



**HAL**  
open science

# Unbiased simulation of stochastic differential equations \*

Pierre Henry-Labordère, Xiaolu Tan, Nizar Touzi

► **To cite this version:**

Pierre Henry-Labordère, Xiaolu Tan, Nizar Touzi. Unbiased simulation of stochastic differential equations \*. 2017. hal-01429548

**HAL Id: hal-01429548**

**<https://hal.science/hal-01429548v1>**

Preprint submitted on 8 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Unbiased simulation of stochastic differential equations <sup>\*</sup>

Pierre Henry-Labordère<sup>†</sup>    Xiaolu Tan<sup>‡</sup>    Nizar Touzi<sup>§</sup>

March 8, 2016

## Abstract

We propose an unbiased Monte-Carlo estimator for  $\mathbb{E}[g(X_{t_1}, \dots, X_{t_n})]$ , where  $X$  is a diffusion process defined by a multi-dimensional stochastic differential equation (SDE). The main idea is to start instead from a well-chosen simulatable SDE whose coefficients are updated 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 Bismu-Elworthy-Li formula from Malliavin calculus, as exploited by Fournié et al. [14] for the simulation of the Greeks in financial applications. In particular, this algorithm can be considered as a variation of the (infinite variance) estimator obtained in Bally and Kohatsu-Higa [3, Section 6.1] as an application of the parametrix method.

**MSC2010.** Primary 65C05, 60J60; secondary 60J85, 35K10.

**Key words.** Unbiased 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  $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{M}^d$  be the drift and diffusion coefficients, where  $\mathbb{M}^d$  denotes the collection of all  $d \times d$  dimensional matrices. Under standard assumptions on these

---

<sup>\*</sup>We are grateful to Emmanuel Gobet, Ahmed Kebaier and two anonymous referees for valuable remarks and suggestions. 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).

<sup>†</sup>Société Générale, Global Market Quantitative Research, pierre.henry-labordere@sgcib.com

<sup>‡</sup>CEREMADE, University of Paris-Dauphine, PSL Research University, tan@ceremade.dauphine.fr

<sup>§</sup>Ecole Polytechnique Paris, Centre de Mathématiques Appliquées, nizar.touzi@polytechnique.edu

coefficients, we consider the process  $X$  defined as the unique strong solution of the multi-dimensional SDE,

$$X_0 = x_0, \quad \text{and} \quad dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t, \quad (1.1)$$

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

$$V_0 := \mathbb{E}[g(X_{t_1}, \dots, X_{t_n})], \quad (1.2)$$

for some function  $g : \mathbb{R}^{d \times n} \rightarrow \mathbb{R}$  and discrete time grid  $0 < t_1 < \dots < t_n = T$ . When  $n = 1$ , the analytic formulation of the problem is obtained by the well-known representation  $V_0 = u(0, X_0)$ , where  $u$  is the solution of the linear PDE

$$\partial_t u + b(t, x) \cdot Du + \frac{1}{2} \text{Tr}[\sigma \sigma^\top(t, x) D^2 u] = 0, \quad u_T = g, \quad (1.3)$$

In practice, the Monte-Carlo method consists in simulating  $N$  independent copies of a discrete-time approximation of  $X$ , and then estimating  $V_0$  by the empirical mean value of the simulations. The corresponding error analysis consists of a statistical error induced by the central limit theorem, and a discretization error which induces a biased Monte Carlo approximation. Under some smoothness conditions, Talay and Tubaro [25] proved that the discretization error for the Euler scheme is controlled by a rate  $C\Delta t$ , where  $\Delta t$  denotes the time step discretization. Since then, many works focused on the analysis of the discretization error under various discretization techniques, see e.g. Kloeden and Platen [23], and Graham and Talay [17] for an overview. However, the statistical error estimate  $N^{-\frac{1}{2}}$  is lost in all cases, as its combination with the discretization error leads to an overall error estimate of the order  $N^{-\frac{1}{2}+\varepsilon}$  for some  $\varepsilon > 0$ .

In the context of one-dimensional homogeneous SDEs with constant volatility coefficient, Beskos and Roberts [6] developed an exact simulation technique for  $X$  by using the Girsanov change measure together with a rejection algorithm, see also Beskos, Paspiliopoulos and Roberts [7], Jourdain and Sbai [22], etc... This technique also applies to more general SDEs by using of the so-called Lamperti transformation which reduces the approximation problem to the unit diffusion case. We also refer to the subsequent active literature of exact simulation of an  $\mathbb{L}^\infty$ -approximation of  $X$ , see. e.g Blanchet, Chen and Dong [5].

An alternative approximation method for  $V_0$  was induced by the multilevel Monte-Carlo (MLMC) algorithm introduced by Giles [15], which generalizes the statistical Romberg method of Kebaier [18]. One of the main advantages of the MLMC algorithm is to control the global error (sum of discretization error and statistical error) with a much better rate w.r.t. the computation complexity. We refer to Giles and Szpruch [16], Alaya and Kebaier [1], Rhee and Glynn [24] for further developments. In particular, Rhee and Glynn [24] obtained an unbiased simulation method for SDEs by using a random level in the estimator.

More recently, Bally and Kohatsu-Higa [3] provided a probabilistic interpretation of the parametrix method for PDEs. In particular, when  $n = 1$ , they obtained a representation formula for  $V_0$  in form

$$\mathbb{E}[g(\widehat{X}_T) \mathcal{W}_T], \quad (1.4)$$

where  $\widehat{X}$  is defined by a Euler scheme of  $X$  on a random discrete-time grid (the time step follows an independent exponential distribution), and  $\mathcal{W}_T$  is a corrective weight function depending on  $\widehat{X}$ . The above representation is formally similar to the stochastic finite element method proposed by Bompis and Gobet [8], where one replaces  $X$  by its Euler scheme solution in (1.2) and then corrects partially the error by some well-chosen weight functions. Notice that in the above representation of Bally and Kohatsu-Higa [3], the process  $\widehat{X}$  can be exactly simulated and hence it may provide an unbiased estimator for  $V_0$ . Nevertheless, the obtained weight function  $\mathcal{W}_T$  is integrable but has infinite variance, and hence the corresponding Monte-Carlo estimator loses the standard central limit error estimate.

In this paper, we provide a representation of  $V_0$  in the spirit of (1.4), but with an alternative weight function for the representation. Our results follow from completely different arguments. More importantly, our unbiased approximation of  $V_0$  has finite variance, and applies for a large class of SDEs.

Our main idea is to consider the Euler scheme solution  $\widehat{X}$  as solution to a regime-switching SDE with some well-chosen coefficients. In order to compensate for the change of the coefficients of the SDE, we introduce some weight functions obtained by the automatic differentiation technique induced by Bismut-Elworthy-Li formula from Malliavin calculus, as exploited by Fournié et al. [14] for the simulation of the Greeks in financial applications.

The technique introduced in the present paper is inspired by the numerical algorithm introduced in [19, 21], for semilinear PDEs of the form

$$\partial_t u + \frac{1}{2} \Delta u + F_0(t, x, u) = 0, \quad u_T = g,$$

for some nonlinearity  $F_0$ . The main idea in [19, 21] is to use an approximation by a branching diffusion representation induced by approximating the nonlinearity  $F_0$  by a polynomial in  $u$ . Namely, given the nature of the linear operator, the representation is obtained by means of a Brownian motion with branching driven by the polynomial approximation of  $F_0$ .

Loosely speaking, the method developed in the present paper follows by reading the PDE part of (1.3) in the following equivalent form:

$$\partial_t u + \frac{1}{2} \Delta u + F_1(t, x, Du, D^2u) = 0,$$

where

$$F_1(t, x, z, \gamma) := b(t, x) \cdot z + \frac{1}{2} \text{Tr}[(\sigma \sigma^\top(t, x) - I)\gamma].$$

However, in contrast with the nonlinearity  $F_0$ , the above function  $F_1$  involves the gradient and the Hessian of the solution  $u$ . Consequently the last PDE can not be handled by the existing literature on branching diffusion representation of PDEs. The automatic differentiation technique introduced in the present paper is an important new idea which allows to convert  $Du$  and  $D^2u$  in  $F_1$  into  $u$ . Since no powers of  $u$  are involved in the equation, this leads to a representation by means of a Brownian motion

with exactly one descendent with two different possible types revealed by the weight function corresponding to the order of differentiation.

We believe that the automatic differentiation trick introduced here has very important consequences, beyond the particular application of the present paper. Indeed, in our paper [20], we provide a significant extension of the branching diffusion representation to a general class of semilinear PDEs.

The paper is organized as follows. In Section 2, we consider the SDE with constant diffusion coefficient, and propose an unbiased estimator for  $V_0$  for both Markovian case and path-dependent case. Then in Section 3, we consider the SDE with general diffusion coefficient function, and obtain a similar representation formula for  $V_0$ , which is integrable but of infinite variance. Next, in Section 4, we provide some numerical examples. Finally, we complete some technical proofs in Section 5. In particular, an easy example is studied in Section 5.1 to illustrate the main idea of the technical proofs.

## 2 Unbiased simulation of SDE with constant diffusion coefficient

In this section, we will restrict to the constant diffusion coefficient case, and propose an unbiased estimator for  $V_0$  having finite variance.

### 2.1 The Markovian case

Let us start by the Markovian case, where the diffusion process  $X$  is defined by

$$X_0 = x_0, \quad dX_t = \mu(t, X_t) dt + \sigma_0 dW_t, \quad (2.1)$$

for some matrix  $\sigma_0 \in \mathbb{M}^d$ , and our objective is to compute

$$V_0 = \mathbb{E}[g(X_T)]. \quad (2.2)$$

for some function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ . We impose the following conditions on  $\mu$  and  $\sigma_0$ :

**Assumption 2.1.** *The diffusion coefficient  $\sigma_0$  is non-degenerate, 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  $L > 0$ ,*

$$\left| \mu(t, x) - \mu(s, y) \right| \leq L \left( \sqrt{|t - s|} + |x - y| \right), \quad \forall (s, x), (t, y) \in [0, T] \times \mathbb{R}^d. \quad (2.3)$$

#### 2.1.1 The unbiased simulation algorithm

To introduce our unbiased simulation algorithm, let us first introduce a random discrete time grid. 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. We define

$$T_k := \left( \sum_{i=1}^k \tau_i \right) \wedge T, \quad k \geq 0, \quad \text{and} \quad N_t := \max \{ k : T_k < t \}. \quad (2.4)$$

Then  $(N_t)_{0 \leq t \leq T}$  is a Poisson process with intensity  $\beta$  and arrival times  $(T_k)_{k > 0}$ . We denote also  $T_0 := 0$  and  $\Delta T_{k+1} := T_{k+1} - T_k$ .

