

MULTILEVEL PATH BRANCHING FOR DIGITAL OPTIONS

BY MICHAEL B. GILES^{1,a}  AND ABDUL-LATEEF HAJI-ALI^{2,b} 

¹MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, ^aMIKE.GILES@MATHS.OX.AC.UK

²MAXWELL INSTITUTE, SCHOOL OF MATHEMATICAL AND COMPUTER SCIENCES, HERIOT-WATT UNIVERSITY,
^bA.HAJIALI@HW.AC.UK

We propose a new Monte Carlo-based estimator for digital options with assets modelled by a stochastic differential equation (SDE). The new estimator is based on repeated path splitting and relies on the correlation of approximate paths of the underlying SDE that share parts of a Brownian path. Combining this new estimator with Multilevel Monte Carlo (MLMC) leads to an estimator with a computational complexity that is similar to the complexity of a MLMC estimator when applied to options with Lipschitz payoffs.

1. Introduction. In its simplest form, the Multilevel Monte Carlo (MLMC) path simulation method [6] considers a scalar SDE

$$(1) \quad dX_t = a(X_t, t) dt + \sigma(X_t, t) dW_t,$$

for $t \in [0, 1]$ with a sequence of approximate paths $\{(\bar{X}_{\ell,t})_{t \in [0,1]}\}_{\ell \in \{0,1,\dots\}}$ using uniform timesteps of size $h_\ell = h_0 M^{-\ell}$ for some $h_0 \in \mathbb{R}_+$ and $M \in \mathbb{Z}_+$. If we are interested in estimating $\mathbb{E}[f(X_1)] \approx \mathbb{E}[f(\bar{X}_{L,1})]$ for some function f and we define $\Delta P_\ell := f(\bar{X}_{\ell,1}) - f(\bar{X}_{\ell-1,1})$ with $\Delta P_0 := f(\bar{X}_{0,1})$, we have the telescoping summation

$$\mathbb{E}[P_L] = \sum_{\ell=0}^L \mathbb{E}[\Delta P_\ell].$$

The MLMC estimator is then

$$(2) \quad \sum_{\ell=0}^L \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \Delta P_\ell^{(i)},$$

with the coarse and fine paths within $\Delta P_\ell^{(i)}$ based on the same driving Brownian path. If there are constants α, β, γ such that the cost of a level ℓ sample ΔP_ℓ is $W_\ell \sim 2^{\gamma\ell}$, its variance is $V_\ell := \text{Var}[\Delta P_\ell] \sim 2^{-\beta\ell}$, and the weak error is $|\mathbb{E}[f(\bar{X}_{L,1}) - f(X_1)]| \sim 2^{-\alpha L}$, then an optimal number of levels, L , and an optimal number of samples per level, $\{N_\ell\}_{\ell=0}^L$, can be chosen to achieve a root-mean-square accuracy of ε with a computational complexity which is $\mathcal{O}(\varepsilon^{-2})$ if $\beta > \gamma$, $\mathcal{O}(\varepsilon^{-2} |\log \varepsilon|^2)$ if $\beta = \gamma$, and $\mathcal{O}(\varepsilon^{-2-(\gamma-\beta)/\alpha})$ if $\beta < \gamma$ [7].

If the function f is globally Lipschitz, with constant L_f , then

$$V_\ell \leq L_f^2 \mathbb{E}[(\bar{X}_{\ell,1} - \bar{X}_{\ell-1,1})^2].$$

In the case of the Euler-Maruyama discretization when the SDE coefficients, a and σ , are Lipschitz and grow linearly in x and are $1/2$ -Hölder continuous in t , this results in $V_\ell = \mathcal{O}(h_\ell)$ [18, Theorem 10.2.2] along with $W_\ell = \mathcal{O}(h_\ell^{-1})$, so $\beta = \gamma$ and the computational

MSC2020 subject classifications: Primary 65C05, 65C30; secondary 65B99, 60J85.

Keywords and phrases: Monte Carlo, Multilevel, Path splitting, Computational complexity, Branching processes.

complexity is $\mathcal{O}(\varepsilon^{-2}|\log \varepsilon|^2)$. When using a first-order Milstein discretization, and under additional differentiability assumptions on σ , the variance is reduced to $V_\ell = \mathcal{O}(h_\ell^2)$ and the complexity is improved to $\mathcal{O}(\varepsilon^{-2})$. A limitation of the first-order Milstein discretization is that it often requires the simulation of Lévy areas for multi-dimensional SDEs. To avoid this, Giles & Szpruch [12] developed an antithetic, truncated Milstein estimator which omits these Lévy area terms and still achieves an MLMC variance V_ℓ which is $\mathcal{O}(h_\ell^2)$ when f is smooth, and $\mathcal{O}(h_\ell^{3/2})$ when f is Lipschitz and piecewise smooth; both are sufficient for the computational complexity of MLMC to be $\mathcal{O}(\varepsilon^{-2})$.

In this article, we are concerned with the more difficult case in which f is a discontinuous function such as $f(x) = \mathbb{I}_{x>K}$; in computational finance this is referred to as a digital option. In this case ΔP_ℓ is nonzero only if the final values of the fine and coarse path approximations $\bar{X}_{\ell,1}$ and $\bar{X}_{\ell-1,1}$ within ΔP_ℓ are on opposite sides of K . Speaking loosely (we will be precise later), in the case of using Euler-Maruyama discretization, this only happens if $X_1, \bar{X}_{\ell,1}, \bar{X}_{\ell-1,1}$ are all within $\mathcal{O}(h_\ell^{1/2})$ of K , and the probability of that is $\mathcal{O}(h_\ell^{1/2})$. Hence $V_\ell \approx \mathcal{O}(h_\ell^{1/2})$ and so $\beta \approx \gamma/2$, resulting in a computational complexity which is approximately $\mathcal{O}(\varepsilon^{-5/2})$ since standard weak convergence results give $\alpha = 2\beta$. With the Milstein discretization, $V_\ell \approx \mathcal{O}(h_\ell)$ and the complexity is improved to $\mathcal{O}(\varepsilon^{-2}|\log \varepsilon|^2)$, but with the antithetic Milstein estimator V_ℓ remains $\mathcal{O}(h_\ell^{1/2})$.

The challenge of discontinuous functions such as this has been tackled in previous research. In the context of the first-order Milstein approximation, a conditional expectation with respect to the Brownian increment for the final timestep, conditional on the Brownian path up to that point, has been used to decrease the variance V_ℓ from $\mathcal{O}(h_\ell)$ to $\mathcal{O}(h_\ell^{3/2})$ [9]. In simple cases the conditional expectation can be evaluated analytically [9], while in harder cases a change of measure or path splitting can be used [7]. Unfortunately, none of these approaches work with the Euler-Maruyama discretization. One method which is effective for a subset of cases with particularly simple functions f is “pre-integration”, a variant of conditional expectation or conditional sampling based on the final value of the driving Brownian path. Originally developed to improve the effectiveness of Quasi-Monte Carlo integration [1, 15, 14], it also works well with MLMC [2]. Another effective method uses adaptive refinement of paths which lie close to the discontinuity [17]. When used for Euler-Maruyama or Milstein schemes, adaptive refinement methods recover the convergence rates of the variance, V_ℓ , that are observed for Lipschitz functions without substantially increasing the cost per sample. However, these methods lead to estimators with high kurtosis which can cause difficulties for MLMC algorithms that rely on variance estimates. Additionally, adaptive refinement does not recover the improved variance convergence rates of antithetic Milstein. See also [8] for a more thorough discussion of existing methodologies.

Inspired in part by the literature on dyadic Branching Brownian Motion [4, 20], the idea that we develop in the current article, as illustrated in Fig. 1, builds on path splitting where each MLMC sample, instead of corresponding to a single pair of fine and coarse paths, is an average of the difference ΔP_ℓ from many particles, each of which is a pair of fine and coarse paths. The branching process to generate the particles is similar to the process of dyadic Branching Brownian Motion, except that the branching times are deterministic not exponential random times. Fig. 1a illustrates the logical structure of the particle generation. If there are 2^ℓ timesteps for the fine path approximation $\bar{X}_\ell$, and $2^{\ell-1}$ timesteps for the coarse path $\bar{X}_{\ell-1}$, then in the simplest version of the algorithm the first branching from 1 to 2 particles is after $2^{\ell-1}$ fine timesteps, the second branch from 2 to 4 particles is after another $2^{\ell-2}$ fine timesteps, and so on, until there is only one coarse timestep left, at which there is a final branching into $2^{\ell-1}$ particles. This gives the following number of particles at different stages of the calculation:

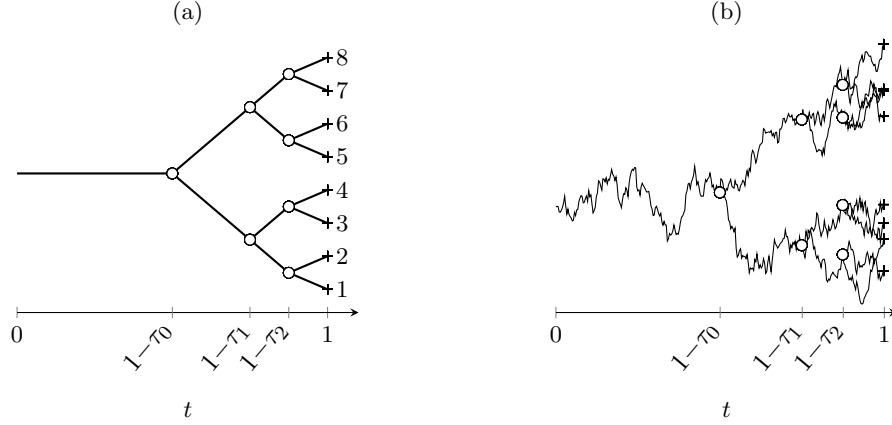


Fig 1: An illustration of the branching estimator $\Delta\mathcal{P}_4$ defined in Definition 2.2 for $\tau_{\ell'} = 2^{-\ell'-1}$ and $h_\ell = 2^{-\ell}$. (a) shows the logical structure ending up in the eight correlated samples of $\Delta\mathcal{P}_4$. (b) shows the eight underlying, correlated Brownian paths.

1	particle for first	$2^{\ell-1}$ fine timesteps
2	particles for next	$2^{\ell-2}$ fine timesteps
$4 = 2^2$	particles for next	$2^{\ell-3}$ fine timesteps
$\vdots$	$\vdots$	$\vdots$
$2^{\ell-2}$	particles for next	2 fine timesteps
$2^{\ell-1}$	particles for final	2 fine timesteps

so that the total cost (i.e. total number of particle-timesteps) is $(\ell+1)2^{\ell-1}$ which is not much more than the usual 2^ℓ cost per sample. The MLMC sample value would be an average of the outputs from the $2^{\ell-1}$ particles:

$$\Delta\mathcal{P}_\ell := \frac{1}{2^{\ell-1}} \sum_{i=1}^{2^{\ell-1}} \Delta P_\ell^{(i)},$$

i.e. this $\Delta\mathcal{P}_\ell$ counts as a single sample within an MLMC estimator similar to (2).

The claim is that with the Euler-Maruyama discretization we obtain $\text{Var}[\Delta\mathcal{P}_\ell] \approx \mathcal{O}(h_\ell)$ so that approximately we have $\beta \approx 1, \gamma \approx 1$. We present here a heuristic analysis which we make rigorous later. Suppose two particles (i) and (j) share a common driving Brownian path up to time $1-\tau$. Conditional on $X_{1-\tau}$, the distribution of X_1 is approximately Normal with a standard deviation of $\mathcal{O}(\tau^{1/2})$ and peak probability density of $\mathcal{O}(\tau^{-1/2})$. For both particles to finish within $\mathcal{O}(h_\ell^{1/2})$ of K (by which we mean that both the coarse and fine path approximations end within $\mathcal{O}(h_\ell^{1/2})$ of K) requires that $X_{1-\tau}$ lies within $\mathcal{O}(\tau^{1/2})$ of K , which occurs with probability $\mathcal{O}(\tau^{1/2})$, and conditional on this the probability that each particle finishes within $\mathcal{O}(h_\ell^{1/2})$ of K is $\mathcal{O}(h_\ell^{1/2}\tau^{-1/2})$. Hence, the probability that both particles finish within $\mathcal{O}(h_\ell^{1/2})$ of K is

$$\mathcal{O}\left(\tau^{1/2} (h_\ell^{1/2}\tau^{-1/2})^2\right) = \mathcal{O}(h_\ell\tau^{-1/2}),$$

and therefore

$$\mathbb{E}\left[|\Delta P_\ell^{(i)}| |\Delta P_\ell^{(j)}|\right] = \mathcal{O}(h_\ell\tau^{-1/2}).$$

There are $2^{2(\ell-1)}$ possible pairs (i, j) , and for each i the number of particle pairs (i, j) with $j \neq i$ and different τ values are:

$$\begin{array}{ll} 2^{\ell-2} & \text{with } \tau = 2^{-1} \\ 2^{\ell-3} & \text{with } \tau = 2^{-2} \\ 2^{\ell-4} & \text{with } \tau = 2^{-3} \\ \vdots & \vdots \\ 2 & \text{with } \tau = 2^{-(\ell-2)} \\ 1 & \text{with } \tau = 2^{-(\ell-1)} \end{array}$$

In addition there is the particle pair (i, i) for which $\mathbb{E} \left[\left(\Delta P_\ell^{(i)} \right)^2 \right] = \mathcal{O}(h_\ell^{1/2})$, as discussed previously. Together, these give

$$V_\ell \leq \mathbb{E}[\Delta \mathcal{P}_\ell^2] = \mathcal{O} \left(2^{-(\ell-1)} h_\ell^{1/2} + 2^{-(\ell-1)} \sum_{\ell'=1}^{\ell-1} 2^{\ell-1-\ell'} h_\ell 2^{\ell'/2} \right) = \mathcal{O}(h_\ell),$$

with the largest contribution coming from the $\tau=1/2$ branch, the most common branching point. This last observation suggests that the variance is well modelled by

$$\begin{aligned} V_\ell &\approx \text{Var} \left[\mathbb{E} \left[f(X_1) \mid X_{1/2} = \bar{X}_{\ell,1/2} \right] - \mathbb{E} \left[f(X_1) \mid X_{1/2} = \bar{X}_{\ell-1,1/2} \right] \right] \\ &\approx \text{Var} \left[(\bar{X}_{\ell,1/2} - \bar{X}_{\ell-1,1/2}) \cdot \nabla_x \mathbb{E} [f(X_1) \mid X_{1/2} = x] \Big|_{x=\bar{X}_{\ell-1,1/2}} \right], \end{aligned}$$

and we will later follow a similar approach in analyzing the branching estimator based on the antithetic Milstein approximation.

One final point for this introduction concerns optimization of the branching times. If $1-\tau_{\ell'}$ is the ℓ' -th branching time, then the total cost of $\Delta \mathcal{P}_\ell$ is of order

$$2^\ell \left(1 + \sum_{\ell'=1}^{\ell} 2^{\ell'} \tau_{\ell'} \right),$$

and the variance bound is of order

$$h_\ell \sum_{\ell'=1}^{\ell} 2^{-\ell'} \tau_{\ell'}^{-1/2}.$$

Optimizing $\tau_{\ell'}$ to minimize the cost for a fixed variance gives $\tau_{\ell'} \propto 2^{-4\ell'/3}$ which is slightly different to the initial choice of $\tau_{\ell'}=2^{-\ell'}$ and eliminates the additional linear factor in the cost. Hence, our main analysis will consider branching times $\tau_{\ell'}=2^{-\eta\ell'}$ for some constant η . These branching times may not coincide with discretization timesteps. There are two ways to handle this in an implementation. One is to round the times to the nearest coarse path timestep, and the other is to keep the times as specified in which case there is a common Brownian increment for the first part of the timestep, and then independent Brownian increments for the branched paths for the second part of the timestep.

