The Annals of Applied Probability
2017, Vol. 27, No. 6, 3305–3341
https://doi.org/10.1214/17-AAP1281

UNBIASED SIMULATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

BY PIERRE HENRY-LABORDÈRE, XIAOLU TAN[1] AND NIZAR TOUZI[1]

Société Générale, University of Paris-Dauphine and École Polytechnique Paris

We propose an unbiased Monte 3 estimator for $\mathbb{E}[g(X_{t_1}, \dots, X_{t_n})]$, where X is a diffusion process defined by a multidimensional 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 the Bismut–Elworthy–Li formula from Malliavin calculus, as exploited by Fournié et al. [*Finance Stoch.* **3** (1999) 391–412] 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 [*Ann. Appl. Probab.* **25** (2015) 3095–3138, Section 6.1] as an application of the parametrix method.

1. Introduction. Let $d \geq 1$, $T > 0$ and W be a d-dimensional Brownian motion, $\mu : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{M}^d$ 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 coefficients, we consider the process X defined as the unique strong solution of the multidimensional SDE

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

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

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

for some function $g : \mathbb{R}^{d\times n} \to \mathbb{R}$ and discrete time grid $0 < t_1 < \cdots < t_n = T$. 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

$$\partial_t u + \mu(t, x) \cdot Du + \frac{1}{2}\sigma\sigma^\top(t, x) : D^2 u = 0, \qquad u_T = g. \tag{1.3}$$

Received March 2016; revised November 2016.

[1]Supported by 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).

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

Key words and phrases. Unbiased simulation of SDEs, regime switching diffusion, linear parabolic PDEs.

Here, $A : B := \mathrm{Tr}(AB^\top)$ for any two $d \times d$ dimensional matrices $A, B \in \mathbb{M}^d$. 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 [21] proved that the discretization error for the Euler scheme is controlled by a rate $C\Delta t$, where Δt denotes the time step discretization. Since then, many works focused on the analysis of the discretization error under various discretization techniques; see, for example, Kloeden and Platen [19], and Graham and Talay [13] 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, Papaspiliopoulos and Roberts [4], Jourdain and Sbai [17], 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, for example, Blanchet, Chen and Dong [7].

An alternative approximation method for V_0 was induced by the multilevel Monte Carlo (MLMC) algorithm introduced by Giles [11], 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 [12], Alaya and Kebaier [3], Rhee and Glynn [20] for further developments. In particular, Rhee and Glynn [20] proposed a random level technique in the MLMC algorithm and obtained a simulatable random with expectation V_0, thus inducing an unbiased simulation method.

The unbiased simulation of a functional of a SDE has been investigated by many people. Wagner [22, 23] provided an unbiased estimator for a class of functionals of a $\mathbb{R}^d$-valued Markov process Z with known transition function, where a key step is to expand an exponential term in a power series. Using similar expansion techniques, Beskos, Papaspiliopoulos, Roberts and Fearnhead [5] obtained an unbiased method for a larger class of functionals of solution of the SDE. More recently, Bally and Kohatsu-Higa [2] provided a probabilistic interpretation of the parametrix method for PDEs. In particular, when $n = 1$, they obtained a representation formula for V_0 of the form

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

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 [2], the process $\widehat{X}$ can be exactly simulated, and hence it may provide an unbiased estimator for V_0. Nevertheless, the weight function $\mathcal{W}_T$ has infinite variance, and hence the corresponding Monte Carlo estimator loses the standard central limit error estimate. A recent improvement, to obtain a finite variance estimator, is provided by Andersson and Kohatsu-Higa [1].

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 similar but 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 the Bismut–Elworthy–Li formula from Malliavin calculus, as exploited by Fournié et al. [10] for the simulation of the Greeks in financial applications.

The technique introduced in the present paper is inspired by the numerical algorithm introduced in [14, 16], for semilinear PDEs of the form

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

for some nonlinearity F_0. The main idea in [14, 16] 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:

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

where

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

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 cannot 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 of 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 [15], 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 a constant diffusion coefficient, and propose an unbiased estimator for V_0 for the Markovian and path-dependent case. In Section 3, we consider the SDE with a general diffusion coefficient function, and obtain a similar representation formula for V_0, which is integrable but of infinite variance. Section 4 reports 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 the 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 with finite variance.

2.1. *The Markovian case.* Let us start by the Markovian case, where the diffusion process X is defined by

$$X_0 = x_0, \qquad dX_t = \mu(t, X_t)\,dt + \sigma_0\,dW_t, \tag{2.1}$$

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

$$V_0 = \mathbb{E}\big[g(X_T)\big], \tag{2.2}$$

for some function $g : \mathbb{R}^d \to \mathbb{R}$. We impose the following conditions on μ and σ_0.

ASSUMPTION 2.1. The diffusion coefficient σ_0 is nondegenerate, the drift function $\mu(t, x)$ is bounded and continuous in (t, x), uniformly $\frac{1}{2}$-Hölder in t and uniformly Lipschitz in x, that is, for some constant $L > 0$,

$$\begin{aligned} &\big|\mu(t, x) - \mu(s, y)\big| \\ &\quad \le L\big(\sqrt{|t - s|} + |x - y|\big) \qquad \forall (s, x), (t, y) \in [0, T] \times \mathbb{R}^d. \end{aligned} \tag{2.3}$$

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

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

Then $(N_t)_{0\le t\le T}$ is a Poisson process with intensity β 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

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

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

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

Then our estimator is given by

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

with

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

THEOREM 2.2. *Suppose that Assumption 2.1 holds true, and g is globally Lipschitz. Then*

$$\mathbb{E}\big[\widehat{\psi}^2\big] < \infty \quad \textit{and} \quad V_0 = \mathbb{E}[\widehat{\psi}].$$

PROOF. (i) We first show that $\mathbb{E}[\widehat{\psi}^2] < \infty$. For simplicity, we denote $\Delta\widehat{X}_{T_k} := \widehat{X}_{T_k} - \widehat{X}_{T_{k-1}}$ 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 nondegeneracy of σ_0. Notice that $T_{N_T+1} = T$ from its definition in (2.4), then using Assumption 2.1, it follows by direct computation that

$$\begin{aligned}
\big|e^{-\beta T}\widehat{\psi}\big| &\le \big(\big|g(x_0)\big| + L_g|\Delta\widehat{X}_{T_1}|\big)\mathbf{1}_{\{N_T=0\}} \\
&\quad + L_g|\Delta\widehat{X}_{T_{N_T+1}}|\prod_{k=1}^{N_T}\frac{L(\sqrt{\Delta T_k} + |\Delta\widehat{X}_{T_k}|)}{\beta\Delta T_{k+1}}\big|(\sigma_0^\top)^{-1}\Delta W_{T_{k+1}}\big|\mathbf{1}_{\{N_T>0\}} \\
&\le L_g\left(\frac{|g(x_0)|}{L_g} + \sqrt{\Delta T_1} + |\Delta\widehat{X}_{T_1}|\right) \\
&\quad \times\prod_{k=1}^{N_T}\frac{L(\sqrt{\Delta T_{k+1}} + |\Delta\widehat{X}_{T_{k+1}}|)}{\beta\Delta T_{k+1}}\big|(\sigma_0^\top)^{-1}\Delta W_{T_{k+1}}\big|.
\end{aligned}$$

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

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

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

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

We therefore get the following upper bound:

$$\mathbb{E}[\widehat{\psi}^2] \le Ce^{2\beta T}e^{-\beta T + \frac{\gamma L^2 T}{\beta}} \quad \text{where } C := L_g^2\mathbb{E}\bigg[\bigg(\frac{|g(0)|}{L_g} + \sqrt{\Delta T_1} + |\Delta \widehat{X}_{T_1}|\bigg)^2\bigg]. \tag{2.8}$$

(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. □

2.1.2. *On the choice of β.* 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 and the computational effort:

