitz constant of the function g , and set $L_0 := |(\sigma_0 \sigma_0^T)^{-1}| > 0$ by the non-degeneracy of σ_0 . Then using Assumption 3.1, it follows by direct computation that

$$|e^{-\beta T} \widehat{\psi}| \leq L_g \left(|g(x_0)| + \Delta T_1 + |\Delta \widehat{X}_{T_1}| \right) \prod_{k=1}^{N_T} \frac{L(\sqrt{\Delta T_{k+1}} + |\Delta \widehat{X}_{T_{k+1}}|)}{\beta \Delta T_{k+1}} \left| (\sigma_0^T)^{-1} \Delta W_{T_{k+1}} \right|.$$

Denote $|\mu|_\infty := \sqrt{\sum_{i=1}^d |\mu_i|_0^2}$ and $Z := \frac{\Delta W_{T_{k+1}}}{\sqrt{\Delta T_{k+1}}}$, we have

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{\sqrt{\Delta T_{k+1}} + |\Delta \widehat{X}_{T_{k+1}}|}{\Delta T_{k+1}} \right|^2 \left| (\sigma_0^T)^{-1} \Delta W_{T_{k+1}} \right|^2 \middle| \widehat{X}_{T_k}, \Delta T_{k+1} \right] \\ & \leq \mathbb{E} \left[(1 + |\mu|_\infty \sqrt{T} + |\sigma_0 Z|)^2 \left| (\sigma_0^T)^{-1} Z \right|^2 \right] \\ & \leq 2(1 + |\mu|_\infty \sqrt{T})^2 \mathbb{E} \left[\left| (\sigma_0^T)^{-1} Z \right|^2 \right] + 2\mathbb{E} \left[|\sigma_0 Z|^2 \left| (\sigma_0^T)^{-1} Z \right|^2 \right] \\ & = 2(1 + |\mu|_\infty \sqrt{T})^2 \text{Tr}((\sigma_0 \sigma_0^T)^{-1}) + 2(3d + d(d-1)). \end{aligned}$$

Next, denote $C := L_g^2 \mathbb{E}[(|g(0)| + \Delta T_1 + |\Delta X_1|)^2]$, we obtain an upper bound:

$$\begin{aligned} \mathbb{E}[(\widehat{\psi})^2] & \leq C e^{2\beta T} e^{-\beta T + \frac{L'T}{\beta}}, \\ \text{with } L' & = L^2 (2(1 + |\mu|_\infty \sqrt{T})^2 \text{Tr}((\sigma_0 \sigma_0^T)^{-1}) + 2(3d + d(d-1))). \end{aligned} \quad (3.7)$$

(ii) To prove that $V_0 = \mathbb{E}[\widehat{\psi}]$, we use Theorem 2.3. First, it is clear that Assumption 2.1 holds true with Malliavin weight functions in (3.4). Next, by the above variance analysis, it is clear that the uniform integrability condition (Condition (ii) in Theorem 2.3) holds true.

Finally, suppose that $g \in C_b^2(\mathbb{R}^d)$, then $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ so that all the conditions in Theorem 2.3 are satisfied and hence $V_0 = u(0, x_0) = \mathbb{E}[\widehat{\psi}]$. For general Lipschitz function $g(\cdot)$, we can always approximate g by some bounded smooth function $g^\varepsilon \in C_b^\infty(\mathbb{R}^d)$ such that $g^\varepsilon(\cdot) \rightarrow g(\cdot)$ locally uniformly as $\varepsilon \rightarrow 0$. Then the corresponding value function $u^\varepsilon \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ and $u^\varepsilon(t, x) \rightarrow u(t, x)$. Moreover, the corresponding estimator $\widehat{\psi}^\varepsilon$ (as in (3.5)) converges to $\widehat{\psi}$, and uniformly bounded by $\sqrt{C}e^{\beta T}(L')^{N_T/2}$, with the same constant C and L' defined in (3.7). We then conclude the proof by using the dominated convergence theorem. \square

Remark 3.3 (Lamperti's transformation). *We also notice that in some cases, the SDE (2.2) may be reduced into the constant diffusion coefficient case (3.1), by the so-called the Lamperti transformation.*

(i) When $d = 1$ and $\sigma(t, x) > 0$, let us define a function $h : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ by

$$h(t, x) := \int_0^x \frac{1}{\sigma(t, y)} dy.$$

Notice that for fixed $t \in [0, T]$, $x \mapsto h(t, x)$ is strictly increasing, we denote $h^{-1}(t, \cdot)$ its inverse function. Then by Ito's formula, it is easy to obtain that $Y_t := h(t, X_t)$ satisfies the SDE

$$dY_t = \left(\partial_t h(t, h^{-1}(t, Y_t)) + \frac{\mu(t, h^{-1}(t, Y_t))}{\sigma(t, h^{-1}(t, Y_t))} - \frac{1}{2} \partial_x \sigma(t, h^{-1}(t, Y_t)) \right) dt + dW_t,$$

whose diffusion coefficient is a constant as in SDE (3.1).

(ii) When $d > 1$, σ is non-degenerate and satisfies some further compatibility conditions, one can also obtain a similar transformation to reduce SDE (2.2) to the constant diffusion coefficient case.

3.2 A sub-optimal choice of β

Notice that the estimator $\widehat{\psi}$ defined by (3.5) induces an exact simulation Monte-Carlo method to compute V_0 by Theorem 3.2. Indeed, one needs to simulate only a sequence of Gaussian variables and a sequence of exponential random variables of distribution $\mathcal{E}(\beta)$. Here, the constant $\beta > 0$ could be chosen arbitrarily, let us discuss how to choose the value β in a sub-optimal way.

We denote the upper bound estimation (3.7) of the second order moment of the estimator by

$$F(\beta) := Ce^{-\beta T + L'T/\beta},$$

with L' defined in (3.7). Notice also that the computation effort is also proportional to the number N_T , whose expectation is given by $\mathbb{E}[N_T] = \beta T$. Then a sub-optimal

choice of the constant $\beta > 0$ can be obtained by solving

$$\min_{\beta > 0} \frac{F(\beta)}{\beta T} \iff \min_{\beta > 0} f(\beta), \quad \text{with } f(\beta) := \frac{1}{\beta T} \exp\left(T\left(\beta + \frac{L'}{\beta}\right)\right).$$

Notice that $\lim_{\beta \searrow 0} f(\beta) = \lim_{\beta \rightarrow \infty} f(\beta) = \infty$, then by direct computation, it follows that $f'(\beta) = 0$ has a unique solution on $(0, \infty)$, given by

$$\beta^* := \sqrt{L' + T^2/4} + \frac{T}{2},$$

which provides a sub-optimal choice of β for the exact simulation estimator (3.5).

4 One-dimensional driftless SDE

In this section, we consider a one-dimensional ($d = 1$) SDE, with zero drift coefficient, so that SDE (2.2) reduces to

$$dX_t = \sigma(t, X_t) dW_t, \quad (4.1)$$

with initial condition $X_0 = x_0$. Our objective is again to compute

$$V_0 := \mathbb{E}[g(X_T)], \quad \text{for some function } g : \mathbb{R} \rightarrow \mathbb{R}.$$

Assumption 4.1. *The diffusion coefficient $\sigma(\cdot)$ satisfies $\sigma(t, x) \geq \varepsilon > 0$ for all $(t, x) \in [0, T] \times \mathbb{R}$, $\sigma(t, x)$ is bounded and Lipschitz in (t, x) , $\partial_x \sigma(t, x)$ is bounded continuous in (t, x) and uniformly Lipschitz in x . Further, the terminal condition function $g(\cdot) \in C_b^2(\mathbb{R})$.*

4.1 The algorithm

To introduce the algorithm in the context of Theorem 2.3, we propose to choose

$$\hat{\mu}(\cdot) \equiv 0 \quad \text{and} \quad \hat{\sigma}(s, y, t, x) = \sigma(s, y) + \partial_x \sigma(s, y)(x - y).$$

Then \hat{X} in (2.6) turns to be $\hat{X}_0 = x_0$,

$$d\hat{X}_t = \left(\sigma(T_k, \hat{X}_{T_k}) + \partial_x \sigma(T_k, \hat{X}_{T_k})(\hat{X}_t - \hat{X}_{T_k}) \right) dW_t, \quad \text{on } [T_k, T_{k+1}], \quad (4.2)$$

for $k = 0, 1, \dots, N_T$, where $(T_k)_{k \geq 0}$ is defined from a sequence of i.i.d. $\mathcal{E}(\beta)$ -exponential distributed random variables $(\tau_i)_{i \geq 0}$ in (2.5).