In the remainder of the article, we consider the SDE (1) in d -dimensions for $t \in [0, 1]$:

$$(3) \quad dX_t = a(X_t, t) dt + \sigma(X_t, t) dW_t,$$

where $a : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d \times \mathbb{R}^{d'}$ are Borel-measurable functions and W is a d' -dimensional Wiener process and denote its natural filtration by $(\mathcal{F}_t)_{0 \leq t \leq 1}$. We will again refer to a corresponding sequence of approximations $\{\bar{X}_{\ell,t}\}$, $\ell = 0, 1, \dots$

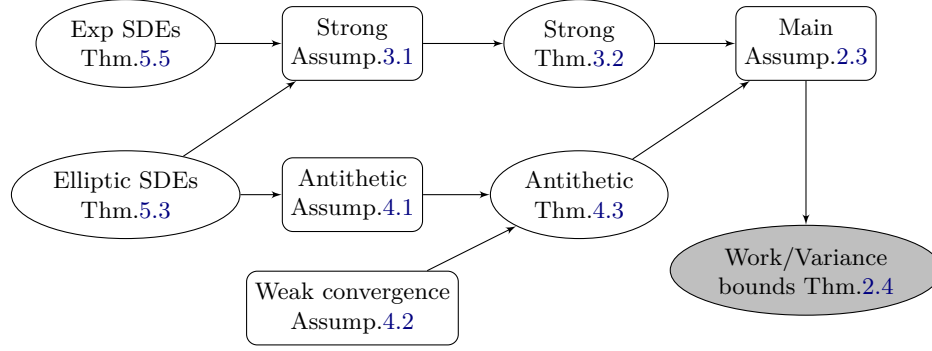


Fig 2: Outline of the analysis presented in the current work. Rectangles are assumptions while ellipses are lemmas and theorems. An arrow indicates implication under sufficient but not necessary conditions.

using uniform timesteps $h_\ell = h_0 M^{-\ell}$ for some $h_0 \in \mathbb{R}_+$ and $M \in \mathbb{Z}_+$. We assume that a and σ satisfy at least the necessary conditions (measurability, linear growth and global Lipschitzness in x) for existence and uniqueness of X_t in the strong sense [18]. Our goal is to estimate $\mathbb{P}[X_1 \in S] = \mathbb{E}[\mathbb{I}_{X_1 \in S}]$ for some closed set $S \subset \mathbb{R}^d$ with boundary $\partial S =: K$.

The article is organized as follows; see Fig. 2 for an outline of the assumptions/analysis carried out in the current work. In Section 2, we present the new branching estimator for a given underlying estimator ΔP_ℓ . We also bound the work and variance of the branching estimator in Theorem 2.4 under the main Assumption 2.3 on the underlying estimator ΔP_ℓ and show the improved computational complexity of MLMC when using the branching estimator in Corollary 2.7. In Section 3, we consider $\Delta P_\ell \equiv \mathbb{I}_{\bar{X}_{\ell,1} \in S} - \mathbb{I}_{\bar{X}_{\ell-1,1} \in S}$, and prove that under Assumption 3.1, “strong” approximations such as the Euler-Maruyama or Milstein numerical schemes satisfy Assumption 2.3. We conclude Section 3 with a numerical verification of the results in that section. In Section 4, we consider the antithetic estimator that was proposed in [12]. We again show in Theorem 4.3 that an antithetic estimator under Assumptions 4.1 and 4.2 satisfies Assumption 2.3 and conclude the section with a numerical verification of the presented theoretical results. In Section 5, we consider elliptic SDEs, i.e., SDEs whose coefficients are bounded and their diffusion coefficient is elliptic. In Theorem 5.3 we prove that the solutions to such SDEs satisfy Assumptions 3.1 and 4.1 under mild assumptions on K . Then in Theorem 5.5 we prove that exponentials of those solutions also satisfy Assumption 3.1 under different assumptions on K .

In what follows, for $\ell \in \mathbb{N} \equiv \{0, 1, \dots\}$, we use the notation $A_\ell \lesssim B_\ell$ to denote $A_\ell \leq cB_\ell$, and $A_\ell \simeq B_\ell$ to denote $c'B_\ell \leq A_\ell \leq cB_\ell$ for some constant, deterministic $c, c' > 0$ that are independent of the index, ℓ , the accuracy tolerance, ε , and other parameters which will be specified. For $u \in \{-1, 1\}^\ell$, let $|u|_0 := \ell$ and, for $\ell \geq 1$, let $\langle u \rangle := (u_1, u_2, \dots, u_{\ell-1}) \in \{-1, 1\}^{\ell-1}$, using the convention $\{-1, 1\}^0 := \{\emptyset\}$. For $\lambda \in (0, 1), k \in \mathbb{N}, l \in \mathbb{N}$, let $C_b^{k+\lambda, l}$ be the space of continuously differentiable and bounded functions $(x, t) \mapsto f(x, t)$ for $(x, t) \in \mathbb{R}^d \times [0, T]$ with uniformly continuous and bounded x -derivatives (resp. t -derivatives) up to order k (resp. l) and with x -derivatives of order k being λ -Hölder continuous in x uniformly with respect to $t \in [0, 1]$. When f is a vector- or a matrix-valued function, $f \in C_b^{k, l}$ means that all function components are in $C_b^{k, l}$. In addition, all vector and matrix norms are Euclidean ℓ^2 norms.

2. Branching Estimator. We begin by giving a formal definition to our branching estimator, using random discrete trees and branching processes [19].

Definition 2.1 (Branching Brownian Motion) . Given $\tau_0 \in (0, 1), \eta \in \mathbb{R}_+$, let $\tau_{\ell'} := \tau_0 2^{-\eta \ell'}$ for all $\ell' \in \mathbb{N}$. Let $\{W^u\}_{u \in \bigcup_{\ell'=0}^{\infty} \{-1, 1\}^{\ell'}}$ be mutually independent Wiener processes and let $\tilde{B}_t^{\varnothing} := W_t^{\varnothing}$ for $t \in [0, 1]$. Then for any $u \in \bigcup_{\ell'=1}^{\infty} \{-1, 1\}^{\ell'}$, define the u 'th branch of a Branching Brownian Motion as

$$\tilde{B}_t^u := \begin{cases} \tilde{B}_t^{\langle u \rangle} & t \in [0, 1 - \tau_{\langle u \rangle|_0}], \\ \tilde{B}_{1-\tau_{\langle u \rangle|_0}}^{\langle u \rangle} + W_{t-1+\tau_{\langle u \rangle|_0}}^u & t \in (1 - \tau_{\langle u \rangle|_0}, 1]. \end{cases}$$

Definition 2.2 (Branching estimator) . Given $\ell \in \mathbb{N}$, let $\hat{\ell} := \lfloor \log_2((h_{\ell}/\tau_0)^{-1})/\eta \rfloor$, such that $\tau_{\hat{\ell}-1} \simeq h_{\ell}$. Let $\{\bar{X}_{\ell,1}^u, \bar{X}_{\ell-1,1}^u\}_{u \in \{-1, 1\}^{\ell}}$ be approximations of the SDE path in (3) with timesteps of sizes $\{h_{\ell}, h_{\ell-1}\}$, respectively, and for a given a Branching Brownian Motion, $\{\tilde{B}^u\}_{u \in \{-1, 1\}^{\ell}}$, as in Definition 2.1, and let $\Delta P_{\ell}^u := \mathbb{I}_{\bar{X}_{\ell,1}^u \in S} - \mathbb{I}_{\bar{X}_{\ell-1,1}^u \in S}$. Finally, define the branching estimator as

$$(4) \quad \Delta \mathcal{P}_{\ell} := \frac{1}{2^{\hat{\ell}}} \sum_{u \in \{-1, 1\}^{\hat{\ell}}} \Delta P_{\ell}^u.$$

See Fig. 1 for an illustration of the path branching involved in $\Delta \mathcal{P}_{\ell}$. Note that a single branch of a Branching Brownian motion is itself a Brownian Motion. Hence the distribution of ΔP_{ℓ}^u for $u \in \{-1, 1\}^{\hat{\ell}}$ is independent of u . We will refer to a generic ΔP_{ℓ} and the filtration of its underlying Brownian motion, $(\mathcal{F}_t)_{0 \leq t \leq 1}$, when the dependence on u is not relevant. We will also refer to the cost of $\Delta \mathcal{P}_{\ell}$, which we define as the total number of Brownian increments needed to compute $\Delta \mathcal{P}_{\ell}$ based on an Euler-Maruyama approximation, a Milstein approximation which does not require simulations of Lévy areas or a truncated Milstein scheme as described in Section 4. We denote the cost of $\Delta \mathcal{P}_{\ell}$ by $\text{Work}(\Delta \mathcal{P}_{\ell})$. Note that $\mathbb{E}[\Delta \mathcal{P}_{\ell}] = \mathbb{E}[\Delta P_{\ell}] = \mathbb{E}[\mathbb{I}_{\bar{X}_{\ell,1} \in S} - \mathbb{I}_{\bar{X}_{\ell-1,1} \in S}]$ and hence we can use $\Delta \mathcal{P}_{\ell}$ instead of ΔP_{ℓ} in the MLMC setup (2). However, under some conditions, we will see in this section that the cost of $\Delta \mathcal{P}_{\ell}$ is not significantly larger than the cost of ΔP_{ℓ} while the variance is significantly smaller, leading to a better computational complexity of MLMC. In particular, even though the cost of each sample of ΔP_{ℓ} is h_{ℓ}^{-1} , many of the samples share an underlying Brownian path up to some branching point and hence the total cost for the $2^{\hat{\ell}}$ samples is greatly reduced as we will show in Theorem 2.4. We make the following general assumptions which we will relate, in Sections 3 and 4, to assumptions on the SDE (3) and $K \equiv \partial S$:

Assumption 2.3 (Estimator assumptions) . There exist $\beta_c \geq \beta_d > 0, p \geq 0, \tau_0 \in (0, 1)$ such that

$$(5a) \quad \mathbb{E}[(\Delta P_{\ell})^2] \lesssim h_{\ell}^{\beta_d},$$

$$(5b) \quad \text{and} \quad \mathbb{E}[(\mathbb{E}[\Delta P_{\ell} | \mathcal{F}_{1-\tau}])^2] \lesssim h_{\ell}^{\beta_c} / \tau^p,$$

for all $\ell \in \mathbb{N}$ and $\tau \in [h_{\ell}, \tau_0]$. We also assume that the estimator satisfies the following bias constraint for some $\alpha \geq \beta_d/2$

$$(6) \quad \left| \mathbb{E}[\mathbb{I}_{\bar{X}_{\ell,1} \in S} - \mathbb{I}_{X_1 \in S}] \right| \lesssim h_{\ell}^{\alpha}.$$

Typical approximate values for β_c, β_d and p are in Table 1. The assumption (6) is shown to hold for the Euler-Maruyama scheme for $\alpha = 1$ in [13, Theorem 2.5], when the SDE (3) is uniformly elliptic, $a, \sigma \in C_b^{3,1}$ and $\frac{\partial \sigma}{\partial t} \in C_b^{1,0}$.

	β_d	β_c	p
Euler-Maruyama	1/2	1	1/2
Milstein	1	2	1/2
Antithetic Milstein	1/2	2	3/2

TABLE 1

Limiting values for β_d, β_c and p in Assumption 2.3 for estimators ΔP_ℓ based on Euler-Maruyama, Milstein or antithetic Milstein discretizations. These are proved later in Sections 3 and 4.

Theorem 2.4 (Work and variance) . For any $\ell \in \mathbb{N}$, the estimator ΔP_ℓ in Definition 2.2 satisfies

$$\text{Work}(\Delta P_\ell) \lesssim \begin{cases} h_\ell^{-\max\{1, 1/\eta\}} & \eta \neq 1, \\ \ell h_\ell^{-1} & \eta = 1, \end{cases}$$

and, under Assumption 2.3,

$$\text{Var}[\Delta P_\ell] \lesssim \begin{cases} h_\ell^{\beta_d+1/\eta} + h_\ell^{\beta_c-\max\{0, p-1/\eta\}} & \eta p \neq 1, \\ h_\ell^{\beta_d+1/\eta} + \ell h_\ell^{\beta_c} & \eta p = 1. \end{cases}$$

PROOF. The proof is a slight generalization of the argument in the Introduction.

Work: If the branching points coincide with the discretization grid specified by h_ℓ , each path on the time interval $[1-\tau_{\ell'-1}, 1-\tau_{\ell'}]$ contains $h_\ell^{-1}(\tau_{\ell'-1} - \tau_{\ell'})$ fine timesteps; if they do not coincide then at worst each path segment requires $\lfloor h_\ell^{-1}(\tau_{\ell'-1} - \tau_{\ell'}) \rfloor + 2$ Brownian increments. Accordingly, the total work is bounded by

$$\begin{aligned} & h_\ell^{-1} \left((1 - \tau_0) + \sum_{\ell'=1}^{\hat{\ell}-1} 2^{\ell'} (\tau_{\ell'-1} - \tau_{\ell'}) + 2^{\hat{\ell}} \tau_{\hat{\ell}-1} \right) + 2 \sum_{\ell'=0}^{\hat{\ell}-1} 2^{\ell'} \\ &= h_\ell^{-1} \left((1 - \tau_0) + \tau_0 \sum_{\ell'=1}^{\hat{\ell}-1} 2^{\ell'} (2^{-\eta(\ell'-1)} - 2^{-\eta\ell'}) + \tau_0 2^{\hat{\ell}} 2^{-\eta(\hat{\ell}-1)} \right) + 2 \sum_{\ell'=0}^{\hat{\ell}-1} 2^{\ell'} \\ &\lesssim \begin{cases} h_\ell^{-1} 2^{\max\{1-\eta, 0\}\hat{\ell}} + 2^{\hat{\ell}} & \eta \neq 1, \\ \hat{\ell} h_\ell^{-1} + 2^{\hat{\ell}} & \eta = 1. \end{cases} \end{aligned}$$

and noting that $2^{\hat{\ell}} \lesssim h_\ell^{-1/\eta}$ and $\hat{\ell} \lesssim \ell$ we obtain the desired result.