- By the central limit theorem, the error induced by the Monte Carlo estimator based on the representation $\hat{\psi}$ 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 β by minimizing the product of the variance bound (2.8) and the mean computational effort. This minimization problem is obviously independent of the constants C, C', and reduces to

$$\min_{\beta>0} f(\beta) \qquad \text{where } f(\beta) := \beta T \exp\bigg(T\bigg(\beta + \frac{\gamma L^2}{\beta}\bigg)\bigg).$$

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

$$\beta^* := \sqrt{\gamma L^2 T + \frac{1}{4T^2}} - \frac{1}{2T}.$$

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

2.2. *The path-dependent case.* We next provide an extension of the above estimator $\widehat{\psi}$ in (2.6) to the path-dependent case. In particular, the present setting covers the setting of delayed SDEs. Let $n > 0$, $0 = t_0 < t_1 < \cdots < t_n = T$, $\sigma_0 \in \mathbb{M}^d$ be a nondegenerate matrix, and $\mu : [0, T] \times \mathbb{R}^{d \times n} \to \mathbb{R}^d$ be a continuous function, Lipschitz in the space variable. Let X be the unique solution of the SDE, with initial condition $X_0 = x_0$,

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

and the objective is to compute the value,

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

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

REMARK 2.3. It is clear that the value $\widetilde{V}_0$ defined above can be characterized by a parabolic PDE system. Namely, for every $k = 1, \dots, n$ and $(x_1, \dots, x_{k-1}) \in \mathbb{R}^{d \times (k-1)}$, we define

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

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

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

with terminal conditions

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

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

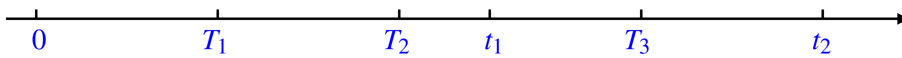
2.2.1. *The algorithm.* The unbiased simulation algorithm of $\widetilde{V}_0$ can be obtained by an iteration of the estimator (2.6) on each time interval $[t_k, t_{k+1}]$. One should just be careful about the integrability issue. Let us first introduce the algorithm.

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 $N = (N_s)_{0 \le s \le t}$ 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 $\tilde{T}_j^k := t_k \wedge T_{N_{t_{k-1}}+j}$,

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

EXAMPLE 2.4. Let us illustrate the last notation in the case $n=2$ in a context of $\tilde{N}^1 = 2$ jump arrivals on $[0, t_1)$, and $\tilde{N}^2 = 1$ jump arrivals on $[t_1, t_2)$. The 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 $\tilde{T}_3^1 = t_1$; $\widetilde{W}_0^1 = 0$, $\widetilde{W}_1^1 = W_{T_1}$, $\widetilde{W}_2^1 = W_{T_2}$ and $\widetilde{W}_3^1 = W_{t_1}$. For $k=2$, we have $\tilde{T}_0^2 = t_1$, $\tilde{T}_1^2 = T_3$, $\tilde{T}_2^2 = t_2$, and $\widetilde{W}_0^2 = W_{t_1}$, $\widetilde{W}_1^2 = W_{T_3}$ and $\widetilde{W}_2^2 = W_{t_2}$.



We next introduce a process $(\widetilde{X}_j^{k,\mathbf{x}})$, $\forall j = 0, 1, \dots, N_k + 1$, 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 $\widetilde{X}_0^{k,\mathbf{x}} := x_{k-1}$ and

$$\widetilde{X}_{j+1}^{k,\mathbf{x}} := \widetilde{X}_j^{k,\mathbf{x}} + \mu_k(\tilde{T}_j^k, \widetilde{X}_j^{k,\mathbf{x}}) \Delta \tilde{T}_{j+1}^k + \sigma_0 \Delta \widetilde{W}_{j+1}^k. \tag{2.13}$$

Similarly, for every $j = 1, \dots, \tilde{N}^k$, we define an automatic differentiation weight, with μ_k defined by (2.11),

$$\widetilde{\mathcal{W}}_j^{k,x} := \frac{(\mu_k(\tilde{T}_j^k, \widetilde{X}_j^{k,\mathbf{x}}) - \mu_k(\tilde{T}_{j-1}^k, \widetilde{X}_{j-1}^{k,\mathbf{x}})) \cdot (\sigma_0^\top)^{-1} \Delta \widetilde{W}_{j+1}^k}{\Delta \tilde{T}_{j+1}^k}. \tag{2.14}$$

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

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

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

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

$$\tilde{\psi}_k^{\mathbf{x}} := e^{\beta(t_k - t_{k-1})} \bigl(\tilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x}}} - \tilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x},0}} \mathbf{1}_{\{\tilde{N}^k > 0\}}\bigr) \beta^{-\tilde{N}^k} \prod_{j=1}^{\tilde{N}^k} \widetilde{\mathcal{W}}_j^{k,x}. \tag{2.15}$$

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

$$\tilde{\psi} := \tilde{\psi}_1^{x_0}. \tag{2.16}$$

EXAMPLE 2.5. In the context of Example 2.4:

- On $[0, t_1]$, we simulate the Brownian motion W on the discrete grid $0 < T_1 < T_2 < t_1$ and let $\widetilde{W}_0^1 = W_0 = 0$, $\widetilde{W}_1^1 := W_{T_1}$, $\widetilde{W}_2^1 := W_{T_2}$, $\widetilde{W}_3^1 := W_{t_1}$. We next define $(\widetilde{X}_j^{1,x_0})_{0 \le j \le 3}$ on $0 < T_1 < T_2 < t_1$ by (2.13), and obtain two variables $\widetilde{\mathcal{W}}_1^{1,x_0}$ and $\widetilde{\mathcal{W}}_2^{1,x_0}$ by (2.14).

- On $[t_1, t_2]$, given the initial value W_{t_1}, we simulate the Brownian motion W on the discrete grid $t_1 < T_3 < t_2$ and let $\widetilde{W}_0^2 = W_{t_1}$, $\widetilde{W}_1^2 := W_{T_3}$, $\widetilde{W}_2^2 := W_{t_2}$. We next define two processes $(\widetilde{X}_j^{2,\widetilde{X}_2^1})_{0\le j\le 2}$ and $(\widetilde{X}_j^{2,\widetilde{X}_3^1})_{0\le j\le 2}$ by (2.13) with two different initial conditions: $\widetilde{X}_0^{2,\widetilde{X}_2^1} := \widetilde{X}_2^1$ and $\widetilde{X}_0^{2,\widetilde{X}_3^1} := \widetilde{X}_3^1$. Moreover, the two different processes induce two different variables $\widetilde{\mathcal{W}}_1^{2,\widetilde{X}_2^1}$ and $\widetilde{\mathcal{W}}_1^{2,\widetilde{X}_3^1}$ by (2.14).
- The two different processes on $[t_1, t_2]$ induce two different variables:

$$\widetilde{\psi}_2^{\widetilde{X}_3^1} := e^{\beta(t_2-t_1)}\big(g\big(\widetilde{X}_2^1, \widetilde{X}_2^{2,\widetilde{X}_3^1}\big) - g\big(\widetilde{X}_2^1, \widetilde{X}_1^{2,\widetilde{X}_3^1}\big)\big)\beta^{-1}\widetilde{\mathcal{W}}_1^{2,\widetilde{X}_3^1}$$

  and

$$\widetilde{\psi}_2^{\widetilde{X}_2^1} := e^{\beta(t_2-t_1)}\big(g\big(\widetilde{X}_2^1, \widetilde{X}_2^{2,\widetilde{X}_2^1}\big) - g\big(\widetilde{X}_2^1, \widetilde{X}_1^{2,\widetilde{X}_2^1}\big)\big)\beta^{-1}\widetilde{\mathcal{W}}_1^{2,\widetilde{X}_2^1}.$$

- With $\widetilde{\psi}_2^{\widetilde{X}_2^1}$, $\widetilde{\psi}_2^{\widetilde{X}_3^1}$ and the variables on $[0, t_1]$, we obtain the variable

$$\widetilde{\psi} := \widetilde{\psi}_1^{x_0} = e^{\beta t_1}\big(\widetilde{\psi}_2^{\widetilde{X}_3^1} - \widetilde{\psi}_2^{\widetilde{X}_2^1}\big)\beta^{-2}\widetilde{\mathcal{W}}_1^{1,x_0}\widetilde{\mathcal{W}}_2^{1,x_0}.$$

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 $\widetilde{\psi}$ in (2.16) is integrable, it is not surprising to obtain the representation $\widetilde{V}_0 = \mathbb{E}[\widetilde{\psi}]$ as a consequence of Theorem 2.2. However, because of the renormalization term $(\widetilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x}}} - \widetilde{\psi}_{k+1}^{\mathbf{X}^{k,\mathbf{x},0}}\mathbf{1}_{\{\tilde{N}^k>0\}})$ in (2.15), the variance analysis is less obvious. We provide here a sufficient condition to ensure that $\widetilde{\psi}$ has finite variance.