Let  $W$  be a  $d$ -dimensional Brownian motion independent of  $(\tau_i)_{i > 0}$ , we introduce

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

and a process  $\widehat{X}$  as the Euler scheme of  $X$  on the random discrete grid  $(T_k)_{k \geq 0}$ , i.e.  $\widehat{X}_0 = x_0$  and

$$\widehat{X}_{T_{k+1}} := \widehat{X}_{T_k} + \mu(T_k, \widehat{X}_{T_k}) \Delta T_{k+1} + \sigma_0 \Delta W_{T_{k+1}}, \quad k = 0, 1, \dots, N_T. \quad (2.5)$$

Then our estimator is given by

$$\widehat{\psi} := e^{\beta T} \left[ g(\widehat{X}_T) - g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right] \beta^{-N_T} \prod_{k=1}^{N_T} \overline{\mathcal{W}}_k^1, \quad (2.6)$$

with

$$\overline{\mathcal{W}}_k^1 := \frac{(\mu(T_k, \widehat{X}_{T_k}) - \mu(T_{k-1}, \widehat{X}_{T_{k-1}})) \cdot (\sigma_0^\top)^{-1} \Delta W_{T_{k+1}}}{\Delta T_{k+1}}. \quad (2.7)$$

**Theorem 2.2.** *Suppose that Assumption 2.1 holds true, and  $g$  is Lipschitz. Then*

$$\mathbb{E}[(\widehat{\psi})^2] < \infty \quad \text{and} \quad V_0 = \mathbb{E}[\widehat{\psi}].$$

**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  $L_0 := |(\sigma_0 \sigma_0^\top)^{-1}| > 0$  by the non-degeneracy of  $\sigma_0$ . Then using Assumption 2.1, it follows by direct computation that

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

Then denoting  $\widehat{\mathbb{E}}_{T_k} := \mathbb{E}[\cdot | \widehat{X}_{T_k}, \Delta T_{k+1}]$ , we have

$$\widehat{\mathbb{E}}_{T_k} \left[ \left| \frac{\sqrt{\Delta T_{k+1}} + |\Delta \widehat{X}_{T_{k+1}}|}{\Delta T_{k+1}} (\sigma_0^\top)^{-1} \Delta W_{T_{k+1}} \right|^2 \right] \leq \mathbb{E} \left[ (1 + |\mu|_\infty \sqrt{T} + |\sigma_0 Z|)^2 |(\sigma_0^\top)^{-1} Z|^2 \right],$$

where  $|\mu|_\infty := \sqrt{\sum_{i=1}^d |\mu_i|_0^2}$ ,  $|\mu_i|_0 := \sup_{t,x} |\mu_i(t, x)|$ , and  $Z$  is a standard centered normal distribution in  $\mathbb{R}^d$ . This provides

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

We therefore get the following upper bound:

$$\mathbb{E}[\widehat{\psi}^2] \leq C e^{2\beta T} e^{-\beta T + \frac{\gamma L^2 T}{\beta}}, \quad \text{where } C := L_g^2 \mathbb{E} \left[ (|g(0)| + \Delta T_1 + |\Delta X_1|)^2 \right]. \quad (2.8)$$

(ii) The equality  $V_0 = \mathbb{E}[\widehat{\psi}]$  will be proved in Section 5, with illustration of the main idea in Section 5.1.  $\square$

### 2.1.2 On the choice of $\beta$

Notice that the random variable  $\widehat{\psi}$  in (2.6) can be exactly simulated from a sequence of Gaussian  $\mathcal{N}(0,1)$  and exponential  $\mathcal{E}(\beta)$  random variables. Then the integrability and representation results in Theorem 2.2 induce an unbiased simulation Monte-Carlo method to approximate  $V_0$ , with error induced by the standard central limit theorem.

We next observe that the constant  $\beta > 0$  may be chosen so as to minimize the approximation error relative to the computational effort:

- By the central limit theorem, the error induced by the Monte Carlo estimator based on the representation  $\widehat{\psi}$  is characterized by the variance of  $\widehat{\psi}$ . For tractability reasons, we shall instead replace it by the bound (2.8).
- The computation effort is proportional to the number  $N_T$  of arrivals of the Poisson process before the maturity  $T$ , and is thus given by  $C'\mathbb{E}[N_T] = C'\beta T$ .

In view of this, we shall choose  $\beta$  by minimizing the ratio of the variance bound (2.8) to the mean computational effort. This minimization problem is obviously independent of the constants  $C, C'$ , and reduces to:

$$\min_{\beta > 0} f(\beta), \quad \text{where} \quad f(\beta) := \frac{1}{\beta T} \exp\left(T\left(\beta + \frac{\gamma L^2}{\beta}\right)\right).$$

Direct computation shows that the equation  $f'(\beta) = 0$  has a unique solution on  $(0, \infty)$  given by

$$\beta^* := \sqrt{\gamma L^2 + T^2/4} + \frac{T}{2}.$$

As  $\lim_{\beta \searrow 0} f(\beta) = \lim_{\beta \rightarrow \infty} f(\beta) = \infty$ , this shows that  $\beta^*$  is the minimizer of the above defined criterion, and will be taken as our “best sub-optimal” choice of  $\beta$  for the unbiased estimator  $\widehat{\psi}$ .

## 2.2 The path-dependent case

In this part, we would like to provide an extension of the above estimator  $\widehat{\psi}$  in (2.6) to the path-dependent case. Let  $n > 0$ ,  $0 = t_0 < t_1 < \dots < t_n = T$ ,  $\sigma_0 \in \mathbb{M}^d$  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  $X_0 = x_0$ ,

$$dX_t = \mu(t, X_{t_1 \wedge t}, \dots, X_{t_n \wedge t}) dt + \sigma_0 dW_t; \quad (2.9)$$

and the objective is to compute the value,

$$\widetilde{V}_0 := \mathbb{E}[g(X_{t_1}, \dots, X_{t_n})], \quad (2.10)$$

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

**Remark 2.3.** It is clear that the value  $\tilde{V}_0$  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

$$\mu_k(t, x) := \mu(t, x_1, \dots, x_{k-1}, x, \dots, x), \quad \forall (t, x) \in [t_{k-1}, t_k] \times \mathbb{R}^d. \quad (2.11)$$

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 a solution (at least in the viscosity sense) of

$$\partial_t u_k + \frac{1}{2} \sigma_0 \sigma_0^\top : D^2 u_k + \mu_k \cdot D u_k = 0, \quad (2.12)$$

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  $\tilde{V}_0 = u_1(0, x_0)$ .

## 2.2.1 The algorithm

The unbiased simulation algorithm of  $\tilde{V}_0$  can be obtained by an iteration of the estimator (2.6) on each time interval  $[t_k, t_{k+1}]$ . One should just be careful on the integrability issue. Let us first introduce the algorithm.

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 (2.4). Define further for every  $k = 1, \dots, n$ ,  $\tilde{N}^k := N_{t_k} - N_{t_{k-1}}$  the number of jump arrivals on  $[t_{k-1}, t_k)$ , and  $\tilde{T}_0^k := t_{k-1}$  and  $\tilde{T}_j^k := T_{N_{t_{k-1}} + j} \wedge t_k$ ,

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

**Example 2.4.** We give below an example for the case  $n = 2$ . In the following example, the number of jump arrivals on  $[0, t_1)$  is  $\tilde{N}^1 = 2$ , that on  $[t_1, t_2)$  is  $\tilde{N}^2 = 1$ , and total number of jump arrivals 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  $\tilde{T}_3^1 = t_1$ ;  $\tilde{W}_0^1 = 0$ ,  $\tilde{W}_1^1 = W_{T_1}$ ,  $\tilde{W}_2^1 = W_{T_2}$  and  $\tilde{W}_3^1 = W_{t_1}$ . For  $k = 2$ , we have  $\tilde{T}_0^2 = t_1$ ,  $\tilde{T}_1^2 = T_3$ ,  $\tilde{T}_2^2 = t_2$ , and  $\tilde{W}_0^2 = W_{t_1}$ ,  $\tilde{W}_1^2 = W_{T_3}$  and  $\tilde{W}_2^2 = W_{t_2}$ .



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

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

Similarly, for every  $j = 1, \dots, N_k$ , we define a automatic differentiation weight, with  $\mu_k$  defined by (2.11),

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



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

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

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

$$\tilde{\psi}_k^{\mathbf{x}} := e^{\beta(t_k - t_{k-1})} \left( \tilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x}}} - \tilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x},0}} \mathbf{1}_{\{\tilde{N}^k > 0\}} \right) \beta^{-\tilde{N}^k} \prod_{j=1}^{\tilde{N}^k} \tilde{\mathcal{W}}_j^k. \quad (2.13)$$

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

$$\tilde{\psi} := \tilde{\psi}_1^{x_0}. \quad (2.14)$$

## 2.2.2 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 suggested by the PDEs (2.12) in Remark (2.3). When the random variable  $\tilde{\psi}$  in (2.14) is integrable, it is not surprising to obtain the representation  $\tilde{V}_0 = \mathbb{E}[\tilde{\psi}]$  as a consequence of Theorem 2.2. 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 (2.13)), the variance analysis becomes less obvious. We provide here a sufficient condition to ensure that  $\tilde{\psi}$  has finite variance.

**Theorem 2.5.** *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 differentiable up to the order  $n$ , with bounded derivatives. Then*

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

We will prove the integrability result here, and leave the proof of the representation result  $\tilde{V}_0 := \mathbb{E}[\tilde{\psi}]$  in Section 5. As preparation, let us 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

$$X_{k+1}^{\pi,x} := X_k^{\pi,x} + \bar{\mu}(s_k, X_k^{\pi,x}) \Delta s_{k+1} + W_{s_{k+1}} - W_{s_k}. \quad (2.15)$$

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 2.6.** *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 (2.15) with initial condition  $X_0^{\pi,x} = x$ . Then  $x \mapsto X_k^{\pi,x}$  is differentiable up to order  $n$  and there is a constant  $C$  independent of the partition  $\pi$  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  $\pi$ . By induction, it is easy to deduce that for  $\ell = 2, \dots, n$ ,

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

where  $P_\ell$  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  $\pi$ . Hence  $\partial_{x^\ell} X_k^{\pi,x}$  is also bounded by some constant  $C_\ell$  independent of  $k = 1, \dots, m$  and the partition  $\pi$ .  $\square$

**Lemma 2.7.** Let  $(\tilde{\psi}_k^{\mathbf{x}})_{1 \leq k \leq n+1}$  be defined by (2.13). 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  $x_{k-1} \mapsto \tilde{\psi}_k^{\mathbf{x}}$  has derivatives up to order  $k - 1$  and

$$\max_{1 \leq \ell \leq k-1} \left| \partial_{x_{k-1}, i}^\ell \tilde{\psi}_k^{\mathbf{x}} \right| \leq C \prod_{j=k}^n (\tilde{N}^j + 1)^{j-1}. \quad (2.16)$$

**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  $|\partial_{x_n}^\ell \tilde{\psi}^{\mathbf{x}}| \leq C$  for some constant  $C$  and for every  $\ell = 1, \dots, n$ .

Next, suppose that (2.16) holds true for  $\tilde{\psi}_{k+1}^{\mathbf{x}}$ , we know from (2.13) that

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

Then using the estimation in Lemma 2.6, we see that (2.16) is also true for  $\tilde{\psi}_k^{\mathbf{x}}$ , and we hence conclude the proof.  $\square$

**Proof of Theorem 2.5 (i).** By Lemma 2.7, we know that  $x \mapsto \tilde{\psi}_2^{x,x}$  is differentiable 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  $\tilde{N}^k \leq N_T$  has 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  $\tilde{\psi}$  has finite variance.  $\square$

### 3 Unbiased simulation of general SDEs

Let us now consider the SDE (1.1) with general diffusion coefficient function, i.e. with drift and diffusion coefficients  $\mu : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{M}^d$ :

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

Our objective of study in this section is

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

We will provide a representation result of  $V_0$  in the same spirit of that in Section 2.

**Remark 3.1** (Lamperti's transformation). *We also notice that in some cases, the above SDE (1.1) may be reduced to the constant diffusion coefficient case (2.1), 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 Itô'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 (2.1).*

(ii) *When  $d > 1$ ,  $\sigma$  is non-degenerate and satisfies some further compatibility conditions, one can also obtain a similar transformation to reduce SDE (5.8) to the constant diffusion coefficient case.*

#### 3.1 An estimator of infinite variance for general SDEs

Let us impose the following conditions on coefficient functions  $\mu$  and  $\sigma$ .