Variance:

$$\text{Var}[\Delta P_\ell] = \frac{1}{2^{2\hat{\ell}}} \sum_{u \in \{-1, 1\}^{\hat{\ell}}} \mathbb{E}[(\Delta P_\ell^u)^2] + \frac{1}{2^{2\hat{\ell}}} \sum_{u \in \{-1, 1\}^{\hat{\ell}}} \sum_{\substack{v \in \{-1, 1\}^{\hat{\ell}} \\ u \neq v}} \mathbb{E}[\Delta P_\ell^u \Delta P_\ell^v].$$

Using (5a), we have that

$$2^{-2\hat{\ell}} \sum_{u \in \{-1, 1\}^{\hat{\ell}}} \mathbb{E}[(\Delta P_\ell^u)^2] = 2^{-\hat{\ell}} \mathbb{E}[(\Delta P_\ell)^2] \lesssim 2^{-\hat{\ell}} h_\ell^{\beta_d} \lesssim h_\ell^{\beta_d+1/\eta}.$$

Let $|u \wedge v|_0 = \max\{\ell' \leq \min\{|u|_0, |v|_0\} : u_i = v_i \text{ for all } i \in \{1, \dots, \ell'\}\}$, and note that two payoff differences ΔP_ℓ^u and ΔP_ℓ^v share a path up to time $1-\tau_{|u \wedge v|_0}$ and then the paths are independent and identically distributed after that. Hence, using (5b), we have

$$\mathbb{E}[\Delta P_\ell^u \Delta P_\ell^v] = \mathbb{E} \left[\left(\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau_{|u \wedge v|_0}}] \right)^2 \right] \lesssim h_\ell^{\beta_c} (\tau_{|u \wedge v|_0})^{-p}.$$

We can then evaluate the double sum as

$$\begin{aligned}
\frac{1}{2^{2\hat{\ell}}} \sum_{u \in \{-1,1\}^{\hat{\ell}}} \sum_{\substack{v \in \{-1,1\}^{\hat{\ell}} \\ u \neq v}} \mathbb{E}[\Delta P_{\ell}^u \Delta P_{\ell}^v] &\lesssim \frac{h_{\ell}^{\beta_c}}{2^{2\hat{\ell}}} \sum_{u \in \{-1,1\}^{\hat{\ell}}} \sum_{\substack{v \in \{-1,1\}^{\hat{\ell}} \\ u \neq v}} \tau_{|u \wedge v|_0}^{-p} \\
&= \frac{h_{\ell}^{\beta_c}}{2^{2\hat{\ell}}} \sum_{\ell'=0}^{\hat{\ell}-1} 2^{2\hat{\ell}-\ell'-1} \tau_{\ell'}^{-p} \\
&\lesssim h_{\ell}^{\beta_c} \left(\sum_{\ell'=0}^{\hat{\ell}-1} 2^{(\eta p-1)\ell'} \right).
\end{aligned}$$

Bounding the sum based on the sign of $\eta p-1$ concludes the proof. $\square$

Remark 2.5 (Optimal η) . *Theorem 2.4 shows that the choice of η in the estimator $\mathcal{P}_{\ell}$ in Definition 2.2 compared to p is crucial to improving the variance convergence rate of the new estimator compared to β_d , the variance convergence rate of the simple estimator ΔP , without substantially increasing the cost of the new estimator. We can optimize the value of η , by noting that work has the term $\sum_{\ell'=0}^{\hat{\ell}-1} 2^{\ell'} \tau_{\ell'}$, and the variance has the term $\sum_{\ell'=0}^{\hat{\ell}-1} 2^{-\ell'} \tau_{\ell'}^{-p}$. Hence, minimizing the work subject to a constrained variance yields the optimal value of $\tau_{\ell'} \propto 2^{-\frac{2}{p+1}\ell'}$ and the optimal value of η is $\frac{2}{p+1}$.*

Remark 2.6 (Number of branches) . *In the estimator outlined above, two branches are created at each branching point, $1 - \tau_{\ell'}$ for $\ell' = 0, \dots, \hat{\ell} - 1$. The method and analysis can be easily extended to allow for a different number of branches at every branching point. However, after adjusting η to keep the total work constant this would not improve the variance convergence rate or the subsequent computational complexities that we later derive.*

Corollary 2.7 . *Under Assumption 2.3, an MLMC estimator with an MSE ε^2 based on the branching estimator $\Delta \mathcal{P}_{\ell}$ in Definition 2.2 with $p < \max\{2\alpha, 1\}$ and $\eta = \frac{2}{p+1}$ and $h_{\ell} = h_0 M^{-\ell}$ for $M \in \mathbb{N}_+$ has the following computational complexity*

$$\begin{cases} \mathcal{O}(\varepsilon^{-2-\max\{0, (1-p)/2-\beta_d\}/\alpha} |\log \varepsilon|^{2\mathbb{I}_{2\beta_d=1-p}}) & p < \min\{1, 2(\beta_c - \beta_d) - 1\}, \\ \mathcal{O}(\varepsilon^{-2-\max\{0, 1-\beta_c\}/\alpha} |\log \varepsilon|^{2\mathbb{I}_{\beta_c=1}}) & 2(\beta_c - \beta_d) - 1 \leq p < 1, \\ \mathcal{O}(\varepsilon^{-2-\max\{0, 1-\beta_c\}/\alpha} |\log \varepsilon|^{2\mathbb{I}_{\beta_c=1}+2\mathbb{I}_{\beta_c \leq 1}}) & p = 1, \\ \mathcal{O}(\varepsilon^{-2-\max\{0, p-\beta_c\}/\alpha} |\log \varepsilon|^{2\mathbb{I}_{\beta_c=p}}) & p > 1. \end{cases}$$

Since $\beta_c \geq \beta_d$ as in Assumption 2.3, the computational complexity of an MLMC estimator based on $\Delta \mathcal{P}_{\ell}$ is lower than that of an MLMC estimator based on ΔP_{ℓ} , the latter being $\mathcal{O}(\varepsilon^{-2-\max\{1-\beta_d, 0\}/\alpha} |\log(\varepsilon)|^{2\mathbb{I}_{\beta_d=1}})$, whenever $p < 1 + \beta_c - \beta_d$.

PROOF. Recall that the MLMC estimator is defined as

$$\sum_{\ell=0}^L \frac{1}{N_{\ell}} \sum_{m=1}^{N_{\ell}} \Delta \mathcal{P}_{\ell}^{(m)},$$

where $\Delta \mathcal{P}_{\ell}^{(m)}$ are independent samples of $\Delta \mathcal{P}_{\ell}$. For $V_{\ell} := \text{Var}[\Delta \mathcal{P}_{\ell}]$ and $W_{\ell} := \text{Work}(\Delta \mathcal{P}_{\ell})$ the total work of the MLMC estimator for a fixed L is $\sum_{\ell=0}^L W_{\ell} N_{\ell}$ while

the total variance is $\sum_{\ell=0}^L V_\ell/N_\ell$. Minimizing the work while constraining the variance to be less than $\varepsilon^2/2$ leads to the optimal choice of number of samples on level ℓ [7],

$$N_\ell = \left\lceil 2\varepsilon^{-2}(V_\ell/W_\ell)^{1/2} \left(\sum_{\ell=0}^L (W_\ell V_\ell)^{1/2} \right) \right\rceil,$$

so that the total work is bounded by

$$2\varepsilon^{-2} \left(\sum_{\ell=0}^L (W_\ell V_\ell)^{1/2} \right)^2 + \sum_{\ell=0}^L W_\ell.$$

The sum $\sum_{\ell=0}^L W_\ell \lesssim M^{L(1+\max\{p-1,0\}/2)} L^{\mathbb{I}_{p=1}}$ is dominated by the first term when setting $L \propto \frac{1}{\alpha} \log \varepsilon^{-1}$, such that the bias in (6) is $\mathcal{O}(\varepsilon)$, and when $p < \max\{2\alpha, 1\}$ and for sufficiently small ε . By Theorem 2.4, for $\eta = \frac{2}{p+1}$, we have

$$W_\ell V_\ell \lesssim \begin{cases} h_\ell^{\beta_d+(p-1)/2} + h_\ell^{\beta_c-1} & p < 1, \\ \ell h_\ell^{\beta_d} + \ell^2 h_\ell^{\beta_c-1} & p = 1, \\ h_\ell^{\beta_d} + h_\ell^{\beta_c-p} & p > 1. \end{cases}$$

Evaluating the sum for $h_L \approx \varepsilon^{1/\alpha}$, which implies $L \propto \frac{1}{\alpha} \log \varepsilon^{-1}$, yields the result. $\square$

Remark 2.8 . *The simple case $\eta = 1$ is optimal only when $p = 1$. In other cases, the computational complexity increases slightly compared to Corollary 2.7:*

$$\begin{cases} \mathcal{O}(\varepsilon^{-2-\max\{1-\beta_c,0\}/\alpha} |\log(\varepsilon)|^{2\mathbb{I}_{\beta_c=1}+\mathbb{I}_{\beta_c \leq 1}}) & p < 1, \\ \mathcal{O}(\varepsilon^{-2-\max\{1-\beta_c,0\}/\alpha} |\log(\varepsilon)|^{2\mathbb{I}_{\beta_c=1}+2\mathbb{I}_{\beta_c \leq 1}}) & p = 1, \\ \mathcal{O}(\varepsilon^{-2-\max\{p-\beta_c,0\}/\alpha} |\log(\varepsilon)|^{2\mathbb{I}_{\beta_c=p}+\mathbb{I}_{\beta_c \leq p}}) & p > 1. \end{cases}$$

3. Strong Analysis. In this section, we consider

$$(7) \quad \Delta P_\ell \equiv \mathbb{I}_{\bar{X}_{\ell,1} \in S} - \mathbb{I}_{\bar{X}_{\ell-1,1} \in S},$$

and make well-motivated assumptions on the solution of the SDE in (3) and its numerical approximation $\{\bar{X}_{\ell,t}\}_{0 \leq t \leq 1}$, and then show that our main Assumption 2.3 follows from these. We then present the results of several numerical experiments. For a set $J \subset \mathbb{R}^d$, define the distance of x to J as

$$(8) \quad d_J(x) := \inf_{y \in J} \|y - x\|.$$

Assumption 3.1 . *Assume that for some $\delta_0 > 0$ and all $0 < \delta \leq \delta_0$ and $\tau \in (0, 1]$, there is a constant C independent of δ, τ and $\mathcal{F}_{1-\tau}$ such that*

$$(9) \quad \mathbb{E} \left[(\mathbb{P}[d_K(X_1) \leq \delta | \mathcal{F}_{1-\tau}])^2 \right] \leq C \frac{\delta^2}{\tau^{1/2}}.$$

Assumption 3.1 is fairly generic and depends on the set $K \equiv \partial S$ and the conditional density of X_1 given the filtration $\mathcal{F}_{1-\tau}$. It is motivated by the case when $(X_t)_{t \geq 0}$ is a d -dimensional Wiener process and we prove it in Theorem 5.3 (and Theorem 5.5) for solutions (and exponentials of solutions) to uniformly elliptic SDEs.

Theorem 3.2 . *Let Assumption 3.1 hold and assume that there is $q \geq 1$ and $\beta > 0$ such that*

$$(10) \quad \mathbb{E} \left[\|X_1 - \bar{X}_{\ell,1}\|^q \right]^{1/q} \lesssim h_\ell^{\beta/2}.$$

Then for ΔP_ℓ in (7) and all $\tau \in (0, 1)$

$$(11a) \quad \mathbb{E}[(\Delta P_\ell)^2] \lesssim h_\ell^{\beta(1-1/(q+1))/2}$$

$$(11b) \quad \text{and} \quad \mathbb{E}[(\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}])^2] \lesssim h_\ell^{\beta(1-2/(q+2))} / \tau^{1/2}.$$

This theorem shows that Assumption 2.3 is satisfied with $\beta_c = \beta(1-2/(q+2))$, $\beta_d = \beta(1-1/(q+1))/2$ and $p=1/2$. Under standard conditions on the coefficients of the SDE, assumption (10) is satisfied for the Euler-Maruyama and Milstein numerical schemes for $\beta = 1$ and $\beta = 2$, respectively, and any $q \geq 1$ [18].

PROOF. The proof of (11a) is similar to the proof in [10], and is included here for completeness. We start by noting that

$$\mathbb{E}[(\Delta P_\ell)^2] \leq \mathbb{E}[\|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell-1,1} \in S}\|] + \mathbb{E}[\|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell,1} \in S}\|],$$

and for any $\delta > 0$

$$\begin{aligned} \|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell,1} \in S}\| &\leq \mathbb{I}_{\|X_1 - \bar{X}_{\ell,1}\| > d_K(X_1)} \\ &\leq \mathbb{I}_{\|X_1 - \bar{X}_{\ell,1}\| > d_K(X_1)} \mathbb{I}_{d_K(X_1) \leq \delta} + \mathbb{I}_{\|X_1 - \bar{X}_{\ell,1}\| > d_K(X_1) > \delta} \\ &\leq \mathbb{I}_{d_K(X_1) \leq \delta} + \mathbb{I}_{\|X_1 - \bar{X}_{\ell,1}\| > \delta} \\ &\leq \mathbb{I}_{d_K(X_1) \leq \delta} + (h_\ell^{-\beta/2} \delta)^{-q} (h_\ell^{-\beta/2} \|X_1 - \bar{X}_{\ell,1}\|)^q. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E}[\|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell,1} \in S}\|] &\leq \mathbb{P}[d_K(X_1) \leq \delta] + (h_\ell^{\beta/2} \delta)^{-q} \mathbb{E}[(h_\ell^{\beta/2} \|X_1 - \bar{X}_{\ell,1}\|)^q] \\ &\leq C \delta + (h_\ell^{-\beta/2} \delta)^{-q} \mathbb{E}[(h_\ell^{-\beta/2} \|X_1 - \bar{X}_{\ell,1}\|)^q], \end{aligned}$$

where we used Assumption 3.1 for $\tau = 1$ and the fact that $\mathbb{E}[(h_\ell^{-\beta/2} \|X_1 - \bar{X}_{\ell,1}\|)^q]$ is bounded by (10). A similar bound is obtained for $\mathbb{E}[\|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell-1,1} \in S}\|]$ and then we select $\delta = h_\ell^{\frac{q}{q+1}\beta/2}$ to obtain (11a).

Proving (11b) follows the same steps by similarly noting that

$$\begin{aligned} \mathbb{E}[(\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}])^2] &\leq 2 \mathbb{E}[(\mathbb{E}[\|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell,1} \in S}\| | \mathcal{F}_{1-\tau}])^2] \\ &\quad + 2 \mathbb{E}[(\mathbb{E}[\|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell-1,1} \in S}\| | \mathcal{F}_{1-\tau}])^2], \end{aligned}$$

and for any $\delta > 0$

$$\|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell,1} \in S}\| \leq \mathbb{I}_{d_K(X_1) \leq \delta} + (h_\ell^{-\beta/2} \delta)^{-q/2} (h_\ell^{-\beta/2} \|X_1 - \bar{X}_{\ell,1}\|)^{q/2},$$

so that

$$\begin{aligned} \mathbb{E}[(\mathbb{E}[\|\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell,1} \in S}\| | \mathcal{F}_{1-\tau}])^2] &\leq 2 \mathbb{E}[(\mathbb{E}[\mathbb{I}_{d_K(X_1) \leq \delta} | \mathcal{F}_{1-\tau}])^2] \\ &\quad + 2 (h_\ell^{-\beta} \delta^2)^{-q/2} \mathbb{E}[(h_\ell^{-\beta/2} \|X_1 - \bar{X}_{\ell,1}\|)^q]. \end{aligned}$$

Here $\mathbb{E}\left[\left(h_\ell^{-\beta/2}\|X_1 - \bar{X}_{\ell,1}\|\right)^q\right]$ is bounded by (10), and using Assumption 3.1 we have

$$\begin{aligned} \mathbb{E}\left[\left(\mathbb{E}[\mathbb{I}_{X_1 \in S} - \mathbb{I}_{\bar{X}_{\ell,1} \in S} | \mathcal{F}_{1-\tau}]\right)^2\right] &\lesssim \delta^2/\tau^{1/2} + (h_\ell^{-\beta}\delta^2)^{-q/2} \\ &\lesssim \delta^2/\tau^{1/2} + (h_\ell^{-\beta}\delta^2)^{-q/2}/\tau^{1/2}, \end{aligned}$$

for $\tau < 1$ and we choose $\delta^2 = h_\ell^{\frac{q}{q+2}\beta}$ to obtain the result. $\square$

As a direct implication of Corollary 2.7 and Theorem 3.2, we have the following result

Corollary 3.3 (MLMC Computational Complexity) . *Let Assumption 3.1 and (10) be satisfied for all $q \geq 2$. When $\beta \leq 1$, assume further that the bias is bounded according to (6) for some α . Then, the computational complexity of a MLMC estimator with MSE ε^2 based on $\Delta\mathcal{P}_\ell$ and $h_\ell = h_0 M^{-\ell}$ for $M \in \mathbb{N}_+$ and $\eta = 4/3$ is*

$$\begin{cases} \mathcal{O}(\varepsilon^{-2-\frac{1-\beta}{\alpha}-\nu}) & \beta \leq 1, \\ \mathcal{O}(\varepsilon^{-2}) & \beta > 1, \end{cases}$$

for any $\nu > 0$.