THEOREM 2.6. *Suppose that $\mu : [0, T] \times \mathbb{R}^{d\times n} \to \mathbb{R}^d$ and $g : \mathbb{R}^{d\times n} \to \mathbb{R}$ are differentiable up to the order n, with bounded derivatives. Then*

$$\mathbb{E}\big[\widetilde{\psi}^2\big] < \infty \quad \textit{and} \quad \widetilde{V}_0 := \mathbb{E}[\widetilde{\psi}].$$

We will prove the integrability result here, and defer the proof of the representation result $\widetilde{V}_0 := \mathbb{E}[\widetilde{\psi}]$ to Section 5. As preparation, we start with two technical lemmas. Let $\pi = (0 = s_0 < s_1 < \cdots < s_m = T)$ be an arbitrary partition of the interval $[0, T]$, $\bar{\mu} : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ a $\mathbb{R}^d$−valued function. We define $X^{\pi,x}$ by $X_0^{\pi,x} := x$ and

$$X_{k+1}^{\pi,x} := X_k^{\pi,x} + \bar{\mu}\big(s_k, X_k^{\pi,x}\big)\Delta s_{k+1} + W_{s_{k+1}} - W_{s_k}. \tag{2.17}$$

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

LEMMA 2.7. *Suppose that $x \mapsto \bar{\mu}(t,x)$ is differentiable up to order n with uniformly bounded derivatives, and let $X^{\pi,x}$ be defined by* (2.17) *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 π such that*

$$\max_{1\le \ell\le n}\max_{i\in\{1,\dots,d\}^\ell}\max_{0\le k\le m}\big|\partial^\ell_{x,i} X_k^{\pi,x}\big| \le C.$$

PROOF. For simplicity, we consider the one-dimensional $d = 1$ case, the multidimensional can follows from the same arguments. First, for $\ell = 1$, we have

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

which implies that

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

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 $1 \le k \le m$ and the partition π. Denote by $\partial_x^i X_k^{\pi,x}$ the ith derivative of $x \mapsto X_k^{\pi,x}$. By induction, it is easy to deduce that for $\ell = 2,\dots,n$,

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

where P_ℓ is a polynomial. By induction, it then follows that $\partial_x^\ell X_k^{\pi,x}$ is also bounded by some constant C_ℓ independent of $k = 1,\dots,m$ and the partition π. □

LEMMA 2.8. *Let $(\widetilde{\psi}_k^{\mathbf{x}})_{1\le k\le n+1}$ be defined by* (2.15). *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 $x_{k-1} \mapsto \widetilde{\psi}_k^{\mathbf{x}}$ has derivatives up to order $k-1$ and*

$$\max_{1\le \ell\le k-1}\big|\partial^\ell_{x_{k-1}}\widetilde{\psi}_k^{\mathbf{x}}\big| \le C\prod_{j=k}^{n}\big(\tilde{N}^j + 1\big)^{j-1}. \tag{2.18}$$

PROOF. We proceed by induction. First, for $k = n+1$, we have $\widetilde{\psi}_{n+1}^{\mathbf{x}} := g(x, x_1,\dots,x_n)$ and hence $|\partial^\ell_{x_n}\widetilde{\psi}^{\mathbf{x}}| \le C$ for some constant C and for every $\ell = 1,\dots,n$.

Next, suppose that (2.18) holds true for $\widetilde{\psi}_{k+1}^{\mathbf{x}}$, we know from (2.15) that

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

Then using the estimation in Lemma 2.7 we see that (2.18) is also true for $\widetilde{\psi}_k^{\mathbf{x}}$. □

PROOF OF THEOREM 2.6 (I). By Lemma 2.8, we know that $x \mapsto \widetilde{\psi}_2^{x,x}$ 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 $\widetilde{\psi}_1^{x_0}$ falls into the Markovian case $n=1$, but with terminal condition $x \mapsto \widetilde{\psi}_2^{x,x}$. Notice that $\tilde{N}^k \le N_T$ has a Poisson distribution: $\mathbb{P}(N_T = m) = e^{-\beta T}\frac{(\beta T)^m}{m!}$. It follows that, for some constant $C > 0$,

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

which implies that $\widetilde{\psi}$ has finite variance. □

3. Unbiased simulation of general SDEs. Let us now consider the SDE (1.1) with general diffusion coefficient function, that is, with drift and diffusion coefficients $\mu : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{M}^d$:

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

Our objective of study in this section is

$$V_0 = \mathbb{E}\big[g(X_T)\big] \qquad \text{for some function } g : \mathbb{R}^d \to \mathbb{R}.$$

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

REMARK 3.1 (Lamperti's transformation). In some cases, the 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 σ is positive an C^1, the function $h(t, x) := \int_0^x \frac{1}{\sigma(t,y)}\,dy$, $(t, x) \in [0, T] \times \mathbb{R}$ is $C^{1,2}$ and strictly increasing in x, with inverse function denoted $h^{-1}(t, \cdot)$. By Itô's formula, it follows that $Y_t := h(t, X_t)$ satisfies the SDE

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

whose diffusion coefficient is a constant as in the SDE (2.1).

(ii) When $d > 1$, the last transformation is also possible when the diffusion field σ is positive and its inverse $\sigma^{-1} = Dh$ for some $C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$-function h. In particular, this requires that $\partial_{x_j}(\sigma^{-1})_{k,i} = \partial_{x_i}(\sigma^{-1})_{k,j}$ for all $k, i, j = 1, \dots, d$. Then it follows from an immediate application of Itô's formula that $Y := h(\cdot, X)$ is a Markov diffusion with unit diffusion.

3.1. *An estimator with infinite variance for general SDEs.* Let us impose the following conditions on coefficient functions μ and σ.

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

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

for all $(t,x),(s,y)\in[0,T]\times\mathbb{R}^d$; moreover $\sigma(t,x)$ is uniformly elliptic, that is, for some constant $\varepsilon_0>0$,

$$a(t,x):=\frac{1}{2}\sigma\sigma^\top(t,x)\ge\varepsilon_0 I_d \qquad \text{for all } (t,x)\in[0,T]\times\mathbb{R}^d.$$

Recall that $(T_k)_{k\ge 0}$ 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

$$\begin{aligned}\widehat{X}_{T_{k+1}}:=&\ \widehat{X}_{T_k}+\mu(T_k,\widehat{X}_{T_k})\Delta T_{k+1}\\&+\sigma(T_k,\widehat{X}_{T_k})\Delta W_{T_{k+1}},\qquad k=0,\dots,N_T.\end{aligned} \tag{3.2}$$

We then introduce a representation formula by

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

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

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

and

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

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

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

PROOF. (i) The random vectors $\xi^1_k:=\frac{\Delta W_{T_k}}{\sqrt{\Delta T_k}}$ and $\xi^2_k:=\frac{\Delta W_{T_k}\Delta W^\top_{T_k}-\Delta T_k I_d}{\Delta T_k}$, $k=1,\dots,N_T+1$, are independent of ΔT_k conditional on $\{\Delta T_k>0\}=\{N_T\ge k-1\}$, and 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 σ is uniformly bounded from below above zero. Then, for each $k = 1, \dots, N_T$,

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

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

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

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

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

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

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

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

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

for some constant $C > 0$, where we have also used the independence of the ξ_k^i's and the boundedness of their second-order moments. The integrability of $\widehat{\psi}$ is now a direct consequence of Lemma A.1.