By denoting $c_1^k := \sigma(T_k, \hat{X}_{T_k}) - \partial_x \sigma(T_k, \hat{X}_{T_k})\hat{X}_{T_k}$ and $c_2^k := \partial_x \sigma(T_k, \hat{X}_{T_k})$, then the above linear SDE (4.2) admits an explicit solution which is given by

$$\hat{X}_{T_{k+1}} = \hat{X}_{T_k} + \sigma(T_k, \hat{X}_{T_k})\Delta W_{T_{k+1}}, \quad \text{if } c_2^k = 0,$$

and

$$\begin{aligned} \hat{X}_{T_{k+1}} = & -\frac{c_1^k}{c_2^k} + \frac{c_1^k}{c_2^k} \exp\left(-\frac{(c_2^k)^2}{2}\Delta T_{k+1} + c_2^k \Delta W_{T_{k+1}}\right) \\ & + \hat{X}_{T_k} \exp\left(-\frac{(c_2^k)^2}{2}\Delta T_{k+1} + c_2^k \Delta W_{T_{k+1}}\right), \quad \text{if } c_2^k \neq 0. \end{aligned}$$

The estimator $\widehat{\psi}$ in (2.8) is then given by

$$\widehat{\psi} := e^{\beta T} \left[g(\widehat{X}_T) - g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right] \beta^{-N_T} \prod_{k=1}^{N_T} \overline{\mathcal{W}}_k^2, \quad (4.3)$$

where the Malliavin weight is given by (see Lemma 4.4 below)

$$\overline{\mathcal{W}}_k^2 := \frac{a(T_k, \widehat{X}_{T_k}) - \tilde{a}_k}{2a(T_k, \widehat{X}_{T_k})} \left(-\partial_x \sigma(T_k, \widehat{X}_{T_k}) \frac{\Delta W_{T_{k+1}}}{\Delta T_{k+1}} + \frac{\Delta W_{T_{k+1}}^2 - \Delta T_{k+1}}{\Delta T_{k+1}^2} \right), \quad (4.4)$$

with $a(\cdot) := \frac{1}{2}\sigma^2(\cdot)$, $\tilde{a}_k := \frac{1}{2}\tilde{\sigma}_k^2$ and $\tilde{\sigma}_k := \sigma(T_{k-1}, \widehat{X}_{T_{k-1}}) + \partial_x \sigma(T_{k-1}, \widehat{X}_{T_{k-1}})(\widehat{X}_{T_k} - \widehat{X}_{T_{k-1}})$.

As discussed in Remark 2.4, we can observe that $\widehat{\psi}$ defined by (4.3) is in fact integrable but of infinite variance in general. Motivated by this, we now introduce an alternative estimator using an antithetic variable. Let \widehat{X}_T^- be an antithetic variable of \widehat{X}_T defined by

$$\widehat{X}_T^- := \widehat{X}_{T_{N_T}} - \sigma(T_{N_T}, \widehat{X}_{T_{N_T}}) \Delta W_{T_{N_T}}, \quad \text{if } c_2^{N_T} = 0,$$

and

$$\begin{aligned} \widehat{X}_T^- &= -\frac{c_1^{N_T}}{c_2^{N_T}} + \frac{c_1^{N_T}}{c_2^{N_T}} \exp \left(-\frac{(c_2^{N_T})^2}{2} \Delta T_{N_T+1} - c_2^{N_T} \Delta W_{T_{N_T+1}} \right) \\ &\quad + \widehat{X}_{T_{N_T}} \exp \left(-\frac{(c_2^{N_T})^2}{2} \Delta T_{N_T+1} - c_2^{N_T} \Delta W_{T_{N_T+1}} \right), \quad \text{if } c_2^{N_T} \neq 0. \end{aligned}$$

Denote $\overline{\mathcal{W}}_k^- := \overline{\mathcal{W}}_k^2$ for $k = 1, \dots, N_T - 1$ and

$$\overline{\mathcal{W}}_{N_T}^- := \frac{a(T_{N_T}, \widehat{X}_{N_T}) - \tilde{a}_{N_T}}{2a(T_{N_T}, \widehat{X}_{N_T})} \left(\partial_x \sigma(T_{N_T}, \widehat{X}_{N_T}) \frac{\Delta W_{T_{N_T+1}}}{\Delta T_{N_T+1}} + \frac{\Delta W_{T_{N_T+1}}^2 - \Delta T_{N_T+1}}{\Delta T_{N_T+1}^2} \right).$$

We then introduce

$$\overline{\psi} := \frac{\widehat{\psi} + \widehat{\psi}^-}{2} \quad \text{with} \quad \widehat{\psi}^- := e^{\beta T} \left[g(\widehat{X}_T^-) - g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right] \beta^{-N_T} \prod_{k=1}^{N_T} \overline{\mathcal{W}}_k^-. \quad (4.5)$$

Notice that the Brownian motion is symmetric, thus $\widehat{\psi}^-$ has exactly the same distribution as $\widehat{\psi}$, and it serves as an antithetic variable.

Theorem 4.2. *Suppose that Assumption 4.1 holds true. Then*

$$\mathbb{E}[\widehat{\psi}] + \mathbb{E}[\overline{\psi}] < \infty; \quad \text{and} \quad V_0 = \mathbb{E}[\widehat{\psi}] = \mathbb{E}[\overline{\psi}]. \quad (4.6)$$

Remark 4.3. *For a general SDE with drift function and/or $d > 0$, we can also consider a similar choice of $(\hat{\mu}, \hat{\sigma})$, which leads to $\hat{\mu}(t, x) = c_1 + c_2 x$ and $\hat{\sigma}(t, x) = c_3 + c_4 x$ and a linear SDE*

$$d\widehat{X}_t = (c_1 + c_2 \widehat{X}_t) dt + (c_3 + c_4 \widehat{X}_t) dW_t, \quad (4.7)$$

where $c_1 \in \mathbb{R}^d$, $c_2, c_3 \in \mathbb{S}^d$ and c_4 is linear operator from \mathbb{R}^d to \mathbb{S}^d . However, it is not known how to simulate exactly the solution to SDE (4.7), as well as the associated Malliavin weight as in (4.4) (see also Lemma 4.4 below).

4.2 Proof of Theorem 4.2

Before providing the proof of Theorem 4.2, we first give a lemma which justifies our choice of the Malliavin weight function $\overline{\mathcal{W}}_k^2$ in (4.4), as well as some related estimations. Let $c_1, c_2, x \in \mathbb{R}$ be constants such that $c_1 + c_2x \neq 0$, we denote by $\overline{X}^{0,x}$ solution of the SDE

$$\overline{X}_0 = x, \quad d\overline{X}_t = (c_1 + c_2\overline{X}_t)dW_t, \quad (4.8)$$

whose solution is given explicitly by

$$\overline{X}_t^{0,x} = \begin{cases} -\frac{c_1}{c_2} + \left(\frac{c_1}{c_2} + x\right) \exp\left(-\frac{c_2^2}{2}t + c_2W_t\right), & \text{if } c_2 \neq 0, \\ x + c_1W_t, & \text{if } c_2 = 0. \end{cases} \quad (4.9)$$

Consider also its antithetic variable \tilde{X}_t^x defined by

$$\tilde{X}_t^{0,x} = \begin{cases} -\frac{c_1}{c_2} + \left(\frac{c_1}{c_2} + x\right) \exp\left(-\frac{c_2^2}{2}t - c_2W_t\right), & \text{if } c_2 \neq 0, \\ x - c_1W_t, & \text{if } c_2 = 0. \end{cases}$$

Lemma 4.4. *Let $x \in \mathbb{R}$, $(c_1, c_2) \in \mathbb{R}^2$ be two constants such that $c_1 + c_2x \neq 0$, $\phi : \mathbb{R} \rightarrow \mathbb{R}$ a bounded continuous function.*

(i) *Then for all $t \in (0, T]$,*

$$\partial_{xx}^2 \mathbb{E}[\phi(\overline{X}_t^{0,x})] = \mathbb{E}\left[\phi(\overline{X}_t^{0,x}) \frac{1}{(c_1 + c_2x)^2} \left(-c_2 \frac{W_t}{t} + \frac{W_t^2 - t}{t^2}\right)\right]. \quad (4.10)$$

