Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation

Michael B. Giles  Kristian Debrabant  Andreas Rößler
October 24, 2013

Abstract
The multilevel Monte Carlo path simulation method introduced by Giles (Operations Research, 56(3):607-617, 2008) exploits strong convergence properties to improve the computational complexity by combining simulations with different levels of resolution. Previous research has analysed its efficiency when using the Euler-Maruyama discretisation, and also demonstrated its improved efficiency using the Milstein discretisation with its improved strong convergence. In this paper we analyse its efficiency for scalar SDEs using the Milstein discretisation, bounding the order of convergence of the variance of the multilevel estimator, and hence determining the computational complexity of the method.

Keywords: Multilevel, Monte Carlo, stochastic differential equations, numerical analysis.

Mathematical Subject Classification: MSC 60H10, MSC 60H35, MSC 65C05, MSC 65C30.

1 Introduction
In computational finance, Monte Carlo methods are used to estimate the expected value $E[P]$ of a discounted payoff function which depends on the solution of an SDE of the generic form

$$dS = a(S,t)\, dt + b(S,t)\, dW, \quad 0 \leq t \leq T,$$

subject to specified initial data $S(0) = S_0$.

Using a simple Monte Carlo method with a numerical discretisation with first order weak convergence, to achieve a root-mean-square error of $O(\varepsilon)$ would require $O(\varepsilon^{-2})$ independent paths, each with $O(\varepsilon^{-1})$ timesteps, giving a computational complexity which is $O(\varepsilon^{-3})$. However, Giles introduced a new multilevel Monte Carlo (MLMC) approach [10, 9] which reduces the cost to $O(\varepsilon^{-2})$ under certain circumstances. This multilevel approach, which is related to the two-level method of Kebaier [17], and Heinrich’s multilevel approach for parametric integration [15], combines the results of simulations with different numbers of timesteps.
The key identity underlying the method is

\[ E[\hat{P}_L] = E[\hat{P}_0] + \sum_{\ell=1}^{L} E[\hat{P}_\ell - \hat{P}_{\ell-1}]. \]  

(2)

This expresses the expectation on the finest level of resolution, using \(2^{-L}\) uniform timesteps, as the sum of the expected value on level 0, using just one timestep of size \(T\), plus a sum of expected corrections between levels \(\ell\) and \(\ell-1\). The quantity \(E[\hat{P}_\ell - \hat{P}_{\ell-1}]\) can be estimated using \(N_\ell\) independent samples by

\[ \hat{Y}_\ell = N_\ell^{-1} \sum_{i=1}^{N_\ell} (\hat{P}^{(i)}_\ell - \hat{P}^{(i)}_{\ell-1}). \]  

(3)

Note that the difference \(\hat{P}^{(i)}_\ell - \hat{P}^{(i)}_{\ell-1}\) comes from two discrete approximations with different timesteps but the same Brownian path; this difference is small because of the strong convergence properties of the numerical discretisation. The variance of this simple estimator is \(V[\hat{Y}_\ell] = N_\ell^{-1} V_\ell\) where \(V_\ell\) is the variance of a single sample. It is the convergence of \(V_\ell\) as \(\ell \to \infty\) which is the focus of this paper, because of its central role in the following theorem [10] which comes from an optimal choice of the number of samples to be used on each level.

**Theorem 1.1** Let \(P\) denote a functional of the solution of stochastic differential equation (1) for a given Brownian path \(W\), and let \(\hat{P}_\ell\) denote the corresponding approximation using a numerical discretisation with timestep \(h_\ell = 2^{-\ell} T\).

If there exist independent estimators \(\hat{Y}_\ell\) based on \(N_\ell\) Monte Carlo samples, and positive constants \(\alpha \geq \frac{1}{2}, \beta, c_1, c_2, c_3\) such that

i) \(E[|\hat{P}_\ell - P|] \leq c_1 h_\ell^\alpha\)

ii) \(E[\hat{Y}_\ell] = \begin{cases} E[\hat{P}_0], & \ell = 0 \\ E[\hat{P}_\ell - \hat{P}_{\ell-1}], & \ell > 0 \end{cases}\)

iii) \(V[\hat{Y}_\ell] \leq c_2 N_\ell^{-1} h_\ell^\beta\)

iv) \(C_\ell\), the computational complexity of \(\hat{Y}_\ell\), is bounded by

\[ C_\ell \leq c_3 N_\ell h_\ell^{-1}, \]

then there exists a positive constant \(c_4\) such that for any \(\varepsilon < e^{-1}\) there are values \(L\) and \(N_\ell\) for which the multilevel estimator

\[ \hat{Y} = \sum_{\ell=0}^{L} \hat{Y}_\ell, \]

has a mean-square-error with bound

\[ MSE \equiv E \left[ \left( \hat{Y} - E[P] \right)^2 \right] < \varepsilon^2 \]
Table 1: Orders of convergence for $V_L$ as observed numerically and proved analytically for both the Euler and Milstein discretisations; $\delta$ can be any strictly positive constant.