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

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 has infinite variance. Therefore, the representation result through $\widehat{\psi}$ does not induce the standard dimension-free rate of convergence of the classical central limit theorem. Nevertheless, we believe that it is still interesting as an alternative to the representation of Bally and Kohatsu-Higa [2], 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 a higher order approximation $\widehat{X}$ of X, and obtain an estimator of finite variance for the one-dimensional ($d = 1$) driftless SDE of the form

$$X_0 = x_0, \qquad dX_t = \sigma(t, X_t)\, dW_t. \tag{3.5}$$

Our objective is to compute

$$V_0 := \mathbb{E}\big[g(X_T)\big] \qquad \text{for some function } g : \mathbb{R} \to \mathbb{R}.$$

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 $\widehat{X}_0 = x_0$,

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

for $k = 0, 1, \ldots, N_T$. Denoting

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

we write the solution of the above linear SDE (3.6) as

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

and

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

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

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

where the automatic differentiation weight is given by (see Lemma 5.8 below)

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

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

Similar to the discussion at the end of Section 3.1 (see also Remark 5.7 below), the variable $\widehat{\psi}$ in (3.10) is integrable but has infinite variance in general. In order to bypass this problem, we introduce antithetic variable $\widehat{X}_T^-$ of $\widehat{X}_T$ defined by

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

and

$$\widehat{X}_T^- = -\frac{c_1^{N_T}}{c_2^{N_T}} + \frac{c_1^{N_T}}{c_2^{N_T}} \exp\Bigl(-\frac{(c_2^{N_T})^2}{2}\Delta T_{N_T+1} - c_2^{N_T}\Delta W_{T_{N_T+1}}\Bigr)$$
$$+ \widehat{X}_{T_{N_T}} \exp\Bigl(-\frac{(c_2^{N_T})^2}{2}\Delta T_{N_T+1} - c_2^{N_T}\Delta W_{T_{N_T+1}}\Bigr) \qquad \text{if } c_2^{N_T} \neq 0.$$

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

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

We then introduce

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

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 and 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*

$$\mathbb{E}\bigl[|\widehat{\psi}|\bigr] + \mathbb{E}\bigl[|\overline{\psi}|^2\bigr] < \infty; \quad \textit{and} \quad V_0 = \mathbb{E}[\widehat{\psi}] = \mathbb{E}[\overline{\psi}]. \tag{3.13}$$

The proof is reported in Section 5.4.

REMARK 3.6. (i) As $\overline{\psi}$ has finite variance, we may use the representation of Theorem 3.5 to built 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 $\hat{\mu}(t,x) = c_1 + c_2 x$ and $\hat{\sigma}(t,x) = c_3 + c_4 x$ and a linear SDE

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

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 a linear SDE (3.14) in high dimensional case, as well as the associated automatic differentiation (Malliavin) weight as in (3.11) (see also Lemma 5.8 below), is still an open question.

4. Numerical examples. Notice that our estimator $\widehat{\psi}$ given by (2.6) [resp., $\widetilde{\psi}$ given by (2.15) and (2.16)] is an unbiased estimator for V_0 in (2.2) [resp., $\widetilde{V}_0$ 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., $\widetilde{V}_0$) is of order $O(\varepsilon^{-2})$, thus avoiding of the dependence on the discretization error.

By combining different levels of simulations, the MultiLevel Monte Carlo (MLMC) method proposed by Giles [11] 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 [20] 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 [11].

4.1. *Two one-dimensional SDEs.* Let W be a one-dimensional standard Brownian motion, we consider the SDE

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

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),

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

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

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

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 2.6. 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 [11], and we compare the computation time (in seconds) of the two methods. More precisely, the statistical error of the unbiased simulation method is given by $\sqrt{\mathrm{Var}[\widehat{\psi}]/N}$, where $\mathrm{Var}[\widehat{\psi}]$ denotes the estimated variance of $\widehat{\psi}$. For the implementation of MLMC, we choose $M = 4$, $N_L = 10^4$ and use equation (10) in [11] as criteria to stop the loop in MLMC (see more details in Section 5 of [11] 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

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

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

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

We implement the MLMC algorithm and our unbiased simulation method with different value of β, but with a given fixed error $\varepsilon = 0.0002$. The two methods provide very close estimation of value V_0, so we give a comparison on the computation time in Figure 1. We can observe that β in the unbiased simulation method

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 seconds

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

	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

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.

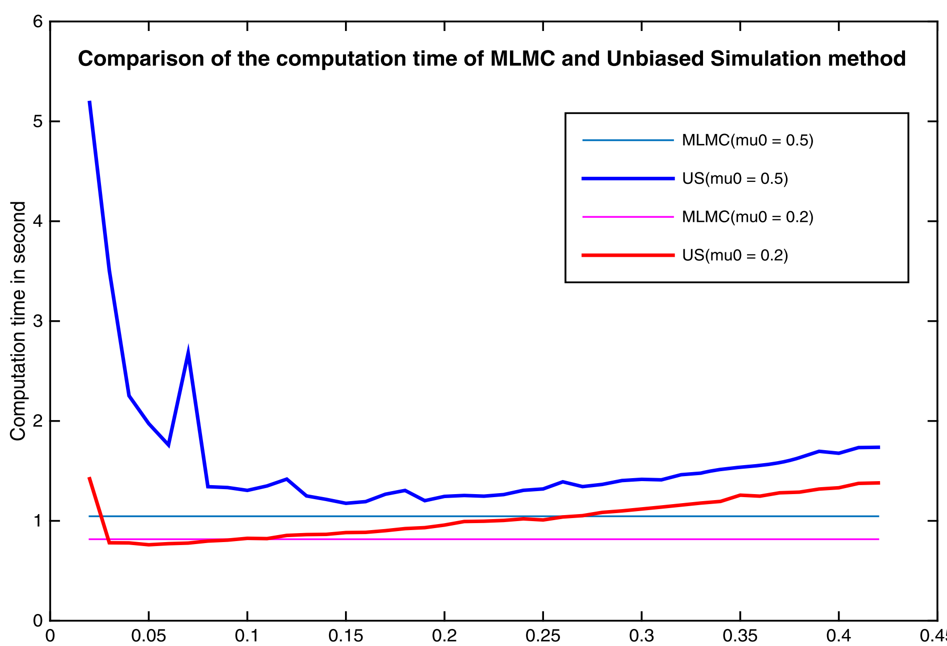


FIG. 1. *Comparison of the computation time of MLMC method and unbiased simulation method for problem (4.3), with the same given error.*

4.2. *A multidimensional SDE.* Let $W = (W^1, \dots, W^4)^\top$ be a 4-dimensional standard Brownian motion, and σ_0 the 4×4 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

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

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

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

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. Then we use the errors, obtained from our unbiased simulation method, in the MLMC algorithm in Section 5 of Giles [11], and we compare the computation time (in seconds) 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.

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 seconds

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

	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

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

$$V_0 = \mathbb{E}\big[(X_T - K)_+\big] \qquad \text{with } X_0 = 1, \qquad dX_t = \frac{2\sigma}{1 + X_t^2}\, dW_t. \tag{4.5}$$

We perform the following three implementations:

– the standard Euler scheme with time step $\Delta t = 1/10$ and simulation number $N = 10^6$,
– the unbiased simulation method (3.12) with $\beta = 0.1$ and simulation number $N = 10^6$,
– the MLMC scheme using the statistical error obtained from the unbiased simulation method.

The results displayed in Table 5 show that all three methods provide very similar estimation of V_0. In particular, the unbiased simulation method has a significant advantage.