**Assumption 3.2.** *The function  $(\mu, \sigma) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{M}^d$  and  $a := \frac{1}{2} \sigma \sigma^\top : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{M}^d$  are uniformly bounded, and are uniformly Hölder in the time variable, uniformly Lipschitz in the space variable, i.e. for some constant  $L$ ,*

$$|(\mu, \sigma, a)(t, x) - (\mu, \sigma, a)(s, y)| \leq L(\sqrt{|t-s|} + |x-y|), \quad (3.1)$$

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

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

Recall that  $(T_k)_{k \geq 0}$  are defined by (2.4) with a sequence of i.i.d.  $\mathcal{E}(\beta)$ -exponential random variables, and  $W$  is a Brownian motion; the increment of the Brownian motion are defined by  $\Delta W_{t_k} := W_{t_k} - W_{t_{k-1}}$ , and  $\Delta T_k := T_k - T_{k-1}$ . As in (2.5), we introduce  $\widehat{X}$  as solution of the Euler scheme on discrete grid by  $\widehat{X}_0 := x_0$  and

$$\widehat{X}_{T_{k+1}} := \widehat{X}_{T_k} + \mu(T_k, \widehat{X}_{T_k}) \Delta T_{k+1} + \sigma(T_k, \widehat{X}_{T_k}) \Delta W_{T_{k+1}} \quad k = 0, \dots, N_T. \quad (3.2)$$

We then introduce a representation formula by

$$\widehat{\psi} := e^{\beta T} [g(\widehat{X}_T) - g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}}] \beta^{-N_T} \prod_{k=1}^{N_T} (\overline{\mathcal{W}}_k^1 + \overline{\mathcal{W}}_k^2), \quad (3.3)$$

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

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

and

$$\begin{aligned} \overline{\mathcal{W}}_k^2 &:= [a(T_k, \widehat{X}_{T_k}) - a(T_{k-1}, \widehat{X}_{T_{k-1}})] \\ &: \left[ (\sigma^\top(T_k, \widehat{X}_{T_k}))^{-1} \frac{\Delta W_{T_{k+1}} \Delta W_{T_{k+1}}^\top - \Delta T_{k+1} I_d}{\Delta T_{k+1}^2} \sigma(T_k, \widehat{X}_{T_k})^{-1} \right]. \end{aligned} \quad (3.4)$$

**Theorem 3.3.** *Suppose that Assumption 3.2 holds true, and  $g$  is Lipschitz. Then*

$$\mathbb{E}[\widehat{\psi}] < \infty \quad \text{and} \quad V_0 = \mathbb{E}[\widehat{\psi}].$$

**Proof.** (i) Consider the random vectors  $\xi_k^1 := \frac{\Delta W_{T_k}}{\sqrt{\Delta T_k}}$  and  $\xi_k^2 := \frac{\Delta W_{T_k} \Delta W_{T_k}^\top - \Delta T_k I_d}{\Delta T_k}$ , for all  $k = 1, \dots, N_T + 1$ , which are independent of  $\Delta T_k$  conditional on  $\{\Delta T_k > 0\} = \{N_T \geq k - 1\}$ , and which have finite second order moment. Notice that  $\mu(t, x)$  and  $a(t, x)$  are uniformly bounded, and 1/2-Hölder-continuous in  $t$  and Lipschitz in  $x$ , and  $\sigma$  is uniformly bounded from below above zero. Then, for each  $k = 1, \dots, N_T$ ,

$$|\overline{\mathcal{W}}_k^1| \leq C(\sqrt{\Delta T_k} + |\widehat{X}_{T_k} - \widehat{X}_{T_{k-1}}|) \left| \frac{\Delta W_{T_{k+1}}}{\Delta T_{k+1}} \right| \leq C(1 + \sqrt{\Delta T_k} + |\xi_k^1|) |\xi_{k+1}^1| \sqrt{\frac{\Delta T_k}{\Delta T_{k+1}}},$$

where the constant  $C > 0$  may vary from term by term but is uniformly bounded for all  $k$ . Similarly, one obtains that

$$|\overline{\mathcal{W}}_k^2| \leq C(1 + \sqrt{\Delta T_k} + |\xi_k^1|) |\xi_{k+1}^2| \sqrt{\frac{\Delta T_k}{\Delta T_{k+1}}} \frac{1}{\sqrt{\Delta T_{k+1}}}.$$

As  $\Delta T_k \leq T$ , it follows that

$$|\overline{\mathcal{W}}_k^1| + |\overline{\mathcal{W}}_k^2| \leq C(1 + |\xi_k^1|) (|\xi_{k+1}^1| + |\xi_{k+1}^2|) \sqrt{\frac{\Delta T_k}{\Delta T_{k+1}}} \frac{1}{\sqrt{\Delta T_{k+1}}},$$

for some constant  $C > 0$  independent of  $k$ . In addition, we have by the Lipschitz condition on  $g$  that

$$\mathbb{E} \left[ |g(\widehat{X}_T) - g(\widehat{X}_{T_{N_T}})| \mid \Delta T_{N_T+1} \right] < C \sqrt{\Delta T_{N_T+1}}.$$

Then, it follows from the expression of  $\widehat{\psi}$  in (3.3) that

$$\begin{aligned}\mathbb{E}[|\widehat{\psi}|] &\leq C\mathbb{E}\left[\prod_{k=1}^{N_T}\frac{C}{\sqrt{\Delta T_{k+1}}}\mathbf{1}_{\{N_T\geq 1\}}\right]+C\mathbb{E}\left[|g(\widehat{X}_T)|\mathbf{1}_{\{N_T=0\}}\right] \\ &\leq C\mathbb{E}\left[\prod_{k=1}^{N_T}\frac{C}{\sqrt{\Delta T_{k+1}}}\right]+C\mathbb{E}\left[|g(x_0+\mu(0,x_0)T+\sigma(0,x_0)W_T)|\right]\end{aligned}$$

for some constant  $C > 0$ , where we have also used the independence of the  $\xi_k^i$ 's and their the boundedness of their second order moments. The integrability of  $\widehat{\psi}$  is now a direct consequence of Lemma A.2.

(ii) The proof of the equality  $V_0 = \mathbb{E}[\widehat{\psi}]$  will be completed in Section 5.  $\square$

To conclude, we notice that the variable  $\widehat{\psi}$  is of order  $\prod_{k=1}^{N_T} 1/\sqrt{\Delta T_{k+1}}$  in general cases, and the latter is integrable but of infinite variance. Therefore,  $\widehat{\psi}$  is not a good estimator for Monte-Carlo method. Nevertheless, it should still have some theoretical value as an alternative representation formula obtained by Bally and Kohatsu-Higa [3, Section 6.1].

## 3.2 An estimator for one-dimensional driftless SDE

To overcome the problem of variance explosion of the estimator (3.3), we will consider the higher order approximation  $\widehat{X}$  of  $X$ , and obtain an estimator of finite variance for the one dimensional ( $d = 1$ ) driftless SDE of form

$$X_0 = x_0, \quad dX_t = \sigma(t, X_t) dW_t, \quad (3.5)$$

Our objective is to compute

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

Recall  $(T_k)_{k \geq 0}$  has been introduced in (2.4) from a sequence of i.i.d. exponential random variables, independent of the Brownian motion  $W$ . We next define  $\widehat{X}$  by  $\widehat{X}_0 = x_0$ ,

$$d\widehat{X}_t = \left( \sigma(T_k, \widehat{X}_{T_k}) + \partial_x \sigma(T_k, \widehat{X}_{T_k})(\widehat{X}_t - \widehat{X}_{T_k}) \right) dW_t, \quad \text{on } [T_k, T_{k+1}], \quad (3.6)$$

for  $k = 0, 1, \dots, N_T$ . By denoting

$$c_1^k := \sigma(T_k, \widehat{X}_{T_k}) - \partial_x \sigma(T_k, \widehat{X}_{T_k})\widehat{X}_{T_k} \quad \text{and} \quad c_2^k := \partial_x \sigma(T_k, \widehat{X}_{T_k}), \quad (3.7)$$

then the above linear SDE (3.6) has an explicit solution which is given by

$$\widehat{X}_{T_{k+1}} = \widehat{X}_{T_k} + \sigma(T_k, \widehat{X}_{T_k})\Delta W_{T_{k+1}}, \quad \text{if } c_2^k = 0, \quad (3.8)$$

and

$$\begin{aligned}\widehat{X}_{T_{k+1}} &= -\frac{c_1^k}{c_2^k} + \frac{c_1^k}{c_2^k} \exp\left(-\frac{(c_2^k)^2}{2}\Delta T_{k+1} + c_2^k \Delta W_{T_{k+1}}\right) \\ &\quad + \widehat{X}_{T_k} \exp\left(-\frac{(c_2^k)^2}{2}\Delta T_{k+1} + c_2^k \Delta W_{T_{k+1}}\right), \quad \text{if } c_2^k \neq 0. \quad (3.9)\end{aligned}$$

We then define  $\widehat{\psi}$  by

$$\widehat{\psi} := e^{\beta T} \left[ g(\widehat{X}_T) - g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right] \beta^{-N_T} \prod_{k=1}^{N_T} \overline{\mathcal{W}}_k^2, \quad (3.10)$$

where the automatic differentiation weight is given by (see Lemma 5.9 below)

$$\overline{\mathcal{W}}_k^2 := \frac{a(T_k, \widehat{X}_{T_k}) - \tilde{a}_k}{2a(T_k, \widehat{X}_{T_k})} \left( -\partial_x \sigma(T_k, \widehat{X}_{T_k}) \frac{\Delta W_{T_{k+1}}}{\Delta T_{k+1}} + \frac{\Delta W_{T_{k+1}}^2 - \Delta T_{k+1}}{\Delta T_{k+1}^2} \right), \quad (3.11)$$

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

Similarly to the discussion at the end of Section 3.1 (see also Remark 5.8 below), the variable  $\widehat{\psi}$  in (3.10) is integrable but of infinite variance in general. To make the variance finite, we 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

$$\overline{\psi} := \frac{\widehat{\psi} + \widehat{\psi}^-}{2} \quad \text{with} \quad \widehat{\psi}^- := e^{\beta T} \left[ g(\widehat{X}_T^-) - g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right] \beta^{-N_T} \prod_{k=1}^{N_T} \overline{\mathcal{W}}_k^-. \quad (3.12)$$

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.

**Assumption 3.4.** *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  $g(\cdot) \in C_b^2(\mathbb{R})$ .*

**Theorem 3.5.** *Suppose that Assumption 3.4 holds true. Then*

$$\mathbb{E}[\widehat{\psi}] + \mathbb{E}[\overline{\psi}^2] < \infty; \quad \text{and} \quad V_0 = \mathbb{E}[\widehat{\psi}] = \mathbb{E}[\overline{\psi}]. \quad (3.13)$$

We will complete the proof in Section 5.4.

**Remark 3.6.** (i) As  $\bar{\psi}$  has finite variance, we may use the representation of Theorem 3.5 to build an unbiased Monte-Carlo estimator of  $V_0$ . However, given the assumed regularity conditions, and the restriction to the one-dimensional setting, such a Monte-Carlo approximation is not competitive with the corresponding PDE based approximation methods. However, we believe that the present methodology is open to potential improvements, and we hope to improve our results in some future work so as to address the higher dimensions.

(ii) For a general SDE with drift function and/or  $d \geq 1$ , we can also consider a similar choice of  $(\hat{\mu}, \hat{\sigma})$ , which leads to  $\hat{\mu}(t, x) = c_1 + c_2x$  and  $\hat{\sigma}(t, x) = c_3 + c_4x$  and a linear SDE

$$d\hat{X}_t = (c_1 + c_2\hat{X}_t)dt + (c_3 + c_4\hat{X}_t)dW_t, \quad (3.14)$$

where  $c_1 \in \mathbb{R}^d$ ,  $c_2, c_3 \in \mathbb{M}^d$  and  $c_4$  is linear operator from  $\mathbb{R}^d$  to  $\mathbb{M}^d$ . However, to the best of our knowledge, the exact simulation of linear SDE (3.14) in high dimensional case, as well as the associated automatic differentiation (Malliavin) weight as in (3.11) (see also Lemma 5.9 below), is still an open question.

