MULTILEVEL PATH BRANCHING FOR DIGITAL OPTIONS

BY MICHAEL B. GILES\textsuperscript{1,a} and ABDUL-LAEEF HAJI-ALI\textsuperscript{2,b,c}

\textsuperscript{1}Mathematical Institute, University of Oxford, \textsuperscript{a}mike.giles@maths.ox.ac.uk
\textsuperscript{2}Maxwell Institute, School of Mathematical and Computer Sciences, Heriot-Watt University, \textsuperscript{b}a.haji@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\cite{giles2015Multilevel} considers a scalar SDE

\[ dX_t = a(X_t, t) \, dt + \sigma(X_t, t) \, dW_t, \]

for $t \in [0, 1]$ with a sequence of approximate paths $\{ (\overline{X}_{\ell,t})_{t \in [0,1]} \}_{\ell \in \{0,1,\ldots\}^\infty}$ 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 $E[f(X_1)] \approx E[f(\overline{X}_{L,1})]$ for some function $f$ and we define $\Delta P_\ell := f(\overline{X}_{\ell,1}) - f(\overline{X}_{\ell-1,1})$ with $\Delta P_0 := f(\overline{X}_{0,1})$, we have the telescoping summation

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

The MLMC estimator is then

\[ \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \Delta P_\ell^{(n)}, \]

with the coarse and fine paths within $\Delta P_\ell^{(n)}$ based on the same driving Brownian path. If there are constants $\alpha, \beta, \gamma$ such that the cost of a level $\ell$ sample $\Delta P_\ell$ is $W_\ell \sim 2^{\alpha \ell}$, its variance is $V_\ell := \text{Var}[\Delta P_\ell] \sim 2^{\beta \ell}$, and the weak error is $|E[f(\overline{X}_{L,1})] - f(X_1)| \sim 2^{-\alpha L}$, 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 $\varepsilon$ 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 $\beta < \gamma$\cite{giles2015Multi}. If the function $f$ is globally Lipschitz, with constant $L_f$, then

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

In the case of the Euler-Maruyama discretization when the SDE coefficients, $a$ and $\sigma$, are Lipschitz and grow linearly in $x$ and are $1/2$-Hölder continuous in $t$, this results in $V_\ell = \mathcal{O}(h_\ell)$\cite[Theorem 10.2.2]{giles2015Multi} along with $W_\ell = \mathcal{O}(h_\ell^{-1})$, so $\beta = \gamma$ and the computational complexity grows linearly in $\varepsilon^{-1}$.

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

Keywords and phrases: Monte Carlo, Multilevel, Path splitting, Computational complexity, Branching processes.
complexity is $O(\varepsilon^{-2}\log \varepsilon^2)$. When using a first-order Milstein discretization, and under additional differentiability assumptions on $\sigma$, the variance is reduced to $V_t = O(h_t^2)$ and the complexity is improved to $O(\varepsilon^{-2})$. 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 $V_t$ which is $O(h_t^2)$ when $f$ is smooth, and $O(h_t^{3/2})$ when $f$ is Lipschitz and piecewise smooth; both are sufficient for the computational complexity of MLMC to be $O(\varepsilon^{-2})$.

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 $\Delta P_t$ is nonzero only if the final values of the fine and coarse path approximations $\overline{X}_{t,1}$ and $\overline{X}_{t-1,1}$ within $\Delta P_t$ 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 $X_1, \overline{X}_{t,1}, \overline{X}_{t-1,1}$ are all within $O(h_t^{1/2})$ of $K$, and the probability of that is $O(h_t^{1/2})$. Hence $V_t \approx O(h_t^{1/2})$ and so $\beta \approx \gamma/2$, resulting in a computational complexity which is approximately $O(\varepsilon^{-5/2})$ since standard weak convergence results give $\alpha = 2\beta$. With the Milstein discretization, $V_t \approx O(h_t)$ and the complexity is improved to $O(\varepsilon^{-2}\log \varepsilon^2)$, but with the antithetic Milstein estimator $V_t$ remains $O(h_t^{1/2})$.

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 $V_t$ from $O(h_t)$ to $O(h_t^{3/2})$ [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, $V_t$, 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 $\Delta P_t$ 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 $2^\ell$ timesteps for the fine path approximation $\overline{X}_{\ell}$, and $2^{\ell-1}$ timesteps for the coarse path $\overline{X}_{\ell-1}$, then in the simplest version of the algorithm the first branching from 1 to 2 particles is after $2^\ell$ 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 $\Delta P_4$ 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 $\Delta P_4$. (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
$2^{\ell-2}$ 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 $2^\ell$ cost per sample. The MLMC sample value would be an average of the outputs from the $2^{\ell-1}$ particles:

$$\Delta P_\ell := \frac{1}{2^{\ell-1}} \sum_{i=1}^{2^{\ell-1}} \Delta P_\ell^{(i)};$$

i.e. this $\Delta P_\ell$ counts as a single sample within an MLMC estimator similar to (2).

The claim is that with the Euler-Maruyama discretization we obtain $\text{Var}[\Delta P_\ell] \approx O(h_\ell)$ 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 $O(\tau^{1/2})$ and peak probability density of $O(\tau^{-1/2})$. For both particles to finish within $O(h_\ell^{1/2})$ of $K$ (by which we mean that both the coarse and fine path approximations end within $O(h_\ell^{1/2})$ of $K$) requires that $X_{1-\tau}$ lies within $O(\tau^{1/2})$ of $K$, which occurs with probability $O(\tau^{1/2})$, and conditional on this the probability that each particle finishes within $O(h_\ell^{1/2})$ of $K$ is $O(h_\ell^{1/2}\tau^{-1/2})$. Hence, the probability that both particles finish within $O(h_\ell^{1/2})$ of $K$ is

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

and therefore

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

- $2^{\ell-2}$ with $\tau = 2^{-1}$
- $2^{\ell-3}$ with $\tau = 2^{-2}$
- $2^{\ell-4}$ with $\tau = 2^{-3}$

\[ \vdots \]
- $2$ with $\tau = 2^{-(\ell-2)}$
- $1$ with $\tau = 2^{-(\ell-1)}$

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

\[
V_\ell \leq E[\Delta P_\ell^2] = O\left(2^{-(\ell-1)} h_\ell^{1/2} + 2^{-(\ell-1)} \sum_{\ell'=1}^{\ell-4} 2^{\ell'-1} \tau_{\ell'} h_\ell 2^{\ell'/2}\right) = O(h_\ell),
\]

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

\[
V_\ell \approx \text{Var}\left[\mathbb{E}\left[f(X_{1/2}) \mid X_{1/2} = \overline{X}_{\ell,1/2}\right] - \mathbb{E}\left[f(X_{1/2}) \mid X_{1/2} = \overline{X}_{\ell-1,1/2}\right]\right]
\]

\[
\approx \text{Var}\left[ (\overline{X}_{\ell,1/2} - \overline{X}_{\ell-1,1/2}) \cdot \nabla_x \mathbb{E}[f(X_{1/2}) \mid X_{1/2} = x] \right]_{x = \overline{X}_{\ell-1,1/2}},
\]

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 $\ell'$-th branching time, then the total cost of $\Delta P_\ell$ 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 $\eta$. 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 $t \in [0, 1]$:

\[
\text{d}X_t = a(X_t, t) \text{d}t + \sigma(X_t, t) \text{d}W_t,
\]

where $a : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d \times \mathbb{R}^d$ 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 $\{\overline{X}_{\ell,t}\}, \ell = 0, 1, \ldots$.
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 \( \sigma \) 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 \( \mathbb{P}[X_1 \in S] = \mathbb{E}[1_{X_1 \in S}] \) 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 \( \Delta P_\ell \). 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 \( \Delta P_\ell \) and show the improved computational complexity of MLMC when using the branching estimator in Corollary 2.7. In Section 3, we consider \( \Delta P_\ell = \mathbb{I}_{u_1 \in S} - \mathbb{I}_{u_{-1} \in S} \), 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 \( \ell \in \mathbb{N} \equiv \{0, 1, \ldots \} \), we use the notation \( A_\ell \leq B_\ell \) to denote \( A_\ell \leq c B_\ell \), and \( A_\ell \simeq B_\ell \) to denote \( c B_\ell \leq A_\ell \leq c B_\ell \) for some constant, deterministic \( c, c' > 0 \) that are independent of the index, \( \ell \), the accuracy tolerance, \( \varepsilon \), and other parameters which will be specified. For \( u \in \{-1, 1\}^\ell \), let \( |u|_0 := \ell \) and, for \( \ell \geq 1 \), let \( \langle u \rangle := \langle u_1, u_2, \ldots, u_{\ell-1} \rangle \in \{-1, 1\}^{\ell-1} \), using the convention \( \{-1, 1\}^0 := \{0\} \). For \( k \in \mathbb{N}, l \in \mathbb{N}, \) let \( C_{\ell+1}^{k,l} \) be the space of continuously differentiable bounded functions \( (x,t) \mapsto f(x,t) \) for \( (x,t) \in \mathbb{R}^d \times [0,T] \) with uniformly bounded derivatives with respect to \( x \) (resp. with respect to \( t \)) up to order \( k \) (resp. \( l \)). When \( f \) is a vector- or a matrix-valued function, \( f \in C_{\ell+1}^{k,l} \) means that all function components are in \( C_{\ell+1}^{k,l} \). In addition, all vector and matrix norms are Euclidean \( \ell^2 \) 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 \( \tau_\ell := \tau_0 2^{-p\ell} \) 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 \( \tilde{B}^u_t := W^u_t \) 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