(ii) *Suppose in addition that $\phi(\cdot) \in C_b^2(\mathbb{R})$. Then there is some constant C independent of (t, x) such that, for all $(t, x) \in [0, T] \times \mathbb{R}^d$,*

$$\begin{aligned} & \mathbb{E}\left[\left(\phi(\overline{X}_t^{0,x}) - \phi(x)\right)^2 \left(\frac{W_t}{t}\right)^2\right] + \mathbb{E}\left[\left(\phi(\overline{X}_t^{0,x}) - 2\phi(x) + \phi(\tilde{X}_t^{0,x})\right)^2 \left(\frac{W_t^2 - t}{t^2}\right)^2\right] \\ & \leq C(c_1 + c_2x)^2. \end{aligned}$$

Proof. (i) First, when $c_2 = 0$, it is clear that result is correct (see e.g. Lemma 2.1 of Fahim, Touzi and Warin [9]). Next, when $c_2 \neq 0$, denote $v(x) := \mathbb{E}[\phi(\overline{X}_t^{0,x})]$, then with the expression of $\overline{X}_t^{0,x}$ in (4.9), it follows that

$$v(x) = \int_{\mathbb{R}} \phi\left(-\frac{c_1}{c_2} + \left(\frac{c_1}{c_2} + x\right)e^{-c_2^2 t/2 + c_2 \sqrt{t}y}\right) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Suppose that $\phi(\cdot) \in C_b^2(\mathbb{R})$, then using integration by parts, it follows that

$$\begin{aligned} v'(x) &= \int_{\mathbb{R}} \phi'\left(-\frac{c_1}{c_2} + \left(\frac{c_1}{c_2} + x\right)e^{-c_2^2 t/2 + c_2 \sqrt{t}y}\right) e^{-c_2^2 t/2 + c_2 \sqrt{t}y} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= \int_{\mathbb{R}} \phi\left(-\frac{c_1}{c_2} + \left(\frac{c_1}{c_2} + x\right)e^{-c_2^2 t/2 + c_2 \sqrt{t}y}\right) \frac{1}{c_1 + c_2x} \frac{y}{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= \mathbb{E}\left[\phi(\overline{X}_t^{0,x}) \frac{1}{c_1 + c_2x} \frac{W_t}{t}\right]. \end{aligned}$$

Similarly, still using integration by parts, and by direct computation, we obtain

$$v''(x) = \mathbb{E}\left[\phi(\bar{X}_t^{0,x}) \frac{1}{(c_1 + c_2 x)^2} \left(-c_2 \frac{W_t}{t} + \frac{W_t^2 - t}{t^2}\right)\right].$$

When $\phi(\cdot)$ is only a bounded continuous function, one can approximate $\phi(\cdot)$ by a sequence of smooth function $\phi_\varepsilon(\cdot)$ which converges to $\phi(\cdot)$ uniformly, and ϕ'_ε and ϕ''_ε are bounded continuous. We then obtain

$$v_\varepsilon(x) := \mathbb{E}[\phi_\varepsilon(\bar{X}_t^{0,x})] \rightarrow v(x).$$

Moreover, the limit $\lim_{\varepsilon \rightarrow 0} v'_\varepsilon(x)$, $\lim_{\varepsilon \rightarrow 0} v''_\varepsilon(x)$ exist, thus $v''(x)$ also exists and

$$v''(x) = \lim_{\varepsilon \rightarrow 0} v''_\varepsilon(x) = \mathbb{E}\left[\phi(\bar{X}_t^{0,x}) \frac{1}{(c_1 + c_2 x)^2} \left(-c_2 \frac{W_t}{t} + \frac{W_t^2 - t}{t^2}\right)\right].$$

(ii) When $c_2 = 0$, the estimation in (ii) of the statement is clear true since ϕ' and ϕ'' are uniformly bounded.

When $c_2 \neq 0$, by direct computation, we obtain

$$\begin{aligned} \mathbb{E}\left[\left(\phi(\bar{X}_t^{0,x}) - \phi(x)\right)^2 \left(\frac{W_t}{t}\right)^2\right] &\leq |\phi'|_\infty \mathbb{E}\left[(\bar{X}_t^{0,x} - x)^2 \frac{W_t^2}{t^2}\right] \\ &= |\phi'|_\infty \mathbb{E}\left[(c_1 + c_2 x)^2 \left(\frac{e^{-c_2^2 t/2 + c_2 W_t} - 1}{c_2 W_t - c_2^2 t/2}\right)^2 \frac{W_t^2 (c_2 W_t - c_2^2 t/2)^2}{t^2}\right], \end{aligned}$$

which is clearly uniformly bounded by $C(c_1 + c_2 x)^2$ for some constant C independent of $(t, x) \in [0, T] \times \mathbb{R}^d$.

Next, denote $\ell(y) := (x + \frac{c_1}{c_2})(e^{-c_2^2 t/2 + c_2 y} - 1)$, and define $\varphi(y) := \phi(x + \ell(y))$. Then

$$\varphi''(y) = \phi''(x + \ell(y))(c_2 + c_1 x)^2 e^{-c_2^2 t/2 + c_2 y} + \phi'(x + \ell(y))(c_2 + c_1 x) c_2 e^{-c_2^2 t/2 + c_2 y}. \quad (4.11)$$

It follows by the definition of φ as well as its derivative, together with direct computation, that

$$\begin{aligned} &\mathbb{E}\left[\left(\phi(\bar{X}_t^{0,x}) - 2\phi(x) + \phi(\tilde{X}_t^{0,x})\right)^2 \left(\frac{W_t^2 - t}{t^2}\right)^2\right] \\ &= \mathbb{E}\left[(\varphi(W_t) + \varphi(-W_t) - 2\varphi(0))^2 \left(\frac{W_t^2 - t}{t^2}\right)^2\right] + \mathbb{E}\left[2(\varphi(0) - \phi(x))^2 \left(\frac{W_t^2 - t}{t^2}\right)^2\right] \\ &\leq \mathbb{E}\left[\left(\frac{W_t^2 (W_t^2 - t)}{t^2}\right)^2 \sup_{|z| \leq |W_t|} \varphi''(z)\right] \\ &\quad + \mathbb{E}\left[2\left(\phi\left(x + \frac{c_1 + c_2 x}{c_2} (e^{-c_2^2 t/2} - 1)\right) - \phi(x)\right)^2 \left(\frac{W_t^2 - t}{t^2}\right)^2\right], \end{aligned}$$

which is also uniformly bounded by $C(c_1 + c_2 x)^2$ for some constant $C > 0$, \square

Proof of Theorem 4.2. (i) Let us first prove that $\mathbb{E}[\bar{\psi}^2] < \infty$ for $\bar{\psi}$ defined by (4.5). First, notice that $\widehat{\mathcal{W}}_k^- = \widehat{\mathcal{W}}_k^2$ for all $k = 1, \dots, N_T - 1$, and $g \in C_b^2(\mathbb{R})$, and

by considering the conditional expectation over $(\hat{X}_{T_{N_T}}, \Delta T_{N_T+1})$ using items (ii) of Lemma 4.4, we have $\mathbb{E}[|\bar{\psi}|^2]$ is bounded by

$$C\mathbb{E}\left[\beta^{-2N_T} \prod_{k=2}^{N_T} \left\{ \frac{a(T_k, \hat{X}_{T_k}) - \tilde{a}_k}{2a(T_k, \hat{X}_{T_k})} \left(-\partial_x \sigma(T_k, \hat{X}_{T_k}) \frac{\Delta W_{T_k}}{\Delta T_k} + \frac{\Delta W_{T_k}^2 - \Delta T_k}{\Delta T_k^2} \right) \right\}^2 \right],$$

for some constant C . Further, by denoting $\Delta \hat{X}_{T_k} := \hat{X}_{T_k} - \hat{X}_{T_{k-1}}$, one has

$$\begin{aligned} |a(T_k, \hat{X}_{T_k}) - \tilde{a}_k| &\leq \left(|\sigma|_0 + |\partial_x \sigma(T_{k-1}, \hat{X}_{T_{k-1}}) \Delta \hat{X}_{T_k}|/2 \right) \\ &\quad \left(|\partial_t \sigma|_0 \Delta T_k + |\partial_{xx}^2 \sigma|_0 (\Delta \hat{X}_{T_k})^2 \right). \end{aligned}$$