## 4 Numerical examples

Notice that our estimator  $\hat{\psi}$  given by (2.6) (resp.  $\tilde{\psi}$  given by (2.13) and (2.14)) is an unbiased estimator for  $V_0$  in (2.2) (resp.  $\tilde{V}_0$  in (2.10)). Then the error analysis of the Monte-Carlo approximation reduces to the statistical error. Hence the computation cost to achieve the accuracy  $O(\varepsilon)$  for the approximation of  $V_0$  (resp.  $\tilde{V}_0$ ) is of order  $O(\varepsilon^{-2})$ , thus avoiding of the dependence on the discretization error.

By combining different level of simulations, the MultiLevel Monte Carlo (MLMC) method proposed by Giles [15] achieves a computation cost of order  $O(\varepsilon^{-2}(\log \varepsilon)^2)$  or  $O(\varepsilon^{-2})$  depending on the strong discretization error rate. In particular, by considering a randomization of the level, Rhee and Glynn [24] obtained an unbiased estimator. In the following, we provide some numerical results and comparisons between our unbiased simulation method with the Euler based MLMC method proposed by [15].

### 4.1 Two one-dimensional SDEs

Let  $W$  be a one-dimensional standard Brownian motion, we consider the SDE given by

$$S_0 = 1, \quad dS_t = 0.1(\sqrt{M \wedge S_t} - 1)S_t dt + \frac{1}{2}S_t dW_t,$$

where  $M$  is a large constant introduced in order to guarantee the Lipschitz property of the drift coefficient (in our numerical implementation, we have observed that the value of  $M$  is not relevant for large  $M$ , and that the numerical finding are not changed by taking  $M = \infty$ ; this hints that our results may be extended beyond the case of Lipschitz coefficients). Applying Lemperti's transformation  $X_t := \log(S_t)$ , we reduce the above SDE to the constant diffusion coefficient case, in form of (2.1),

$$X_0 = 0, \quad dX_t = \left(0.1(\sqrt{M \wedge e^{X_t}} - 1) - 1/8\right)dt + \frac{1}{2}dW_t. \quad (4.1)$$

	Mean value	Statistical error	Computation time
US ( $N = 10^5$ )	0.204864	0.00140709	0.016814
MLMC	0.204993	0.000949166	0.032017
US ( $N = 10^6$ )	0.205396	0.000444462	0.171835
MLMC	0.205602	0.000308634	0.234526
US ( $N = 10^7$ )	0.20552	0.000142554	1.63013
MLMC	0.205648	0.0001	1.96197
US ( $N = 10^8$ )	0.205641	4.52282e-05	16.2189
MLMC	0.205638	3.18855e-05	18.3833

Table 1: Numerical results for  $V_0$  in (4.2) (case  $d = 1$ ), US denotes our unbiased simulation algorithm with  $\beta = 0.1$ , the computation times are expressed in second.

We implement our unbiased simulation method for the two following expectations:

$$V_0 := \mathbb{E}[(S_T - K)_+], \quad \text{and} \quad \tilde{V}_0 := \mathbb{E}\left[\left(\frac{1}{n} \sum_{k=1}^n S_{t_k} - K\right)_+\right], \quad (4.2)$$

where we choose  $K = 1$ ,  $T = 1$ ,  $n = 10$  and  $t_k := \frac{k}{n}T$ . Notice that the path-dependent example does not satisfy the differentiability sufficient condition in Theorem 2.5. However, our numerical findings do not show any numerical difficulty in the present setting.

Using different numbers  $N$  of simulations, we obtain the standard deviation as (statistical) error of our estimator. Next, using the errors obtained by our unbiased simulation method, we implement the MLMC algorithm in Section 5 of Giles [15], and we compare the computation time (in second) of the two methods. More precisely, the statistical error of the unbiased simulation method is given by  $\sqrt{\text{Var}[\hat{\psi}]/N}$ , where  $\text{Var}[\hat{\psi}]$  denotes the estimated variance of  $\hat{\psi}$ . For the implementation of MLMC, we choose  $M = 4$ ,  $N_L = 10^4$  and use equation (10) in [15] as criteria to stop the loop in MLMC (see more details in Section 5 of [15] for the meaning of  $M$  and  $N_L$ ).

The numerical results are given in Tables 1 and 2. We observe that with the same Monte-Carlo error, both methods have very close performance. In the present particular example, the computational time of our methods is slightly smaller. However, the conclusion may change depending on the nature of the example. Let us consider the problem

$$V_0 := \mathbb{E}[\sin(X_T)], \quad (4.3)$$

where  $X$  is defined by SDE, for some constant  $\mu_0 \in \mathbb{R}$ ,

$$X_0 = 0, \quad dX_t = \mu_0 \cos(X_t)dt + \frac{1}{2}dW_t.$$

We implement the MLMC algorithm and our unbiased simulation method with different value of  $\beta$ , but with a given fixed error  $\varepsilon = 0.0002$ . The two methods provide



	Mean value	Statistical error	Computation time
US ( $N = 10^5$ )	0.127032	0.000762635	0.144998
MLMC	0.127053	0.000536248	0.323337
US ( $N = 10^6$ )	0.126363	0.000241231	1.40843
MLMC	0.126747	0.000169842	1.8194
US ( $N = 10^7$ )	0.126703	7.6418e-05	13.9005
MLMC	0.126643	5.37691e-05	16.7499

Table 2: Numerical results for  $\tilde{V}_0$  in (4.2) (case  $d = 1$ ), US denotes our unbiased simulation algorithm with  $\beta = 0.05$ , the computation times are expressed in second.

very close estimation of value  $V_0$ , so we give a comparison on the computation time in Figure 1. We can observe that  $\beta$  in the unbiased simulation method should not be too big nor too small, to minimize the computation effort. When  $\mu_0 = 0.2$ , the computation time of MLMC method is slightly longer than the US method with  $\beta \approx 0.05$ . However, when  $\mu_0 = 0.5$ , the computation time MLMC method is always smaller than the US method for any choice of  $\beta > 0$ . This shows that, in the context of the present example, the performance of our unbiased simulation method is of the order of that of the multilevel Monte Carlo method.

## 4.2 A multi-dimensional SDE

We next consider a  $d$ -dimensional SDE with  $d = 4$ . Let  $W = (W^1, \dots, W^4)^\top$  be a 4-dimensional standard Brownian motion, and  $\sigma_0$  the  $4 \times 4$  be the lower triangular matrix such that

$$\sigma_0 \sigma_0^\top = \begin{pmatrix} 1 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

We consider the SDE

$$dX_t = \mu(t, X_t)dt + \sigma_0 dW_t, \quad X_0^i = 0, \quad i = 1, \dots, 4,$$

with drift function  $\mu(t, x) = (\mu_i(t, x), i = 1, \dots, 4)$  be given by  $\mu_i(t, x_1, \dots, x_4) = 0.1 \left( \sqrt{\frac{3}{4} \exp(x_i)} + \frac{1}{4} \overline{\exp(x)} - 1 \right) - \frac{1}{8}$ , where  $\overline{\exp(x)} := (e^{x_1} + \dots + e^{x_4})/4$ . We then consider two problems:

$$V_0 := \mathbb{E} \left[ \left( \frac{1}{4} \sum_{i=1}^4 e^{M \wedge X_T^i} - K \right)_+ \right], \quad \text{and} \quad \tilde{V}_0 := \mathbb{E} \left[ \left( \frac{1}{4n} \sum_{k=1}^n \sum_{i=1}^4 e^{M \wedge X_{t_k}^i} - K \right)_+ \right], \quad (4.4)$$

where we choose  $K = 1$ ,  $T = 1$ ,  $n = 10$  and  $t_k := \frac{k}{n}T$  and  $M$  is a large number so as to ensure that the terminal condition is Lipschitz. As in the one-dimensional case, we implement our unbiased simulation method using different sample sizes  $N$ .

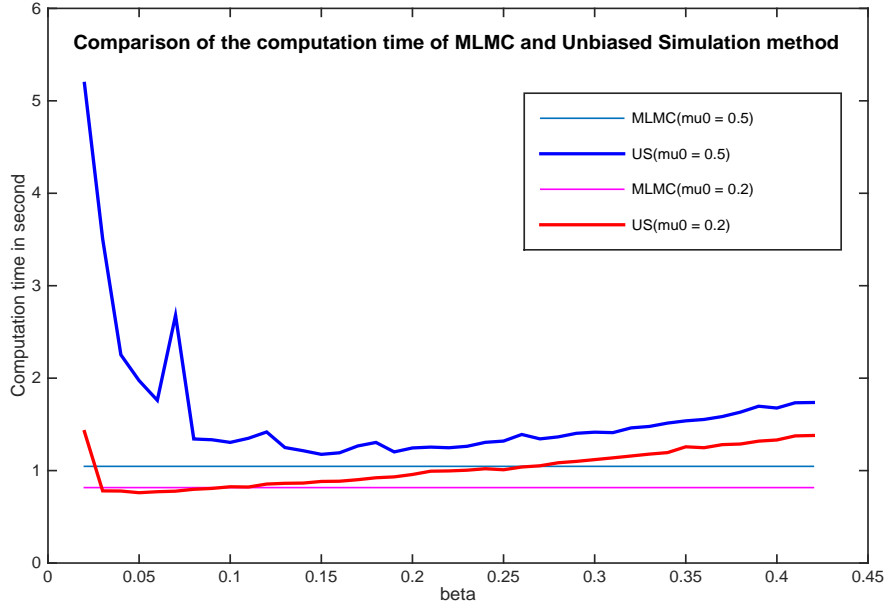


Figure 1: Comparison of the computation time of MLMC method and unbiased simulation method for problem (4.3), with the same given error.

Then, we use the errors, obtained from our unbiased simulation method, in the MLMC algorithm in Section 5 of Giles [15], and we compare the computation time (in second) of the two methods.

The numerical results are given in Tables 3 and 4. We observe that both methods have very similar performance, with a slightly small advantage for our method. However, similar to the one-dimensional case, the MLMC algorithm could be better in other examples.

### 4.3 A one-dimensional driftless SDE

Finally, we provide an example of a one-dimensional driftless SDE. We recall that under the assumed regularity in Theorem 3.5, we are not expecting our method to be competitive with the PDE based approximations. Instead, our objective is to study numerically the performance of the estimator (3.12).

Let us consider the SDE

$$X_0 = 1, \quad dX_t = \frac{2\sigma}{1 + X_t^2} dW_t, \quad (4.5)$$

and we aim to compute

$$V_0 = \mathbb{E}[(X_T - K)_+]. \quad (4.6)$$

We implemented the simple Euler scheme with time step  $\Delta t = 1/10$  and simulation number  $N = 10^6$ , and next the unbiased simulation method (3.12) with  $\beta = 0.1$  and

	Mean value	Statistical error	Computation time
US ( $N = 10^5$ )	0.739374	0.00921078	0.109151
MLMC	0.732707	0.00568921	0.136884
US ( $N = 10^6$ )	0.735745	0.00239613	1.06639
MLMC	0.733539	0.00176862	1.15886
US ( $N = 10^7$ )	0.73659	0.000831597	10.6957
MLMC	0.737087	0.000578058	12.171

Table 3: Numerical results for  $V_0$  in (4.4) (case  $d = 4$ ), US denotes our unbiased simulation algorithm with  $\beta = 0.5$ , the computation times are expressed in second.

	Mean value	Statistical error	Computation time
US ( $N = 10^5$ )	0.382186	0.00247547	0.769847
MLMC	0.381071	0.00167112	2.07589
US ( $N = 10^6$ )	0.382846	0.000762393	7.65796
MLMC	0.383107	0.000535905	10.8444
US ( $N = 10^7$ )	0.383282	0.000244861	85.0265
MLMC	0.383653	0.00017245	104.223

Table 4: Numerical results for  $\tilde{V}_0$  in (4.4) (case  $d = 4$ ), US denotes our unbiased simulation algorithm with  $\beta = 0.05$ , the computation times are expressed in second.