\[
\tilde{B}^u_t := \begin{cases} 
\tilde{B}^{(u)}_{t} & t \in [0, 1 - \tau_{(u)_0}], \\
\tilde{B}^{(u)}_{t - \tau_{(u)_0}} + W^u_{t - 1 + \tau_{(u)_0}} & t \in (1 - \tau_{(u)_0}, 1].
\end{cases}
\]

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

\[
\Delta P_\ell := \frac{1}{2^{\ell}} \sum_{u \in \{-1, 1\}^\ell} \Delta P^u_\ell.
\]

See Fig. 1 for an illustration of the path branching involved in \( \Delta P_\ell \). Note that a single branch of a Branching Brownian motion is itself a Brownian Motion. Hence the distribution of \( \Delta P^u_\ell \) for \( u \in \{-1, 1\}^\ell \) is independent of \( u \). We will refer to a generic \( \Delta P_\ell \) 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 \( \Delta P_\ell \), which we define as the total number of Brownian increments needed to compute \( \Delta P_\ell \) 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 \( \Delta P_\ell \) by \( \text{Work}(\Delta P_\ell) \). Note that \( \mathbb{E}[\Delta P_\ell] = \mathbb{E}[\Delta P_\ell'] = \mathbb{E}[\mathbb{1}_{X_{\tilde{\ell}, 1} \in S} - \mathbb{1}_{X_{\tilde{\ell} - 1,1} \in S}] \) and hence we can use \( \Delta P_\ell \) instead of \( \Delta P_\ell' \) in the MLMC setup (2). However, under some conditions, we will see in this section that the cost of \( \Delta P_\ell \) is not significantly larger than the cost of \( \Delta P_\ell' \) while the variance is significantly smaller, leading to a better computational complexity of MLMC. In particular, even though the cost of each sample of \( \Delta P_\ell \) is \( h_\ell^{-1} \), many of the samples share an underlying Brownian path up to some branching point and hence the total cost for the \( 2^\ell \) 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

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

and

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

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 \( \alpha \geq \beta_d / 2 \)

\[
\mathbb{E}[\mathbb{1}_{X_{\tilde{\ell}, 1} \in S} - \mathbb{1}_{X_{\tilde{\ell} - 1,1} \in S}] \lesssim h_\ell^\alpha. \quad (6)
\]

Typical approximate values for \( \beta_c, \beta_d \) 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, \( a, \sigma \in C^{3,1}_b \) and \( \frac{\partial a}{\partial x} \in C^{1,0}_b \).
and
\[ Var(\Delta P_{\ell}) \lesssim \begin{cases} \ell h_{\ell}^{\beta_{\ell}+1/\eta} + \ell h_{\ell}^{\beta_{\ell}-\max\{0,p-1/\eta\}} & \text{if } \eta \neq 1, \\ \ell h_{\ell}^{\beta_{\ell}+1/\eta} + \ell h_{\ell}^{\beta_{\ell}} & \text{if } \eta = 1. \end{cases} \]

**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 \( h_{\ell} \), 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 \( h_{\ell}^{-1}(\tau_{\ell-1} - \tau_{\ell}) + 2 \) Brownian increments. Accordingly, the total work is bounded by

\[ h_{\ell}^{-1} \left( (1 - \tau_{0}) + \sum_{\ell' = 1}^{\ell-1} 2^{\ell'} (\tau_{\ell'-1} - \tau_{\ell'}) + 2^{\ell} \tau_{\ell-1} \right) + 2 \sum_{\ell' = 0}^{\ell-1} 2^{\ell'} \]

\[ = h_{\ell}^{-1} \left( (1 - \tau_{0}) + \tau_{0} \sum_{\ell' = 1}^{\ell-1} 2^{\ell'} (2^{-\eta(\ell'-1)} - 2^{-\eta\ell'}) + \tau_{0} 2^{\ell} 2^{-\eta(\ell-1)} \right) + 2 \sum_{\ell' = 0}^{\ell-1} 2^{\ell'} \]

\[ \lesssim \begin{cases} h_{\ell}^{-1/2} \max\{1-\eta,0\} \ell + 2^{\ell} & \text{if } \eta \neq 1, \\ \ell h_{\ell}^{-1/2} + 2^{\ell} & \text{if } \eta = 1. \end{cases} \]

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

**Variance:**

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

Using (5a), we have that

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

Let \( |u \wedge v|_{0} = \max\{\ell' \leq \min\{\|u\|_{0},|v|_{0}\} : u_i = v_i \text{ for all } i \in \{1,\ldots,\ell'\} \} \), and note that two payoff differences \( \Delta P_{\ell}^{u} \) and \( \Delta P_{\ell}^{v} \) 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

\[ \mathbb{E}[\Delta P_{\ell}^{u} \Delta P_{\ell}^{v}] = \mathbb{E}\left[\left(\mathbb{E}[\Delta P_{\ell}^{u} | F_{1-\tau_{|u \wedge v|_{0}}}]\right)^{2}\right] \lesssim h_{\ell}^{-\beta_{\ell}} (\tau_{|u \wedge v|_{0}})^{-p}. \]
We can then evaluate the double sum as

$$\frac{1}{2^2}\sum_{u\in\{-1,1\}^\ell} \sum_{v\in\{-1,1\}^\ell} \sum_{u
ot\equiv v} E[\Delta P^u_{\ell} \Delta P^v_{\ell}] \lesssim \frac{h^2}{2^2} \sum_{u\in\{-1,1\}^\ell} \sum_{v\in\{-1,1\}^\ell} \tau_{\mid u\land v\mid}^{-p} \lesssim \frac{h^2}{2^2} \sum_{\ell'=0}^{\ell-1} 2^{\ell'-\ell-1}\tau_{\ell'}^{-p} \lesssim \frac{h^2}{2^2} \left( \sum_{\ell'=0}^{\ell-1} 2^{\ell'-\ell-1}\tau_{\ell'}^{-p} \right).$$

Here, we evaluated the double sum by noting that $|u \land v| \in \{0, \ldots, \ell - 1\}$ when $u, v \in \{-1,1\}^\ell$ and $u \not\equiv v$. Then summing over these possible values, indexing by $\ell'$, we count the number of possibilities of having $u = v$ up to the $\ell'$ index (this is $2^{\ell'}$) the $\ell'$ being not equal (this is 2), and then the rest of the indices ($\ell - \ell'$ being arbitrary (this is $2^{\ell' + 1}$ for $u$ and similar for $v$). The result is

$$\sum_{u\in\{-1,1\}^\ell} \sum_{v\in\{-1,1\}^\ell} \sum_{u
ot\equiv v} \left( \tau_{|u\land v|} \right)^{-p} \lesssim \sum_{\ell'=0}^{\ell-1} 2^{\ell'+1} 2^{\ell'-\ell-1} 2^{\ell'-1} \tau_{\ell'}^{-p} \lesssim \sum_{\ell'=0}^{\ell-1} 2^{\ell'-\ell-1} \tau_{\ell'}^{-p}.$$

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

$$\sum_{\ell'=1}^\ell 2^{(\eta p - 1)\ell'} \lesssim \begin{cases} \mathcal{O}(1) & \eta p < 1 \\ \ell & \eta p = 1 \\ 2^{(\eta p - 1)\ell} & \eta p > 1 \end{cases} \lesssim \begin{cases} \mathcal{O}(1) & \eta p < 1 \\ \frac{1}{\eta} \log_2(h^{-1}_\ell) & \eta p = 1 \\ h^{-p + 1/\eta} & \eta p > 1. \end{cases}$$

\[ \square \]