TABLE 5
Numerical results for V_0 in (4.5) (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 error even smaller than that of the Euler scheme

	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

5. Proofs.

5.1. *A toy example.* Before completing the technical part of the proofs for Theorems 2.2, 2.6 and 3.3, we would like to illustrate the main idea by studying a simplified example in the one-dimensional case with unit diffusion. Consider the processes X defined by the drift coefficients $\mu : \mathbb{R}_+ \times \mathbb{R} \longrightarrow \mathbb{R}$, and $\widehat{X}$ with constant drift function $b \in \mathbb{R}$:

$$X_t = x_0 + \int_0^t \mu(s, X_s)\,ds + W_t, \quad \text{and} \quad \widehat{X}_t := x_0 + bt + W_t, \qquad t \geq 0.$$

For $\beta > 0$, let $(\tau_k)_{k\geq 1}$ be a sequence of i.i.d. random variable with distribution $\mathcal{E}(\beta)$, define $(T_k)_{k\geq 1}$ by (2.4), $\Delta T_{k+1} := T_{k+1} - T_k$, $\Delta W_{k+1} := W_{T_{k+1}} - W_{T_k}$, and let

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

PROPOSITION 5.1. *Let $\mu(\cdot,\cdot)$ and $g(\cdot)$ be bounded smooth functions in C_b^2. Then for all constants $b \in \mathbb{R}$ and $\beta > 0$:*

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

PROOF. Since μ and g are uniformly bounded, and for some constant $C > 0$, the conditional expectation $\mathbb{E}_{\Delta T_{k+1}}[|\Delta W_{T_{k+1}}|] \leq C/\sqrt{\Delta T_{k+1}}$, the integrability of ψ follows from Lemma A.1. In order to prove that $\mathbb{E}[g(X_T)] = \mathbb{E}[\psi]$, we introduce

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

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.1.

(i) Since μ and g are smooth functions, it follows from the Feynman–Kac formula that $\mathbb{E}[g(X_T)] = u(0, x_0)$, where $u \in C_b^\infty([0, T] \times \mathbb{R})$ is a solution of the PDE

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

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

$$-\partial_t u(t,x) - b\partial_x u(t,x) - \frac{1}{2}\partial^2_{xx} u(t,x) = \big(\mu(t,x) - b\big)\partial_x u(t,x),$$

we may also obtain the Feynman–Kac representation formula

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

where the second equality follows from the fact that $T_1 = T\wedge\tau_1$, and τ_1 is a random variable independent of $\widehat{X}$, with density function $\beta e^{-\beta t}\mathbf{1}_{\{t\ge 0\}}$.

(ii) For bounded and continuous function ϕ_0, and $t>0$, we obtain by direct differentiation of the gaussian marginal density function of the Brownian motion that

$$
\partial_x\mathbb{E}\big[\phi_0(x+bt+W_t)\big] = \mathbb{E}\Big[\phi_0(x+bt+W_t)\frac{W_t}{t}\Big]. \tag{5.3}
$$

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.2 that

$$
\begin{aligned}
\partial_x u(0,x_0) &= \mathbb{E}\Big[e^{\beta\Delta T_1}\frac{\Delta W_{T_1}}{\Delta T_1}\Big(g(\widehat{X}_T)\mathbf{1}_{\{T\le T_1\}} \\
&\qquad + \frac{\mu-b}{\beta}\partial_x u(T_1,\widehat{X}_{T_1})\mathbf{1}_{\{T_1<T\}}\Big)\Big].
\end{aligned}
\tag{5.4}
$$

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

$$
\begin{aligned}
\partial_x u(T_1,\widehat{X}_{T_1}) &= \mathbb{E}\Big[e^{\beta\Delta T_2}\frac{\Delta W_{T_2}}{\Delta T_2}\Big(g(\widehat{X}_T)\mathbf{1}_{\{T\le T_2\}} \\
&\qquad + \frac{\mu-b}{\beta}\partial_x u(T_2,\widehat{X}_{T_2})\mathbf{1}_{\{T_2<T\}}\Big)\Big| T_1,\widehat{X}_{T_1}\Big].
\end{aligned}
$$

Plugging the above expression of $\partial_x u(T_1,\widehat{X}_{T_1})$ into the right-hand side of (5.2), and using the fact that $T\le T_2$ is equivalent to $N_{T-}\le 1$, and $\mathbb{P}[\{N_{T-}\le 1\}\setminus\{N_T\le 1\}]=0$, 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 ψ_1, it follows that $u(0,x_0)=\mathbb{E}[\psi_2]$. Repeating the procedure, we have for all $n\ge 0$,

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

Finally sending $n\to\infty$, and using Lemma A.1 together with the dominated convergence theorem, it follows that $\mathbb{E}[g(X_T)] = \mathbb{E}[\lim_{n\to\infty}\psi_n] = \mathbb{E}[\psi]$. $\square$

REMARK 5.2. We may also interpret formally the representation ψ 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

$$\begin{aligned}\mathbb{E}[g(X_T)] &= \mathbb{E}\Bigg[\sum_{k=0}^{\infty}\frac{(\mu_0 T)^k}{k!}g^{(k)}(W_T)\Bigg] \\ &= \mathbb{E}\Bigg[e^{\beta T}g(W_T)\prod_{k=1}^{N_T}\bigg(\frac{\mu_0 \Delta W_{k+1}}{\beta \Delta T_{k+1}}\bigg)\Bigg],\end{aligned} \tag{5.5}$$

where the second equality follows by the fact that

$$\mathbb{P}[N_T = k] = e^{\beta T}\frac{(\beta T)^k}{k!}, \quad \text{and}$$

$$\mathbb{E}\Bigg[g^{(k)}\bigg(\sum_{i=0}^{k}\Delta W_{i+1}\bigg)\Bigg] = \mathbb{E}\Bigg[g\bigg(\sum_{i=0}^{k}\Delta W_{i+1}\bigg)\prod_{i=1}^{k}\frac{\Delta W_{i+1}}{\Delta T_{i+1}}\Bigg].$$

In particular, the right-hand side of (5.5) is exactly $\mathbb{E}[\psi]$ defined by (5.1) in this case.

Notice that ψ in (5.1) is integrable, but has infinite variance in general. In the next subsection, we exploit the arbitrariness of the constant b, choosing it adaptively at each time T_k, so as to restore square integrability of the modified estimator.

5.2. *A regime switching diffusion representation.* For $d \geq 1$, $T > 0$, let $(\mu, \sigma) : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{M}^d$ be bounded and continuous functions satisfying

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

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

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

with terminal condition $u(T,x) = g(x)$, where $a(\cdot) := \frac{1}{2}\sigma\sigma^\top(\cdot)$, $A : B := \mathrm{Tr}(AB^\top)$ 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

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

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

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

Then $(N_t)_{0\le t\le T}$ is a Poisson process with intensity β and arrival times $(T_k)_{k>0}$, and $T_0 = 0$. We also introduce, for all $k > 0$, $\Delta W^k_t := W_{(T_{k-1}+t)\wedge T_k} - W_{T_{k-1}}$. It is clear that the sequence of processes $(\Delta W^k_\cdot)_{k>0}$ are mutually independent.