Numerical Experiments. In this section, we consider the SDE for the Geometric Brownian Motion (GBM)

$$(12) \quad dX_{i,t} = \mu_i X_{i,t} dt + \sigma_i X_{i,t} \left(\rho dW_{i,t} + (1 - \rho^2)^{1/2} dW_{0,t} \right),$$

for $i = 1, \dots, d$. Here $\{(W_{i,t})_{t \geq 0}\}_{i=0}^d$ are independent Wiener processes. The processes $\{(W_{i,t})_{t \geq 0}\}_{i=1}^d$ model the idiosyncratic noise in the d -dimensional system while $(W_{0,t})_{t \geq 0}$ models the systematic noise in the system. As an example, we compute $\mathbb{P}[\frac{1}{d} \sum_{i=1}^d X_{i,1} \leq 1]$. We set $\rho = 0.7$ and $X_{i,0} = 1$, $\mu_i = 0.05$, $\sigma_i = 0.2$ for all $i \in \{1, \dots, d\}$. We approximate the path of $\{X_{i,t}\}_{i=1}^d$ using the Euler-Maruyama or the Milstein numerical schemes [18] and set the time step size at level ℓ as $h_\ell = 2^{-\ell-1}$ and use the new branching estimator in Definition 2.2 with $\eta = 1$ along with a traditional estimator without branching. Note that this sequence of time steps sizes is not optimal and other choices such as $h_\ell \propto 4^{-\ell}$ would lead to better computational performance for both the branching and non-branching estimators, see [6, 16] for further analysis. We choose this sub-optimal sequence as it produces more data points in the plots below and makes inferring the computational complexity and convergence trends easier.

Fig. 3 shows the convergence of $\mathbb{E}[(\mathbb{E}[\Delta\mathcal{P}_\ell | \mathcal{F}_{1-\tau}])^2]$ for an Euler-Maruyama approximation for $d=1, 2, 3$, verifying Assumption 2.3 for $p = 1/2$ as shown by Theorem 3.2. Figures 4a and 4b confirm the claims of Theorem 2.4. We only show the results for $d=1$ as the numerical results for $d>1$ show similar rates for the work and variance convergence when using Euler-Maruyama. Recall that for the example in (12), by Theorem 3.2, when (10) and Assumption 3.1 are satisfied as we argued above, we have $\beta_d \approx \beta/2$ and $\beta_c \approx \beta$, hence $\text{Var}[\Delta\mathcal{P}_\ell] \approx \mathcal{O}(h_\ell^{\min\{\beta, \beta/2+1\}})$ where $\beta=1$ for Euler-Maruyama and $\beta=2$ for the Milstein numerical scheme. Figure 4d shows the total work estimate of a MLMC sampler based on $\Delta\mathcal{P}_\ell$. For the previous values of β_c, β_d and β and $p = 1/2, \eta = 1$, the computational complexity for the MLMC sampler based on $\Delta\mathcal{P}_\ell$ is $\mathcal{O}(\varepsilon^{-2}|\log \varepsilon|^3)$ when using Euler-Maruyama and $\mathcal{O}(\varepsilon^{-2})$ when using Milstein, c.f., Remark 2.8. As discussed in Remark 2.5 and Theorem 2.4, in theory choosing $\eta=4/3$ when $p = 1/2$ leads to

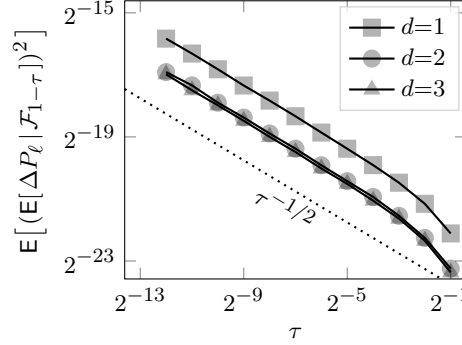


Fig 3: Numerical verification for (5b) with $p=1/2$ and $h_\ell=2^{-14}$ for the GBM example in (12) when using Euler-Maruyama.

smaller computational complexity than $\eta=1$; in particular the computational complexity of MLMC when using the branching estimator with an Euler-Maruyama scheme is $\mathcal{O}(\varepsilon^{-2}|\log \varepsilon|^2)$. In practice, we observed that the difference in computational complexity is not significant for reasonable tolerances due to the additional branching cost when using $\eta=4/3$ where the branching times do not align with the time-stepping scheme. In any case, these results are an improvement over computational complexity for the MLMC sampler based on ΔP_ℓ , labelled “Without branching”, which is approximately $\mathcal{O}(\varepsilon^{-5/2})$ for Euler-Maruyama, and $\mathcal{O}(\varepsilon^{-2}|\log(\varepsilon)|^2)$ for Milstein.

The kurtosis of ΔP_ℓ grows approximately in proportion to $h_\ell^{-1/2}$ when using Euler-Maruyama or h_ℓ^{-1} when using Milstein; recall that in both cases $\mathbb{E}[(\Delta P_\ell)^2] = \mathbb{E}[(\Delta P_\ell)^4]$. This leads to difficulties when estimating $\text{Var}[\Delta P_\ell]$ for sufficiently large ℓ and determining the optimal number of samples in MLMC for these levels becomes difficult. On the other hand, Figure 4c illustrates another benefit of our branching estimator: $\Delta \mathcal{P}_\ell$ has a bounded kurtosis and hence an MLMC algorithm that relies on variance estimates is more stable when using $\Delta \mathcal{P}_\ell$ than when using ΔP_ℓ . See also Appendix A for a proof of the boundedness of the kurtosis of $\Delta \mathcal{P}_\ell$.

4. Antithetic Estimator. As discussed in the previous section, the Milstein numerical scheme has faster variance convergence than Euler-Maruyama leading to lower computational complexity of MLMC estimators. However, for multi-dimensional SDEs, evaluating the Milstein scheme requires expensive sampling of Lévy areas in most cases. In [12], an antithetic estimator was introduced which has the same MLMC variance convergence rate as a Milstein approximation for smooth payoff functions $f(x)$, but without requiring sampling of Lévy areas. In this section, we analyse the corresponding branching sampler for such an antithetic estimator.

Let $\{(\bar{X}_{\ell,t})_{t=0}^1, (\bar{X}_{\ell,t}^{(a)})_{t=0}^1\}$ be an antithetic pair which are identically distributed. For example, [12] presents such an antithetic estimator for a Clark-Cameron SDE which is derived from a truncated Milstein discretization by setting the Lévy areas to zero. A similar branching estimator to Definition 2.2 can be defined by considering the triplet of approximate paths, $(\bar{X}_{\ell,t}^u, \bar{X}_{\ell,t}^{(a),u}, \bar{X}_{\ell-1,t}^u)_{t=0}^1$ for the same branch, $\tilde{B}^u$, of a Branching Brownian Motion. The antithetic estimator can then be defined as in (4) for

$$(13) \quad \Delta P_\ell^u = \begin{cases} \mathbb{I}_{\bar{X}_{0,1}^u \in S} & \ell = 0, \\ \frac{1}{2}(\mathbb{I}_{\bar{X}_{\ell,1}^u \in S} + \mathbb{I}_{\bar{X}_{\ell,1}^{(a),u} \in S}) - \mathbb{I}_{\bar{X}_{\ell-1,1}^u \in S} & \ell > 0. \end{cases}$$

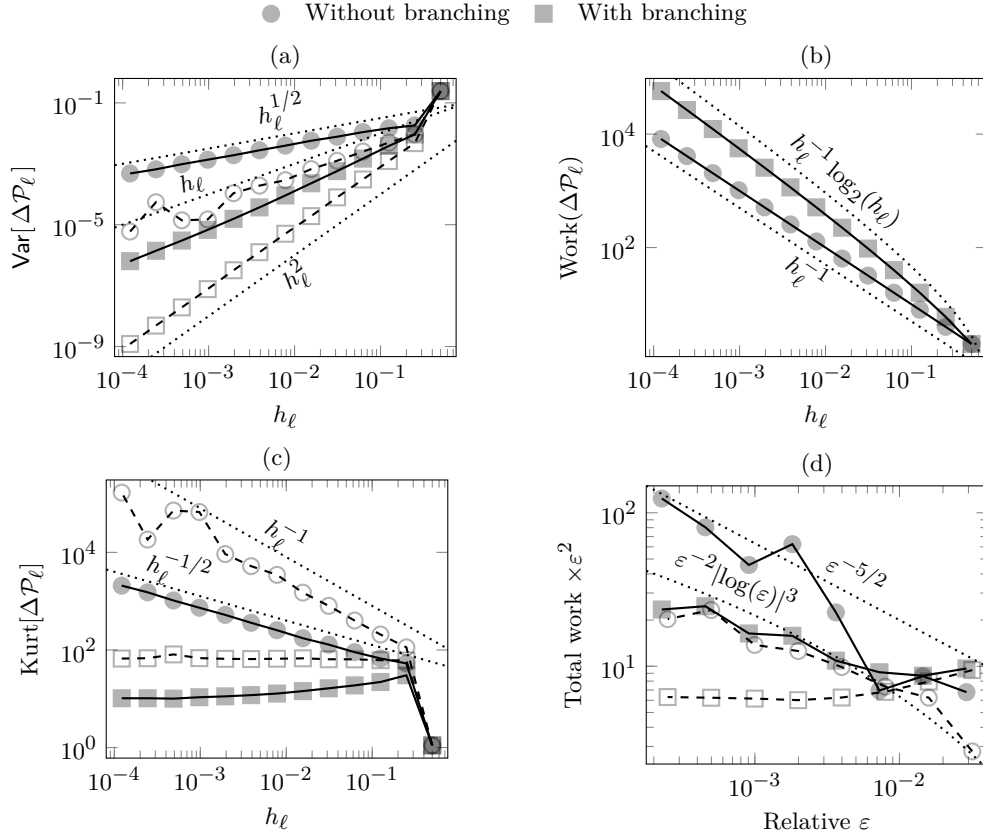


Fig 4: The GBM example in (12) for $d=1$ when using Euler-Maruyama (*solid*) and Milstein (*dashed*) in the traditional, ΔP_ℓ , and branching, $\Delta \mathcal{P}_\ell$, estimators. (a) shows numerical verification of the variance convergence of $\Delta \mathcal{P}_\ell$. (b) The work estimate per sample, based on the number of generated samples from the standard normal distribution. The work estimates when using the Milstein scheme are identical. (c) The kurtosis of $\Delta \mathcal{P}_\ell$. (d) The total work estimate of MLMC for different tolerances. This figure illustrates the improved computational complexity of MLMC when using the new branching estimator.

Since the cost of sampling $\bar{X}_{\ell,t}^{(a),u}$ is the same as $\bar{X}_{\ell,t}^u$, Theorem 2.4 still applies if Assumption 2.3 is satisfied for (13).

In this section, we will impose well-motivated assumptions that allow us to prove that Assumption 2.3 is satisfied for

$$(14) \quad \Delta P_\ell := \begin{cases} \mathbb{I}_{\bar{X}_{\ell,1} \in S} & \ell = 0, \\ \frac{1}{2}(\mathbb{I}_{\bar{X}_{\ell,1} \in S} + \mathbb{I}_{\bar{X}_{\ell,1}^{(a)} \in S}) - \mathbb{I}_{\bar{X}_{\ell-1,1} \in S} & \ell > 0. \end{cases}$$

which has the same distribution as ΔP_ℓ^u in (13). In what follows, define

$$g_t(\xi) := \mathbb{P}[X_1 \in S \mid X_t = \xi]$$

$$\text{and} \quad \bar{g}_{\ell,t}(\xi) := \mathbb{P}[\bar{X}_{\ell,1} \in S \mid \bar{X}_{\ell,t} = \xi].$$

We will make the following assumptions

Assumption 4.1 . Let ∇g_t and H_{g_t} be the Gradient and Hessian of g_t , respectively. We assume that there exist constants $c_1, c_2 > 0$ such that for all $\xi \in \mathbb{R}^d$ and $\tau \in (0, 1)$,

$$(15a) \quad \|\nabla g_{1-\tau}(\xi)\| \leq \frac{c_1}{\tau^{1/2}} \exp\left(-c_2 \frac{d_K^2(\xi)}{\tau}\right)$$

$$(15b) \quad \text{and} \quad \|H_{g_{1-\tau}}(\xi)\| \leq \frac{c_1}{\tau} \exp\left(-c_2 \frac{d_K^2(\xi)}{\tau}\right),$$

and for any $\tau \in (0, 1]$,

$$(16) \quad \mathbb{P}[d_K(X_{1-\tau}) < \delta] \leq c_1 \delta.$$

This assumption is motivated by the case when $(X_t)_{t \geq 0}$ is a d -dimensional Wiener process, i.e., $X_1 \sim \mathcal{N}(\xi, \tau)$ and we prove it in Theorem 5.3 for solutions of uniformly elliptic SDEs.

Assumption 4.2 . There exists a constant $c_1 > 0$ such that for any $\ell \in \mathbb{N}$, $\tau \in (0, 1)$ and $\xi \in \mathbb{R}^d$,

$$(17) \quad |\bar{g}_{\ell, 1-\tau}(\xi) - g_{1-\tau}(\xi)| \leq \frac{c_1 h_\ell}{\tau^{1/2}}.$$

This assumption is motivated by a result that was proved in [13, Theorem 2.3] for an Euler-Maruyama scheme. In particular, letting $\Gamma(\cdot, t; \xi, s)$ and $\Gamma_\ell(\cdot, t; \xi, s)$ be the densities of X_t and the Euler-Maruyama approximation, $\bar{X}_{\ell, t}$, respectively, given $X_s = \bar{X}_{\ell, s} = \xi$, the authors in [13] proved that when the SDE is uniformly elliptic and the coefficients $a, \sigma \in C_b^{3,1}$ and $\frac{\partial \sigma}{\partial t} \in C_b^{1,0}$ then for all $\xi, x, y \in \mathbb{R}^d$, $0 \leq s \leq t \leq 1$, and some constants c and C

$$|\Gamma_\ell(x, t; \xi, s) - \Gamma(x, t; \xi, s)| \leq \frac{C h_\ell}{(t-s)^{\frac{d+1}{2}}} \exp\left(-\frac{c \|x-\xi\|^2}{t-s}\right).$$

By setting $t=1, s=1-\tau$ and integrating with respect to x , (17) follows. Note however that the SDE in the numerical example below does not satisfy the uniform ellipticity condition of [13, Theorem 2.3] and a slightly different numerical scheme is used in our case, namely the truncated Milstein scheme without Lévy areas which was proposed in [12].