**Remark 2.5** (Optimal $\eta$). Theorem 2.4 shows that the choice of $\eta$ in the estimator $P_{\ell}$ in Definition 2.2 compared to $p$ is crucial to improving the variance convergence rate of the new estimator compared to $\beta_d$, the variance convergence rate of the simple estimator $\Delta P$, without substantially increasing the cost of the new estimator. We can optimize the value of $\eta$, by noting that work has the term $\sum_{\ell'=0}^{\ell-1} 2^{\ell'} \tau_{\ell'}$, and the variance has the term $\sum_{\ell'=0}^{\ell-1} 2^{-\ell'} \tau_{\ell'}^{-p}$. Hence, minimizing the work subject to a constrained variance yields the optimal value of $\tau_{\ell'} \propto 2^{\frac{\eta p - 1}{\eta p + 1}}$ and the optimal value of $\eta$ 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 $\ell' = 0, \ldots, \ell - 1$. The method and analysis can be easily extended to allow for a different number of branches at every branching point. However, after adjusting $\eta$ 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 $\varepsilon^2$ based on the branching estimator $\Delta P_\ell$ 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
\[
\begin{cases}
    O(\varepsilon^{-2\max\{0, (1-p)/2\}/\alpha} \log \varepsilon^{2l_{\beta_1}^1 = 1-p}) & p < \min\{1, 2(\beta_c - \beta_d) - 1\}, \\
    O(\varepsilon^{-2\max\{0, (1-p)/\alpha\} \log \varepsilon^{2l_{\beta_1}^1 = 1}) & 2(\beta_c - \beta_d) - 1 \leq p < 1, \\
    O(\varepsilon^{-2\max\{0, (1-p)/\alpha\} \log \varepsilon^{2l_{\beta_1}^1 + 2l_{\beta_c}^1 \leq 1}) & p = 1, \\
    O(\varepsilon^{-2\max\{0, (1-p)/\alpha\} \log \varepsilon^{2l_{\beta_c}^1 = p}) & p > 1.
\end{cases}
\]

Since $\beta_c \geq \beta_4$ as in Assumption 2.3, the computational complexity of an MLMC estimator based on $\Delta P_\ell$ is lower than that of an MLMC estimator based on $\Delta P_\ell$, the latter being $O(\varepsilon^{-2\max\{1-\beta_4, 0\}/\alpha} \log (\varepsilon)^{2l_{\beta_1}^1 = 1})$, whenever $p < 1 + \beta_c - \beta_4$.

**Proof.** Recall that the MLMC estimator is defined as
\[
\sum_{\ell=0}^L \frac{1}{N_\ell} \sum_{m=1}^{N_\ell} \Delta P_\ell^{(m)},
\]
where $\Delta P_\ell^{(m)}$ are independent samples of $\Delta P_\ell$. For $V_\ell := \text{Var}(\Delta P_\ell)$ and $W_\ell := \text{Work}(\Delta P_\ell)$ 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 $\ell$ [7],
\[
N_\ell = \left\lfloor 2\varepsilon^{-2}(V_\ell/W_\ell)^{1/2} \left(\sum_{\ell=0}^L (W_\ell V_\ell)\right)^{1/2} \right\rfloor,
\]
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 $\sum_{\ell=0}^L W_\ell \lesssim M L^{p-1} \log \varepsilon^{-1}$ is dominated by the first term when setting $L \propto \frac{1}{\alpha} \log \varepsilon^{-1}$, such that the bias in (6) is $O(\varepsilon)$, and when $p < \max\{2\alpha, 1\}$ and for sufficiently small $\varepsilon$. By Theorem 2.4, for $\eta = \frac{2}{p+1}$, we have
\[
W_\ell \lesssim \begin{cases}
    h_\ell^{-\max\{p-1, 0\}/2} & p \neq 1 \\
    \ell h_\ell^{-1} & p = 1
\end{cases},
\]
while
\[
V_\ell \lesssim \begin{cases}
    h_\ell^{\beta_4 + (p-1)/2} + h_\ell^{-\max\{p-1, 0\}/2} & p \neq 1 \\
    h_\ell^{\beta_4 + 1} + \ell h_\ell^{\beta_c} & p = 1
\end{cases}.
\]
Hence,
\[
W_\ell V_\ell \lesssim \begin{cases}
    h_\ell^{\beta_4 + (p-1)/2} + h_\ell^{\beta_c - 1} & p < 1, \\
    \ell h_\ell^{\beta_4} + \ell^2 h_\ell^{\beta_c - 1} & p = 1, \\
    h_\ell^{\beta_4} + h_\ell^{\beta_c - p} & p > 1.
\end{cases}
\]
Evaluating the sum for $h_L \approx \varepsilon^{1/\alpha}$, which implies $L \propto \frac{1}{\alpha} \log \varepsilon^{-1}$, yields the result. Using that $h_L < h_0$ and $\beta_4 > 0$,
\[
\left(\sum_{\ell=0}^L \sqrt{W_\ell V_\ell}\right)^2 \lesssim \begin{cases}
    L^{2\beta_4^1 - 1 - p} h_L^{\beta_4 + (p-1)/2} + L^{2\beta_4^1 = 1} h_L^{\beta_c - 1} & p < 1, \\
    L^{2\beta_4^1 + 2l_{\beta_4}^1, \beta_c \leq 1} h_L^{\beta_c - 1} & p = 1, \\
    L^{2\beta_c^1 = p} h_L^{\beta_c - p} & p > 1.
\end{cases}
\]
where the hidden constant is independent of $L$. Setting $L \propto \log(\varepsilon)/\alpha$

\[
\left( \sum_{\ell=0}^{L} \sqrt{W_{\ell}V_{\ell}} \right)^2 \lesssim \begin{cases} 
\varepsilon^{-\max\{0,1-\beta_{c}\}} |\log \varepsilon|^{2L_{\beta_{c}=1}} & p < 1 \\
\varepsilon^{-\max\{0,p-\beta_{c}\}} |\log \varepsilon|^{2L_{\beta_{c}=p}+2L_{\beta_{c}=p}} & p \geq 1 
\end{cases}
\]

The result is

\[
\begin{cases} 
\varepsilon^{-\max\{0,1-\beta_{c}\}} |\log \varepsilon|^{2L_{\beta_{c}=1}} & p < \min\{1,2(\beta_{c} - \beta_{d}) - 1\} \\
\varepsilon^{-\max\{0,1-\beta_{c}\}} |\log \varepsilon|^{2L_{\beta_{c}=1}+2L_{\beta_{c}=1}} & 2(\beta_{c} - \beta_{d}) - 1 \leq p < 1 \\
\varepsilon^{-\max\{0,p-\beta_{c}\}} |\log \varepsilon|^{2L_{\beta_{c}=p}+2L_{\beta_{c}=p}} & p = 1 \\
\varepsilon^{-\max\{0,p-\beta_{c}\}} |\log \varepsilon|^{2L_{\beta_{c}=p}+2L_{\beta_{c}=p}} & p > 1 
\end{cases}
\]

Note: When $p = 1$, we get an extra 2 in the log only when $\beta_{c} = 1$.

\[
\therefore
\]

**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:

\[
\begin{cases} 
\mathcal{O}(\varepsilon^{-2\max\{1-\beta_{c},0\}/\alpha} |\log \varepsilon|^{2L_{\beta_{c}=1}+1L_{\beta_{c}=1}}) & p < 1, \\
\mathcal{O}(\varepsilon^{-2\max\{1-\beta_{c},0\}/\alpha} |\log \varepsilon|^{2L_{\beta_{c}=1}+2L_{\beta_{c}=1}}) & p = 1, \\
\mathcal{O}(\varepsilon^{-2\max\{p-\beta_{c},0\}/\alpha} |\log \varepsilon|^{2L_{\beta_{c}=p}+1L_{\beta_{c}=p}}) & p > 1.
\end{cases}
\]

**Proof.** $\text{Work}(\Delta P_{\ell}) \lesssim \ell h_{\ell}^{-1}$ and

\[
\text{Var}[\Delta P_{\ell}] \lesssim \begin{cases} 
\ell h_{\ell}^{\beta_{d}+1} + h_{\ell}^{\beta_{d}-\max\{0,p-1\}} & p \neq 1 \\
\ell h_{\ell}^{\beta_{d}+1} + \ell h_{\ell}^{\beta_{c}} & p = 1
\end{cases}
\]