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

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

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

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

EXAMPLE 5.3. (i) Let $(\hat{\mu}, \hat{\sigma})(s, y, t, x) = (\mu, \sigma)(s, y)$, then $\widehat{X}$ is defined as a Euler scheme as in (3.2), that is, $\widehat{X}_0 = x_0$, and

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

(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 the 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 the SDE (5.9). Let $\theta \in [0, T) \times \mathbb{R}^d$ and $(t, x) \in [0, T] \times \mathbb{R}^d$, the process $(\widetilde{X}^{t,x,\theta}_s)_{s\in[t,T]}$ is defined by the SDE

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

ASSUMPTION 5.4. There is a pair of measurable functions $(\widehat{\mathcal{W}}^1_\theta(\cdot), \widehat{\mathcal{W}}^2_\theta(\cdot))$, called automatic differentiation weights, taking values in $\mathbb{R}^d \times \mathbb{M}^d$, such that:

- $\widehat{\mathcal{W}}^i_\theta(t, x, s - t, (W_r - W_t)_{r\in[t,s]})$ is integrable, for all $\theta \in [0, T) \times \mathbb{R}^d$, $(t, x) \in [0, T) \times \mathbb{R}^d$, $s > t$, $i = 1, 2$,
- and for all bounded and continuous functions $\phi : \mathbb{R}^d \to \mathbb{R}$,

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

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

Let $a(\cdot) := \frac{1}{2}\sigma\sigma^\top(\cdot)$, $\hat{a}(\cdot) := \frac{1}{2}\hat{\sigma}\hat{\sigma}^\top(\cdot)$, $\widehat{\Theta}_0 = (t, x)$, and $\widehat{\Theta}_k = (T_k, \widehat{X}_{T_k})$,

$$\Delta f_k := \big((\mu, a) - (\hat{\mu}, \hat{a})(\widehat{\Theta}_{k-1}, .)\big)(T_k, \widehat{X}_{T_k}) \quad \text{and}$$

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

for $k \geq 1$. We next define

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

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

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

ASSUMPTION 5.5. (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:

$$\begin{aligned}&\widehat{\mathcal{W}}_\theta^1\big(t,x,\tau_1\wedge(T-t),(W_r-W_t)_{r\in[t,(t+\tau_1)\wedge T]}\big),\\ &\sup_{\varepsilon\in(0,\varepsilon_0]}\frac{1}{\varepsilon}\big[\widehat{\mathcal{W}}_\theta^1\big(t,x+\varepsilon e_i,\tau_1\wedge(T-t),(W_\cdot-W_t)\big)\\ &\qquad -\widehat{\mathcal{W}}_\theta^1\big(t,x,\tau_1\wedge(T-t),(W_\cdot-W_t)\big)\big]\end{aligned}$$

and

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

THEOREM 5.6. *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* 5.4 *and* 5.5 *hold true. Then $\widehat{\psi}$ is integrable and $u(0,x_0)=\mathbb{E}[\widehat{\psi}]$.*

REMARK 5.7. (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 5.6. 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 5.4.

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

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

However, 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 $[0, T]$ and, therefore, $\mathbb{E}[1/\Delta T_{N_T+1}] = \infty$. Consequently, the weight function $\widehat{\mathcal{W}}_k$ is typically of infinity variance, or even not integrable, in general. 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 5.5 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 5.6. (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:

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

Using the Feynmann–Kac formula, it follows that

$$\begin{aligned} u(0, x_0) = \mathbb{E}\bigg[& g\big(\widetilde{X}_T^{0,x_0,\theta}\big) \\ & + \int_0^T \big((\mu - \hat{\mu}_\theta) \cdot Du + (a - \hat{a}_\theta) : D^2 u\big)\big(s, \widetilde{X}_s^{0,x_0,\theta}\big)\, ds\bigg], \end{aligned} \tag{5.14}$$

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 τ_1 is a random variable of density $\beta e^{-\beta s}\mathbf{1}_{\{s\ge 0\}}$ independent of the Brownian motion W. Fixing $\theta = (0, x_0)$, it follows that

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

(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 5.5 with Lemma A.2, and also the fact that $Du(\cdot)$ is continuous, it follows that

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

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 τ_1 to the Brownian motion W, and setting $\theta = (0, x_0)$, it follows that

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

Next, for $D^2u(0,x_0)$, we use again Lemma A.2 together with the integrability condition in Assumption 5.5 and Lipschitz property of $x \mapsto ((\mu - \hat{\mu}) \cdot Du + (a - \hat{a}_\theta) : D^2u)(s,x)$, and the continuity of $D^2u(\cdot)$ that

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

Setting $\theta = (0,x_0)$, it follows from Assumption 5.4, we see that

$$\begin{aligned} D^2u(0,x_0) &= \mathbb{E}\big[\widehat{\psi}_0 \widehat{\mathcal{W}}^2_{(0,x_0)}(0,x_0,T_1,\Delta W^1_\cdot)\big] \\ &= \mathbb{E}\big[\big(\widehat{\psi}_0 - e^{\beta T} g(x_0)\mathbf{1}_{\{N_T=0\}}\big)\widehat{\mathcal{W}}^2_{(0,x_0)}(0,x_0,T_1,\Delta W^1_\cdot)\big], \end{aligned} \tag{5.16}$$

as $\mathbb{E}[\widehat{\mathcal{W}}^2_{(0,x_0)}(0,x_0,T,\Delta W_\cdot)] = 0$.

Changing the initial condition $(0,x_0)$ in (5.15) and (5.16) by $(T_1,\widehat{X}_1)$ (recall that $\widehat{\psi}_0$ 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), we see that

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

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

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

which concludes the proof. $\square$

5.3. *Proof of the representation results in Theorems* 2.2, 2.6 *and* 3.3. Using the results in Theorem 5.6, we can easily complete the proof of the representation results in Theorems 2.2, 2.6 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 5.6 with automatic differentiation weights function

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

In particular, it is clear that Assumption 5.4 holds true with the above choice of automatic differentiation weight functions in (5.17).

(ii) The uniform integrability conditions and integrability conditions in Assumption 5.5 can be easily obtained following the lines in the first part of the proof of Theorems 2.2 and 3.3 using Lemma A.1.

(iii) Now, suppose in addition that μ, σ and g are bounded smooth functions with bounded and continuous derivatives, so that $u \in C_b^{1,3}([0,T] \times \mathbb{R}^d)$. It follows by Theorem 5.6 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 $\varepsilon \to 0$. Let X^ε be the solution of

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

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

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

By Lemma A.1 together with the dominated convergence theorem, it follows that $\mathbb{E}[\widehat{\psi}^\varepsilon] \to \mathbb{E}[\widehat{\psi}]$ as $\varepsilon \to 0$, 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. □

PROOF OF THEOREM 2.6(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. □

5.4. *Proof of Theorem* 3.5. To introduce the algorithm in the context of Theorem 5.6, 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 $\overline{\mathcal{W}}_k^2$ in (3.11), as well as some related estimations. Let $c_1, c_2, x \in \mathbb{R}$ be constants such that $c_1 + c_2 x \neq 0$, we denote by $\overline{X}^{0,x}$ solution of the SDE

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

whose solution is given explicitly by

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

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

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

LEMMA 5.8. *Let $x \in \mathbb{R}$, $(c_1, c_2) \in \mathbb{R}^2$ be two constants such that $c_1 + c_2 x \neq 0$, $\phi : \mathbb{R} \to \mathbb{R}$ a bounded and continuous function.*

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

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

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

$$\begin{aligned} &\mathbb{E}\left[(\phi(\overline{X}_t^{0,x}) - \phi(x))^2\left(\frac{W_t}{t}\right)^2\right] \\ &\qquad + \mathbb{E}\left[(\phi(\overline{X}_t^{0,x}) - 2\phi(x) + \phi(\widetilde{X}_t^{0,x}))^2\left(\frac{W_t^2 - t}{t^2}\right)^2\right] \\ &\quad \le C(c_1 + c_2 x)^2. \end{aligned}$$

PROOF. (i) For $c_2 = 0$, the result reduces to the Gaussian case; see, for example, Lemma 2.1 of Fahim, Touzi and Warin [9]. Next, when $c_2 \neq 0$, denote $v(x) := \mathbb{E}[\phi(\overline{X}_t^{0,x})]$, then with the expression of $\overline{X}_t^{0,x}$ in (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