Theorem 4.3 . Let Assumptions 4.1 and 4.2 hold and assume further that for some $q \geq 1$

$$(18a) \quad \mathbb{E}\left[\left\|X_1 - \bar{X}_{\ell, 1}\right\|^q\right]^{1/q} \lesssim h_\ell^{1/2}$$

$$(18b) \quad \text{and} \quad \mathbb{E}\left[\left\|\frac{1}{2}(\bar{X}_{\ell, 1} + \bar{X}_{\ell, 1}^{(a)}) - \bar{X}_{\ell-1, 1}\right\|^q\right]^{1/q} \lesssim h_\ell.$$

Then for ΔP_ℓ in (14) there exists a constant c_2 such that for all $\tau \in (0, 1)$,

$$(19a) \quad \mathbb{E}[(\Delta P_\ell)^2] \leq c_2 h_\ell^{(1-1/(q+1))/2}$$

$$(19b) \quad \text{and} \quad \mathbb{E}\left[(\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}])^2\right] \leq c_2 h_\ell^{2(1-5/(q+5))/2} / \tau^{3/2}.$$

This theorem shows that Assumption 2.3 is satisfied with $\beta_d = (1-1/(q+1))/2$, $\beta_c = 2(1-5/(q+5))$ and $p=3/2$. Note that (18a) is the same as (10) for $\beta=1$. For example, [12, Theorem 4.13] shows that (18) is satisfied for all $q \geq 2$ and an antithetic pair of estimators of a Clark-Cameron SDE derived from the Milstein discretization by setting the Lévy areas to zeros.

PROOF. The first claim (19a) follows from a similar proof to Theorem 3.2 given (18a) and (16). To prove (19b), we start by defining E for a given τ and some $0 < r < 1$, that we will choose later, to be the set of paths for which

$$\max \left\{ \frac{\|\bar{X}_{\ell,1-\tau} - X_{1-\tau}\|}{h_\ell^{r/2}}, \frac{\|\bar{X}_{\ell,1-\tau}^{(a)} - X_{1-\tau}\|}{h_\ell^{r/2}}, \frac{\|\bar{X}_{\ell-1,1-\tau} - X_{1-\tau}\|}{h_\ell^{r/2}}, \frac{\|\frac{1}{2}(\bar{X}_{\ell,1-\tau} + \bar{X}_{\ell,1-\tau}^{(a)}) - \bar{X}_{\ell-1,1-\tau}\|}{h_\ell^r} \right\} \geq 1.$$

Then since $|\Delta P_\ell| \leq 1$,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}]^2] &= \mathbb{E}[\mathbb{I}_E \mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}]^2] + \mathbb{E}[\mathbb{I}_{E^c} \mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}]^2] \\ &\leq \mathbb{P}[E] + \mathbb{E}[\mathbb{I}_{E^c} (\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}])^2]. \end{aligned}$$

Due to (18a) and the Markov inequality,

$$\mathbb{P}[\|\bar{X}_{\ell,1-\tau} - X_{1-\tau}\| \geq h_\ell^{r/2}] \leq C_q h_\ell^{(1-r)q/2},$$

for some constant C_q with similar bounds for $\|\bar{X}_{\ell,1-\tau} - X_{1-\tau}\|$ and $\|\bar{X}_{\ell,1-\tau}^{(a)} - X_{1-\tau}\|$, and

$$\mathbb{P}[\|\frac{1}{2}(\bar{X}_{\ell,1-\tau} + \bar{X}_{\ell,1-\tau}^{(a)}) - \bar{X}_{\ell-1,1-\tau}\| \geq h_\ell^r] \leq C_q h_\ell^{(1-r)q}.$$

Hence $\mathbb{P}[E] \leq 4C_q h_\ell^{(1-r)q/2}$. For the other term we have

$$\begin{aligned} \mathbb{E}[\mathbb{I}_{E^c} (\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}])^2] &= \mathbb{E} \left[\mathbb{I}_{E^c} \left(\frac{1}{2}(\bar{g}_{\ell,1-\tau}(\bar{X}_{\ell,1-\tau}) + \bar{g}_{\ell,1-\tau}(\bar{X}_{\ell,1-\tau}^{(a)})) - \bar{g}_{\ell-1,1-\tau}(\bar{X}_{\ell-1,1-\tau}) \right)^2 \right] \\ &\leq 3 \mathbb{E} \left[\mathbb{I}_{E^c} \left(\frac{1}{2}(g_{1-\tau}(\bar{X}_{\ell,1-\tau}) + g_{1-\tau}(\bar{X}_{\ell,1-\tau}^{(a)})) - g_{1-\tau}(\bar{X}_{\ell-1,1-\tau}) \right)^2 \right] \\ &\quad + 3 \mathbb{E}[\mathbb{I}_{E^c} (\bar{g}_{\ell,1-\tau}(\bar{X}_{\ell,1-\tau}) - g_{1-\tau}(\bar{X}_{\ell,1-\tau}))^2] \\ &\quad + 3 \mathbb{E}[\mathbb{I}_{E^c} (g_{1-\tau}(\bar{X}_{\ell-1,1-\tau}) - \bar{g}_{\ell-1,1-\tau}(\bar{X}_{\ell-1,1-\tau}))^2]. \end{aligned}$$

Due to (17) in Assumption 4.2, the second of the final three terms of this inequality is bounded by $2c_1^2 h_\ell^2 / \tau$ and the third by $2c_1^2 h_{\ell-1}^2 / \tau$ so both are $\mathcal{O}(h_\ell^2 / \tau)$.

To bound the first term we perform a Taylor series expansion about $\bar{X}_{\ell-1,1-\tau}$ to obtain

$$\frac{1}{2}(g_{1-\tau}(\bar{X}_{\ell,1-\tau}) + g_{1-\tau}(\bar{X}_{\ell,1-\tau}^{(a)})) - g_{1-\tau}(\bar{X}_{\ell-1,1-\tau}) = Y_1 + \frac{1}{4}Y_2 + \frac{1}{4}Y_3,$$

where

$$Y_1 := \left(\frac{1}{2}(\bar{X}_{\ell,1-\tau} + \bar{X}_{\ell,1-\tau}^{(a)}) - \bar{X}_{\ell-1,1-\tau} \right) \cdot \nabla g_{1-\tau}(\bar{X}_{\ell-1,1-\tau}),$$

$$Y_2 := (\bar{X}_{\ell,1-\tau} - \bar{X}_{\ell-1,1-\tau})^T H_{g_{1-\tau}}(\xi_1) (\bar{X}_{\ell,1-\tau} - \bar{X}_{\ell-1,1-\tau}),$$

$$Y_3 := (\bar{X}_{\ell,1-\tau}^{(a)} - \bar{X}_{\ell-1,1-\tau})^T H_{g_{1-\tau}}(\xi_2) (\bar{X}_{\ell,1-\tau}^{(a)} - \bar{X}_{\ell-1,1-\tau}),$$

and where ξ_1 is a positively weighted average of $\bar{X}_{\ell,1-\tau}$ and $\bar{X}_{\ell-1,1-\tau}$ and ξ_2 is a positively weighted average of $\bar{X}_{\ell,1-\tau}^{(a)}$ and $\bar{X}_{\ell-1,1-\tau}$. Hence,

$$\mathbb{E} \left[\mathbb{I}_{E^c} \left(\frac{1}{2}(g_{1-\tau}(\bar{X}_{\ell,1-\tau}) + g_{1-\tau}(\bar{X}_{\ell,1-\tau}^{(a)})) - g_{1-\tau}(\bar{X}_{\ell-1,1-\tau}) \right)^2 \right] \leq 2 \mathbb{E}[\mathbb{I}_{E^c} (Y_1^2 + Y_2^2 + Y_3^2)].$$

Defining $\delta := 2(\tau/(h_\ell))^{1/2} h_\ell^{r/2}$, we now split $\mathbb{I}_{E^c}$ into

$$\mathbb{I}_{E^c} = \mathbb{I}_{E^c} \mathbb{I}_{d_K(X_{1-\tau}) > \delta} + \mathbb{I}_{E^c} \mathbb{I}_{d_K(X_{1-\tau}) < \delta}.$$

Since $h_\ell \leq \tau$ it follows that $\delta > 2h_\ell^{r/2}$ and if $\mathbb{I}_{E^c} \mathbb{I}_{d_K(X_{1-\tau}) > \delta} = 1$ then $d_K(\bar{X}_{\ell,1-\tau}) > \frac{1}{2}\delta$, $d_K(\bar{X}_{\ell,1-\tau}^{(a)}) > \frac{1}{2}\delta$, $d_K(\bar{X}_{\ell-1,1-\tau}) > \frac{1}{2}\delta$, and also $d_K(\xi_1) > \frac{1}{2}\delta$, $d_K(\xi_2) > \frac{1}{2}\delta$. Therefore, by Assumption 4.1, there is a constant C such that $\|\nabla g_t(\bar{X}_{\ell-1,1-\tau})\| \leq C$ and $\|H_{g_t}(\xi_i)\| \leq C$ for $i=1,2$ and all $h_\ell \leq \tau$. Hence,

$$\mathbb{E}[\mathbb{I}_{E^c} \mathbb{I}_{d_K(X_{1-\tau}) > \delta} (Y_1^2 + Y_2^2 + Y_3^2)] \lesssim h_\ell^2.$$

On the other hand, when $\mathbb{I}_{E^c} \mathbb{I}_{d_K(X_{1-\tau}) < \delta} = 1$, Assumption 4.1 implies that $\|\nabla g_{1-\tau}(\xi)\| \leq c_1/\tau^{1/2}$ and $\|H_{g_{1-\tau}}(\xi)\| \leq c_1/\tau$ for any $\xi \in \mathbb{R}^d$ and $\tau \in (0,1)$ leading to

$$\mathbb{E}[\mathbb{I}_{E^c} \mathbb{I}_{d_K(X_{1-\tau}) < \delta} Y_1^2] \leq \frac{c_1^2}{\tau} h_\ell^{2r} \mathbb{P}[d_K(X_{1-\tau}) < \delta] \leq 2 \frac{c_1^3}{\tau^{1/2}} h_\ell^{(5r-1)/2},$$

$$\mathbb{E}[\mathbb{I}_{E^c} \mathbb{I}_{d_K(X_{1-\tau}) < \delta} (Y_2^2 + Y_3^2)] \leq \frac{c_1^2}{\tau^2} h_\ell^{2r} \mathbb{P}[d_K(X_{1-\tau}) < \delta] \leq 2 \frac{c_1^3}{\tau^{3/2}} h_\ell^{(5r-1)/2},$$

for any $r < 1$. Picking r so that $\mathbb{P}[E] \simeq h_\ell^{(5r-1)/2}$ yields $r = (q+1)/(q+5)$ and the final result. $\square$

Corollary 4.4 (MLMC Computational Complexity) . *Under the assumptions of Theorem 4.3, the MLMC method with MSE ε^2 based on $\Delta\mathcal{P}_\ell$ with $\eta \in (1/2, 2)$, $h_\ell = h_0 M^{-\ell}$ for $M \in \mathbb{N}_+$ and the antithetic estimator (14) has a computational complexity $\mathcal{O}(\varepsilon^{-2})$.*

Note that even though the MLMC estimator has the same computational complexity for all values of $\eta \in (1/2, 2)$, in theory the value $\eta \approx 4/5$ minimizes the work and variance of $\Delta\mathcal{P}_\ell$ as discussed in Remark 2.5.

Numerical Experiments. In this section, we consider the Clark-Cameron SDE

$$(20) \quad \begin{aligned} dX_{1,t} &= dW_{1,t}, \\ dX_{2,t} &= X_{1,t} dW_{2,t}. \end{aligned}$$

Here $\{(W_{i,t})_{t \geq 0}\}_{i=1}^2$ are independent Wiener processes. Note that we can sample paths of $X_{1,t} = W_{1,t}$ exactly. To approximate the paths of $X_{2,t}$, we use the Euler-Maruyama numerical scheme [18] as follows

$$\bar{X}_{2,\ell,(n+1)h_\ell} = \bar{X}_{2,\ell,nh_\ell} + W_{1,nh_\ell} \Delta_{\ell,n} W_2,$$

for $n = 0, \dots, (h_\ell^{-1} - 1)$ and

$$\Delta_{\ell,n} W_i := W_{i,(n+1)h_\ell} - W_{i,nh_\ell}$$

is the Brownian increment. We again set the time step size at level ℓ as $h_\ell = 2^{-\ell-1}$ and use the new estimator in Definition 2.2 with $\eta = 1$. We also test the antithetic estimators outlined in [12] which is obtained by setting the Lévy area term in a Milstein discretization to zero. In particular, the ℓ 'th level approximation is defined as

$$(21) \quad \bar{X}_{2,\ell,(n+1)h_\ell} = \bar{X}_{2,\ell,nh_\ell} + W_{1,nh_\ell} \Delta_{\ell,n} W_2 + \frac{1}{2} \Delta_{\ell,n} W_1 \Delta_{\ell,n} W_2.$$

When computing (14) for a given Brownian path, the coarse, $\bar{X}_{i,\ell-1,\cdot}$, and fine, $\bar{X}_{i,\ell,\cdot}$, approximations are constructed according to (21). On the other hand, the antithetic approximation at $(2n+1)h_\ell$ and $(2n+2)h_\ell$ for $n=0, 1, \dots, (h_\ell^{-1}/2 - 1)$ is defined as

$$\begin{aligned}\bar{X}_{2,\ell,(2n+1)h_\ell}^{(a)} &= \bar{X}_{2,\ell,2nh_\ell}^{(a)} + W_{1,2nh_\ell} \Delta_{\ell,2n+1} W_2 + \frac{1}{2} \Delta_{\ell,2n+1} W_1 \Delta_{\ell,2n+1} W_2, \\ \bar{X}_{2,\ell,(2n+2)h_\ell}^{(a)} &= \bar{X}_{2,\ell,(2n+1)h_\ell}^{(a)} + (W_{1,2nh_\ell} + \Delta_{\ell,2n+1} W_1) \Delta_{\ell,2n} W_2 + \frac{1}{2} \Delta_{\ell,2n} W_1 \Delta_{\ell,2n} W_2.\end{aligned}$$

In other words, the roles of $\Delta_{\ell,n} W_j$ and $\Delta_{\ell,n+1} W_j$ for $j=1, 2$ are swapped when computing $\{\bar{X}_{2,\ell,(2n+1)h_\ell}^{(a)}, \bar{X}_{2,\ell,(2n+2)h_\ell}^{(a)}\}$ compared to $\{\bar{X}_{2,\ell,(2n+1)h_\ell}, \bar{X}_{2,\ell,(2n+2)h_\ell}\}$.

Under certain conditions on the coefficients of (3), the assumption (18a) is satisfied when using the Euler-Maruyama scheme [18] and both (18a) and (18b) are satisfied for the antithetic, truncated Milstein estimator [12, Theorem 4.13]. However, we emphasize that the diffusion coefficient in (20) is not bounded, and more importantly, is not elliptic. Hence the results of [13, Theorem 2.3] showing (17) in Assumption 4.2 are not applicable. Nevertheless, we first consider an example where we compute $\mathbb{P}[X_1 \in S]$ where $S = \{x \in \mathbb{R}^d : \min\{x_1, x_2\} \geq 1\}$. For this example, the SDE in (20) is locally elliptic at the boundary of S .