Hence,

\[
W_{\ell}V_{\ell} \lesssim \begin{cases} 
\ell h_{\ell}^{\beta_{d}+1} + \ell h_{\ell}^{\beta_{c} - 1} & p < 1 \\
\ell h_{\ell}^{\beta_{d}} + \ell^{2} h_{\ell}^{\beta_{c} - 1} & p = 1 \\
\ell h_{\ell}^{\beta_{d}} + \ell h_{\ell}^{\beta_{c} - p} & p > 1
\end{cases}
\]

which leads to

\[
\left( \sum_{\ell=0}^{L} \sqrt{W_{\ell}V_{\ell}} \right)^2 \lesssim \begin{cases} 
\varepsilon^{-\max\{1-\beta_{c},0\}/\alpha} |\log \varepsilon|^{2L_{\beta_{c}=1}+1L_{\beta_{c}=1}} & p < 1 \\
\varepsilon^{-\max\{1-\beta_{c},0\}/\alpha} |\log \varepsilon|^{2L_{\beta_{c}=1}+2L_{\beta_{c}=1}} & p = 1 \\
\varepsilon^{-\max\{p-\beta_{c},0\}/\alpha} |\log \varepsilon|^{2L_{\beta_{c}=p}+1L_{\beta_{c}=p}} & p > 1
\end{cases}
\]

\[
\therefore
\]

3. **Strong Analysis.** In this section, we consider

\[
\Delta P_{\ell} \equiv \mathbb{I}_{x_{\ell,1} \in S} - \mathbb{I}_{x_{\ell,1-1} \in S},
\]

and make well-motivated assumptions on the solution of the SDE in (3) and its numerical approximation $\{X_{\ell,1}\}_{0 \leq \ell \leq 1}$, 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

\[
d_J(x) := \inf_{y \in J} \|y - x\|.
\]
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 $\delta, \tau$ and $\mathcal{F}_{1-\tau}$ such that
\[
\mathbb{E}\left(\mathbb{P}[d_K(X_1) \leq \delta | \mathcal{F}_{1-\tau}]^2\right) \leq C \frac{\delta^2}{\tau^{1/2}}.
\]

Assumption 3.1 is fairly generic and depends on the set $\mathcal{F}_1 = \partial \mathcal{S}$ 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
\[
\mathbb{E}\left(\|X_1 - \mathbf{X}_{\ell,1}\|^{q}\right)^{1/q} \lesssim h^{\beta/2}. \tag{10}
\]
Then for $\Delta P_\ell$ in (7) and all $\tau \in (0, 1)$,
\[
\mathbb{E}[\Delta P_\ell]^2 \lesssim h^{\beta(1-1/(q+1))/2} \tag{11a}
\]
and
\[
\mathbb{E}[\mathbb{E}[\Delta P_\ell | \mathcal{F}_{1-\tau}]^2] \lesssim h^{\beta(1-2/(q+2)) / \tau^{1/2}}. \tag{11b}
\]

This theorem shows that Assumption 2.3 is satisfied with $\beta_c = \beta(1-2/(q+2))$, $\beta_a = \beta(1-1/(q+1))/2$ 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]. Note that $2(\beta_c - \beta_a) = \beta(1 - \frac{1}{q+2} + \frac{1}{q+1})$. Hence $\beta_c > \beta_a$ whenever $1 - \frac{1}{q+2} + \frac{1}{q+1} > 0$, i.e., when $q \geq 1$.

Proof. The proof of (11a) is similar to the proof in [10], and is included here for completeness. We start by noting that
\[
\mathbb{E}[\Delta P_\ell]^2 \leq \mathbb{E}\left[\|X_{1} \in \mathcal{S} - \mathbb{I}_{X_{\ell-1} \in \mathcal{S}}\|\right] + \mathbb{E}\left[\|X_{1} \in \mathcal{S} - \mathbb{I}_{X_{\ell,1} \in \mathcal{S}}\|\right],
\]
and for any $\delta > 0$
\[
\|X_{1} \in \mathcal{S} - \mathbb{I}_{X_{\ell,1} \in \mathcal{S}}\| \leq \|X_1 - \mathbf{X}_{\ell,1}\| > d_K(X_1)
\]
\[
\leq \|X_1 - \mathbf{X}_{\ell,1}\| > d_K(X_1) \mathbb{P}[d_K(X_1) \leq \delta] + \|X_1 - \mathbf{X}_{\ell,1}\| > d_K(X_1) > \delta
\]
\[
\leq \|d_K(X_1) \leq \delta + \|X_1 - \mathbf{X}_{\ell,1}\| > \delta
\]
\[
\leq \mathbb{P}[d_K(X_1) \leq \delta] + (h_\ell^{-\beta/2})^{-q} \left(h_\ell^{-\beta/2} \|X_1 - \mathbf{X}_{\ell,1}\|\right)^q.
\]
Hence
\[
\mathbb{E}\left[\|X_{1} \in \mathcal{S} - \mathbb{I}_{X_{\ell,1} \in \mathcal{S}}\|\right] \leq \mathbb{P}[d_K(X_1) \leq \delta] + (h_\ell^{-\beta/2})^{-q} \mathbb{E}\left[\left(h_\ell^{-\beta/2} \|X_1 - \mathbf{X}_{\ell,1}\|\right)^q\right]
\]
\[
\leq C \delta + (h_\ell^{-\beta/2})^{-q} \mathbb{E}\left[\left(h_\ell^{-\beta/2} \|X_1 - \mathbf{X}_{\ell,1}\|\right)^q\right],
\]
where we used Assumption 3.1 for $\tau = 1$ and the fact that $\mathbb{E}\left[\left(h_\ell^{-\beta/2} \|X_1 - \mathbf{X}_{\ell,1}\|\right)^q\right]$ is bounded by (10). A similar bound is obtained for $\mathbb{E}\left[\|X_{1} \in \mathcal{S} - \mathbb{I}_{X_{\ell-1} \in \mathcal{S}}\|\right]$ and then we select $\delta = h_\ell^{-\beta/2}$ to obtain (11a).
Proving (11b) follows the same steps by similarly noting that
\[
E\left[ (E[\Delta P_{\ell} | \mathcal{F}_{1-\tau}] )^2 \right] \leq 2E\left[ (E[\|X_1 \in S - \|X_{t,1} \in S \| | \mathcal{F}_{1-\tau}] )^2 \right] \\
+ 2E\left[ (E[\|X_1 \in S - \|X_{t-1,1} \in S \| | \mathcal{F}_{1-\tau}] )^2 \right],
\]
and for any \( \delta > 0 \)
\[
|\mathbb{I}_{X_1 \in S} - \|X_{t,1} \in S \| | \leq \mathbb{I}_{dK(X_1) \leq \delta} + (h_{\ell}^{-\beta/2} \delta)^{-q/2} (h_{\ell}^{-\beta/2} \|X_1 - \|X_{t,1} \| )^{q/2},
\]
so that
\[
E\left[ (E[\|X_1 \in S - \|X_{t,1} \in S \| | \mathcal{F}_{1-\tau}] )^2 \right] \leq 2E[(E[\mathbb{I}_{dK(X_1) \leq \delta} | \mathcal{F}_{1-\tau}] )^2] \\
+ 2 (h_{\ell}^{-\beta} \delta^2)^{-q/2} E\left[ \left( h_{\ell}^{-\beta/2} \|X_1 - \|X_{t,1} \| \right)^q \right].
\]
Here \( E\left[ \left( h_{\ell}^{-\beta/2} \|X_1 - \|X_{t,1} \| \right)^q \right] \) is bounded by (10), and using Assumption 3.1 we have
\[
E\left[ \left( E[\|X_1 \in S - \|X_{t,1} \in S \| | \mathcal{F}_{1-\tau}] \right)^2 \right] \leq \delta^2/\tau^{1/2} + (h_{\ell}^{-\beta} \delta^2)^{-q/2} \\
\leq \delta^2/\tau^{1/2} + (h_{\ell}^{-\beta} \delta^2)^{-q/2}/\tau^{1/2},
\]
for \( \tau < 1 \) and we choose \( \delta^2 = h_{\ell}^{-\beta+2} \) to obtain the result.

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 \( \alpha \). Then, the computational complexity of a MLMC estimator with MSE \( \varepsilon^2 \) based on \( \Delta P_{\ell} \) and \( h_{\ell} = h_0 M^{-\ell} \) for \( M \in \mathbb{N}_+ \) and \( \eta = 4/3 \) is
\[
\begin{cases}
O(\varepsilon^{-2 - \frac{1 - \beta}{2\alpha - \beta}}) & \beta \leq 1, \\
O(\varepsilon^{-2}) & \beta > 1,
\end{cases}
\]
for any \( \nu > 0 \).