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

We compute similarly that

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

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

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

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

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

(ii) Since ϕ' and ϕ'' are uniformly bounded, we only focus of the nontrivial case $c_2 \neq 0$. Denoting $|\phi'|_\infty := \sup_x |\phi'(x)|$, we obtain by direct computation that

$$\begin{aligned}
&\mathbb{E}\bigg[\big(\phi\big(\overline{X}_t^{0,x}\big) - \phi(x)\big)^2\bigg(\frac{W_t}{t}\bigg)^2\bigg] \\
&\quad \le |\phi'|_\infty \mathbb{E}\bigg[\big(\overline{X}^{0,x} - x\big)^2 \frac{W_t^2}{t^2}\bigg] \\
&\quad = |\phi'|_\infty \mathbb{E}\bigg[(c_1+c_2x)^2\bigg(\frac{e^{-c_2^2t/2+c_2W_t} - 1}{c_2W_t - c_2^2t/2}\bigg)^2 \frac{W_t^2(c_2W_t - c_2^2t/2)^2}{t^2}\bigg],
\end{aligned}$$

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 $\ell(y) := (x + \frac{c_1}{c_2})(e^{-c_2^2t/2+c_2y} - 1)$, and define $\varphi(y) := \phi(x + \ell(y))$. Then

$$\begin{aligned}
\varphi''(y) &= \phi''\big(x+\ell(y)\big)(c_2+c_1x)^2e^{-c_2^2t+2c_2y} \\
&\quad + \phi'\big(x+\ell(y)\big)(c_2+c_1x)c_2e^{-c_2^2t/2+c_2y}.
\end{aligned} \tag{5.21}$$

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

$$\begin{aligned}
&\mathbb{E}\bigg[\big(\phi\big(\overline{X}_t^{0,x}\big) - 2\phi(x) + \phi\big(\widetilde{X}_t^{0,x}\big)\big)^2\bigg(\frac{W_t^2 - t}{t^2}\bigg)^2\bigg] \\
&\quad = \mathbb{E}\bigg[\big(\varphi(W_t) + \varphi(-W_t) - 2\varphi(0)\big)^2\bigg(\frac{W_t^2-t}{t^2}\bigg)^2\bigg] \\
&\qquad + \mathbb{E}\bigg[2\big(\varphi(0) - \phi(x)\big)^2\bigg(\frac{W_t^2-t}{t^2}\bigg)^2\bigg] \\
&\quad \le \mathbb{E}\bigg[\bigg(\frac{W_t^2(W_t^2-t)}{t^2}\bigg)^2 \sup_{|z|\le|W_t|}\varphi''(z)\bigg] \\
&\qquad + \mathbb{E}\bigg[2\bigg(\phi\bigg(x + \frac{c_1+c_2x}{c_2}\big(e^{-c_2^2t/2} - 1\big)\bigg) - \phi(x)\bigg)^2\bigg(\frac{W_t^2-t}{t^2}\bigg)^2\bigg],
\end{aligned}$$

which is also uniformly bounded by $C(c_1+c_2x)^2$ 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 5.8, we have $\mathbb{E}[|\overline{\psi}|^2]$ is bounded by

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

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

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

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

$$\begin{aligned}&\prod_{k=2}^{N_T}\Bigg[C\big(C+\big|\partial_x\sigma(T_{k-1},\widehat{X}_{T_{k-1}})\Delta\widehat{X}_{T_k}\big|\big)^2\bigg(C+\frac{\Delta\widehat{X}_{T_k}^2}{\Delta T_k}\bigg)^2\\ &\qquad\times\bigg(C|\Delta W_{T_k}|+\frac{\Delta W_{T_k}^2}{\Delta T_k}+1\bigg)^2\Bigg]\end{aligned}\tag{5.22}$$

is finite. Similar to the computation in item (ii) of Lemma 5.8, we have

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

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

$$\begin{aligned}&\mathbb{E}\Bigg\{\Bigg[\big(C+|\partial_x\sigma(T_{k-1},\widehat{X}_{T_{k-1}})\Delta\widehat{X}_{T_k}|\big)\bigg(C+\frac{\Delta\widehat{X}_{T_k}^2}{\Delta T_k}\bigg)\bigg(C|\Delta W_{T_k}|+\frac{\Delta W_{T_k}^2}{\Delta T_k}+1\bigg)\Bigg]^2\Bigg|\\ &\qquad\widehat{X}_{T_{k-1}},T_{k-1},\Delta T_k\Bigg\}\le C',\end{aligned}$$

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 $\overline{\psi}$ in (3.12) has finite variance.

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

$$\mathbb{E}[\widehat{\psi}|N_T,\Delta T_1,\dots,\Delta T_{N_T+1}] \le C^{N_T}\frac{1}{\sqrt{\Delta T_{N_T+1}}}\qquad\text{for some } C>0,$$

where the right-hand side is integrable but has infinite variance (see Lemma A.1). Similarly, it is easy to check the uniform integrability condition in Assumption 5.5 for $\widehat{\psi}$ in (3.10).

(iii) Finally, using Lemma 5.8(i), it follows that Assumption 5.4 holds true. Moreover, the regularity conditions on $\sigma(t,x)$ and g in Assumption 3.4 guarantee that $u \in C_b^{1,3}(\mathbb{R})$. We then deduce from Theorem 5.6 that $u(0,x_0) = \mathbb{E}[\widehat{\psi}] = \mathbb{E}[\overline{\psi}]$. $\square$

APPENDIX

In this section, we denote

$$S_0^t = 0, \qquad S_i^t := t \wedge \left(\sum_{j=1}^{i} \tau_j\right) \quad \text{and}$$

$$\Delta S_i^t := S_i^t - S_{i-1}^t \qquad \text{for } i \geq 1 \text{ and } t \geq 0.$$

In particular, for all $i \geq 1$, we have $S^T = T_i$ as defined in (2.4).

LEMMA A.1. *Let $0 = t_0 < t_1 < \cdots < t_n = T < \infty$. Then, for all constants $p \in (0,1)$ and $C > 0$, we have*

$$\mathbb{E}\left[C^{N_T} \prod_{k=1}^{N_T} (\Delta S_{i+1}^T)^{-p}\right] \leq \mathbb{E}\left[C^{N_T} \prod_{k=1}^{n} \prod_{i=N_{t_{k-1}}+1}^{N_{t_k}} (\Delta S_{i+1}^{t_k})^{-p}\right] < \infty.$$

PROOF. Notice that $\Delta S_i^T \geq \Delta S_i^{t_k}$ for all $i \geq 1$, $k = 1, \dots, n$. Then

$$\begin{aligned} \mathbb{E}\left[C^{N_T} \prod_{k=1}^{N_T} (\Delta S_{i+1}^T)^{-p}\right] &\leq \mathbb{E}\left[\prod_{k=1}^{n} \prod_{i=1+N_{t_{k-1}}}^{N_{t_k}} C(\Delta S_{i+1}^{t_k})^{-p}\right] \\ &= \prod_{k=1}^{n} \mathbb{E}\left[\prod_{i=1+N_{t_{k-1}}}^{N_{t_k}} C(\Delta S_{i+1}^{t_k})^{-p}\right] \\ &= \prod_{k=1}^{n} \mathbb{E}\left[\prod_{i=1}^{N_{\Delta t_k}} C(\Delta S_{i+1}^{\Delta t_k})^{-p}\right], \end{aligned} \tag{A.1}$$

where $\Delta t_k := t_k - t_{k-1}$, and the two last equalities follow from the independence and the stationarity of the increments of the Poisson process.