Fig. 5a shows the convergence of $\mathbb{E}[(\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}])^2]$ for an Euler-Maruyama scheme, which as predicted by Theorem 5.3 increases in proportion to $\tau^{-1/2}$, and an antithetic approximation, which as predicted by Theorem 4.3 increases in proportion to $\tau^{-3/2}$ approximately. Figs. 6a and 6b confirm the claims of Theorem 2.4. Figure 6d shows the total work estimate of an MLMC sampler based on $\Delta \mathcal{P}$ when using Euler-Maruyama or the antithetic estimator. The computational complexities of MLMC based on the branching estimator $\Delta \mathcal{P}_\ell$ using Euler-Maruyama and the antithetic estimators are consistent with Corollaries 3.3 and 4.4, respectively. Recall that in this case, the optimal value of η is 4/3 when using Euler-Maruyama, and 4/5 when using an antithetic approximation. However, similar to Section 3, we did not observe a better computational cost when using $\eta \neq 1$ for the considered tolerances because of the additional cost of branching when the branching points do not align with the time-stepping scheme. For MLMC based on ΔP_ℓ , labelled “Without branching”, the computational complexity of MLMC is $\mathcal{O}(\varepsilon^{-5/2})$ for both Euler-Maruyama and the antithetic estimators since $\text{Var}[\Delta P_\ell] \lesssim h_\ell^{1/2}$; see Theorems 3.2 and 4.3. Fig. 6c again illustrates that our branching estimator has bounded kurtosis while the kurtosis of ΔP_ℓ grows approximately in proportion to $h_\ell^{-1/2}$. Hence an MLMC algorithm that relies on variance estimates is more stable when using the branching estimator.

As a second test, we consider $S = \{x \in \mathbb{R}^d : x_1 \geq 1\}$ for which the diffusion coefficient is *not* locally elliptic at the boundary. Fig. 5b shows the convergence of $\mathbb{E}[(\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}])^2]$ and Fig. 7a shows the convergence of $\text{Var}[\Delta \mathcal{P}_\ell]$. The observed convergence rates are slightly worse than those observed for the previous example. Nevertheless, recalling that the work of $\Delta \mathcal{P}$ increases in proportion to $h_\ell^{-1} |\log(h_\ell)|$, the computational complexity of a MLMC estimator is still $\mathcal{O}(\varepsilon^{-2})$ when using the antithetic estimator, as confirmed in Fig. 7b. This is a more difficult problem as Fig. 7c illustrates and the branching estimator has the same increasing kurtosis as ΔP_ℓ .

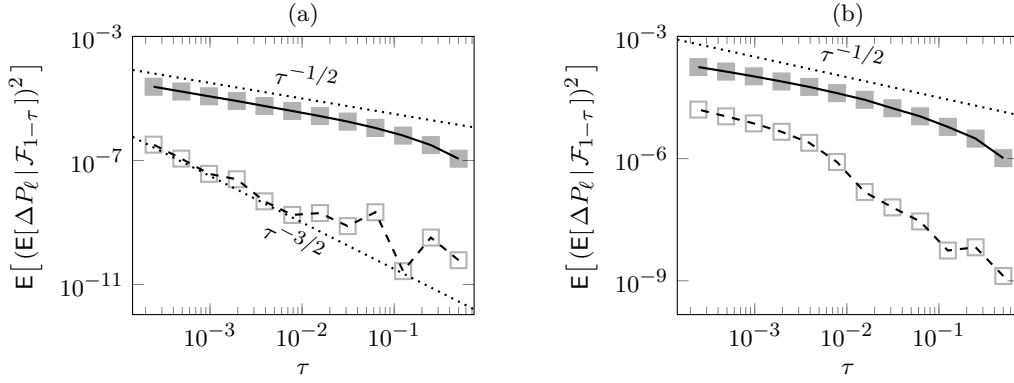


Fig 5: Numerical verification for (5b) with $h_\ell = 2^{-14}$ for the Clark-Cameron example in (20) when using Euler-Maruyama (*solid*) and antithetic Milstein (*dashed*). For (a), $S = \{x \in \mathbb{R}^2 : \min\{x_1, x_2\} \geq 1\}$ while for (b) we choose $S = \{x \in \mathbb{R}^2 : x_2 \geq 1\}$.

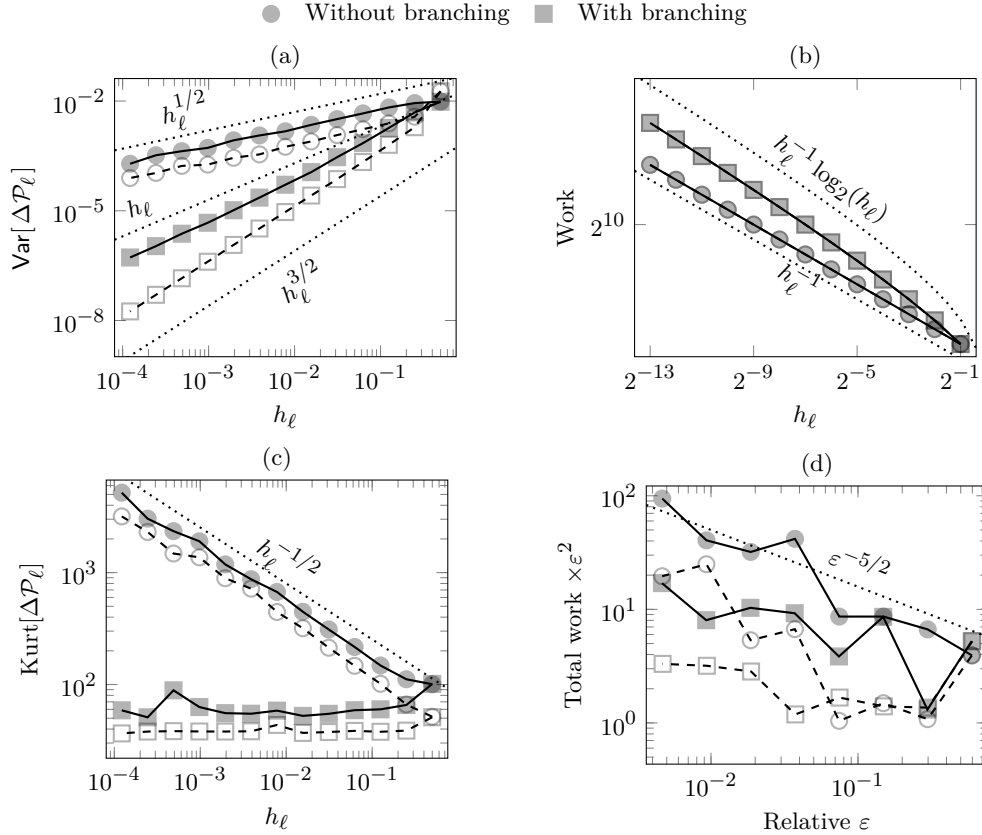


Fig 6: Numerical results for Clark-Cameron example in (20) when $S = \{x \in \mathbb{R}^2 : \min\{x_1, x_2\} \geq 1\}$ and using Euler-Maruyama (*solid*) or antithetic Milstein (*dashed*) in the traditional, ΔP_ℓ , and branching, $\Delta \mathcal{P}_\ell$, estimators. (a) shows numerical verification of the variance convergence of ΔP_ℓ (b) The work estimate per sample based on the number of generated samples from the standard normal distribution. The work estimates when using the Milstein scheme are identical. (c) The kurtosis of $\Delta \mathcal{P}_\ell$. (d) The total work estimate of MLMC for different tolerances.

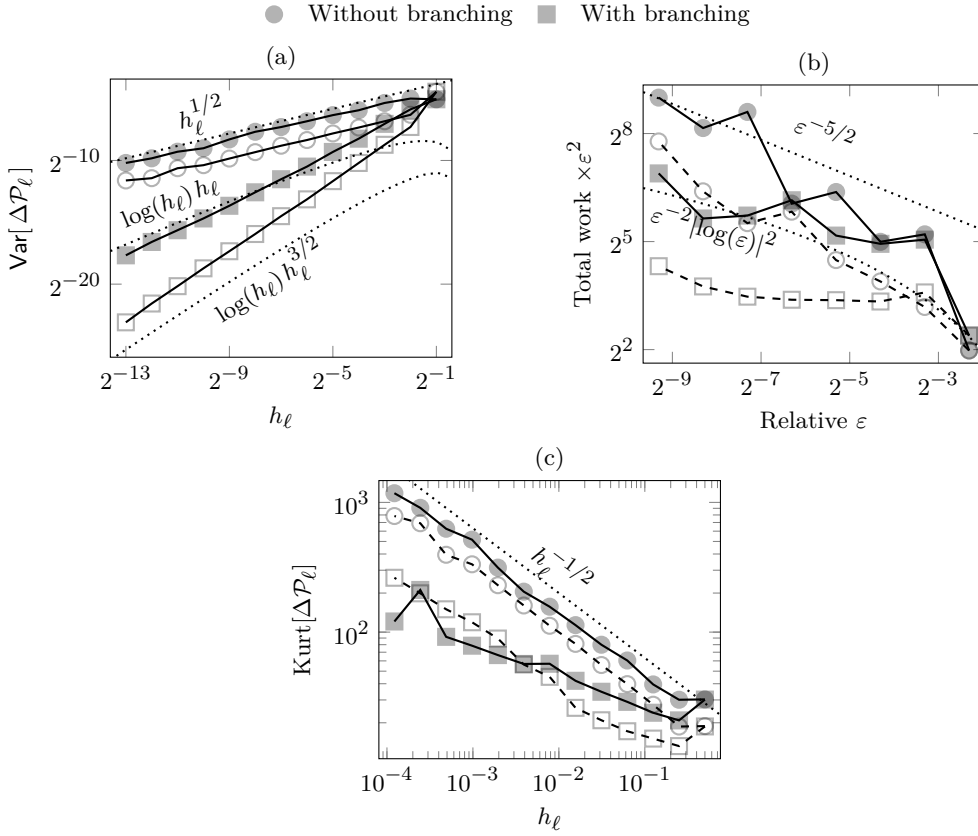


Fig 7: The Clark-Cameron example in (20) for $S = \{x \in \mathbb{R}^2 : x_2 \geq 1\}$ when using Euler-Maruyama (*solid*) or Milstein (*dashed*) in the traditional, ΔP_ℓ , and branching, $\Delta \mathcal{P}_\ell$, estimators. (a) shows numerical verification of the variance convergence of $\Delta \mathcal{P}_\ell$. (b) shows the total work estimate of MLMC for different tolerances. (c) The kurtosis of $\Delta \mathcal{P}_\ell$.

5. Bounds on Solutions of Elliptic SDEs. In this section, we prove Assumptions 3.1 and 4.1 for solutions to SDEs with certain conditions on the SDE coefficients and the boundary K . For any $x \equiv (x_i)_{i=1}^d \in \mathbb{R}^d$, define $x_{-j} \equiv (x_i)_{i=1, i \neq j}^d \in \mathbb{R}^{d-1}$ and define $\exp x$, $\log x$ and x^{-1} component-wise. For $m \in \mathbb{N}^d$, define $|m| = m_1 + \dots + m_d$ and $D_\xi^m \equiv \frac{\partial^{m_1}}{\partial \xi_1^{m_1}} \dots \frac{\partial^{m_d}}{\partial \xi_d^{m_d}}$. Given a set $J \subset \mathbb{R}^d$, define

$$(22) \quad J^\delta \equiv \{y \in \mathbb{R}^d : d_J(y) \leq \delta\},$$

and for a function $u : \mathbb{R}^d \rightarrow \mathbb{R}^d$, let $u(J)$ denote the image of J under the mapping $x \rightarrow u(x)$, i.e., $u(J) \equiv \{u(x) : x \in J\}$.

We now define a class of “Simple” sets which are a particularly simple form of Lipschitz boundaries.

Definition 5.1 ((Si) sets) . *We say that a set $J \subset \mathbb{R}^d$ is an (Si) set if it is the graph of a Lipschitz function. In other words, there exists an index $j \in \{1, \dots, d\}$ and a Lipschitz function $f : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ such that*

$$J = \{x \in \mathbb{R}^d : x_j = f(x_{-j})\}.$$

Lemma 5.2 . *Let $\{Z_i\}_{i=1}^d$ be a set of i.i.d. Gaussian random variables with $\text{Var}[Z_i] = \tau$ for all i and denote $Z = (Z_i)_{i=1}^d$. If $J \subset \mathbb{R}^d$ is an (Si) set, then there exists a constant $C > 0$ such that*

$$\mathbb{P}[d_J(Z) \leq \delta] \leq C \times \frac{\delta}{\tau^{1/2}}.$$

PROOF. For the (Si) set J with corresponding index j and Lipschitz function f with Lipschitz constant L , we first show that $J^\delta \subseteq \widetilde{J}^\delta$ where

$$\widetilde{J}^\delta = \{x \in \mathbb{R}^d : |x_j - f(x_{-j})| \leq (L+1)\delta\}.$$

Letting $y \in J$ and $x \in J^\delta$ such that $y_j = f(y_{-j})$ and $\|x - y\| \leq \delta$. It follows that

$$\begin{aligned} |f(x_{-j}) - x_j| &\leq |f(x_{-j}) - f(y_{-j})| + |x_j - y_j| \\ &\leq L\|x_{-j} - y_{-j}\| + |x_j - y_j| \\ &\leq (L+1)\|x - y\| \\ &\leq (L+1)\delta. \end{aligned}$$

Then

$$\mathbb{P}[Z \in J^\delta] \leq \mathbb{P}[Z \in \widetilde{J}^\delta] = \mathbb{E}[\mathbb{P}[|Z_j - f(Z_{-j})| \leq (L+1)\delta \mid Z_{-j}]].$$

Using standard 1D results on Z_j yields

$$\mathbb{P}[|Z_j - f(Z_{-j})| \leq (L+1)\delta \mid Z_{-j}] \leq \frac{2(L+1)}{(2\pi)^{1/2}} \times \frac{\delta}{\tau^{1/2}},$$

and the result follows. $\square$

The previous lemma can be used to show that Assumption 3.1 is satisfied for a Wiener process, i.e., $X_t = W_t$, and a set S whose boundary $\partial S \equiv K$ is (Si). We next prove a more general result showing both Assumptions 3.1 and 4.1 for sets whose boundary can be covered by (Si) sets and SDEs whose coefficients satisfy certain smoothness conditions.

Theorem 5.3 . For $S \subset \mathbb{R}^d$, assume that $\partial S \equiv K \subseteq \bigcup_{j=1}^n J_j$ for some finite n and (Si) sets $\{J_j\}_{j=1}^n$. Assume that the SDE (3) is uniformly elliptic and that $a, \sigma\sigma^T \in C_b^{\lambda,0}$ for some $\lambda \in (0, 1)$ and let $(X_t)_{t \in [0,1]}$ satisfy the SDE. Then, there exist $c_1 > 0$ such that for all $0 < s < 1$ and all $\delta > 0$ the following holds

$$(23) \quad \mathbb{E} \left[(\mathbb{P}[d_K(X_1) \leq \delta | \mathcal{F}_s])^2 \right] \leq c_1 \frac{\delta^2}{(1-s)^{1/2}}.$$

Assume further that, that $a, \sigma\sigma^T \in C_b^{2,0}$, then there exists $c_1, c_2 > 0$ such that for all $\xi \in \mathbb{R}^d$, $0 < s < 1$ and $m \in \mathbb{N}^d$, $0 \leq |m| \leq 2$,

$$(24) \quad |D_\xi^m \mathbb{P}[X_1 \in S | X_s = \xi]| \leq \frac{c_1}{(1-s)^{|m|/2}} \cdot \exp \left(-c_2 \frac{d_K^2(\xi)}{(1-s)} \right).$$

The assumption on S (or K) is illustrated in Fig. 8. Theorem 5.3 shows that Assumptions 3.1 and 4.1 are satisfied for a solution to a uniformly elliptic SDE assuming that the set K is covered by a finite number of (Si) sets.