	Mean value	Statistical error	Computation time
Euler scheme	0.161483	0.000196733	0.570541
US	0.160362	9.34729e-05	0.201904
MLMC	0.16057	6.61696e-05	6.65799

Table 5: Numerical results for  $V_0$  in (4.6) (case  $d = 1$ ), US denotes the unbiased simulation algorithm (3.12) with  $\beta = 0.1$ , the computation times are expressed in second. Notice also that the unbiased algorithm (3.12) contains implicitly an antithetic variance reduction, which makes its statistical even error smaller than that of the Euler scheme.

simulation number  $N = 10^6$ , and then the MLMC scheme using the statistical error obtained from the unbiased simulation method. The results are given in Table 5, and we can observe that all three methods provides very similar estimation of  $V_0$ . In particular, the unbiased simulation method has a significant advantage.

## 5 Proofs

### 5.1 A toy example

Before completing the technical part of the proofs for Theorems 2.2, 2.5 and 3.3, we would like to illustrate the main idea by studying a simplified example in the one dimensional case with unit diffusion:

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

Let  $b \in \mathbb{R}$  and  $\beta > 0$ , we define a sequence of i.i.d. random variable  $(\tau_k)_{k \geq 1}$  of distribution  $\mathcal{E}(\beta)$ . Then let  $(T_k)_{k \geq 1}$ ,  $\Delta T_{k+1} := T_{k+1} - T_k$  and  $\Delta W_{k+1} := W_{T_{k+1}} - W_{T_k}$  be defined in and below (2.4), we introduce  $\widehat{X}$  by

$$\widehat{X}_t := x_0 + bt + W_t.$$

and then define

$$\psi = e^{\beta T} g(\widehat{X}_T) \prod_{k=1}^{N_T} \frac{(\mu(T_k, \widehat{X}_{T_k}) - b) \Delta W_{T_{k+1}}}{\beta \Delta T_{k+1}}. \quad (5.1)$$

**Proposition 5.1.** *Let  $\mu(\cdot, \cdot)$  and  $g(\cdot)$  be both bounded smooth functions in  $C_b^2$ . Then for all constants  $\beta > 0$  and  $\beta > 0$ , one has*

$$\mathbb{E}[|\psi|] < \infty \quad \text{and} \quad \mathbb{E}[g(X_T)] = \mathbb{E}[\psi].$$

**Proof.** Since  $\mu$  and  $g$  are uniformly bounded, and for some constant  $C > 0$ , the conditional expectation  $\mathbb{E}_{\Delta T_{k+1}}[|\Delta W_{T_{k+1}}|] \leq C/\sqrt{\Delta T_{k+1}}$ , then by Lemma A.2, it is

obvious that  $\psi$  is integral. Then it is enough to prove that  $\mathbb{E}[g(X_T)] = \mathbb{E}[\psi]$ . In preparation, let us introduce

$$\begin{aligned} \psi_n &= e^{\beta T_{n+1}} \prod_{k=1}^{N_T \wedge n} \frac{(\mu(T_k, \widehat{X}_{T_k}) - b) \Delta W_{T_{k+1}}}{\beta \Delta T_{k+1}} \\ &\quad \left( g(\widehat{X}_T) \mathbf{1}_{N_T \leq n} + \left( \frac{\mu - b}{\beta} \partial_x u \right) (T_{n+1}, \widehat{X}_{T_{n+1}}) \mathbf{1}_{N_T > n} \right). \end{aligned}$$

for all  $n \geq 0$ , with the convention  $\prod_{k=1}^0 \equiv 1$ . It is clear that  $(\psi_n)_{n \geq 0}$  are all integrable by Lemma A.2.

(i) Notice that  $\mu$  and  $g$  are both smooth functions, then by Feynmann-Kac formula, we know  $\mathbb{E}[g(X_T)] = u(0, x_0)$ , where  $u \in C_b^\infty([0, T] \times \mathbb{R})$  is a smooth function of PDE

$$\partial_t u(t, x) + \frac{1}{2} \partial_{xx}^2 u(t, x) + \mu(t, x) \partial_x u(t, x) = 0, \quad \text{for all } (t, x) \in [0, T] \times \mathbb{R},$$

with terminal condition  $u(T, x) = g(x)$ . Rewriting the above PDE in the following equivalent way:

$$\partial_t u(t, x) + b \partial_x u(t, x) + \frac{1}{2} \partial_{xx}^2 u(t, x) + (\mu(t, x) - b) \partial_x u(t, x) = 0,$$

it follows from the Feynmann-Kac formula that

$$\begin{aligned} u(0, x_0) &= \mathbb{E} \left[ g(\widehat{X}_T) + \int_0^T (\mu(t, \widehat{X}_t) - b) \partial_x u(t, \widehat{X}_t) dt \right] \\ &= \mathbb{E} \left[ e^{\beta T} g(\widehat{X}_T) \mathbf{1}_{\{T_1 \geq T\}} + \frac{e^{\beta T_1}}{\beta} (\mu(T_1, \widehat{X}_{T_1}) - b) \partial_x u(T_1, \widehat{X}_{T_1}) \mathbf{1}_{\{T_1 < T\}} \right] \quad (5.2) \\ &= \mathbb{E}[\psi_0], \end{aligned}$$

where the second equality follows from the fact that  $T_1 = T \wedge \tau_1$ , and  $\tau_1$  is a random variable independent of  $\widehat{X}$ , with density function  $\beta e^{-\beta t} \mathbf{1}_{\{t \geq 0\}}$ .

(ii) Next, notice that for any  $t$  and bounded continuous function  $\phi_0$ , one has by integration by parts that

$$\partial_x \mathbb{E}[\phi_0(x + bt + W_t)] = \mathbb{E} \left[ \phi_0(x + bt + W_t) \frac{W_t}{t} \right]. \quad (5.3)$$

Notice also that  $\Delta W_{T_1} = W_{T_1}$ ,  $\Delta T_1 = T_1$  and  $\widehat{X}_{T_1} := x_0 + bT_1 + W_{T_1}$ . It follows by Lemma A.3 that

$$\partial_x u(0, x_0) = \mathbb{E} \left[ e^{\beta \Delta T_1} \frac{\Delta W_{T_1}}{\Delta T_1} \left( g(\widehat{X}_T) \mathbf{1}_{\{T \leq T_1\}} + \frac{\mu - b}{\beta} \partial_x u(T_1, \widehat{X}_{T_1}) \mathbf{1}_{\{T_1 < T\}} \right) \right]. \quad (5.4)$$

Changing the initial condition  $(0, x_0)$  to  $(T_1, \widehat{X}_{T_1})$ , one obtains that, whenever  $T_1 < T$ ,

$$\partial_x u(T_1, \widehat{X}_{T_1}) = \mathbb{E} \left[ e^{\beta \Delta T_2} \frac{\Delta W_{T_2}}{\Delta T_2} \left( g(\widehat{X}_T) \mathbf{1}_{\{T \leq T_2\}} + \frac{\mu - b}{\beta} \partial_x u(T_2, \widehat{X}_{T_2}) \mathbf{1}_{\{T_2 < T\}} \right) \middle| T_1, \widehat{X}_{T_1} \right].$$

Plugging the above expression of  $\partial_x u(T_1, \widehat{X}_{T_1})$  into the r.h.s. of (5.2), and using the fact that  $T \leq T_2$  is equivalent to  $N_{T-} \leq 1$ , and  $\mathbb{P}[\{N_{T-} \leq 1\} \setminus \{N_T \leq 1\}] = 0$ , it follows that  $u(0, x_0) = \mathbb{E}[\psi_1]$ .

(iii) Next, changing the initial condition in (5.4) from  $(0, x_0)$  to  $(T_2, \widehat{X}_{T_2})$  when  $T_2 < T$ , and then plugging the corresponding expression of  $\partial_x u(T_2, \widehat{X}_{T_2})$  into  $\psi_1$ , it follows that  $u(0, x_0) = \mathbb{E}[\psi_2]$ . Repeating the procedure, we have for all  $n \geq 0$ ,

$$\mathbb{E}[g(X_T)] = u(0, x_0) = \mathbb{E}[\psi_n].$$

Finally sending  $n \rightarrow \infty$ , and using Lemma A.2 together with the dominated convergence theorem, it follows that  $\mathbb{E}[g(X_T)] = \mathbb{E}[\lim_{n \rightarrow \infty} \psi_n] = \mathbb{E}[\psi]$ .  $\square$

**Remark 5.2.** *We can also interpret formally the representation  $\psi$  in (5.1) as the expansion of the diffusion process  $X$  around a Brownian motion. Let  $b = 0$  and  $\mu(t, x) \equiv \mu_0$  for some constant  $\mu_0 \in \mathbb{R}$ , so that  $\widehat{X}_t = W_t$  and  $X_t = \mu_0 t + W_t$ . Using the fact that  $\mathbb{P}(N_T = k) = e^{-\beta T} \frac{(\beta T)^k}{k!}$ ,  $\forall k \geq 0$ , it follows formally that*

$$\mathbb{E}[g(X_T)] = \mathbb{E} \left[ \sum_{k=0}^{\infty} \frac{(\mu_0 T)^k}{k!} g^{(k)}(W_T) \right] = \mathbb{E} \left[ e^{\beta T} g(W_T) \prod_{k=1}^{N_T} \left( \frac{\mu_0 \Delta W_{k+1}}{\beta \Delta T_{k+1}} \right) \right], \quad (5.5)$$

where the second equality follows by the fact that

$$\mathbb{P}[N_T = k] = e^{\beta T} \frac{(\beta T)^k}{k!}, \text{ and } \mathbb{E} \left[ g^{(k)} \left( \sum_{i=0}^k \Delta W_{i+1} \right) \right] = \mathbb{E} \left[ g \left( \sum_{i=0}^k \Delta W_{i+1} \right) \prod_{i=1}^k \frac{\Delta W_{i+1}}{\Delta T_{i+1}} \right].$$

In particular, the r.h.s. of (5.5) is exactly  $\mathbb{E}[\psi]$  defined by (5.1) in this case.

To conclude this part, we notice that  $\psi$  in (5.1) is integrable, but has an infinite variance in general. In the next subsection, we exploit the arbitrariness of the constant  $b$  which is involved in the definition of  $\psi$ . More precisely, we shall choose different constants  $b$  at each time  $T_k$ , in an adaptive way. This will lead to the estimator  $\widehat{\psi}$  in Theorems 2.2 and 3.3.

## 5.2 A regime switching diffusion representation

For  $d \geq 1$ ,  $T > 0$ , let  $(\mu, \sigma) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{M}^d$  be bounded continuous functions satisfying

$$|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq L|x - y|; \quad (t, x, y) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d, \quad (5.6)$$

for some constant  $L > 0$ . We start by considering a linear parabolic PDE

$$\partial_t u + \mu \cdot Du + a : D^2 u = 0, \quad \text{on } [0, T] \times \mathbb{R}^d, \quad (5.7)$$

with terminal condition  $u(T, x) = g(x)$ , where  $a(\cdot) := \frac{1}{2} \sigma \sigma^\top(\cdot)$ ,  $A : B := \text{Tr}(AB^\top)$  for any two  $d \times d$  dimensional matrices  $A, B \in \mathbb{M}^d$ , and  $D, D^2$  denote the gradient and Hessian operators with respect to the space variable  $x$ . Next, let us consider the diffusion process  $(X_s^{0, x_0})_{s \in [0, T]}$  defined as unique strong solution of the SDE

$$X_0 = x_0, \quad \text{and} \quad dX_s = \mu(s, X_s) ds + \sigma(s, X_s) dW_s, \quad s \in [0, T]. \quad (5.8)$$

When PDE (5.7) admits a classical solution in  $C_b^{1,3}([0, T] \times \mathbb{R}^d)$ , i.e. the collection of all functions  $\phi(t, x)$  such that  $\phi, \partial_t \phi, D\phi, D^2 \phi$  and  $D^3 \phi$  all exist and bounded continuous, it follows by Feynmann-Kac formula that  $V_0 := \mathbb{E}[g(X_T^{0, x_0})] = u(0, x_0)$ .