Notice that $\sigma \geq \varepsilon > 0$, σ and $\partial_x \sigma$ are uniformly bounded, then to prove that $\bar{\psi}$ is of finite variance, it is enough to prove that the variance of

$$\prod_{k=2}^{N_T} \left[C \left(C + |\partial_x \sigma(T_{k-1}, \hat{X}_{T_{k-1}}) \Delta \hat{X}_{T_k}| \right) \left(C + \frac{\Delta \hat{X}_{T_k}^2}{\Delta T_k} \right) \left(C |\Delta W_{T_k}| + \frac{\Delta W_{T_k}^2}{\Delta T_k} + 1 \right) \right] \quad (4.12)$$

is finite. Similarly to the computation in item (ii) of Lemma 4.4, we have

$$\begin{aligned} \Delta \hat{X}_{T_k} &= \hat{X}_{T_k} - \hat{X}_{T_{k-1}} \\ &= \sigma(T_{k-1}, \hat{X}_{T_{k-1}}) \frac{\exp(-\partial_x \sigma(T_{k-1}, \hat{X}_{T_{k-1}})^2 \Delta T_k / 2 + \partial_x \sigma(T_{k-1}, \hat{X}_{T_{k-1}}) \Delta W_{T_k}) - 1}{\partial_x \sigma(T_{k-1}, \hat{X}_{T_{k-1}})}. \end{aligned}$$

Notice again that $\sigma(\cdot)$ and $\partial_x \sigma(\cdot)$ are uniformly bounded, it follows that

$$\begin{aligned} \mathbb{E} \left\{ \left[\left(C + |\partial_x \sigma(T_{k-1}, \hat{X}_{T_{k-1}}) \Delta \hat{X}_{T_k}| \right) \left(C + \frac{\Delta \hat{X}_{T_k}^2}{\Delta T_k} \right) \left(C |\Delta W_{T_k}| + \frac{\Delta W_{T_k}^2}{\Delta T_k} + 1 \right) \right]^2 \right. \\ \left. \middle| \hat{X}_{T_{k-1}}, T_{k-1}, \Delta T_k \right\} \leq C', \end{aligned}$$

for some constant $C' > 0$ independent of $\hat{X}_{T_{k-1}}, T_{k-1}, \Delta T_k$. Then the variance of (4.12) is bounded by $C\mathbb{E}[(C')^{N_T}] < \infty$ and hence $\bar{\psi}$ in (4.5) is of finite variance.

(ii) Let us now consider the estimator $\hat{\psi}$. By the same computation, we obtain that

$$\mathbb{E}[\hat{\psi} \mid N_T, \Delta T_1, \dots, \Delta T_{N_T+1}] \leq C^{N_T} \frac{1}{\sqrt{\Delta T_{N_T+1}}}, \quad \text{for some } C > 0,$$

where the r.h.s. is integrable but of infinite variance (see Lemma A.2). Similarly, it is easy to check the uniform integrability condition in item (ii) Theorem 2.3 for $\hat{\psi}$ in (4.3).

(iii) Finally, using item (i) of Lemma 4.4, it follows that Assumption 2.1 holds true. Moreover, with the regularity condition on $\sigma(t, x)$ and g in Assumption 4.1, we know $u \in C_b^{1,2}(\mathbb{R})$. We then conclude the proof of $u(0, x_0) = \mathbb{E}[\hat{\psi}] = \mathbb{E}[\bar{\psi}]$ by Theorem 2.3. \square

5 Numerical examples

5.1 One-dimensional SDE

As a numerical illustration of our algorithm, we consider the following one-dimensional SDE, with $\sigma = 0.4$,

$$X_0 = 1, \quad dX_t = \frac{2\sigma}{1 + X_t^2} dW_t. \quad (5.1)$$

By setting $2\sigma Y_t = \left(X_t - X_0 + \frac{(X_t^3 - X_0^3)}{3}\right)$, this SDE can be transformed into (i.e., by the Lamperti transformation¹)

$$Y_0 = 0, \quad dY_t = \frac{2\sigma X_t}{(1 + X_t^2)^2} dt + dW_t. \quad (5.2)$$

We have computed the functional $V_0(K) := \mathbb{E}[(X_T - K)^+]$ for different values of K ranging from 0.5 to 1.5 and $T = 1$ (i.e. one year). As usual in mathematical finance, the value $V_0(K)$, representing the price of a call option with maturity T and strike K , is quoted in implied volatility, i.e., the constant volatility that must be plugged into the Black-Scholes formula in order to reproduce the value $V_0(K)$. We have computed $V_0(K)$ using two exact algorithms: one based on SDE (5.2) (see section 3) and the second one based on SDE (5.1) (see section 4). We have used $\beta = 0.2$ and check that our results are independent of this value.

Although artificial, these (equivalent) SDEs have been chosen because they require a small timestep discretization in an Euler scheme in order to achieve convergence. More precisely, our (exact) methods have been checked against an Euler discretization scheme with a timestep $\Delta = \{1/10, 1/50, 1/100, 1/400\}$. The values obtained with $\Delta = 1/400$ converge exactly to our exact scheme and is therefore not reported in our figures. In Figure 1, we can observe that the Euler scheme converges towards our exact two methods (which coincide) when $\Delta \leq 1/50$.

Note that in practice, for a fixed number of Monte-Carlo paths, the variance of the algorithm based on (5.2) is smaller than the one based on (5.1) as the Malliavin weight appearing in the stochastic representation (corresponding only to the first-order derivative) has less variance. In table 1, we have shown the standard deviation for an at-the-money call option with $K = 1$, $T = 1$ as a function of the Monte-Carlo paths 2^N using SDE (5.1). We have used here two different values for β : $\beta = 0.1$ and $\beta = 0.2$. The (finite) variance is of the same order as the Euler algorithm.

5.2 Multi-dimensional SDE

We considered the following multi-dimensional SDE:

$$\frac{dX_t^i}{X_t^i} = \frac{1}{2} dW_t^i + 0.1 \left(\sqrt{X_t^i} - 1 \right) dt \quad X_0^i = 1, \quad d\langle W^i, W^j \rangle_t = 0.5dt, \quad i \neq j = 1, \dots, d.$$

¹We do not write the lengthy relation giving X_t as a function of Y_t .

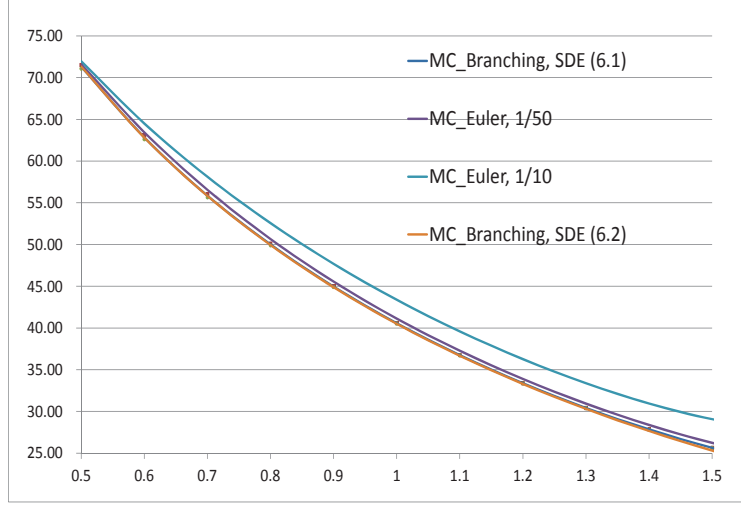


Figure 1: $V_0(K)$ quoted in implied volatility $\times 100$ as a function of K . The dots correspond to the standard deviation error.

N	$\beta = 0.1$	$\beta = 0.2$	Euler
12	0.32	0.34	0.30
14	0.16	0.17	0.15
16	0.08	0.09	0.08
18	0.05	0.04	0.04
20	0.02	0.02	0.02
22	0.01	0.02	0.01
24	0.01	0.01	0.00

Table 1: Standard deviation for an at-the-money call option with $K = 1$, $T =$ one year as a function of the Monte-Carlo paths 2^N .

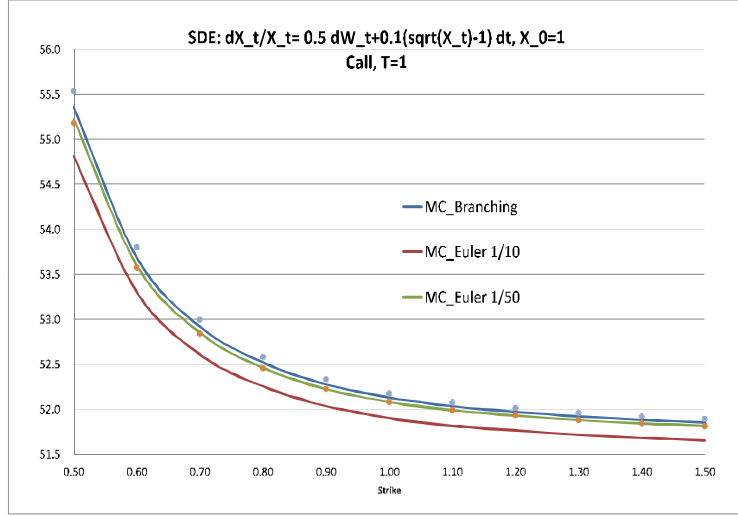


Figure 2: $d = 1$. $V_0(K)$ quoted in implied volatility $\times 100$ as a function of K . The dots correspond to the standard deviation error.