We next use the property that, conditional on $N_t = m$, the distribution of $(S_i^t, i = 1, \dots, m)$ is distributed as the order statistics of the uniform distribution on $[0,t]^m$

with density $t^{-m} m! \mathbf{1}_{\{0 \le s_1 \le \cdots \le s_m \le t\}}$. Then

$$
\begin{aligned}
\mathbb{E}\Bigg[\prod_{i=1}^{N_{\Delta t_k}} C\big(\Delta S_{i+1}^{\Delta t_k}\big)^{-p}\Bigg] &= \sum_{m \ge 1} \mathbb{P}[N_{\Delta t_k} = m](\Delta t_k)^{-mp} C^m G_{m,p} \\
&= e^{-\beta \Delta t_k} \sum_{m \ge 1} \frac{G_{m,p}}{m!} \big[\beta C (\Delta t_k)^{1-p}\big]^m,
\end{aligned} \tag{A.2}
$$

where

$$
\begin{aligned}
G_{m,p} &:= \mathbb{E}\Bigg[\prod_{i=1}^{m} \Delta S_{i+1}^1 \Big| N_1 = m\Bigg] \\
&= m! \int_0^1 \cdots \int_{u_{m-1}}^1 \big[u_1 \cdots (u_m - u_{m-1})\big]^{-p}\, du_1 \cdots du_m \\
&= m! \int_0^1 \cdots \int_{u_{m-2}}^1 \big[u_1 \cdots (u_{m-1} - u_{m-2})\big]^{-p} \\
&\qquad \times \frac{(1-u_{m-1})^{1-p}}{1-p}\, du_1 \cdots du_{m-1} \\
&\le \frac{m}{1-p} G_{m-1,p} \le \cdots \le \frac{m!}{(1-p)^m}.
\end{aligned}
$$

In view of the first inequality in (A.1), we can assume, without loss of generality, that the grid mesh $\max_{k \le n} \Delta t_k$ is sufficiently small so that $\lambda := \frac{\beta C}{1-p} \times (\max_{k \le n} \Delta t_k)^{1-p} < 1$, and the last estimate guarantees that the summation in (A.2) is finite. Then plugging into (A.1), we obtain our final estimate:

$$
\begin{aligned}
\mathbb{E}\Bigg[C^{N_T} \prod_{k=1}^{N_T} \big(\Delta S_{i+1}^T\big)^{-p}\Bigg] &\le \mathbb{E}\Bigg[\prod_{k=1}^{n} \prod_{i=1+N_{t_{k-1}}}^{N_{t_k}} C\big(\Delta S_{i+1}^{t_k}\big)^{-p}\Bigg] \\
&\le e^{-\beta T}(1-\lambda)^{-n}. \qquad \square
\end{aligned}
$$

Let X^x be the solution of the SDE

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

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

LEMMA A.2. *Suppose that for all bounded and continuous functions $\hat{\phi}$: $[0, T] \times \mathbb{R}^d \to \mathbb{R}$, the partial derivatives*

$$
\big(\partial_{x_i} \mathbb{E}[\hat{\phi}(t, X_t^x)], \partial^2_{x_i, x_j} \mathbb{E}[\hat{\phi}(t, X_t^x)]\big)_{i,j=1,\dots,d}
$$

exist, and there are some measurable $\mathbb{R}^d$-valued function $\widehat{\mathcal{W}}^1(x,t,(W_s)_{s\in[0,t]})$ such that

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

(i) *Let $F(dt)$ be a probability measure on $(0,T]$ such that, for each $x\in\mathbb{R}^d$, $i=1,\dots,d$,*

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

and $\phi:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ be a continuous (possibly unbounded) function satisfying

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

Then

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

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

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

and

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

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

$$\partial^2_{x_ix_j}\int_0^T \mathbb{E}[\phi(t,X_t^x)]F(dt) = \int_0^T \partial^2_{x_ix_j}\mathbb{E}[\phi(t,X_t^x)]F(dt), \tag{A.5}$$

where, in particular, the partial derivative at the left-hand side and the integration at the right-hand side 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 [13]), there is some constant C independent of $\varepsilon>0$ and $i=1,\dots d$, such that

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

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

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

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 ϕ 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 ϕ.

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

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

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. $\square$

Acknowledgments. We are grateful to Emmanuel Gobet, Ahmed Kebaier and two anonymous referees for valuable remarks and suggestions.

REFERENCES

[1] ANDERSSON, P. and KOHATSU-HIGA, A. (2017). Unbiased simulation of stochastic differential equations using parametrix expansions. *Bernoulli* **23** 2028–2057. MR3624885

[2] BALLY, V. and KOHATSU-HIGA, A. (2015). A probabilistic interpretation of the parametrix method. *Ann. Appl. Probab.* **25** 3095–3138. MR3404632

[3] BEN ALAYA, M. and KEBAIER, A. (2015). Central limit theorem for the multilevel Monte Carlo Euler method. *Ann. Appl. Probab.* **25** 211–234. MR3297771

[4] BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12** 1077–1098. MR2274855

[5] BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 333–382. MR2278331

[6] BESKOS, A. and ROBERTS, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.* **15** 2422–2444. MR2187299

[7] BLANCHET, J., CHEN, X. and DONG, J. (2017). ε-Strong simulation for multidimensional stochastic differential equations via rough path analysis. *Ann. Appl. Probab.* **27** 275–336. MR3619789

[8] BOMPIS, R. and GOBET, E. (2014). Stochastic approximation finite element method: Analytical formulas for multidimensional diffusion process. *SIAM J. Numer. Anal.* **52** 3140–3164. MR3293453

[9] FAHIM, A., TOUZI, N. and WARIN, X. (2011). A probabilistic numerical method for fully nonlinear parabolic PDEs. *Ann. Appl. Probab.* **21** 1322–1364. MR2857450

[10] FOURNIÉ, E., LASRY, J.-M., LEBUCHOUX, J., LIONS, P.-L. and TOUZI, N. (1999). Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.* **3** 391–412. MR1842285

[11] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Oper. Res.* **56** 607–617. MR2436856

[12] GILES, M. B. and SZPRUCH, L. (2014). Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.* **24** 1585–1620. MR3211005

[13] GRAHAM, C. and TALAY, D. (2013). *Stochastic Simulation and Monte Carlo Methods: Mathematical Foundations of Stochastic Simulation. Stochastic Modelling and Applied Probability* **68**. Springer, Heidelberg. MR3097957

[14] HENRY-LABORDÈRE, P. (2012). Counterparty risk valuation: A marked branching diffusion approach. Preprint. Available at arXiv:1203.2369.

[15] HENRY-LABORDÈRE, P., OUDJANE, N., TAN, X., TOUZI, N. and WARIN, X. (2016). Branching diffusion representation of semilinear PDEs and Monte Carlo approximation. Work in progress.

[16] HENRY-LABORDÈRE, P., TAN, X. and TOUZI, N. (2014). A numerical algorithm for a class of BSDEs via the branching process. *Stochastic Process. Appl.* **124** 1112–1140. MR3138609

[17] JOURDAIN, B. and SBAI, M. (2007). Exact retrospective Monte Carlo computation of arithmetic average Asian options. *Monte Carlo Methods Appl.* **13** 135–171. MR2338086

[18] KEBAIER, A. (2005). Statistical Romberg extrapolation: A new variance reduction method and applications to option pricing. *Ann. Appl. Probab.* **15** 2681–2705. MR2187308

[19] KLOEDEN, P. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations. Applications of Mathematics* **23**. Springer, Berlin.

[20] RHEE, C.-H. and GLYNN, P. W. (2015). Unbiased estimation with square root convergence for SDE models. *Oper. Res.* **63** 1026–1043. MR3422533

[21] TALAY, D. and TUBARO, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8** 483–509. MR1091544

[22] WAGNER, W. (1988). Unbiased multi-step estimators for the Monte Carlo evaluation of certain functional integrals. *J. Comput. Phys.* **79** 336–352. MR0973333

[23] WAGNER, W. (1989). Unbiased Monte Carlo estimators for functionals of weak solutions of stochastic differential equations. *Stoch. Stoch. Rep.* **28** 1–20. MR1013319

P. Henry-Labordère
Global Market Quantitative Research
Société Générale
29 Boulevard Haussmann
75009 Paris
France
E-mail: pierre.henry-labordere@sgcib.com

X. Tan
CEREMADE
PSL Research University
University of Paris-Dauphine
Place du Maréchal De Lattre De Tassigny
75775 Paris Cedex 16
France
E-mail: tan@ceremade.dauphine.fr

N. Touzi
Centre de Mathématiques Appliquées
École Polytechnique Paris
91128 Palaiseau Cedex
France
E-mail: nizar.touzi@polytechnique.edu