**Remark 5.3.** For technical reason, we will assume that  $u \in C_b^{1,3}([0, T] \times \mathbb{R}^d)$  rather than in  $C_b^{1,2}([0, T] \times \mathbb{R}^d)$ . But by approximating the coefficient  $\mu, \sigma$  with smooth functions, one can relax this regularity condition in more concrete context.

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

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

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

Let  $(\hat{\mu}, \hat{\sigma}) : (s, y, t, x) \in [0, T] \times \mathbb{R}^d \times [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{M}^d$  be uniformly bounded, and continuous in  $t$ , Lipschitz in  $x$ , we define  $\hat{X}$  by

$$\hat{X}_0 := x_0 \quad \text{and} \quad d\hat{X}_t = \hat{\mu}(\Theta_t, t, \hat{X}_t)dt + \hat{\sigma}(\Theta_t, t, \hat{X}_t)dW_t, \quad (5.9)$$

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  $k \geq 0$ ,

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

**Example 5.4.** (i) Let  $(\hat{\mu}, \hat{\sigma})(s, y, t, x) = (\mu, \sigma)(s, y)$ , then  $\hat{X}$  is defined as a Euler scheme as in (3.2), i.e.  $\hat{X}_0 = x_0$ , and

$$\hat{X}_{T_{k+1}} = \hat{X}_{T_k} + \mu(T_k, \hat{X}_{T_k})\Delta T_{k+1} + \sigma(T_k, \hat{X}_{T_k})\Delta W_{T_{k+1}}.$$

(ii) When  $\hat{\mu}(\cdot) \equiv 0$  and  $\hat{\sigma}(s, y, t, x) = \sigma(s, y) + \partial_x \sigma(s, y)(x - y)$ , then SDE (5.9) turns to be a linear SDE, whose solution is given explicitly in (3.8).

We first formulate an assumption on the existence of automatic differentiation weights associated to SDE (5.9). Let  $\theta \in [0, T] \times \mathbb{R}^d$  and  $(t, x) \in [0, T] \times \mathbb{R}^d$ , the process  $(\tilde{X}_s^{t,x,\theta})_{s \in [t, T]}$  is defined by SDE

$$\tilde{X}_t^{t,x,\theta} := x, \quad d\tilde{X}_s^{t,x,\theta} = \hat{\mu}(\theta, s, \tilde{X}_s^{t,x,\theta})ds + \hat{\sigma}(\theta, s, \tilde{X}_s^{t,x,\theta})dW_s, \quad (5.10)$$

**Assumption 5.5.** There is a pair of measurable functions  $(\widehat{W}_\theta^1(\cdot), \widehat{W}_\theta^2(\cdot))$ , called automatic differentiation weights, taking values in  $\mathbb{R}^d \times \mathbb{M}^d$ , such that, for all  $\theta \in [0, T] \times \mathbb{R}^d$ ,  $(t, x) \in [0, T] \times \mathbb{R}^d$ ,  $s > t$ , one has  $(\widehat{W}_\theta^i(t, x, s - t, (W_r - W_t)_{r \in [t, s]}))$ ,  $i = 1, 2$  are both integrable. Moreover, for all bounded continuous function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ ,

$$D^i \mathbb{E}[\phi(\tilde{X}_s^{t,x,\theta})] = \mathbb{E}\left[\phi(\tilde{X}_s^{t,x,\theta}) \widehat{W}_\theta^i(t, x, s - t, (W_r - W_t)_{r \in [t, s]})\right], \quad i = 1, 2,$$

where  $D, D^2$  denote the gradient and Hessian operators with respect to the variable  $x$ .

Let  $a(\cdot) := \frac{1}{2}\sigma\sigma^\top(\cdot)$  and  $\hat{a}(\cdot) := \frac{1}{2}\hat{\sigma}\hat{\sigma}^\top(\cdot)$ , we denote

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

and then for  $k > 0$ ,

$$\Delta f_k := (\mu, a)(T_k, \hat{X}_{T_k}) - (\hat{\mu}, \hat{a})(\hat{\Theta}_{k-1}, T_k, \hat{X}_{T_k}) \in \mathbb{R}^d \times \mathbb{M}^d,$$

and for  $k \geq 0$ ,

$$\widehat{\mathcal{W}}_k := (\widehat{\mathcal{W}}_{\hat{\Theta}_k}^1, \widehat{\mathcal{W}}_{\hat{\Theta}_k}^2)(T_k, \hat{X}_{T_k}, T_{k+1}, \Delta W^{k+1}) \in \mathbb{R}^d \times \mathbb{M}^d,$$

with the weight functions  $(\widehat{\mathcal{W}}_\theta^1(\cdot), \widehat{\mathcal{W}}_\theta^2(\cdot))$  given in Assumption 5.5. We then define

$$\widehat{\psi} := e^{\beta T} \left( g(\hat{X}_T) - g(\hat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right) \beta^{-N_T} \prod_{k=1}^{N_T} (\Delta f_k \bullet \widehat{\mathcal{W}}_k), \quad (5.11)$$

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

$$\begin{aligned} \widehat{\psi}_n = e^{\beta T_{n+1}} \prod_{k=1}^{N_T \wedge n} (\beta^{-1} \Delta f_k \bullet \widehat{\mathcal{W}}_k) & \left[ \left( g(\hat{X}_T) - g(\hat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}} \right) \mathbf{1}_{\{N_T \leq n\}} \right. \\ & \left. + \beta^{-1} \left( \Delta f_{n+1} \bullet (Du, D^2u)(T_{n+1}, \hat{X}_{T_{n+1}}) \right) \mathbf{1}_{\{N_T > n\}} \right] \end{aligned} \quad (5.12)$$

**Assumption 5.6.** (i) *The sequence  $(\psi_n)_{n \geq 0}$  is uniformly integrable.*

(ii) *Let  $(e_i)_{i=1, \dots, d}$  denote the canonical basis of  $\mathbb{R}^d$ . There is some  $\varepsilon_0 > 0$ , such that for all  $(t, x) \in [0, T] \times \mathbb{R}^d$  and  $\theta \in [0, T] \times \mathbb{R}^d$ ,  $n \geq 0$  and  $i = 1, \dots, d$ , all the following random vectors is integrable:*

$$\widehat{\mathcal{W}}_\theta^1(t, x, \tau_1 \wedge (T - t), (W_r - W_t)_{r \in [t, (t+\tau_1) \wedge T]}),$$

$$\sup_{\varepsilon \in (0, \varepsilon_0]} \frac{1}{\varepsilon} \left[ \widehat{\mathcal{W}}_\theta^1(t, x + \varepsilon e_i, \tau_1 \wedge (T - t), (W_\cdot - W_t)) - \widehat{\mathcal{W}}_\theta^1(t, x, \tau_1 \wedge (T - t), (W_\cdot - W_t)) \right]$$

and

$$\Delta f_{n+1} \bullet (Du, D^2u)(T_{n+1}, \hat{X}_{T_{n+1}}) \widehat{\mathcal{W}}_n.$$

**Theorem 5.7.** *Suppose that the PDE (5.7) has a classical solution  $u \in C_b^{1,3}([0, T] \times \mathbb{R}^d)$ , suppose in addition that Assumptions 5.5 and 5.6 hold true. Then  $\widehat{\psi}$  is integrable and  $u(0, x_0) = \mathbb{E}[\widehat{\psi}]$ .*

**Remark 5.8.** (i) The condition that  $u \in C_b^{1,3}([0, T] \times \mathbb{R}^d)$  may be relaxed in the concrete applications of Theorem 5.7. This will be indeed performed in Section 3.3 by exploiting the integrability of the automatic differentiation weights  $(\widehat{\mathcal{W}}_\theta^1, \widehat{\mathcal{W}}_\theta^2)$  of Assumption 5.5.

(ii) By definition, the automatic differentiation weight satisfies  $\mathbb{E}[\widehat{\mathcal{W}}_k] = 0$ , then  $\widehat{\psi}$  in (5.11) has the same mean than the estimator

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



However, in practice, the weight function  $\widehat{\mathcal{W}}_k$  is typically of infinity variance, or even not integrable, in general. Indeed, as we will see in the following,  $\widehat{\mathcal{W}}_k$  is generally of order  $\frac{1}{\Delta T_{k+1}} = \frac{1}{T_{k+1} - T_k}$ , where conditioning on  $N_T = n$ ,  $(T_1, \dots, T_{N_T})$  follows the law of statistic order of uniform distribution on  $[0, T]$ . Then by direct computation, one knows  $\mathbb{E}[1/\Delta T_{N_T+1}] = \infty$ . In the definition of  $\widehat{\psi}$  in (5.11), the additional term  $-g(\widehat{X}_{T_{N_T}}) \mathbf{1}_{\{N_T > 0\}}$  can be seen as a control variate so as to guarantee the integrability of  $\widehat{\psi}$ .

(iii) As a consequence of the integrability problems raised in (ii), Assumption 5.6 is in fact implicitly a restriction on the choice of the coefficients  $\hat{\mu}$  and  $\hat{\sigma}$ , and we cannot expect a representation for  $u(t, x)$  with arbitrary  $\hat{\mu}$  and  $\hat{\sigma}$ , see Section 5.3 below.

**Proof of Theorem 5.7.** (i) Recall that  $u \in C_b^{1,3}([0, T] \times \mathbb{R}^d)$  is a classical solution of PDE (5.7). Denote  $(\hat{\mu}_\theta, \hat{a}_\theta)(\cdot) = (\hat{\mu}, \hat{a})(\theta, \cdot)$ , one can rewrite (5.7) in the following equivalent way:

$$-\partial_t u - \hat{\mu}_\theta \cdot Du - \hat{a}_\theta : D^2 u - ((\mu - \hat{\mu}_\theta) \cdot Du + (a - \hat{a}_\theta) : D^2 u) = 0. \quad (5.13)$$

Using Feynmann-Kac formula, it follows that

$$u(0, x_0) = \mathbb{E} \left[ g(\widetilde{X}_T^{0, x_0, \theta}) + \int_0^T ((\mu - \hat{\mu}_\theta) \cdot Du + (a - \hat{a}_\theta) : D^2 u)(s, \widetilde{X}_s^{0, x_0, \theta}) ds \right], \quad (5.14)$$

where  $\widetilde{X}^{0, x_0, \theta}$  is defined by (5.10), which coincides with  $\widehat{X}$  in (5.9) on  $[0, T_1]$  whenever  $\theta = (0, x_0)$ .