We have computed the functional $V_0(K) := \mathbb{E}[(\frac{1}{n} \sum_{i=1}^n X_T^i - K)^+]$ for different values of K ranging from 0.5 to 1.5 and $T =$ one year. The value $V_0(K)$, representing the price of a basket payoff with strike K , is quoted in implied volatility as in the previous section. Our (exact) method has been checked against a (log)-Euler discretization scheme with a timestep $\Delta = \{1/10, 1/50, 1/100\}$:

$$X_{t+\Delta}^\Delta = X_t^\Delta \exp \left(\frac{1}{2} \Delta W_t + \left(0.1 \left(\sqrt{X_t^i} - 1 \right) - \frac{1}{8} \right) \Delta \right).$$

Note that although the Lipschitz condition is not satisfied by this SDE, we show that our algorithm works numerically.

The values obtained with $\Delta = 1/100$ converges exactly to our exact scheme and is therefore not reported in our figures. We have chosen two different values for the dimension d , mainly $d = 1$ (see Fig. 2) and $d = 4$ (see Fig. 3) in order to illustrate that our method is not only applicable to the one-dimensional setup as in Beskos-Roberts's method previously mentioned. In Figures (2, 3), we can observe that the Euler scheme converges towards our exact method when $\Delta > 1/50$.

6 Further discussions

In this section, we would like to provide some further discussions on the exact simulation of SDEs with general drift and diffusion coefficients, and also on the extension of our algorithm to the path-dependent case.

For a multi-dimensional SDE with general drift and diffusion coefficients, we obtain an exact simulation estimator which is integrable but of infinite variance. For a multi-

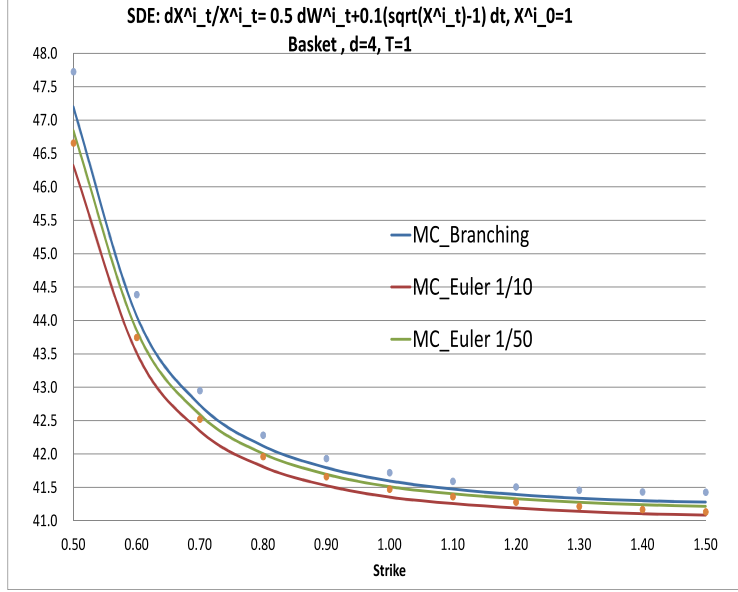


Figure 3: $d = 4$. $V_0(K)$ quoted in implied volatility $\times 100$ as a function of K . The dots correspond to the standard deviation error.

dimensional SDE with constant diffusion coefficient, but path-dependent drift and terminal functions, we obtain an exact simulation estimator of finite variance.

6.1 The general drift and diffusion coefficient case

We now go back to the context of Section 2, where we considered the multi-dimensional SDEs (2.2) on X , with general drift and diffusion coefficients. The objective is to compute $V_0 = \mathbb{E}[g(X_T)]$ as given in (2.3). We will propose an algorithm in the same spirit of that in Section 3. However, in this general context, the estimator is integrable but of infinite variance. This was also the main motivation, in Section 4, to consider a $\hat{\sigma}$ as higher order Taylor expansion of σ .

Let us assume that $(\mu, \sigma) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{S}^d$ and $a := \frac{1}{2}\sigma\sigma^T : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{S}^d$ are uniformly Hölder in the time variable, and uniformly Lipschitz in the space variable, i.e. for some constant L ,

$$|(\mu, \sigma, a)(t, x) - (\mu, \sigma, a)(s, y)| \leq L(\sqrt{|t - s|} + |x - y|), \quad (6.1)$$

for all $(t, x), (s, y) \in [0, T] \times \mathbb{R}^d$. We assume further that $\sigma(t, x)$ is non-degenerate such that, for some constant $\varepsilon_0 > 0$,

$$a(t, x) := \frac{1}{2}\sigma\sigma^T(t, x) \geq \varepsilon_0 I_d, \quad \forall (t, x) \in [0, T] \times \mathbb{R}^d. \quad (6.2)$$

With the choice

$$\hat{\mu}(s, y, t, x) := \mu(s, y) \quad \text{and} \quad \hat{\sigma}(s, y, t, x) := \sigma(s, y),$$

The process \widehat{X} is defined by $\widehat{X}_0 := x_0$ and

$$\widehat{X}_{T_{k+1}} := \widehat{X}_{T_k} + \mu(T_k, \widehat{X}_{T_k})\Delta T_{k+1} + \sigma(T_k, \widehat{X}_{T_k})\Delta W_{T_{k+1}} \quad k = 0, \dots, N_T.$$

And the estimator is given by

$$\widehat{\psi} := e^{\beta T} \left(g(\widehat{X}_T) - g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right) \beta^{-N_T} \prod_{k=1}^{N_T} (\overline{\mathcal{W}}_k^1 + \overline{\mathcal{W}}_k^2), \quad (6.3)$$

where, for each $k = 1, \dots, N_T$,

$$\overline{\mathcal{W}}_k^1 := \left(\mu(T_k, \widehat{X}_{T_k}) - \mu(T_{k-1}, \widehat{X}_{T_{k-1}}) \right) \cdot \frac{(\sigma(T_k, \widehat{X}_{T_k})^T)^{-1} \Delta W_{T_{k+1}}}{\Delta T_{k+1}},$$

and

$$\begin{aligned} \overline{\mathcal{W}}_k^2 &:= \left(a(T_k, \widehat{X}_{T_k}) - a(T_{k-1}, \widehat{X}_{T_{k-1}}) \right) \\ &: \left((\sigma(T_k, \widehat{X}_{T_k})^T)^{-1} \frac{\Delta W_{T_{k+1}} \Delta W_{T_{k+1}}^T - \Delta T_{k+1} I_d}{\Delta T_{k+1}^2} \sigma(T_k, \widehat{X}_{T_k})^{-1} \right). \end{aligned} \quad (6.4)$$

Theorem 6.1. *Suppose that μ and a satisfy the Hölder and Lipschitz condition (6.1), $\sigma(t, x)$ is non-degenerate such that (6.2) holds true. Suppose in addition that g is Lipschitz. Then*

$$\mathbb{E}[\|\widehat{\psi}\|] < \infty; \quad \text{and moreover,} \quad V_0 = u(0, x_0) = \mathbb{E}[\widehat{\psi}].$$

Proof. The proof of Theorem 6.1 is similar to that of Theorem 3.2.

(i) We first prove that $\widehat{\psi}$ is integrable. Notice that $g(\cdot)$ is Lipschitz, $\mu(t, x)$ and $a(t, x)$ are $1/2$ -Hölder in t and Lipschitz in x , and $a(t, x) = \sigma \sigma^T(t, x) \geq \varepsilon_0 I_d$. It follows by direct computation that

$$\mathbb{E}[\widehat{\psi}] \leq C \mathbb{E} \left[\prod_{k=0}^{N_T} \frac{C}{\sqrt{\Delta T_{k+1}}} \right],$$

for some constant $C > 0$. We then have the integrability of $\widehat{\psi}$ by Lemma A.2. Moreover, using the same arguments, it is easy to see that Assumption 2.2 holds true in the above context.

(ii) Next, since the increment $\widehat{X}_{T_{k+1}} - \widehat{X}_{T_k}$, conditional on (T_k, \widehat{X}_{T_k}) , is Gaussian, then Assumption 2.1 also holds true in this context.

(iii) Now, suppose in addition that μ , σ and g are bounded smooth functions with bounded continuous derivatives, so that $u \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$. It follows by Theorem 2.3 that $V_0 = \mathbb{E}[\widehat{\psi}]$.