**Proof.** Substitute the value \( p = 1/2 \) in Corollary 2.7
\[
\begin{cases}
O(\varepsilon^{-2 - \max\{0, 1/4 - \beta_\lambda\}/\alpha \log \varepsilon^{2\mu_{\ell,1} + 1} } & 3/2 < 2(\beta_\lambda - \beta_d) \\
O(\varepsilon^{-2 - \max\{0, 1/4 - \beta_\lambda\}/\alpha \log \varepsilon^{2\mu_{\ell,1} + 1} } & 2(\beta_\lambda - \beta_d) \leq 3/2
\end{cases}
\]

Note that \( 2(\beta_\lambda - \beta_d) = \beta (1 - \frac{1}{q+2} + \frac{1}{q+1} ) \). Hence if \( \beta > 3/2 \), then there is \( q \) large enough such that the first case applies. In this case, for a sufficiently large \( q \), \( \beta_d > 1/4 \) and we arrive at \( \varepsilon^{-2} \) computational complexity. For \( 1 < \beta < 3/2 \) we are in the second case but we can find a sufficiently large \( q \) for which \( \beta_d > 1 \), yielding \( \varepsilon^{-2} \). When \( \beta \leq 1 \), we are in the second case with increased complexity. \( \square \)
\textbf{Numerical Experiments.} In this section, we consider the SDE for the Geometric Brownian Motion (GBM)

\begin{equation}
\label{eq:gbm}
dX_{i,t} = \mu_i X_{i,t} \, dt + \sigma_i X_{i,t} \left( \rho \, dW_{i,t} + \left(1 - \rho^2\right)^{1/2} \, dW_{0,t} \right),
\end{equation}

for \( i = 1, \ldots, d \). 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 \( \mathbb{P}[\sum_{i=1}^d X_{i,1} \leq 1] \). We set \( \rho=0.7 \) and \( X_{i,0}=1, \mu_i=0.05, \sigma_i=0.2 \) for all \( i \in \{1, \ldots, d\} \). 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 \( \ell \) 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 \( h_\ell \propto \ell \) 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 \( \mathbb{E}[\mathbb{E}[\Delta P_{\ell} | \mathcal{F}_{1-\tau}]^2] \) 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 \( \beta_{\ell} \approx \beta \), hence \( \text{Var}[\Delta P_\ell] \approx \mathcal{O}(h_\ell^{\min(\beta, \beta/2+1)}) \) 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 \( \Delta P_\ell \). For the previous values of \( \beta_c, \beta_d \) and \( \beta \) and \( p = 1/2, \eta = 1 \), the computational complexity for the MLMC sampler based on \( \Delta P_\ell \) 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 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 \( \Delta P_\ell \), 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 \( \Delta P_\ell \) grows approximately in proportion to \( \bar{h}_\ell^{-1/2} \) when using Euler-Maruyama or \( \bar{h}_\ell^{-1} \) when using Milstein; recall that in both cases \( \mathbb{E}[\langle (\Delta P_\ell)^2 \rangle] = \mathbb{E}[\langle (\Delta P_\ell)^4 \rangle] \). This leads to difficulties when estimating \( \text{Var}[\Delta P_\ell] \) for sufficiently large \( \ell \) 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: \( \Delta P_\ell \) has a bounded kurtosis and hence an MLMC algorithm that relies on variance estimates is more stable when using \( \Delta P_\ell \) than when using \( \Delta P_\ell \). See also Appendix A for a proof of the boundedness of the kurtosis of \( \Delta P_\ell \).

\section{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
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.

Fig 4: The GBM example in (12) for $d=1$ when using Euler-Maruyama (solid) and Milstein (dashed) in the traditional, $\Delta P_\ell$, and branching, $\Delta P_\ell$, estimators. (a) shows numerical verification of the variance convergence of $\Delta P_\ell$. (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 $\Delta P_\ell$. (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.
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 \( \{(\overline{X}_{\ell,t})_{\ell=1}^{1},(\overline{X}_{\ell,t}^{(a)})_{\ell=0}^{1}\} \) 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, \( (\overline{X}_{\ell,t}^{u},\overline{X}_{\ell,t}^{(a),u},\overline{X}_{\ell-1,t})_{\ell=0}^{1} \) for the same branch, \( \tilde{B}^{u} \), of a Branching Brownian Motion. The antithetic estimator can then be defined as in (4) for

\[
\Delta P_{\ell}^{u} = \begin{cases} 
\left\lfloor \overline{X}_{\ell,1} \in S \right\rfloor - \left\lfloor \overline{X}_{\ell-1,1} \in S \right\rfloor & \ell = 0, \\
\frac{1}{2} \left( \left\lfloor \overline{X}_{\ell,1} \in S \right\rfloor + \left\lfloor \overline{X}_{\ell-1,1}^{(a),u} \in S \right\rfloor \right) & \ell > 0.
\end{cases}
\]

Since the cost of sampling \( \overline{X}_{\ell,t}^{(a),u} \) is the same as \( \overline{X}_{\ell,t}^{u} \), 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

\[
\Delta P_{\ell} := \begin{cases} 
\left\lfloor \overline{X}_{\ell,1} \in S \right\rfloor - \left\lfloor \overline{X}_{\ell-1,1} \in S \right\rfloor & \ell = 0, \\
\frac{1}{2} \left( \left\lfloor \overline{X}_{\ell,1} \in S \right\rfloor + \left\lfloor \overline{X}_{\ell-1,1}^{(a),u} \in S \right\rfloor \right) & \ell > 0.
\end{cases}
\]

which has the same distribution as \( \Delta P_{\ell}^{u} \) in (13). In what follows, define

\[
g_{t}(\xi) := P\left[ X_{1} \in S \mid X_{t} = \xi \right]
\]

and

\[
\overline{g}_{t,\ell}(\xi) := P\left[ \overline{X}_{\ell,1} \in S \mid \overline{X}_{\ell,t} = \xi \right].
\]

We will make the following assumptions

**Assumption 4.1.** Let \( \nabla g_{t} \) 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 \( \tau \in (0, 1) \),

\[
\|\nabla g_{1-\tau}(\xi)\| \leq \frac{c_{1}}{\tau^{1/2}} \exp\left( -c_{2} \frac{d_{K}^{2}(\xi)}{\tau} \right)
\]

and

\[
\|H_{g_{1-\tau}}(\xi)\| \leq \frac{c_{1}}{\tau} \exp\left( -c_{2} \frac{d_{K}^{2}(\xi)}{\tau} \right),
\]

and for any \( \tau \in (0, 1) \),

\[
P[d_{K}(X_{1-\tau}) < \delta] \leq c_{4}\delta.
\]

This assumption is motivated by the case when \( (X_{t})_{t\geq0} \) 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 \( \xi \in \mathbb{R}^{d} \),

\[
|\overline{g}_{t,1-\tau}(\xi) - g_{1-\tau}(\xi)| \leq \frac{c_{1}h_{\ell}}{\tau^{1/2}}.
\]
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, $X_{\ell,t}$, respectively, given $X_s=X_{\ell,s}=\xi$, the authors in [13] proved that when the SDE is uniformly elliptic and the coefficients $a,\sigma \in C_b^{3,1}$ and $\frac{\partial}{\partial x} \in C^{1,0}_b$ then for all $\xi, x, y \in \mathbb{R}^d$, $0 \leq s \leq t \leq 1$, and some constants $c$ and $C$

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

By setting $\ell=1$, $s=1-\tau$ and integrating with respect to $x$, (17) follows.

$$\left|\mathcal{P}_{1,1-\tau}(\xi) - g_{1-\tau}(\xi)\right| \leq \int_S |\Gamma_\ell(y,1;\xi,1-\tau) - \Gamma(y,1;\xi,1-\tau)| \, dy \leq C \frac{h_\ell}{\tau^{1/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{c \|\xi-y\|^2}{\tau}\right) \, dy = C (2c\pi)^{d/2} h_\ell^{1/2}.$$

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) \[ E\left[ \left\| X_1 - \overline{X}_{\ell,1} \right\|^q \right]^{1/q} \lesssim h_\ell^{1/2} \]

(18b) \[ E\left[ \left\| \frac{1}{2}(X_{\ell,1} + X_{\ell,1}^{(q)}) - \overline{X}_{\ell-1,1} \right\|^q \right]^{1/q} \lesssim h_\ell. \]

Then for $\Delta \mathcal{P}_\ell$ in (14) there exists a constant $c_2$ such that for all $\tau \in (0,1)$,

(19a) \[ E[(\Delta \mathcal{P}_\ell)^2] \leq c_2 h_\ell^{2(1-1/(q+1))/2} \]