Recall that  $T_1 = \tau_1 \wedge T$ , where  $\tau_1$  is a random variable of density  $\beta e^{-\beta s} \mathbf{1}_{\{s \geq 0\}}$  independent of the Brownian motion  $W$ . Fixing  $\theta = (0, x_0)$ , it follows that

$$u(0, x_0) = \mathbb{E} \left[ e^{\beta T_1} \left( g(\widehat{X}_T) \mathbf{1}_{\{N_T = 0\}} + \beta^{-1} \Delta f_1 \bullet (Du, D^2 u)(T_1, \widehat{X}_{T_1}) \mathbf{1}_{\{N_T > 0\}} \right) \right] = \mathbb{E}[\widehat{\psi}_0].$$

(ii) Let us now go back to the expression (5.14), and derive an expression for the derivatives  $Du(0, x_0)$  and  $D^2 u(0, x_0)$ . First, for  $Du(0, x_0)$ , we use the integrability condition in Assumption 5.6 with Lemma A.3, and also the fact the  $Du(\cdot)$  is continuous, it follows that

$$\begin{aligned} Du(0, x_0) &= \mathbb{E} \left[ g(\widetilde{X}_T^{0, x_0, \theta}) \widehat{\mathcal{W}}_\theta^1(x_0, T) \right. \\ &\quad \left. + \int_0^T ((\mu - \hat{\mu}_\theta) \cdot Du + (a - \hat{a}_\theta) : D^2 u)(s, \widetilde{X}_s^{0, x_0, \theta}) \widehat{\mathcal{W}}_\theta^1(x_0, s) ds \right], \end{aligned}$$

where we simplify the notation  $\widehat{\mathcal{W}}_\theta^1(0, x_0, s, (W_r - W_t)_{r \in [0, s]})$  to  $\widehat{\mathcal{W}}_\theta^1(x_0, s)$ . Then by the independence of  $\tau_1$  to the Brownian motion  $W$ , and setting  $\theta = (0, x_0)$ , it follows that

$$Du(0, x_0) = \mathbb{E} \left[ \widehat{\psi}_0 \widehat{\mathcal{W}}_{(0, x_0)}^1(0, x_0, T_1, \Delta W^1) \right]. \quad (5.15)$$

Next, for  $D^2 u(0, x_0)$ , we use again Lemma A.3 together with the integrability condition in Assumption 5.6 and Lipschitz property of  $x \mapsto ((\mu - \hat{\mu}) \cdot Du + (a - \hat{a}) : D^2 u)(s, x)$ , and the continuity of  $D^2 u(\cdot)$  that

$$D^2 u(0, x_0) = D_{x_0}^2 \mathbb{E} [g(\widetilde{X}_T^{0, x_0, \theta})] + \int_0^T D_{x_0}^2 \mathbb{E} \left[ ((\mu - \hat{\mu}_\theta) \cdot Du + (a - \hat{a}_\theta) : D^2 u)(s, \widetilde{X}_s^{0, x_0, \theta}) \right] ds.$$

Setting  $\theta = (0, x_0)$  and using Assumption 5.5, it leads to

$$D^2u(0, x_0) = \mathbb{E}\left[\widehat{\psi}_0 \widehat{\mathcal{W}}_{(0, x_0)}^2(0, x_0, T_1, \Delta W^1)\right].$$

Recall that  $\mathbb{E}[\widehat{\mathcal{W}}_{(0, x_0)}^2(0, x_0, T, \Delta W)] = 0$ , we then obtain

$$D^2u(0, x_0) = \mathbb{E}\left[\left(\widehat{\psi}_0 - e^{\beta T} g(x_0) \mathbf{1}_{\{N_T=0\}}\right) \widehat{\mathcal{W}}_{(0, x_0)}^2(0, x_0, T_1, \Delta W^1)\right]. \quad (5.16)$$

(iii) Changing the initial condition  $(0, x_0)$  in (5.15) and (5.16) by  $(T_1, \widehat{X}_1)$  (remember  $\widehat{\psi}_0$  dependent also on the initial condition  $(0, x_0)$ ), then plugging the expression of  $D^1u(T_1, \widehat{X}_1)$  and  $D^2u(T_1, \widehat{X}_1)$  into the definition of  $\widehat{\psi}_0$  in (5.12), it follows by identifying the term that

$$u(0, x_0) = \mathbb{E}[\psi_1].$$

(iv) Repeating the arguments by replacing the initial condition  $(0, x_0)$  by  $(T_{n+1}, \widehat{X}_{n+1})$  in (5.15) and (5.16) and then plugging the corresponding expression into the definition of  $\psi_n$ , etc., we obtain that  $u(0, x_0) = \mathbb{E}[\widehat{\psi}_n]$  for all  $n \geq 0$ . Then letting  $n \rightarrow \infty$ , we obtain

$$u(0, x_0) = \lim_{n \rightarrow \infty} \mathbb{E}[\widehat{\psi}_n] = \mathbb{E}\left[\lim_{n \rightarrow \infty} \widehat{\psi}_n\right] = \mathbb{E}[\widehat{\psi}],$$

which concludes the proof.  $\square$

### 5.3 Proof of the representation results in Theorems 2.2, 2.5 and 3.3.

Using the results in Theorem 5.7, we can easily complete the proof of the representation results in Theorems 2.2, 2.5 and 3.3.

**Proof of Theorems 2.2 (ii) and 3.3 (ii).** (i) In the context of Theorems 2.2 and 3.3, the increment  $\widehat{X}_{T_{k+1}} - \widehat{X}_{T_k}$ , conditional on  $(T_k, \widehat{X}_{T_k})$ , is Gaussian. And the estimator  $\widehat{\psi}$  corresponds to the estimator in Theorem 5.7 with automatic differentiation weights function

$$\widehat{\mathcal{W}}_{\theta}^1(\cdot, \delta t, \delta w) := (\sigma_0^\top)^{-1} \frac{\delta w}{\delta t} \quad \text{and} \quad \widehat{\mathcal{W}}_{\theta}^2(\cdot, \delta t, \delta w) := (\sigma_0^\top)^{-1} \frac{\delta w \delta w^\top - \delta t I_d}{\delta t^2} \sigma_0^{-1}. \quad (5.17)$$

In particular, it is clear that Assumption 5.5 holds true with the above choice of automatic differentiation weight functions in (5.17).

(ii) Besides, the uniform integrability conditions and integrability conditions in Assumption 5.6 can be easily obtained following the lines in the first part of the proof of Theorems 2.2 and 2.2, using Lemma A.2.

(iii) Now, suppose in addition that  $\mu$ ,  $\sigma$  and  $g$  are bounded smooth functions with bounded continuous derivatives, so that  $u \in C_b^{1,3}([0, T] \times \mathbb{R}^d)$ . It follows by Theorem 5.7 that  $V_0 = \mathbb{E}[\widehat{\psi}]$ .

(iv) Finally, when  $\mu(\cdot)$  and  $\sigma(\cdot)$  satisfy the Lipschitz condition (3.1) 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  $X^\varepsilon$  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 (3.3) associated to the coefficient  $(\mu_\varepsilon, \sigma_\varepsilon, g_\varepsilon)$ . We then conclude the proof.  $\square$

**Proof of Theorem 2.5 (ii).** For the path-dependent case, it is enough to use the same arguments as in Theorem 2.2, together with the PDE system (2.12) in Remark 2.3.  $\square$

## 5.4 Proof of Theorem 3.5

To introduce the algorithm in the context of Theorem 5.7, we propose to choose

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

Before providing the proof of Theorem 3.5, we first give a lemma which justifies our choice of the automatic differentiation weight function  $\overline{W}_k^2$  in (3.11), 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

$$\overline{X}_0 = x, \quad d\overline{X}_t = (c_1 + c_2\overline{X}_t)dW_t, \quad (5.18)$$

whose solution is given explicitly by

$$\overline{X}_t^{0,x} = \begin{cases} -\frac{c_1}{c_2} + \left(\frac{c_1}{c_2} + x\right) \exp\left(-\frac{c_2^2}{2}t + c_2W_t\right), & \text{if } c_2 \neq 0, \\ x + c_1W_t, & \text{if } c_2 = 0. \end{cases} \quad (5.19)$$

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

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

**Lemma 5.9.** *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  $t \in (0, T]$ ,*

$$\partial_{xx}^2 \mathbb{E}[\phi(\overline{X}_t^{0,x})] = \mathbb{E}\left[\phi(\overline{X}_t^{0,x}) \frac{1}{(c_1 + c_2x)^2} \left(-c_2 \frac{W_t}{t} + \frac{W_t^2 - t}{t^2}\right)\right]. \quad (5.20)$$

(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  $(t, x) \in [0, T] \times \mathbb{R}^d$ ,

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

**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 [13]). Next, when  $c_2 \neq 0$ , denote  $v(x) := \mathbb{E}[\phi(\bar{X}_t^{0,x})]$ , then with the expression of  $\bar{X}_t^{0,x}$  in (5.19), 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

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

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

$$v''(x) = \mathbb{E} \left[ \phi(\bar{X}_t^{0,x}) \frac{1}{(c_1 + c_2x)^2} \left( -c_2 \frac{W_t}{t} + \frac{W_t^2 - t}{t^2} \right) \right].$$

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  $\phi'_\varepsilon$  and  $\phi''_\varepsilon$  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

$$v''(x) = \lim_{\varepsilon \rightarrow 0} v''_\varepsilon(x) = \mathbb{E} \left[ \phi(\bar{X}_t^{0,x}) \frac{1}{(c_1 + c_2x)^2} \left( -c_2 \frac{W_t}{t} + \frac{W_t^2 - t}{t^2} \right) \right].$$

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

When  $c_2 \neq 0$ , denote  $|\phi'|_0 := \sup_x |\phi'(x)|$ , we obtain, by direct computation, that

$$\begin{aligned} & \mathbb{E} \left[ \left( \phi(\bar{X}_t^{0,x}) - \phi(x) \right)^2 \left( \frac{W_t}{t} \right)^2 \right] \leq |\phi'|_0 \mathbb{E} \left[ (\bar{X}_t^{0,x} - x)^2 \frac{W_t^2}{t^2} \right] \\ & = |\phi'|_0 \mathbb{E} \left[ (c_1 + c_2x)^2 \left( \frac{e^{-c_2^2 t/2 + c_2 W_t} - 1}{c_2 W_t - c_2^2 t/2} \right)^2 \frac{W_t^2 (c_2 W_t - c_2^2 t/2)^2}{t^2} \right], \end{aligned}$$

which is clearly uniformly bounded by  $C(c_1 + c_2x)^2$  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

$$\varphi''(y) = \phi''(x + \ell(y))(c_2 + c_1 x)^2 e^{-c_2^2 t/2 + c_2 y} + \phi'(x + \ell(y))(c_2 + c_1 x)c_2 e^{-c_2^2 t/2 + c_2 y}. \quad (5.21)$$

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

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

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

**Proof of Theorem 3.5.** (i) Let us first prove that  $\mathbb{E}[\overline{\psi}^2] < \infty$  for  $\overline{\psi}$  defined by (3.12). First, we notice that  $\widehat{\mathcal{W}}_k^- = \widehat{\mathcal{W}}_k^2$  for all  $k = 1, \dots, N_T - 1$ ,  $g \in C_b^2(\mathbb{R})$ , and with the choice of  $c_1^k$  and  $c_2^k$  in (3.7), one has  $c_1^k + c_2^k \widehat{X}_{T_k} = \sigma(T_k, \widehat{X}_{T_k})$ , which is uniformly bounded. By considering the conditional expectation over  $(\widehat{X}_{T_{N_T}}, \Delta T_{N_T+1})$  using items (ii) of Lemma 5.9, we have  $\mathbb{E}[\overline{\psi}^2]$  is bounded by

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

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

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

where  $|\sigma|_0 := \sup_{t,x} |\sigma(t, x)|$ . Notice that  $\sigma \geq \varepsilon > 0$ ,  $\sigma$  and  $\partial_x \sigma$  are uniformly bounded, then to prove that  $\overline{\psi}$  is of finite variance, it is enough to prove that, for some  $C > 0$  large enough, the expectation of

$$\prod_{k=2}^{N_T} \left[ C \left( C + |\partial_x \sigma(T_{k-1}, \widehat{X}_{T_{k-1}}) \Delta \widehat{X}_{T_k}| \right)^2 \left( C + \frac{\Delta \widehat{X}_{T_k}^2}{\Delta T_k} \right)^2 \left( C |\Delta W_{T_k}| + \frac{\Delta W_{T_k}^2}{\Delta T_k} + 1 \right)^2 \right] \quad (5.22)$$

is finite. Similarly to the computation in item (ii) of Lemma 5.9, we have

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

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

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

for some constant  $C' > 0$  independent of  $\widehat{X}_{T_{k-1}}, T_{k-1}, \Delta T_k$ . Then the variance of (5.22) is bounded by  $C \mathbb{E}[(C')^{N_T}] < \infty$  and hence  $\widehat{\psi}$  in (3.12) is of finite variance.