(iv) When $\mu(\cdot)$ and $\sigma(\cdot)$ satisfy the Lipschitz condition (6.1) and g is Lipschitz, we can find a sequence of bounded smooth functions $(\mu_\varepsilon(\cdot), \sigma_\varepsilon(\cdot), g_\varepsilon(\cdot))$ which converges locally uniformly to $(\mu(\cdot), \sigma(\cdot), g(\cdot))$ as $\varepsilon \rightarrow 0$.

Let X^ε be the solution of

$$dX_t^\varepsilon = \mu_\varepsilon(t, X_t^\varepsilon)dt + \sigma_\varepsilon(t, X_t^\varepsilon)dW_t.$$

Then by the stability of SDEs together with dominated convergence theorem, it follows that

$$V_0^\varepsilon := \mathbb{E}[g_\varepsilon(X_T^\varepsilon)] \longrightarrow V_0 := \mathbb{E}[g(X_T)], \quad \text{as } \varepsilon \rightarrow 0.$$

Moreover, by Lemma A.2 together with dominated convergence theorem, it is easy to prove that $\mathbb{E}[\widehat{\psi}^\varepsilon] \rightarrow \mathbb{E}[\widehat{\psi}]$ as $\varepsilon \rightarrow 0$, where $\widehat{\psi}^\varepsilon$ denotes the estimator of the algorithm (6.3) associated to the coefficient $(\mu_\varepsilon, \sigma_\varepsilon, g_\varepsilon)$. We then conclude the proof. \square

Remark 6.2. (i) *In general, the estimator $\widehat{\psi}$ is of order $\Pi_{k=0}^{N_T} \frac{C}{\sqrt{\Delta T_{k+1}}}$, which is integrable but of infinite variance. Therefore, $\widehat{\psi}$ is not a good Monte-Carlo estimator.*

(ii) *The above estimator in (6.3) can be compared to the representation in Bally and Kohatsu-Higa [2, Section 6.1] as an application of their parametrix method for SDEs. The definition of \widehat{X} is exactly the same, but the associated weight functions are different.*

6.2 Extension to the path-dependent case

In this part, we would like to provide an extension of the algorithm in Section 3 in the path-dependent case. Let $n > 0$, $0 = t_0 < t_1 < \dots < t_n = T$, $\sigma_0 \in \mathbb{S}^d$ be a non-degenerate matrix, and $\mu : [0, T] \times \mathbb{R}^{d \times n} \rightarrow \mathbb{R}^d$ be a continuous function, Lipschitz in the space variable. Let X be the unique solution of SDE, with initial condition $X_0 = x_0$,

$$dX_t = \mu(t, X_{t_1 \wedge t}, \dots, X_{t_n \wedge t}) dt + \sigma_0 dW_t; \quad (6.5)$$

and the objective is to compute the value,

$$\widetilde{V}_0 := \mathbb{E}[g(X_{t_1}, \dots, X_{t_n})], \quad (6.6)$$

for some Lipschitz function $g : \mathbb{R}^{d \times n} \rightarrow \mathbb{R}$.

Remark 6.3. (i) *It is clear that the value \widetilde{V}_0 defined above can be characterized by a parabolic PDE system. Namely, for every $k = 1, \dots, n$ and $(x_1, \dots, x_{k-1}) \in \mathbb{R}^{d \times (k-1)}$, we define*

$$\mu_k(t, x) := \mu(t, x_1, \dots, x_{k-1}, x, \dots, x), \quad \forall (t, x) \in [t_{k-1}, t_k] \times \mathbb{R}^d. \quad (6.7)$$

Suppose that $(u_k)_{k=1, \dots, n}$ is a family of functions such that u_k is defined on $[t_{k-1}, t_k] \times \mathbb{R}^{d \times k}$ and $x \mapsto u_k(t, x_1, \dots, x_{k-1}, x)$ is the solution of

$$\partial_t u_k + \frac{1}{2} \sigma_0 \sigma_0^T : D^2 u_k + \mu_k \cdot D u_k = 0, \quad (6.8)$$

with terminal conditions

$$u_k(t_k, x_1, \dots, x_k) = u_{k+1}(t_k, x_1, \dots, x_k, x_k), \quad \text{for } k = 1, \dots, n-1,$$

and $u_n(t_n, x_1, \dots, x_n) = g(x_1, \dots, x_n)$. Then we have $\overline{V}_0 = u_1(0, x_0)$.

(ii) One can also use the notion of path-dependent PDE introduced by Ekren, Keller, Touzi and Zhang [7, 8]. Then \bar{V}_0 can be characterized by the linear path-dependent PDE

$$\partial_t u + \frac{1}{2} \sigma_0 \sigma_0^T : \partial_{\omega\omega}^2 u + \mu \cdot \partial_\omega u = 0,$$

with terminal condition $u(T, \omega) := g(\omega_{t_1}, \dots, \omega_{t_n})$, $\forall \omega \in C([0, T], \mathbb{R}^d)$ (the canonical space of all continuous path on $[0, T]$), where $\mu(t, \omega) := \mu(t, \omega_{t_1 \wedge t}, \dots, \omega_{t_n \wedge t})$, and the derivative ∂_ω and $\partial_{\omega\omega}^2$ are defined in sense of Dupire [6].

6.2.1 The algorithm

The algorithm can be obtained by an iteration of the algorithm (3.5) on every time interval $[t_k, t_{k+1}]$. One should just be careful on the integrability issue.

Recall that W be a standard d -dimensional Brownian motion, $(\tau_i)_{i>0}$ is a sequence of i.i.d. $\mathcal{E}(\beta)$ -exponential random variables independent of W . Then $N = (N_s)_{0 \leq s \leq t}$ and $(T_i)_{i>0}$ are defined in (2.5). Define further for every $k = 1, \dots, n$, $\tilde{N}^k := N_{t_k} - N_{t_{k-1}}$ the number of default on $[t_{k-1}, t_k)$, and $\tilde{T}_0^k := t_{k-1}$ and $\tilde{T}_j^k := T_{N_{t_{k-1}}+j} \wedge t_k$,

$$\Delta \tilde{T}_j^k := \tilde{T}_j^k - \tilde{T}_{j-1}^k, \quad \widetilde{W}_j^k := W_{\tilde{T}_j^k}, \quad \Delta \widetilde{W}_j^k := \widetilde{W}_j^k - \widetilde{W}_{j-1}^k, \quad \forall j = 1, \dots, \tilde{N}^k + 1.$$

Example 6.4. We give below an example for the case $n = 2$. In the following example, the number of default on $[0, t_1)$ is $\tilde{N}^1 = 2$, that on $[t_1, t_2)$ is $\tilde{N}^2 = 1$, and total default number is $N_T = 3$.

For $k = 1$, we have $\tilde{T}_0^1 = 0$, $\tilde{T}_1^1 = T_1$, $\tilde{T}_2^1 = T_2$ and $\tilde{T}_3^1 = t_1$; $\widetilde{W}_0^1 = 0$, $\widetilde{W}_1^1 = W_{T_1}$, $\widetilde{W}_2^1 = W_{T_2}$ and $\widetilde{W}_3^1 = W_{t_1}$. For $k = 2$, we have $\tilde{T}_0^2 = t_1$, $\tilde{T}_1^2 = T_3$, $\tilde{T}_2^2 = t_2$, and $\widetilde{W}_0^2 = W_{t_1}$, $\widetilde{W}_1^2 = W_{T_3}$ and $\widetilde{W}_2^2 = W_{t_2}$.