(19b) \[ E\left[ \left( E[\Delta \mathcal{P}_\ell | \mathcal{F}_{1-\tau}] \right)^2 \right] \leq c_2 h_\ell^{2(1-5/(q+5))/2}. \]

This theorem shows that Assumption 2.3 is satisfied with $\beta_1=(1-1/(q+1))/2$, $\beta_\epsilon=2(1-5/(q+5))$ 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 $\tau$ and some $0 < r < 1$, that we will choose later, to be the set of paths for which

$$\max\left\{ \frac{\|X_{\ell,1-\tau} - X_{1-\tau}\|}{h_\ell^{r/2}}, \frac{\|X_{\ell,1-\tau}^{(q)} - X_{1-\tau}\|}{h_\ell^{r/2}}, \frac{\|X_{\ell-1,1-\tau} - X_{1-\tau}\|}{h_\ell^{r/2}}, \frac{\|\frac{1}{2}(X_{\ell,1-\tau} + X_{\ell,1-\tau}^{(q)}) - X_{\ell-1,1-\tau}\|}{h_\ell^{r/2}} \right\} \geq 1.$$
Then since $|\Delta P| \leq 1$,
\[
E[ E[ \Delta P_t \mid \mathcal{F}_{t-1}]^2 ] = E[ E[ \Delta P_t \mid \mathcal{F}_{t-1}]^2\big|\mathcal{F}_{t-1}\big] + E[ E[ \Delta P_t \mid \mathcal{F}_{t-1}]^2]
\leq P[ E] + E[ E[ E[ \Delta P_t \mid \mathcal{F}_{t-1}]^2]].
\]

Due to (18a) and the Markov inequality,
\[
P\left( \|X_{t,1-1} - X_{t-1}\| \geq h_t^{(1-r)q/2} \right) \leq C_q h_t^{(1-r)q/2},
\]
for some constant $C_q$ with similar bounds for $\|X_{t,1} - X_{t-1}\|$ and $\|X_{t,1}^{(a)} - X_{t-1}\|$.

Hence $P[ E] \leq 4C_q h_t^{(1-r)q/2}$. For the other term we have
\[
E[ E[ E[ \Delta P_t \mid \mathcal{F}_{t-1}]^2 ] = E[ E[ E[ \Delta P_t \mid \mathcal{F}_{t-1}]^2]]
\]

Due to (47) in Assumption 4.2, the second of the final three terms of this inequality is bounded by $2c_2^2 h_{t-1}^2 / \tau$ and the third by $2c_2^2 h_{t-1}^2 / \tau$ so both are $O(h_t^2 / \tau)$.

To bound the first term we perform a Taylor series expansion about $X_{t-1,1-1}$ to obtain
\[
\frac{1}{2}(g_1_{t-1}(X_{t,1-1}) + g_{t-1}(X_{t,1-1}) - g_{t-1}(X_{t-1,1-1})) = Y_1 + \frac{1}{4}Y_2 + \frac{1}{4}Y_3,
\]
where
\[
Y_1 := \left( \frac{1}{2}(X_{t,1-1} + X_{t,1-1}) - X_{t-1,1-1} \right) \nabla g_{t-1}(X_{t-1,1-1}),
Y_2 := (X_{t,1-1} - X_{t-1,1-1})^T H_{g_{t-1}}(\xi_1)(X_{t,1-1} - X_{t-1,1-1}),
Y_3 := (X_{t,1-1} - X_{t-1,1-1})^T H_{g_{t-1}}(\xi_2)(X_{t,1-1} - X_{t-1,1-1}),
\]
and where $\xi_1$ is a positively weighted average of $X_{t,1-1}$ and $X_{t-1,1-1}$ and $\xi_2$ is a positively weighted average of $X_{t,1-1}$ and $X_{t-1,1-1}$. Hence,
\[
E\left[ \mathbb{E}[ E[ E[ \Delta P_t \mid \mathcal{F}_{t-1}]^2]] \right] \leq 2 E\left[ \mathbb{E}[ E[ E[ \Delta P_t \mid \mathcal{F}_{t-1}]^2]] \right].
\]

Defining $\delta := 2(\tau / (h_t))^1/2 h_t^{r/2}$, we now split $\mathbb{E}[ E]$ into
\[
\mathbb{E}[ E] = \mathbb{E}[ E][d_K(X_{t,1-1}) > \delta] + \mathbb{E}[ E][d_K(X_{t,1-1}) < \delta].
\]

Since $h_t \leq \tau$ it follows that $\delta > 2h_t^{r/2}$ and if $\mathbb{E}[ E][d_K(X_{t,1-1}) > \delta] = 1$ then $d_K(X_{t,1-1}) > \tau / 2 \delta$, $d_K(X_{t,1-1}) > \tau / 2 \delta$, and also $d_K(\xi_1) > \tau / 2 \delta$, $d_K(\xi_2) > \tau / 2 \delta$. Therefore, by Assumption 4.1, there is a constant $C$ such that $\|\nabla g_{t-1}(X_{t-1,1-1})\| \leq C$ and $\|H_{g_{t-1}}(\xi_i)\| \leq C$. 

...
for $i=1,2$ and all $h_{\ell} \leq \tau$.

$$
\|\nabla g_1(\mathcal{X}_{\ell-1,1})\| \leq \frac{c_1}{\tau^{1/2}} \exp \left( -c_2 \frac{d_2^2(\mathcal{X}_{\ell-1,1})}{\tau} \right)
$$

$$
\leq \frac{c_1}{h_{\ell}^{1/2}} \exp \left( -c_2 \frac{\delta^2}{4\tau} \right)
$$

$$
= \frac{c_1}{h_{\ell}^{1/2}} \exp \left( -c_2 \frac{(\tau/h_{\ell}) h_{\ell}^r}{\tau} \right)
$$

$$
\leq c_1 h_{\ell}^{-1/2} \exp \left( -c_2 h_{\ell}^{-1} \right)
$$

$$
\leq c_1 \times \left( \frac{\exp(-1)}{2c_2(1-r)} \right)^{1/(2(1-r))}
$$

similarly $\|H_\eta(\xi)\| \leq c_1 \times \left( \frac{\exp(-1)}{c_2(1-r)} \right)^{1/(2(1-r))}$. Hence,

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

On the other hand, when $\mathbb{I}_{E^*} \mathcal{I}_{d_K(X_{\ell-1})} < \delta = 1$, 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

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

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

for any $r < 1$. Picking $r$ so that $\mathbb{P}[E] \approx h_{\ell}^{(5r-1)/2}$ 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 $\varepsilon^2$ based on $\Delta P_{\ell}$ 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 $O(\varepsilon^{-2})$.

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 $\Delta P_{\ell}$ as discussed in Remark 2.5.

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

$$
dX_{1,t} = dW_{1,t},
$$

$$
dX_{2,t} = X_{1,t} dW_{2,t}.
$$

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

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

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

$$
\Delta_{\ell,n} W_{i} := W_{i,(n+1)h_{\ell}} - W_{i,nh_{\ell}}
$$
shows the convergence of
and
\(X\) increases in proportion to
are consistent with Corollaries
which is obtained by setting the Lévy area term in a Milstein discretization to zero. In particular, the \(\ell\)th level approximation is defined as
\[
X_{2, \ell, (n+1)h_\ell} = X_{2, \ell, nh_\ell} + W_{1, nh_\ell} \Delta_{\ell, n} W_2 + \frac{1}{2} \Delta_{\ell, n} W_1 \Delta_{\ell, n} W_2.
\] (21)

When computing (14) for a given Brownian path, the coarse, \(X_{1, \ell-1, \cdot}\), and fine, \(X_{\ell, \cdot}\), approximations are constructed according to (21). On the other hand, the antithetic approximation at \((2n + 1)h_\ell\) and \((2n + 2)h_\ell\) for \(n = 0, 1, \ldots, (h_\ell^{-1}/2 - 1)\) is defined as
\[
\bar{X}_{2, \ell, (2n+1)h_\ell}^{(a)} = \bar{X}_{2, \ell, 2nh_\ell} + W_{1, 2nh_\ell} \Delta_{\ell, 2n+1} W_2 + \frac{1}{2} \Delta_{\ell, 2n+1} W_1 \Delta_{\ell, 2n+1} W_2,
\]
\[
\bar{X}_{2, \ell, (2n+2)h_\ell}^{(a)} = \bar{X}_{2, \ell, 2(n+1)h_\ell} + (W_{1, 2h_\ell} + \Delta_{\ell, 2n+1} W_1) \Delta_{\ell, 2n+1} W_2 + \frac{1}{2} \Delta_{\ell, 2n+1} W_1 \Delta_{\ell, 2n+1} W_2.
\]