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

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

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 Assumption 5.6 for  $\widehat{\psi}$  in (3.10).

(iii) Finally, using Lemma 5.9 (i), it follows that Assumption 5.5 holds true. Moreover, with the regularity condition on  $\sigma(t, x)$  and  $g$  in Assumption 3.4, we know  $u \in C_b^{1,3}(\mathbb{R})$ . We then conclude the proof of  $u(0, x_0) = \mathbb{E}[\widehat{\psi}] = \mathbb{E}[\overline{\psi}]$  by Theorem 5.7.  $\square$

## A Appendix

We first provide an estimation on the order statistics of uniform distribution on  $[0, 1]$ , which induces an estimation on a functional of the arrival times  $(T_k)_{k>0}$  of the Poisson process. We next provide a technical result on the automatic differentiation function related to a SDE.

**Lemma A.1.** *Let  $p \in (0, 1)$ ,  $(U_k)_{k=1, \dots, m}$  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*

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

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

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

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

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

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

Let  $N = (N_s)_{s \geq 0}$  be a Poisson process with arrival times  $(T_k)_{k > 0}$ , denote  $\Delta T_{k+1} := T_{k+1} - T_k$ . Let  $0 = t_0 < t_1 < \dots < t_n = T < \infty$  be a discrete time grid, we define further

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

and

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

**Lemma A.2.** *Let  $p \in (0, 1)$ , then for every constant  $C > 0$ , one has*

$$\mathbb{E} \left[ \prod_{k=1}^{N_T} \frac{C}{(\Delta \tilde{T}_{k+1})^p} \right] < \infty.$$

**Proof.** (i) Notice that we 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  $t_k - t_{k-1} < (2\beta C)^{p-1}$  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

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

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  $k = 1, \dots, n$ ,

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

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

Let  $X^x$  be the solution of SDE

$$X_0^x = x, \quad dX_t^x = \mu(t, X_t^x)dt + \sigma(t, X_t^x)dW_t,$$

where  $(\mu, \sigma) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{M}^d$  is continuous and in addition Lipschitz continuous in  $x$ .

**Lemma A.3.** *Suppose that for all  $t \in [0, T]$  and bounded continuous function  $\phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ , the derivatives  $(\partial_{x_i} \mathbb{E}[\phi(t, X_t^x)], \partial_{x_i x_j}^2 \mathbb{E}[\phi(t, X_t^x)])_{i,j=1,\dots,d}$  exist; and there is some measurable  $\mathbb{R}^d$ -valued function  $\widehat{\mathcal{W}}^1(x, t, (W_s)_{s \in [0,t]})$  such that for*

$$\partial_{x_i} \mathbb{E}[\phi(t, X_t^x)] = \mathbb{E}[\phi(t, X_t^x) \widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0,t]})], \quad i = 1, \dots, d.$$

Let  $F(dt)$  be some probability measure on  $[0, T]$ .

(i) Suppose that for each  $x \in \mathbb{R}^d$ ,  $i = 1, \dots, d$ ,

$$\int_0^T \mathbb{E} \left[ \left| \widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0,t]}) \right| \right] F(dt) < \infty,$$

and the continuous function  $\phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  (which may be unbounded) satisfies

$$\int_0^T \mathbb{E} \left[ \left| \phi(t, X_t^x) \widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0,t]}) \right| \right] F(dt) < \infty.$$

Then

$$\partial_{x_i} \int_0^T \mathbb{E}[\phi(t, X_t^x)] F(dt) = \int_0^T \mathbb{E}[\phi(t, X_t^x) \widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0,t]})] F(dt). \quad (\text{A.3})$$

(ii) Suppose in addition that  $\phi(t, x)$  is bounded continuous function and Lipschitz in  $x$ , and for each  $x \in \mathbb{R}^d$ ,  $i, j = 1, \dots, d$ ,

$$\int_0^T \sqrt{\mathbb{E} \left[ \left| \widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0,t]}) \right|^2 \right]} F(dt) < \infty,$$

and

$$\int_0^T \sup_{\varepsilon \in [0, \varepsilon_0]} \left| \frac{1}{\varepsilon} \mathbb{E} \left[ \widehat{\mathcal{W}}_i^1(x + \varepsilon e_j, t, (W_s)_{s \in [0,t]}) - \widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0,t]}) \right] \right| F(dt) < \infty, \quad (\text{A.4})$$

for some  $\varepsilon_0 > 0$ , where  $(e_j)_{j=1,\dots,d}$  denotes the canonical basis of  $\mathbb{R}^d$ . Then

$$\partial_{x_i x_j}^2 \int_0^T \mathbb{E}[\phi(t, X_t^x)] F(dt) = \int_0^T \partial_{x_i x_j}^2 \mathbb{E}[\phi(t, X_t^x)] F(dt), \quad (\text{A.5})$$

where, in particular, the partial derivative at the l.h.s. and the integration at the r.h.s. are well defined.

**Proof.** (i) First, let us notice that  $(t, x) \mapsto (\mu, \sigma)(t, x)$  is Lipschitz in  $x$ , then by standard analysis (see e.g. Chapter 7.8 of [17]), there is some constant  $C$  independent of  $\varepsilon > 0$  and  $i = 1, \dots, d$ , such that

$$\mathbb{E} \left[ \left| \frac{X^{x+\varepsilon e_i} - X^x}{\varepsilon} \right|^2 \right] \leq C(1 + e^{Ct}). \quad (\text{A.6})$$

(ii) Suppose that  $\phi(t, x)$  is bounded continuous and Lipschitz in  $x$ . It follows that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \int_0^T \frac{1}{\varepsilon} \mathbb{E} \left[ (\phi(t, X_t^{x+\varepsilon e_i}) - \phi(t, X_t^x)) \right] F(dt) \\ &= \int_0^T \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E} \left[ (\phi(t, X_t^{x+\varepsilon e_i}) - \phi(t, X_t^x)) \right] F(dt) \\ &= \mathbb{E} \left[ \int_0^T \phi(t, X_t^x) \widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0,t]}) F(dt) \right], \end{aligned}$$



where the first equality follows by the Lipschitz property of  $x \mapsto \phi(t, x)$  and (A.6). We hence proved (A.3) when  $x \mapsto \phi(t, x)$  is Lipschitz.

(iii) When  $\phi$  is only continuous, it is enough to approximate it by a sequence  $(\phi_n)_{n \geq 1}$  which are all bounded, and Lipschitz in  $x$ . Then by the integrability of  $\widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0, t]})$  as well as that of  $\phi(t, X_t^x) \widehat{\mathcal{W}}_i^1(x, t, (W_s)_{s \in [0, t]})$  under  $\mathbb{P}(d\omega) \times F(dt)$ , it follows that (A.3) holds true for continuous function  $\phi$ .

(iv) To prove (A.5), let us use (A.3) and obtain that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left[ \partial_{x_j} \int_0^T \mathbb{E}[\phi(t, X_t^{x+\varepsilon e_i})] F(dt) - \partial_{x_j} \int_0^T \mathbb{E}[\phi(t, X_t^x)] F(dt) \right] \\ &= \lim_{\varepsilon \rightarrow 0} \int_0^T \frac{1}{\varepsilon} \mathbb{E} \left[ \phi(t, X_t^{x+\varepsilon e_i}) \widehat{\mathcal{W}}_j^1(x + \varepsilon e_i, t, \cdot) - \phi(t, X_t^x) \widehat{\mathcal{W}}_j^1(x, t, \cdot) \right] F(dt) \\ &= \int_0^T \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E} \left[ \phi(t, X_t^{x+\varepsilon e_i}) \widehat{\mathcal{W}}_j^1(x + \varepsilon e_i, t, \cdot) - \phi(t, X_t^x) \widehat{\mathcal{W}}_j^1(x, t, \cdot) \right] F(dt). \end{aligned} \quad (\text{A.7})$$

where the first equality follows by the Lipschitz property of  $x \mapsto \phi(t, x)$  and the estimation (A.6) together with (A.4), and in particular, the integrable in the last term of (A.7) is well defined, and hence the limit of the first term of (A.7) exists.  $\square$

## References

- [1] Alaya, M. B., and Kebaier, A. *Central limit theorem for the multilevel Monte Carlo Euler method*. The Annals of Applied Probability, 25(1):211-234, 2015.
- [2] Alanko, S., Avellaneda, M. : *Reducing variance in the numerical solution of BSDEs*, Comptes Rendus Mathématique Vol. 351, n 3-4:135-138, 2013.
- [3] Bally, V., Kohatsu-Higa, A. : *A probabilistic interpretation of the parametrix method*, preprint, 2014.
- [4] Bally, V., Talay, D. : *The law of the Euler scheme for stochastic differential equations, I: convergence rate of the distribution function*, Probability Theory and Related Fields, 104(1):43-60, 1995.
- [5] Blanchet, J., Chen, X. and Dong, J. :  *$\varepsilon$ -Strong Simulation for Multidimensional Stochastic Differential Equations via Rough Path Analysis*, arXiv preprint arXiv:1403.5722, 2014.
- [6] Beskos, A., Roberts, G. O. : *Exact simulation of diffusions*, Ann. Appl. Probab. Volume 15, Number 4, pp. 2422-2444, 2005.
- [7] Beskos, A., Papaspiliopoulos, O., Roberts, G. O. : *Retrospective exact simulation of diffusion sample paths*, Bernoulli 12(6), December 2006.
- [8] Bompis, R., Gobet, E. : *Stochastic Approximation Finite Element method for analytical approximation of multidimensional diffusion*, SIAM Journal on Numerical Analysis 52(6):3140-3164, 2014.
- [9] Broadie, M., Glasserman, P.: *Estimating Security Price Derivatives Using Simulation*, Management Science, 269-285, 1996.

- [10] Dupire, B. : *Functional Itô calculus*, SSRN, 2009.
- [11] Ekren, I., Keller, C., Touzi, N. and Zhang, J. : *On Viscosity Solutions of Path Dependent PDEs*, Anna. Proba. 2014.
- [12] Ekren, I., Touzi, N. and Zhang, J. : *Viscosity Solutions of Fully Nonlinear Path Dependent PDEs: Part I*, Anna. Proba. to appear.
- [13] 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.
- [14] 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.
- [15] Giles, M. B. : *Multilevel Monte Carlo path simulation*, Oper. Res., 56(3):607-617, 2008.
- [16] Giles, M. B., and Szpruch, L. *Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation*. The Annals of Applied Probability, 24(4):1585-1620, 2014.
- [17] Graham, C. and Talay, D. *Stochastic Simulation and Monte Carlo Methods*, Mathematical Foundations of Stochastic Simulation. Vol. 68. Springer Science & Business Media, 2013.
- [18] Kebaier, A. : *Statistical Romberg extrapolation: A new variance reduction method and applications to option pricing*, Annals of Applied probability, 15(4):2681-2705, 2005.
- [19] Henry-Labordère, P. : *Counterparty Risk Valuation: A Marked Branching Diffusion Approach*, Risk magazine (Jul. 2012).
- [20] Henry-Labordère, P., Oudjane, N., Tan, X., Touzi, N. and Warin, X. : *Branching diffusion representation of semilinear PDEs and Monte Carlo approximation*, Work in progress.
- [21] 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.
- [22] Jourdain, B., Sbai, M. : *Exact retrospective Monte Carlo computation of arithmetic average Asian options*, Monte Carlo Methods and Applications 13, 2:135-171, (2007).
- [23] Kloeden, P., Platen, E. : *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics, Volume 23, Springer, 1992.
- [24] Rhee, C. H., Glynn, P. W. : *Unbiased estimation with square root convergence for SDE models*, Oper. Res., 63(5):1026-1043, 2015.
- [25] Talay, D. and Tubaro, L. *Expansion of the global error for numerical schemes solving stochastic differential equations*. Stochastic Analysis and Applications, 8:483-509, 1990.