PROOF. We have the following bound on $\Gamma(\cdot, 1; \xi, s)$, the density of X_1 given $X_s = \xi$, for $m \in \mathbb{N}^d$ and some $C_m, c_m > 0$,

$$(25) \quad |D_\xi^m \Gamma(x, 1; \xi, s)| \leq \frac{C_m}{(1-s)^{(d+|m|)/2}} \exp \left(-c_m \frac{\|x - \xi\|^2}{1-s} \right),$$

for $|m| = 0$ when $a, \sigma\sigma^T \in C_b^{\lambda,0}$ [5, Chapter 9, Theorem 2] and for $0 \leq |m| \leq 2$ when $a, \sigma\sigma^T \in C_b^{2,0}$ [5, Chapter 9, Theorem 7]. Hence

$$\begin{aligned} \mathbb{P}[d_K(X_1) \leq \delta | \mathcal{F}_s] &= \mathbb{P}[d_K(X_1) \leq \delta | X_s] \\ &= \int_{\mathbb{R}^d} \mathbb{I}_{d_K(x) \leq \delta} \Gamma(x, 1; X_s, s) dx \\ &\leq \int_{\mathbb{R}^d} \mathbb{I}_{d_K(x) \leq \delta} \left(\frac{C_0}{(1-s)^{d/2}} \exp \left(-c_0 \frac{\|x - X_s\|^2}{1-s} \right) \right) dx \\ &= C_0 (2\pi/c_0)^{d/2} \mathbb{P}[d_K(Z) \leq \delta | X_s], \end{aligned}$$

where Z is a multivariate Normal random variable with mean X_s with variance $((1-s)/c_0)I_d$ where I_d is the $d \times d$ identity matrix. Then noting

$$\mathbb{P}[d_K(Z) \leq \delta | X_s] \leq \sum_{j=1}^n \mathbb{P}[d_{J_j}(Z) \leq \delta | X_s],$$

and using Lemma 5.2 we can conclude that there is a constant $\tilde{C}$

$$\mathbb{P}[d_K(X_1) \leq \delta | \mathcal{F}_s] \leq n \tilde{C} \frac{\delta}{(1-s)^{1/2}}.$$

Hence

$$\begin{aligned} \mathbb{E}[(\mathbb{P}[d_K(X_1) \leq \delta | \mathcal{F}_s])^2] &\leq n \tilde{C} \frac{\delta}{(1-s)^{1/2}} \mathbb{E}[\mathbb{P}[d_K(X_1) \leq \delta | \mathcal{F}_s]] \\ &\leq n^2 \tilde{C}^2 \frac{\delta^2}{(1-s)^{1/2}}. \end{aligned}$$

To prove (24), we distinguish between two cases

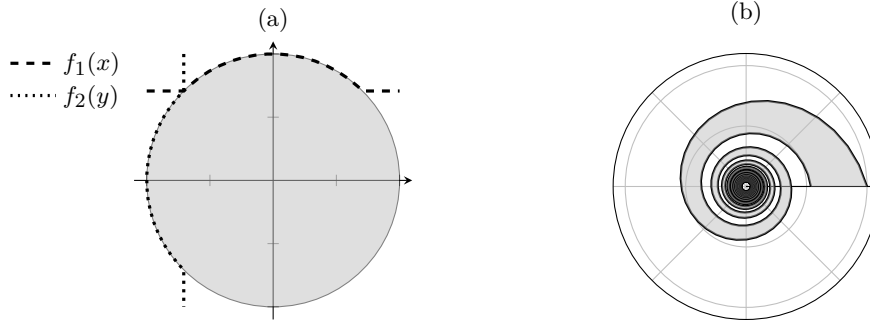


Fig 8: (a) The set $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ whose boundary $\partial S \equiv K$ satisfies the assumptions of Theorem 5.3. We split the circle on the boundary into four parts and we show two of them here. (b) The set $S = \{(r, \theta) \in \mathbb{R}_+^2 : (2 + \theta/\pi)^{-0.9} \leq r \leq (1 + \theta/\pi)^{-0.9}\}$, in polar coordinates, whose boundary does not satisfy the assumptions of Theorem 5.3.

1. $\xi \notin S$, then

$$\begin{aligned}
 & |D_\xi^m \mathbf{P}[X_1 \in S \mid X_s = \xi]| \\
 & \leq \int_S |D_\xi^m \Gamma(x, 1; \xi, s)| dx \\
 & \leq \int_S \frac{C_m}{(1-s)^{(d+|m|)/2}} \times \exp\left(-c_m \frac{\|x - \xi\|^2}{1-s}\right) dx \\
 & \leq \frac{1}{(1-s)^{|m|/2}} \exp\left(-c_m \frac{\inf_{x \in S} \|x - \xi\|^2}{2(1-s)}\right) \int_S \frac{C_m}{(1-s)^{d/2}} \times \exp\left(-c_m \frac{\|x - \xi\|^2}{2(1-s)}\right) dx \\
 & \leq \frac{1}{(1-s)^{|m|/2}} \exp\left(-c_m \frac{\inf_{x \in S} \|x - \xi\|^2}{2(1-s)}\right) ((4\pi)^{d/2} C_m),
 \end{aligned}$$

and we conclude with

$$\inf_{x \in S} \|x - \xi\|^2 = \inf_{x \in \partial S} \|x - \xi\|^2 = d_K^2(\xi),$$

since $\xi \notin S$.

2. $\xi \in S$, then, for S^c being the compliment of S , we have

$$|D_\xi^m \mathbf{P}[X_1 \in S \mid X_s = \xi]| = |D_\xi^m (\mathbf{P}[X_1 \in S^c \mid X_s = \xi])|,$$

and we can use the first step since $\xi \notin S^c$ and $\partial S^c = \partial S$.

□

As an example of a set that does not satisfy the assumptions of Theorem 5.3, consider $K = \{1/n^b : n \in \mathbb{N}\}$ for $0 < b \leq 1$. Then we can show that Assumption 3.1 is not satisfied for a standard Normal random variable Z and any $\delta \leq b$,

$$\begin{aligned}
 \mathbf{P}\left[\min_{n \in \mathbb{N}} |Z - n^{-b}| \leq \delta\right] &= 2 \int_0^\infty \mathbb{I}_{\min_{n \in \mathbb{N}} |y - n^{-b}| \leq \delta} \phi(y) dy \\
 &\geq \frac{2 \exp(-1/2)}{(2\pi)^{1/2}} \int_0^{(\delta/b)^{b/(b+1)}} \mathbb{I}_{\min_{n \in \mathbb{N}} |y - n^{-b}| \leq \delta} dy \\
 &= \frac{2 \exp(-1/2)}{(2\pi)^{1/2}} (\delta/b)^{b/(b+1)},
 \end{aligned}$$

where $\phi(\cdot)$ is the density of standard normal random variable. Similarly, consider the two dimensional set in polar coordinates $K = \{(r, \theta) \in \mathbb{R}_+ \times [0, 2\pi] : r = (n + \theta/\pi)^{-b}, n \in \mathbb{N}\}$ (see Fig. 8b) for some $b \in (0, 1)$. Using a similar calculation to before we can show that for a 2D standard normal random variable, Z , any $\delta < b$

$$\mathbb{P} \left[\min_{y \in K} \|Z - y\| \leq \delta \right] \geq \frac{1}{2} \exp(-1/2) (\delta/b)^{2b/(b+1)}.$$

The Geometric Brownian Motion (GBM) does not satisfy the conditions of Theorem 5.3 since the diffusion coefficient of the SDE of a GBM is not bounded nor uniformly elliptic. To deal with this important case, we first prove a similar result to Lemma 5.2 for log-normal random variables.

Lemma 5.4 . *Let $\{Z_i, Y_i\}_{i=1}^d$ be two sets of independent Gaussian random variables with $\text{Var}[Z_i] = \tau$ for all i and denote $Z = (Z_i)_{i=1}^d, Y = (Y_i)_{i=1}^d$. Let $J \subset \mathbb{R}^d$ be an (Si) set. There exists $\delta_0 > 0$ and $C > 0$ such that for all $0 < \delta < \delta_0$*

$$\mathbb{E}[\mathbb{P}[d_{\exp J}(\exp(Z + Y)) < \delta \mid Y]^2] \leq C \frac{\delta^2}{\tau^{1/2}}.$$

PROOF. If $v \equiv \exp x \in (\exp J)^\delta$ then there exists $w \equiv \exp y \in \exp J$ such that $\|v - w\| \leq \delta$. If, in addition, $\|x\| < R_\delta \equiv |\log \delta|^{3/4}$ then for sufficiently small δ_0 , $2\delta < v_i < (2\delta)^{-1}$ for all $\delta < \delta_0$ and all $i = 1, \dots, d$, and hence $\|w^{-1}\| < 2\|v^{-1}\|$, from which it follows that

$$\|x - y\| = \|\log v - \log w\| \leq 2\|v^{-1}\| \delta.$$

If J has corresponding index j and Lipschitz function f with Lipschitz constant L , then

$$(26) \quad |x_j - f(x_{-j})| \leq |f(x_{-j}) - f(y_{-j})| + |x_j - y_j| \leq (L+1)\|x - y\| \leq 2(L+1)\|v^{-1}\| \delta.$$

Since $\|x\| < R_\delta$, $\exp(\|x\|) \delta \leq \frac{1}{2}$, and so $|x_j - f(x_{-j})| \leq L+1$, and therefore

$$|x_j| \leq L+1 + |f(x_{-j})| \leq L+1 + |f(0)| + L\|x_{-j}\|.$$

Hence, since $\|x\| \leq \|x_{-j}\| + |x_j|$ and $\|v^{-1}\| \leq \exp(\|x\|)$,

$$\begin{aligned} |x_j - f(x_{-j})| &\leq 2(L+1) \exp(\|x_{-j}\| + |x_j|) \delta \\ &\leq 2(L+1) \exp(L+1+|f(0)|) \exp((L+1)\|x_{-j}\|) \delta. \end{aligned}$$

The conclusion is that if $\exp x \in (\exp J)^\delta$ and $\|x\| < R_\delta$ then $x \in \widetilde{J}^\delta$ where $\widetilde{J}^\delta$ is defined as

$$\widetilde{J}^\delta = \{x \in \mathbb{R}^d : |x_j - f(x_{-j})| \leq c \exp((L+1)\|x_{-j}\|) \delta\},$$

with $c \equiv 2(L+1) \exp(L+1+|f(0)|)$.

Letting $U \equiv Z + Y$, to bound $\mathbb{E}[(\mathbb{E}[\mathbb{I}_{\exp U \in (\exp J)^\delta} \mid Y])^2]$, we start by noting that $\mathbb{I}_{\exp U \in (\exp J)^\delta} \leq \mathbb{I}_{\exp U \in (\exp J)^\delta} \mathbb{I}_{\|U\| < R_\delta} + \mathbb{I}_{\|U\| \geq R_\delta}$, and also that $\mathbb{E}[\mathbb{I}_{\|U\| \geq R_\delta} \mid Y] < 1$. Hence,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[\mathbb{I}_{\exp U \in (\exp J)^\delta} \mid Y]^2] &\leq 2\mathbb{E}[\mathbb{E}[\mathbb{I}_{\exp U \in (\exp J)^\delta} \mathbb{I}_{\|U\| < R_\delta} \mid Y]^2] \\ &\quad + 2\mathbb{E}[\mathbb{E}[\mathbb{I}_{\|U\| \geq R_\delta} \mid Y]^2] \\ &\leq 2\mathbb{E}[\mathbb{E}[\mathbb{I}_{U \in \widetilde{J}^\delta} \mid Y]^2] + 2\mathbb{P}[\|U\| \geq R_\delta]. \end{aligned}$$

Then, by normality of U and

$$\mathbb{E}[\mathbb{E}[\mathbb{I}_{U \in \widetilde{J}^\delta} \mid Y]^2] \leq \mathbb{E}[\mathbb{E}[\mathbb{E}[\mathbb{I}_{U \in \widetilde{J}^\delta} \mid Y, Z_{-j}]^2 \mid Y_{-j}, Z_{-j}]].$$

Using standard 1D results on Z_j ,

$$\mathbb{E} \left[\mathbb{E} \left[\mathbb{I}_{U \in \tilde{J}^\delta} \mid Y, Z_{-j} \right]^2 \mid Y_{-j}, Z_{-j} \right] \leq c^2 \exp(2(L+1)\|U_{-j}\|) \frac{\delta^2}{\tau^{1/2}},$$

and we can conclude

$$\mathbb{E} \left[\mathbb{E} \left[\mathbb{I}_{U \in \tilde{J}^\delta} \mid Y \right]^2 \right] \leq c^2 \mathbb{E} [\exp(2(L+1)\|U_{-j}\|)] \frac{\delta^2}{\tau^{1/2}},$$

where U_{-j} is a $(d-1)$ -dimensional Normal random variable for which $\mathbb{E}[\exp(2(L+1)\|U_{-j}\|)]$ is finite. The final result is obtained by noting that $\mathbb{P}[\|U\| \geq R_\delta] = o(\delta^2)$ due to the definition of R_δ and standard asymptotic results for a d -dimensional Normal random variable. $\square$

The previous lemma can be used to show that Assumption 3.1 is satisfied for a process $X_t = \exp(W_t)$ where $(W_t)_{t \geq 0}$ is a Wiener process and a set S whose boundary is K and the set $\log(K)$ is (Si). We next prove a more general result showing Assumption 3.1 for sets whose boundary can be covered by exponentials of (Si) sets and processes that can be written as exponentials of solutions of uniformly elliptic SDEs.