We next introduce a process $(\tilde{X}_j^{k,\mathbf{x}})$, $\forall j = 0, 1, \dots, N_k + 1$, for each $k = 1, \dots, n$ and initial condition $\mathbf{x} = (x_0, x_1, \dots, x_{k-1}) \in \mathbb{R}^{d \times k}$ by $\tilde{X}_0^{k,\mathbf{x}} := x_{k-1}$ and

$$\tilde{X}_{j+1}^{k,\mathbf{x}} := \tilde{X}_j^{k,\mathbf{x}} + \mu_k(\tilde{T}_j^k, \tilde{X}_j^{k,\mathbf{x}}) \Delta \tilde{T}_{j+1}^k + \sigma_0 \Delta \widetilde{W}_{j+1}^k.$$

Similarly, for every $j = 1, \dots, N_k$, we define a Malliavin weight, with μ_k defined by (6.7),

$$\widetilde{W}_j^k := \frac{(\mu_k(\tilde{T}_j^k, \tilde{X}_j^{k,\mathbf{x}}) - \mu_k(\tilde{T}_{j-1}^k, \tilde{X}_{j-1}^{k,\mathbf{x}})) \cdot (\sigma_0^T)^{-1} \Delta \widetilde{W}_{j+1}^k}{\Delta \tilde{T}_{j+1}^k}.$$

We now introduce the algorithm for the path-dependent case, in a recursive way. First, for $\mathbf{x} = (x_0, x_1, \dots, x_n) \in \mathbb{R}^{d \times (n+1)}$, set $\tilde{\psi}_{n+1}^{\mathbf{x}} := g(x_1, \dots, x_n)$. Next, for $k = 1, \dots, n$, denote

$$\mathbf{X}^{k,\mathbf{x}} := (x_0, x_1, \dots, x_{k-1}, \tilde{X}_{\tilde{N}^k+1}^{k,\mathbf{x}}) \text{ and } \mathbf{X}^{k,\mathbf{x},0} := (x_0, x_1, \dots, x_{k-1}, \tilde{X}_{\tilde{N}^k}^{k,\mathbf{x}} \mathbf{1}_{\{\tilde{N}^k > 0\}}).$$

Then given $\tilde{\psi}_{k+1}$, we define

$$\tilde{\psi}_k^{\mathbf{x}} := e^{\beta(t_k - t_{k-1})} \left(\tilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x}}} - \tilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x},0}} \mathbf{1}_{\{\tilde{N}^k > 0\}} \right) \beta^{-\tilde{N}^k} \prod_{j=1}^{\tilde{N}^k} \tilde{\mathcal{W}}_j^k. \quad (6.9)$$

We finally obtain the numerical algorithm of the path-dependent case:

$$\tilde{\psi} := \tilde{\psi}_1^{x_0}. \quad (6.10)$$

6.2.2 The integrability and representation result

We notice that the algorithm in the path-dependent case is nothing else than an iterative algorithm of the Markovian case, as the PDE system (6.8) in Remark (6.3). When the random variable $\tilde{\psi}$ in (6.10) is integrable, it is not surprising that $\tilde{V}_0 = \mathbb{E}[\tilde{\psi}]$. However, because of the renormalization term (i.e. $(\tilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x}}} - \tilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x},0}} \mathbf{1}_{\{\tilde{N}^k > 0\}})$ in (6.9)), the variance analysis becomes less obvious. We provide here a sufficient condition to ensure that $\tilde{\psi}$ admits finite variance.

Theorem 6.5. *Suppose that $\mu : [0, T] \times \mathbb{R}^{d \times n} \rightarrow \mathbb{R}^d$ and $g : \mathbb{R}^{d \times n} \rightarrow \mathbb{R}$ are all differentiable up to the order n , and every derivative of any order up to n is uniformly bounded. Then*

$$\mathbb{E}[(\tilde{\psi})^2] < \infty; \text{ and moreover } \tilde{V}_0 := \mathbb{E}[\tilde{\psi}].$$

In preparation of the proof of Theorem 6.5, we first provide two technical lemmas. Let $\pi = (0 = s_0 < s_1 < \dots < s_m = T)$ be an arbitrary partition of the interval $[0, T]$, $\bar{\mu} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ a \mathbb{R}^d -valued function. We define $X^{\pi,x}$ by $X_0^{\pi,x} := x$ and

$$X_{k+1}^{\pi,x} := X_k^{\pi,x} + \bar{\mu}(s_k, X_k^{\pi,x}) \Delta s_{k+1} + W_{s_{k+1}} - W_{s_k}. \quad (6.11)$$

Further, let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ be a smooth function, $\ell > 0$ and $i = (i_1, \dots, i_\ell) \in \{1, \dots, d\}^\ell$, we denote $\partial_{x,i}^\ell \varphi(x) := \partial_{x_{i_1} \dots x_{i_\ell}}^\ell \varphi(x)$.

Lemma 6.6. *Suppose that $x \mapsto \bar{\mu}(t, x)$ is differentiable up to order n with uniformly bounded derivatives, and $X^{\pi,x}$ is defined by (6.11) with initial condition $X_0^{\pi,x} = x$. Then $x \mapsto X_k^{\pi,x}$ is differential up to order n and there is a constant C independent of the partition π such that*

$$\max_{1 \leq \ell \leq n} \max_{i \in \{1, \dots, d\}^\ell} \max_{0 \leq k \leq m} |\partial_{x,i}^\ell X_k^{\pi,x}| \leq C.$$

Proof. For simplicity, we consider the one dimensional $d = 1$ case, while the multi-dimensional can be deduced by almost the same arguments. First, let $\ell = 1$, we have

$$\partial_x X_{k+1}^{\pi,x} = \partial_x X_k^{\pi,x} + \partial_x \bar{\mu}(s_k, X_k^{\pi,x}) \partial_x X_k^{\pi,x} \Delta s_{k+1},$$

which implies that

$$\partial_x X_{k+1}^{\pi,x} = \prod_{j=1}^{k+1} \left(1 + \partial_x \bar{\mu}(s_k, X_k^{\pi,x}) \Delta s_{k+1} \right).$$

Since $\partial_x \bar{\mu}(t, x)$ is uniformly bounded, it follows that $\partial_x X_k^{\pi, x}$ is bounded by some constant C_1 independent of $1 \leq k \leq m$ and the partition π . By induction, it is easy to deduce that for $\ell = 2, \dots, n$,

$$\partial_{x^\ell}^\ell X_{k+1}^{\pi, x} = \partial_{x^\ell}^\ell X_k^{\pi, x} + P_\ell(\partial_{x^i}^i \bar{\mu}(s_k, X_k^{\pi, x}), \partial_{x^i}^i X_k^{\pi, x}, i = 1, \dots, \ell - 1) \Delta s_{k+1},$$

where P_ℓ is a Polynomial on $\partial_{x^i}^i \bar{\mu}(s_k, X_k^{\pi, x})$ and $\partial_{x^i}^i X_k^{\pi, x}$ for $i = 1, \dots, \ell - 1$, which is uniformly bounded by some constant independent of $k = 1, \dots, m$ and the partition π . Hence $\partial_{x^\ell}^\ell X_k^{\pi, x}$ is also bounded by some constant C_ℓ independent of $k = 1, \dots, m$ and the partition π . \square

Lemma 6.7. *Let $(\tilde{\psi}_k^{\mathbf{x}})_{1 \leq k \leq n+1}$ be defined by (6.9). Then for every $k = 2, \dots, n+1$, and every $\mathbf{x} = (x_0, x_1, \dots, x_{k-1}) \in \mathbb{R}^{d \times k}$, the map $x_{k-1} \mapsto \tilde{\psi}_k^{\mathbf{x}}$ admits derivatives up to order $k-1$ and*

$$\max_{1 \leq \ell \leq k-1} \left| \partial_{x_{k-1}, i}^\ell \tilde{\psi}_k^{\mathbf{x}} \right| \leq C \prod_{j=k}^n (\tilde{N}^j + 1)^{j-1}. \quad (6.12)$$

Proof. We will prove it by induction. First, let $k = n+1$, then $\tilde{\psi}_{n+1}^{\mathbf{x}} := g(x, x_1, \dots, x_n)$ and hence $|\partial_{x_n}^\ell \tilde{\psi}_k^{\mathbf{x}}| \leq C$ for some constant C and for every $\ell = 1, \dots, n$.

Next, suppose that (6.12) holds true for $\tilde{\psi}_{k+1}^{\mathbf{x}}$, we know from (6.9) that

$$\tilde{\psi}_k^{\mathbf{x}} := \left(\tilde{\psi}_{k+1}^{\mathbf{X}^{k, \mathbf{x}}} - \tilde{\psi}_{k+1}^{\mathbf{X}^{k, \mathbf{x}, 0}} \mathbf{1}_{\{\tilde{N}^k > 0\}} \right) \prod_{j=1}^{\tilde{N}^k} \frac{\mu_k(\tilde{T}_j^k, \tilde{X}_j^{k, \mathbf{x}}) - \mu_k(\tilde{T}_{j-1}^k, \tilde{X}_{j-1}^{k, \mathbf{x}})}{\beta \Delta \tilde{T}_{j+1}^k} \cdot (\sigma_0^T)^{-1} \Delta \tilde{W}_{j+1}^k.$$

Then using the estimation in Lemma 6.6, we see that (6.12) is also true for $\tilde{\psi}_k^{\mathbf{x}}$, and we hence conclude the proof. \square

Proof of Theorem 6.5. (i) By Lemma 6.7, we know that $x \mapsto \tilde{\psi}_2^{x, x}$ is differential and in particular uniformly Lipschitz with coefficient bounded by $2C \prod_{j=2}^n (\tilde{N}^j + 1)^{j-1}$. Then the definition of $\tilde{\psi}_1^{x_0}$ falls into the Markovian case $n = 1$, but with terminal condition $x \mapsto \tilde{\psi}_2^{x, x}$. Notice that $\tilde{N}^k \leq N_T$ which admits a Poisson distribution: $\mathbb{P}(N_T = m) = e^{-\beta T} \frac{(\beta T)^m}{m!}$. It follows that, for some constant $C > 0$,

$$\mathbb{E}[|\tilde{\psi}_1^{x_0}|^2] \leq \mathbb{E}\left[C^{\tilde{N}^k} 4C^2 \prod_{j=2}^n (\tilde{N}^j + 1)^{2(j-1)}\right] \leq \mathbb{E}\left[4C^2 C^{N_T} (N_T + 1)^{n(n-1)}\right] < \infty,$$

which implies that $\tilde{\psi}$ admits finite variance.