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 \(P[X_{1} \in S]\) 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 \(\mathbb{E}\left[\mathbb{E}[\Delta P_{\ell} | F_{i-1-\cdot}]^{2}\right]\) 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 \(\Delta P_{\ell}\) when using Euler-Maruyama or the antithetic estimator. The computational complexities of MLMC based on the branching estimator \(\Delta P_{\ell}\) 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 \(\eta\) 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 \(\Delta P_{\ell}\), labelled “Without branching”, the computational complexity of MLMC is \(O(\varepsilon^{-5/2})\) 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 \(\Delta P_{\ell}\) 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 \(\mathbb{E}\left[\mathbb{E}[\Delta P_{\ell} | F_{i-\cdot}]^{2}\right]\) and Fig. 7a shows the convergence of \(\text{Var}[\Delta P_{\ell}]\). The observed convergence rates are slightly worse than those observed for the previous example. Nevertheless, recalling that the work of \(\Delta P_{\ell}\) increases in proportion to \(h_{\ell}^{-1}\log(h_{\ell})\), the computational complexity of a MLMC estimator is still \(O(\varepsilon^{-2})\) 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 \(\Delta P_{\ell}\).
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, $\Delta P_\ell$, and branching, $\Delta P_\ell$, estimators. (a) shows numerical verification of the variance convergence of $\Delta P_\ell$ (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 $\Delta P_\ell$. (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, \( \Delta P_\ell \), and branching, \( \Delta P_\ell' \), estimators. (a) shows numerical verification of the variance convergence of \( \Delta P_\ell \). (b) shows the total work estimate of MLMC for different tolerances. (c) The kurtosis of \( \Delta P_\ell \).
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 $|m| = m_1 + \ldots + m_d$ and $D_x^m \equiv \frac{\partial^{m_1}}{\partial x_1^{m_1}} \ldots \frac{\partial^{m_d}}{\partial x_d^{m_d}}$. Given a set $J \subset \mathbb{R}^d$, define

\begin{equation}
J^\delta \equiv \{ y \in \mathbb{R}^d : d_J(y) \leq \delta \},
\end{equation}

and for a function $u : \mathbb{R}^d \to \mathbb{R}^d$, let $u(J)$ denote the image of $J$ under the mapping $x \to u(x)$, 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 $j \in \{1,\ldots,d\}$ and a Lipschitz function $f : \mathbb{R}^{d-1} \to \mathbb{R}$ 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

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

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

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

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

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

Then

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

Using standard 1D results on $Z_j$ yields

$P[|Z_j - f(Z_{-j})| \leq (L+1)\delta | Z_{-j}] \leq \frac{2(L+1)}{(2\pi)^{1/2}} \times \frac{\delta}{\tau^{1/2}}$, and the result follows. \hfill \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 \( a, \sigma \sigma^T \) are \( \lambda \)-Hölder continuous in space uniformly with respect to time 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

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

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

\[
|D^m_{\xi} \Phi[X_{s} = \xi]| \leq \frac{c_1}{(1-s)^{|m|/2}} \exp\left(-c_2 \frac{d^2_K(\xi)}{(1-s)}\right).
\]

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 \( C_m, c_m > 0 \),

\[
|D^m_{\xi} \Gamma(x, 1; \xi, s)| \leq \frac{C_m}{(1-s)^{(d+|m|)/2}} \exp\left(-c_m \frac{\|x - \xi\|^2}{1 - s}\right),
\]

when the SDE (3) is uniformly elliptic and, for \( |m| = 0 \), when \( a, \sigma \sigma^T \) are \( \lambda \)-Hölder continuous in space uniformly with respect to times \([5, \text{Chapter 9, Theorem 2}]\) and for \( 0 \leq |m| \leq 2 \) when \( a, \sigma \sigma^T \in C^{2,0}_b \) \([5, \text{Chapter 9, Theorem 7}]\). Hence

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

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

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

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

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

Hence

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

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
   \[
   |D_x^m \mathbb{P} [X_1 \in S | X_s = \xi]| 
   \leq \int_S |D_x^m \Gamma(x, 1; \xi, s)| \, dx 
   \leq \int_S \frac{C_m}{(1-s)(d+|m|/2)} \times \exp \left( -c_m \frac{\|x - \xi\|^2}{1-s} \right) \, dx 
   \leq \frac{1}{(1-s)^{|m|/2}} \exp \left( -c_m \frac{\inf_{x \in S} \|x - \xi\|^2}{2(1-s)} \right) \int_S \frac{C_m}{(1-s)^{d/2}} \times \exp \left( -c_m \frac{\|x - \xi\|^2}{2(1-s)} \right) \, dx 
   \leq \frac{1}{(1-s)^{|m|/2}} \exp \left( -c_m \frac{\inf_{x \in S} \|x - \xi\|^2}{2(1-s)} \right) (4\pi)^{d/2} C_m, 
   \]
   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
   \[
   |D_x^m \mathbb{P} [X_1 \in S | X_s = \xi]| = |D_x^m (1 - \mathbb{P} [X_1 \in S^c | X_s = \xi])| 
   = |D_x^m (\mathbb{P} [X_1 \in S^c | X_s = \xi])|, 
   \]
   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 $\delta \leq b$,
\[
P\left[ \min_{n \in \mathbb{N}} |Z - n^{-b}| \leq \delta \right] = 2 \int_0^\infty \mathbb{I}_{\min_{n \in \mathbb{N}} |y - n^{-b}| \leq \delta} \phi(y) \, dy
\]
\[
\geq \frac{2 \exp(-1/2)}{(2\pi)^{1/2}} \int_0^1 \mathbb{I}_{\min_{n \in \mathbb{N}} |y - n^{-b}| \leq \delta} \, dy
\]
\[
\geq \frac{2 \exp(-1/2)}{(2\pi)^{1/2}} \int_0^1 \mathbb{I}_{\min_{n \in \mathbb{N}} |y - n^{-b}| \leq \delta} \, dy
\]
\[
= \frac{2 \exp(-1/2)}{(2\pi)^{1/2}} \frac{(\delta/b)^{b/(b+1)}}{(b+1)},
\]
where $\phi(\cdot)$ is the density of standard normal random variable. To justify the last inequality first note that for some $n$, the distance between two points on $K$ is $n^{-b} - (n + 1)^{-b}$. When $\delta$ is larger than that distance, and since subsequent terms have smaller distances, the indicator from 0 to $n^{-b}$ is 1. Hence we need to find the largest $n^{-b}$ (or smallest $n$) such that
\[
n^{-b} - (n + 1)^{-b} \leq \delta
\]
Recall
\[
\left( n^{-b} - (n + 1)^{-b} \right)^{\frac{1}{b+1}} \leq b^{\frac{1}{b+1}} n^{-b}
\]
To show this, simplify the inequality to
\[
n \left( 1 - (1 + 1/n)^{-b} \right) \leq b
\]
then letting $x = 1/n \in [0, 1]$, the function $x^{-1}(1 - (1 + x)^{-b})$ is decreasing over $x \in [0, 1]$. For $x = 0$ the limit is $b$. Hence we pick the smallest $n$ for which
\[
n^{-b} \leq (\delta/b)^{b/(b+1)}
\]
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 $\delta < b$
\[
P\left[ \min_{y \in K} \|Z - y\| \leq \delta \right] \geq \frac{1}{2} \exp(-1/2) \frac{(\delta/b)^{2b/(b+1)}}{2\pi}.
\]
\[
P\left[ \min_{y \in K} \|Z - y\| \leq \delta \right] = \int_0^{2\pi} \int_0^\infty \mathbb{I}_{\mathbb{D} \leq \delta} \phi(r) \, r \, dr \, d\theta
\]
\[
\geq \frac{\exp(-1/2)}{2\pi} \int_0^{2\pi} \int_0^1 \mathbb{I}_{\mathbb{D} \leq \delta} \, r \, dr \, d\theta
\]
For a fixed $\theta$, the distance between two points is $(\theta/\pi + n)^{-b} - (\theta/\pi + n + 1)^{-b}$. When $\delta$ is larger than that distance for some $n$, and since subsequent terms have smaller distances, the indicator from $r = 0$ to $r = (\theta/\pi + n)^{-b}$ for that $n$ is 1. Hence we need to find the largest $(n + \theta/\pi)^{-b}$ (or smallest $n$) such that
\[
(\theta/\pi + n)^{-b} - (\theta/\pi + n + 1)^{-b} \leq \delta
\]
Like before, we impose (for $\theta \in [0, 2\pi]$)
\[
(\theta/\pi + n)^{-b} \leq (\delta/b)^{b/(b+1)}
\]
Hence, since \( \| I \| \leq R \) and \( \| V \| \leq R \), let \( c = 2(\nu + 1) \| f(0) \| \).