Theorem 5.5 . *For $S \subset \mathbb{R}^d$ assume that $\partial S \equiv K \subseteq \bigcup_{j=1}^n \exp J_j$ for some finite n and (Si) sets $\{J_j\}_{j=1}^n$. Assume that the SDE (3) is uniformly elliptic and $a, \sigma \sigma^T \in C_b^{\lambda,0}$ for some $\lambda > 0$ and let $\{Y_t\}_{t \in [0,1]}$ satisfy the SDE. Then, there exists $C > 0$ such that for all $0 < s < 1$ the following holds*

$$(27) \quad \mathbb{E} \left[(\mathbb{P}[d_K(\exp Y_1) \leq \delta \mid \mathcal{F}_s])^2 \right] \leq C \frac{\delta^2}{(1-s)^{1/2}}.$$

PROOF. By the assumptions on the coefficients of (3) and [5, Chapter 9, Theorem 2] the density $\Gamma(\cdot, 1; \xi, s)$ of Y_1 given $Y_s = \xi$ for $s \leq 1$ exists and the upper bound (25) holds for $|m| = 0$. Hence,

$$\begin{aligned} \mathbb{P}[d_K(\exp Y_1) \leq \delta \mid \mathcal{F}_s] &= \mathbb{P}[d_K(\exp Y_1) \leq \delta \mid Y_s] \\ &= \int_{\mathbb{R}^d} \mathbb{I}_{d_K(\exp y) \leq \delta} \Gamma(y, 1; Y_s, s) dy \\ &\leq \int_{\mathbb{R}^d} \mathbb{I}_{d_K(\exp y) \leq \delta} \left(\frac{C_0}{(1-s)^{d/2}} \exp \left(-c_0 \frac{|y - Y_s|^2}{1-s} \right) \right) dy \\ &= C_0 (2\pi/c_0)^{d/2} \mathbb{P}[d_K(\exp(Z + Y_s)) \leq \delta \mid Y_s], \end{aligned}$$

where Z is a multivariate Normal random variable with zero mean and variance $((1-s)/c_0)I_d$ where I_d is the $d \times d$ identity matrix. Similarly, using (25) on the density of Y_s ,

$$\begin{aligned} &\mathbb{E}[(\mathbb{P}[d_K(\exp(Z + Y_s)) \leq \delta \mid Y_s])^2] \\ &\leq C_0 (2\pi/c_0)^{d/2} \mathbb{E}[(\mathbb{P}[d_K(\exp(Z + Y)) \leq \delta \mid Y])^2], \end{aligned}$$

where Y is a multivariate Normal random variable with variance $(s/c_0)I_d$. Then noting

$$\begin{aligned} &\mathbb{E}[(\mathbb{P}[d_K(\exp(Z + Y)) \leq \delta \mid Y])^2] \\ &\leq n \sum_{j=1}^n \mathbb{E}[(\mathbb{P}[d_{\exp J_j}(\exp(Z + Y)) \leq \delta \mid Y])^2], \end{aligned}$$

and using Lemma 5.4 we obtain the result. $\square$

6. Conclusion. In this article we have developed a new Monte Carlo estimator based on the branching of approximate solution paths of the underlying stochastic differential equation. Under certain assumptions, the new estimator, when combined with MLMC, can be used to compute digital options with an improved computational complexity. Future directions for analysis could include extending Theorem 4.3 to the case of exponentials of solutions to uniformly elliptic SDEs, bounding higher moments of the error, particularly for the case of the antithetic estimators similar to Appendix A and extending the analysis to the case of solutions of locally elliptic SDEs to justify the numerical results in Section 4.

There are also many applications that could benefit from the new estimator and the branching ideas presented above. First, instead of computing a single probability, the new estimator can be used to compute multiple probabilities to reconstruct the cumulative (and probability) density functions. This would provide an alternative approach to the smoothing approach used in [11].

When MLMC is used together with the pathwise sensitivity approach (also known as IPA, Infinitesimal Perturbation Analysis) to evaluate financial sensitivities known collectively as Greeks, the loss of smoothness due to differentiation of the payoff function affects the computational complexity [3]; the branching estimator could significantly alleviate this. Similarly, the branching estimator could be used in combination with the finite difference (or “bumping”) approach to computing Greeks to counteract the increase in the variance that results when decreasing the bump magnitude.

A final observation is that branching could also be used when the underlying model is a parabolic stochastic PDE instead of an SDE.

APPENDIX A: BOUNDING THE KURTOSIS OF THE BRANCHING ESTIMATOR

The objective of this appendix is to prove that the kurtosis of the branching estimator for the Euler-Maruyama and Milstein discretisations is $o(h^{-\nu})$ for any $\nu > 0$, for an elliptic SDE with a boundary set K for which there exists a constant C such that

$$P[d_K(X_1) \leq \delta | \mathcal{F}_{1-\tau}] \leq C \delta / \tau^{1/2},$$

see also Assumption 3.1 and Theorem 5.3. If we define $Q = 2^{\hat{\ell}}$, and number the particles as indicated in Fig. 1, then noting that $|\Delta P_\ell^{(i)}|^n = |\Delta P_\ell^{(i)}|$ for $n=2,3,4$, the fourth moment of the branching estimator is bounded by

$$\begin{aligned} \mathbb{E}[(\Delta \mathcal{P}_\ell)^4] &\leq Q^{-4} \sum_{i,j,k,m=1}^Q \mathbb{E}[|\Delta P_\ell^{(i)}| |\Delta P_\ell^{(j)}| |\Delta P_\ell^{(k)}| |\Delta P_\ell^{(m)}|] \\ &= 12 Q^{-4} \sum_{i < j < k < m} \mathbb{E}[|\Delta P_\ell^{(i)}| |\Delta P_\ell^{(j)}| |\Delta P_\ell^{(k)}| |\Delta P_\ell^{(m)}|] \\ &\quad + 36 Q^{-4} \sum_{i < j < k} \mathbb{E}[|\Delta P_\ell^{(i)}| |\Delta P_\ell^{(j)}| |\Delta P_\ell^{(k)}|] \\ &\quad + 14 Q^{-4} \sum_{i < j} \mathbb{E}[|\Delta P_\ell^{(i)}| |\Delta P_\ell^{(j)}|] \\ &\quad + Q^{-4} \sum_i \mathbb{E}[|\Delta P_\ell^{(i)}|]. \end{aligned} \tag{28}$$

To begin with, we focus attention on the case with four distinct indices $i < j < k < m$, as this is the most common case. There are 5 different branching patterns among these,

but in each case through repeated use of

$$\mathbb{E}[\mathbb{I}_{d_K(X_1^{(i)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(j)}) \leq \delta} | \mathcal{F}_{1-\tau_{i,j}}] \leq C \delta \tau_{i,j}^{-1/2} \mathbb{E}[\mathbb{I}_{d_K(X_1^{(i)}) \leq \delta} | \mathcal{F}_{1-\tau_{i,j}}],$$

where $1-\tau_{i,j}$ is the time at which the particles $X_1^{(i)}$ and $X_1^{(j)}$ separate, we obtain

$$\mathbb{E}[\mathbb{I}_{d_K(X_1^{(i)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(j)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(k)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(m)}) \leq \delta}] \leq C^4 \delta^4 \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2}.$$

If we define the extreme set E to be those cases for which

$$\max_{n \in \{i,j,k,m\}} \max \left\{ \|X_1^{(n)} - \bar{X}_{\ell,1}^{(n)}\|, \|X_1^{(n)} - \bar{X}_{\ell-1,1}^{(n)}\| \right\} \geq \delta,$$

for some $\delta > 0$, then

$$\mathbb{P}[E] \leq 4 \left(\mathbb{P}[\|X_1 - \bar{X}_{\ell,1}\| \geq \delta] + \mathbb{P}[\|X_1 - \bar{X}_{\ell-1,1}\| \geq \delta] \right) \lesssim h_\ell^{\beta q/2} \delta^{-q},$$

due to the usual Markov inequality based on the q -th moment of the strong error being bounded. We also have

$$|\Delta P_\ell^{(i)}| |\Delta P_\ell^{(j)}| |\Delta P_\ell^{(k)}| |\Delta P_\ell^{(m)}| \leq \mathbb{I}_{d_K(X_1^{(i)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(j)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(k)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(m)}) \leq \delta} + \mathbb{I}_E.$$

So it follows that

$$\begin{aligned} \mathbb{E}[|\Delta P_\ell^{(i)}| |\Delta P_\ell^{(j)}| |\Delta P_\ell^{(k)}| |\Delta P_\ell^{(m)}|] &\lesssim \delta^4 \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2} + h_\ell^{\beta q/2} \delta^{-q} \\ &\lesssim h_\ell^{2\beta q/(4+q)} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2}, \end{aligned}$$

by choosing $\delta^2 \simeq h_\ell^{\beta q/(4+q)}$. For any fixed i , provided $\eta < 2$,

$$\sum_{j \neq i} \tau_{i,j}^{-1/2} = \tau_0^{-1/2} \sum_{\ell'=0}^{\hat{\ell}-1} 2^{\hat{\ell}-1-\ell'} 2^{\eta \ell'/2} \simeq 2^{\hat{\ell}} = Q.$$

This gives us

$$\begin{aligned} Q^{-4} \sum_{i < j < k < m} \mathbb{E}[|\Delta P_\ell^{(i)}| |\Delta P_\ell^{(j)}| |\Delta P_\ell^{(k)}| |\Delta P_\ell^{(m)}|] &\lesssim h_\ell^{2\beta q/(4+q)} Q^{-4} \sum_{i < j < k < m} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2} \\ &\lesssim h_\ell^{\beta q/(4+q)} Q^{-4} \sum_{i < j < k} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \left(\sum_{m \neq k} \tau_{k,m}^{-1/2} \right) \\ &\lesssim h_\ell^{2\beta q/(4+q)} Q^{-3} \sum_{i < j < k} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \\ &\lesssim h_\ell^{2\beta q/(4+q)} Q^{-3} \sum_{i < j} \tau_{i,j}^{-1/2} \left(\sum_{k \neq j} \tau_{j,k}^{-1/2} \right) \\ &\lesssim h_\ell^{2\beta q/(4+q)} Q^{-2} \sum_{i < j} \tau_{i,j}^{-1/2} \\ &\lesssim h_\ell^{2\beta q/(4+q)} Q^{-2} \sum_i \left(\sum_{j \neq i} \tau_{i,j}^{-1/2} \right) \\ &\lesssim h_\ell^{2\beta q/(4+q)}. \end{aligned}$$

Further analysis following the same approach proves that this is the dominant contribution in (28), and hence

$$\mathbb{E}[(\Delta\mathcal{P}_\ell)^4] \lesssim h_\ell^{2\beta q/(4+q)}.$$

Similar analysis, or referring to Theorems 2.4 and 3.2, shows that the second moment is bounded as follows

$$\mathbb{E}[(\Delta\mathcal{P}_\ell)^2] \lesssim h_\ell^{\beta q/(2+q)}.$$

If we assume the second moment has a lower bound of $\mathcal{O}(h_\ell^\beta)$ then it follows that the kurtosis is bounded by

$$\text{Kurt}[\Delta\mathcal{P}_\ell] \lesssim h_\ell^{-8\beta/(4+q)},$$

for all $q > 2$, and therefore is $o(h_\ell^{-\nu})$ for any $\nu > 0$.

The analysis can be extended to the exponential SDE case by first expanding the extreme set E to include cases in which

$$\max_{n \in \{i,j,k,m\}} \|\log X_1^{(n)}\| \geq R_\delta,$$

where $R_\delta := |\log \delta|^{3/4}$ as defined previously in the proof of Lemma 5.4. Equation (26) in that proof gives us

$$\mathbb{E}[\mathbb{I}_{d_K(X_1) \leq \delta} \mathbb{I}_{\log X_1 \leq R_\delta} \mid \mathcal{F}_{1-\tau}] \lesssim \exp(R_\delta) \delta / \tau^{1/2},$$

so then we obtain

$$\mathbb{E}[\mathbb{I}_{d_K(X_1^{(i)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(j)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(k)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(m)}) \leq \delta} \mathbb{I}_{E^c}] \lesssim \delta^{4-r} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2},$$

for any $r > 0$ and $\mathbb{P}[E]$ remains $\mathcal{O}(h_\ell^{\beta q/2} \delta^{-q})$ as before when δ is as previously chosen. Therefore the final conclusion remains that the kurtosis is $o(h_\ell^{-\nu})$ for any $\nu > 0$.

Acknowledgments. We thank Soeren Wolfers for the helpful discussion regarding Lemma 5.2. MBG gratefully acknowledges research funding from the UK EPSRC (ICONIC programme grant EP/P020720/1), and the Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA).

REFERENCES

- [1] ACHTSIS, N., COOLS, R. and NUYENS, D. (2013). Conditional sampling for barrier option pricing under the LT method. *SIAM Journal on Financial Mathematics* **4** 327–352. doi:10.1137/110855909
- [2] BAYER, C., HAMMOUDA, C. B. and TEMPONE, R. (2020). Numerical smoothing and hierarchical approximations for efficient option pricing and density estimation. [arxiv:2003.05708](https://arxiv.org/abs/2003.05708)
- [3] BURGOS, S. and GILES, M. B. (2012). Computing Greeks Using Multilevel Path Simulation. In *Monte Carlo and Quasi-Monte Carlo Methods 2010* 281–296. Springer Berlin Heidelberg. doi:10.1007/978-3-642-27440-4_13
- [4] ETHERIDGE, A. (2000). *An introduction to superprocesses* **20**. American Mathematical Soc.
- [5] FRIEDMAN, A. (2008). *Partial differential equations of parabolic type*. Courier Dover Publications.
- [6] GILES, M. B. (2008). Multilevel Monte Carlo Path Simulation. *Operations Research* **56** 607–617. doi:10.1287/opre.1070.0496
- [7] GILES, M. B. (2015). Multilevel Monte Carlo methods. *Acta Numerica* **24** 259–328. doi:10.1017/s096249291500001x
- [8] GILES, M. B. (2024). MLMC techniques for discontinuous functions. In *Monte Carlo and Quasi-Monte Carlo Methods 2022* (A. Keller, S. Heinrich and H. Niederreiter, eds.) Springer Berlin Heidelberg.

- [9] GILES, M. B., DEBRABANT, K. and RÖSSLER, A. (2019). Analysis of multilevel Monte Carlo path simulation using the Milstein discretisation. *Discrete & Continuous Dynamical Systems - B* **24** 3881–3903. doi:10.3934/dcdsb.2018335
- [10] GILES, M. B., HIGHAM, D. J. and MAO, X. (2009). Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics* **13** 403–413. doi:10.1007/s00780-009-0092-1
- [11] GILES, M. B., NAGAPETIAN, T. and RITTER, K. (2015). Multilevel Monte Carlo Approximation of Distribution Functions and Densities. *SIAM/ASA Journal on Uncertainty Quantification* **3** 267–295. doi:10.1137/140960086
- [12] GILES, M. B. and SZPRUCH, L. (2014). Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *The Annals of Applied Probability* **24** 1585–1620. doi:10.1214/13-aap957
- [13] GOBET, E. and LABART, C. (2008). Sharp estimates for the convergence of the density of the Euler scheme in small time. *Electronic Communications in Probability* **13** 352–363. doi:10.1214/ECP.v13-1393
- [14] GRIEBEL, M., KUO, F. Y., LEOVEY, H. and SLOAN, I. H. (2018). High dimensional integration of kinks and jumps – smoothing by preintegration. *Journal of Computational and Applied Mathematics* **344** 259–274. doi:10.1016/j.cam.2018.04.009
- [15] GRIEBEL, M., KUO, F. Y. and SLOAN, I. H. (2013). The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition. *Mathematics of Computation* **82** 383–400. doi:10.1090/S0025-5718-2012-02578-6
- [16] HAJI-ALI, A.-L., NOBILE, F., VON SCHWERIN, E. and TEMPONE, R. (2015). Optimization of mesh hierarchies in multilevel Monte Carlo samplers. *Stochastics and Partial Differential Equations Analysis and Computations* **4** 76–112. doi:10.1007/s40072-015-0049-7
- [17] HAJI-ALI, A.-L., SPENCE, J. and TECKENTRUP, A. L. (2022). Adaptive Multilevel Monte Carlo for Probabilities. *SIAM Journal on Numerical Analysis* **60** 2125–2149. doi:10.1137/21m1447064
- [18] KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations. Applications of Mathematics (New York)* **23**. Springer Berlin Heidelberg. doi:10.1007/978-3-662-12616-5 MR1214374
- [19] LE GALL, J.-F. and DUQUESNE, T. (2002). Random Trees, Lévy Processes, and Spatial Branching Processes. *Astérisque* **281** 30. doi:10.24033/ast.545
- [20] MCKEAN, H. P. (1975). Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Communications on pure and applied mathematics* **28** 323–331. doi:10.1002/cpa.3160280302