(ii) Finally, with the above integrability analysis, using the results in Theorem 3.2 together with the PDE system (6.8) in Remark 6.3, we can conclude the proof of $\tilde{V}_0 = \mathbb{E}[\tilde{\psi}]$. \square

A Appendix

We provide an estimation on the order statistics of uniform distribution on $[0, 1]$, which induces an estimation on a functional of the arrival times $(T_k)_{k \geq 0}$ of the Poisson process.

Lemma A.1. Let $(U_k)_{k=1,\dots,m}$ be a sequence of i.i.d. random variable of uniform distribution on $[0, 1]$, and $(U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(m)})$ be the associated order statistics. Then

$$G_m := \mathbb{E} \left[\frac{1}{\sqrt{U_{(1)}}} \frac{1}{\sqrt{U_{(2)} - U_{(1)}}} \cdots \frac{1}{\sqrt{U_{(m)} - U_{(m-1)}}} \right] \leq m! 2^m.$$

Proof. First, we notice that for any $x \in (0, 1)$,

$$\int_x^1 \frac{1}{\sqrt{u-x}} du = 2\sqrt{1-x} \leq 2. \quad (\text{A.1})$$

Then, since the density of the order statistics $(U_{(1)}, \dots, U_{(m)})$ is provided by

$$f(u_1, \dots, u_m) := m! \mathbf{1}_{\{0 < u_1 < u_2 < \dots < u_m < 1\}},$$

it follows by direct computation that

$$G_m = m! \int_0^1 \int_{U_{(1)}}^1 \cdots \int_{U_{(m-1)}}^1 \frac{1}{\sqrt{u_1}} \frac{1}{\sqrt{u_2 - u_1}} \cdots \frac{1}{\sqrt{u_m - u_{m-1}}} du_1 \cdots du_m \leq m! 2^m,$$

where the last inequality is from (A.1). \square

Let $N = (N_s)_{s \geq 0}$ be a Poisson process with arrival times $(T_k)_{k > 0}$, denote $\Delta T_{k+1} := T_{k+1} - T_k$.

Lemma A.2. For every constant $C > 0$, we have

$$\mathbb{E} \left[\prod_{k=1}^{N_T} \sqrt{\frac{C}{\Delta T_{k+1}}} \right] < \infty.$$

Proof. (i) Let $0 = t_0 < t_1 < \dots < t_n = T < \infty$ be a discrete time grid, we define

$$\tilde{T}_k := \min(T_k, t_i), \text{ whenever } T_{k-1} \in [t_{i-1}, t_i) \text{ for some } i = 1, \dots, n,$$

and

$$\Delta \tilde{T}_k := \tilde{T}_k - \tilde{T}_{k-1} \text{ for every } k = 2, \dots, N_T + 1.$$

Notice that $\Delta T_k \geq \Delta \tilde{T}_k$ for all $k = 1, \dots, N_T + 1$. Then we can replace ΔT_{k+1} by $\Delta \tilde{T}_{k+1}$ in the statement of lemma. Moreover, one can always add points into the time grid $0 = t_0 < t_1 < \dots < t_n = T$, which makes $\Delta \tilde{T}_{i+1}$ smaller. Therefore, one can suppose without loss of generality that $t_k - t_{k-1} < \frac{1}{4\beta^2 C}$ for every $k = 1, \dots, n$.

(ii) For every $k = 1, \dots, n$, we denote $N^k := \#\{i : T_i \in [t_{k-1}, t_k)\}$, and $\tilde{T}_i^k := \tilde{T}_{k_i}$ with $k_i := \sum_{j < k} N^j + i$ for $i = 1, \dots, N^k + 1$, and $\Delta \tilde{T}_i^k := \tilde{T}_i^k - \tilde{T}_{i-1}^k$. By the memoryless property of the exponential distribution, it is clear that $(\Delta \tilde{T}_i^1, i = 2, \dots, N^1 + 1), \dots, (\Delta \tilde{T}_i^n, i = 2, \dots, N^n + 1)$ are mutually independent. Moreover, we have

$$\prod_{i=1}^{N_T} \sqrt{\frac{C}{\Delta \tilde{T}_{i+1}}} = \prod_{k=1}^n \left(\prod_{i=1}^{N^k} \sqrt{\frac{C}{\Delta \tilde{T}_{i+1}^k}} \right). \quad (\text{A.2})$$

Next, the law of $(T_i^k, i = 1, \dots, N^k)$ conditioning on $N^k = m$ is the law of order statistics of uniform distribution on $[t_{k-1}, t_k]$. Then it follows by Lemma A.1 that for every $k = 1, \dots, n$,

$$\begin{aligned} \mathbb{E} \left[\prod_{i=1}^{N^k} \sqrt{\frac{C}{\Delta \tilde{T}_{i+1}^k}} \right] &\leq e^{-\beta(t_k - t_{k-1})} \sum_{m=0}^{\infty} \frac{(\beta(t_k - t_{k-1}))^m}{m!} m! 2^m \left(\sqrt{\frac{C}{t_k - t_{k-1}}} \right)^m \\ &= e^{-\beta(t_k - t_{k-1})} \sum_{m=0}^{\infty} \left(2\beta \sqrt{C(t_k - t_{k-1})} \right)^m < \infty, \end{aligned}$$

where the last inequality follows by the fact $t_k - t_{k-1} < \frac{1}{4\beta^2 C}$. We then conclude the proof by (A.2). \square

References

- [1] Alanko, S., Avellaneda, M. : *Reducing variance in the numerical solution of BSDEs*, Comptes Rendus Mathématique Vol. 351, n 3-4:135-138, 2013.
- [2] Bally, V., Kohatsu-Higa, A. : *A probabilistic interpretation of the parametrix method*, preprint, 2014.
- [3] Beskos, A., Roberts, G. O. : *Exact simulation of diffusions*, Ann. Appl. Probab. Volume 15, Number 4, pp. 2422-2444, 2005.
- [4] Beskos, A., Papaspiliopoulos, O., Roberts, G. O. : *Retrospective exact simulation of diffusion sample paths*, Bernoulli 12(6), December 2006.
- [5] Broadie, M., Glasserman, P.: *Estimating Security Price Derivatives Using Simulation*, Management Science, 269-285, 1996.
- [6] Dupire, B. : *Functional Itô calculus*, ssrn, 2009, <http://ssrn.com/abstract=1435551>.
- [7] Ekren, I., Keller, C., Touzi, N. and Zhang, J. : *On Viscosity Solutions of Path Dependent PDEs*, Anna. Proba. 2014.
- [8] Ekren, I., Touzi, N. and Zhang, J. : *Viscosity Solutions of Fully Nonlinear Path Dependent PDEs: Part I*, preprint.
- [9] Fahim, A., Touzi, N., Warin, X. : *A probabilistic numerical method for fully nonlinear parabolic PDEs*, Ann. Appl. Probab. Volume 21, Number 4, 1322-1364, 2011.
- [10] Fournié, E., Lasry, J. M., Lebuchoux, J., Lions, P. L., and Touzi, N. : *Applications of Malliavin calculus to Monte Carlo methods in finance*, Finance Stoch. 3:391-412, 1999.
- [11] Henry-Labordère, P. : *Counterparty Risk Valuation: A Marked Branching Diffusion Approach*, Risk magazine (Jul. 2012).
- [12] Henry-Labordère, P., Touzi, N., Tan, X. : *A numerical algorithm for a class of BSDE via branching process*, Stochastic Processes and their Applications, 124:1112-1140, 2014.

- [13] Jourdain, B., Sbair, M. : *Exact retrospective Monte Carlo computation of arithmetic average Asian options*, Monte Carlo Methods and Applications 13, 2:135-171, (2007).
- [14] Kloeden, P., Platen, E. : *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics, Volume 23, Springer, 1992.