Letting \( U \equiv Z + Y \), to bound \( E( E[ I \text{exp}(U) | Y] ]^2 \), we start by noting that \( I \text{exp}(U) | Y] \leq I \text{exp}(U) | \| U \| \leq R \), and also that \( E[ I \text{exp}(U) | Y] ] \leq 1 \). Hence,

\[
E( E[ I \text{exp}(U) | Y] ]^2 \leq 2 E( E[ I \text{exp}(U) | \| U \| \leq R \}^2 ] \leq 2 E( E[ I \text{exp}(U) | \| U \| \leq R \}^2 ] \leq 2 E( E[ I \text{exp}(U) | Y] ]^2 ] + 2P(U \geq R).$$

**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 \( (S) \) set. There exists \( \delta_0 > 0 \) and \( C > 0 \) such that for all \( \delta \leq \delta_0 \)

\[
E( P[ d_{\text{exp}}(\exp(Z + Y)) < \delta | Y] ]^2 \leq C \frac{\delta^2}{\tau^{1/2}}.
\]
Then, by normality of $U$ and

$$
E[ E[I_{U \in J} \mid Y]^2] = E \left[ \left( E[ E[I_{U \in J} \mid Y, Z_j] \mid Y] \right)^2 \right] \\
= E \left[ E \left( \left( E[I_{U \in J} \mid Y, Z_j] \mid Y \right)^2 \right) \mid Y \right] \\
\leq E \left[ E \left( \left( E[I_{U \in J} \mid Y, Z_j] \mid Y \right)^2 \right) \mid Y \right] \\
\leq E \left[ E \left( \left( E[I_{U \in J} \mid Y, Z_j] \mid Y \right)^2 \right) \mid Y, Z_j \right].
$$

Using standard 1D results on $Z_j$,

$$
E \left[ E[I_{U \in J} \mid Y, Z_j]^2 \mid Y, Z_j \right] \leq c^2 \exp(2(L+1)\|U-J\|) \frac{\delta^2}{\tau^{1/2}},
$$

and we can conclude

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

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

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 $\partial S = \bigcup_{j=1}^n \exp J_j$ for some finite $n$ and (Si) sets $\{J_j\}_{j=1}^n$. Assume that the SDE (3) is uniformly elliptic and $a, \sigma \sigma^T$ are $\lambda$-Hölder continuous in space uniformly with respect to time 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

$$
E \left[ \left( P[d_K(\exp Y_t) \leq \delta \mid F_s] \right)^2 \right] \leq C \frac{\delta^2}{(1-s)^{1/2}}.
$$

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

$$
P[d_K(\exp Y_t) \leq \delta \mid F_s] = P[d_K(\exp Y_1) \leq \delta \mid Y_s]
= \int_{\mathbb{R}^d} I_{d_K(\exp y) \leq \delta} \Gamma(y, 1; Y_s, s) \, dy
\leq \int_{\mathbb{R}^d} I_{d_K(\exp y) \leq \delta} \left( \frac{C_0}{(1-s)^d} \exp \left( -c_0 \frac{|y-Y_s|^2}{1-s} \right) \right) \, dy
= C_0(2\pi/c_0)^{d/2} P[d_K(\exp(Z+Y_s)) \leq \delta \mid Y_s],
$$

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
where \( Y \) is a multivariate Normal random variable with variance \((s/c_0)I_d\). Then noting
\[
E[(P[d_K(\exp(Z + Y)) \leq \delta | Y]^2] 
\leq C_0 (2\pi/c_0)^{d/2} E[(P[d_K(\exp(Z + Y)) \leq \delta | Y]^2] 
\leq n \sum_{j=1}^{n} E[(P[d_{exp,i}(\exp(Z + Y)) \leq \delta | Y]^2],
\]
and using Lemma 5.4 we obtain the result.

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 \( Q = 2^j \), and number the particles as indicated in Fig. 1, then noting that \( |\Delta F_\ell^{(i)}|_n = |\Delta F_\ell^{(i)}| \) for \( n = 2, 3, 4 \), the fourth
moment of the branching estimator is bounded by

$$E[(\Delta P_\ell)^4] \leq Q^{-4} \sum_{i,j,k,m=1}^Q E[|\Delta P_\ell^{(i)}||\Delta P_\ell^{(j)}||\Delta P_\ell^{(k)}||\Delta P_\ell^{(m)}|]$$

$$= 12 Q^{-4} \sum_{i<j<k<m} E[|\Delta P_\ell^{(i)}||\Delta P_\ell^{(j)}||\Delta P_\ell^{(k)}||\Delta P_\ell^{(m)}|]$$

$$+ 36 Q^{-4} \sum_{i<j<k} E[|\Delta P_\ell^{(i)}||\Delta P_\ell^{(j)}||\Delta P_\ell^{(k)}|]$$

$$+ 14 Q^{-4} \sum_{i<j} E[|\Delta P_\ell^{(i)}||\Delta P_\ell^{(j)}|]$$

$$+ Q^{-4} \sum_i E[|\Delta P_\ell^{(i)}|].$$

(28)

In more details: $Q(Q-1)(Q-2)(Q-3)$ quads all different, $6Q(Q-1)(Q-2)$ with 3 different, $3Q(Q-1)(Q-2)$ with 2 pairs, $4Q(Q-1)(Q-2)$ with 3 same, $Q$ all the same.

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

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

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

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

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

$$\max_{n \in \{i,j,k,m\}} \mathbb{I}_{d_K(X_1^{(i)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(j)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(k)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(m)}) \leq \delta} \geq \delta,$$

for some $\delta > 0$, then

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

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

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

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

$$\sum_{j \neq i} \tau_{ij}^{-1/2} = \tau_0^{-1/2} \sum_{\ell' = 0}^{\ell-1} 2^{\ell-1-\ell'} 2^{\eta \ell'/2} \sim 2^\ell = Q.$$
This gives us

\[ Q^{-4} \sum_{i<j<k<m} E[|\Delta P_{\ell}^{(i)}||\Delta P_{\ell}^{(j)}||\Delta P_{\ell}^{(k)}||\Delta P_{\ell}^{(m)}|] \lesssim h_{\ell}^{2\beta q/(4+q)} Q^{-4} \sum_{i<j<k<m} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2} \]

\[ \lesssim h_{\ell}^{2\beta q/(4+q)} Q^{-4} \sum_{i<j<k} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2} \]

\[ \lesssim h_{\ell}^{2\beta q/(4+q)} Q^{-3} \sum_{i<j<k} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \]

\[ \lesssim h_{\ell}^{2\beta q/(4+q)} Q^{-2} \sum_{i<j} \tau_{i,j}^{-1/2} \]

\[ \lesssim h_{\ell}^{2\beta q/(4+q)} Q^{-2} \sum_{i} \left( \sum_{j \neq i} \tau_{i,j}^{-1/2} \right) \]

\[ \lesssim h_{\ell}^{2\beta q/(4+q)}. \]

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

\[ E[|\Delta P_{\ell}|^4] \lesssim h_{\ell}^{2\beta q/(4+q)}. \]

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

\[ E[|\Delta P_{\ell}|^2] \lesssim h_{\ell}^{\beta q/(2+q)}. \]

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

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

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

\[ E\left[ \mathbb{I}_{d_K(X_1) \leq \delta} \mathbb{I}_{\log X_1 \leq R_\delta} \left| F_{1-\tau} \right. \right] \lesssim \exp(R_\delta) \delta^{1/2}, \]

so then we obtain for the non-extreme paths

\[ E[\mathbb{I}_{d_K(X_1^{(i)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(j)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(k)}) \leq \delta} \mathbb{I}_{d_K(X_1^{(m)}) \leq \delta} \mathbb{I}_{E^c}] \lesssim \delta^{4} \exp(3R_\delta) \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2} \]

\[ \lesssim \delta^{4-r} \tau_{i,j}^{-1/2} \tau_{j,k}^{-1/2} \tau_{k,m}^{-1/2} \]

for any \( r>0 \) and \( P[E] \) remains \( \mathcal{O}(h_{\ell}^{3q/2} \delta^{-q}) \) as before when \( \delta \) 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:2003.0578
[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/orep.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., Nagapetyan, 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 multidimensional SDEs without Lévy area simulation. The Annals of Applied Probability 24 1585–1620. doi:10.1214/13-aap957
[13] Goët, 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] Hajj-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] Hajj-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] McKea, 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