| option       | Euler numerical | Euler analysis | Milstein numerical | Milstein analysis |
|--------------|-----------------|----------------|--------------------|-------------------|
| Lipschitz    | $O(h)$          | $O(h)$         | $O(h^2)$           | $O(h^2)$          |
| Asian        | $O(h)$          | $O(h)$         | $O(h^2)$           | $O(h^2)$          |
| lookback     | $O(h)$          | $O(h)$         | $O(h^2)$           | $O(h^2)$          |
| barrier      | $O(h^{1/2})$    | $o(h^{1/2-\delta})$ | $O(h^3/2)$        | $o(h^{3/2-\delta})$ |
| digital      | $O(h^{1/2})$    | $O(h^{1/2} \log h)$ | $O(h^3/2)$        | $o(h^{3/2-\delta})$ |

Table 1 shows the orders of convergence for $V_L$ as observed numerically and proved analytically for both the Euler and Milstein discretisations. $\delta$ can be any strictly positive constant.

For the MLMC method based on the simple Euler discretisation, Giles, Higham and Mao [11] proved that $V_L = O(h)$ for European options (based on the final value of the underlying $S(T)$) with a uniform Lipschitz payoff, Asian options (based on the average value of the underlying) and lookback options (based on the minimum or maximum of the underlying). They also proved that $V_L = o(h^{1/2-\delta})$, for any $\delta > 0$, for barrier options (in which the payoff is zero if the underlying crosses, or fails to cross, a certain level) and digital options (for which the payoff is a discontinuous function of $S(T)$). The final result has been tightened by Avikainen [1] who proved that $V_L = O(h^{1/2} \log h)$.

As summarised in Table 1, numerical results [10] suggest that all of these results are near-optimal.

For the MLMC method based on the Milstein discretisation, numerical results [9] suggest that $V_L = O(h^2)$ for European options with a uniform Lipschitz payoff and for Asian and lookback options, and $V_L = O(h^{3/2})$ for the barrier and digital options. In this paper we aim to establish these orders of convergence analytically, and do so near optimally in each case.

## 2 Previous results

The analysis in this paper builds on a large body of results in the existing literature. They are included here for the sake of completeness.

### 2.1 Solution of the SDE and its Milstein discretisation

We will assume throughout this paper that the SDE [11] is scalar, and the drift function $a \in C^{2,1}(\mathbb{R} \times \mathbb{R}^+)$ and volatility function $b \in C^{3,1}(\mathbb{R} \times \mathbb{R}^+)$ satisfy the following standard conditions in which we use the notation $L_0 \equiv \partial / \partial t + a \partial / \partial S$ and $L_1 \equiv b \partial / \partial S$. 

with a computational complexity $C$ with bound

$$
C \leq \begin{cases} 
    c_4 \varepsilon^{-2}, & \beta > 1, \\
    c_4 \varepsilon^{-2} (\log \varepsilon)^2, & \beta = 1, \\
    c_4 \varepsilon^{-2 - (1-\beta)/\alpha}, & 0 < \beta < 1.
\end{cases}
$$
• A1 (uniform Lipschitz condition): there exists $K_1$ such that
  
  \[ |a(x, t) - a(y, t)| + |b(x, t) - b(y, t)| + |L_1 b(x, t) - L_1 b(y, t)| \leq K_1 |x - y| \]

• A2 (linear growth bound): there exists $K_2$ such that
  
  \[ |a(x, t)| + |L_0 a(x, t)| + |L_1 a(x, t)| + |b(x, t)| + |L_0 b(x, t)| 
  + |L_1 b(x, t)| + |L_0 L_1 b(x, t)| + |L_1 L_1 b(x, t)| \leq K_2 (1 + |x|) \]

• A3 (additional Lipschitz condition): there exists $K_3$ such that
  
  \[ |b(x, t) - b(x, s)| \leq K_3 (1 + |x|) \sqrt{|t - s|} \]

Under these conditions, we have the following result for the analytic solution to the SDE [18].

**Theorem 2.1** Provided the assumptions A1-A3 are satisfied, then for all positive integers $m$

\[ \mathbb{E} \left[ \sup_{0 < t < T} |S(t)|^m \right] < \infty. \]

With initial data $\hat{S}_0 = S(0)$, the Milstein discretisation of equation (1) using a uniform timestep of size $h$ is

\[ \hat{S}_{n+1} = \hat{S}_n + a_n h + b_n \Delta W_n + \frac{1}{2} b'_n b_n \left( (\Delta W_n)^2 - h \right), \quad (4) \]

where $b' \equiv \partial b/\partial S$, the subscript $n$ denotes the timestep index, and $a_n$, $b_n$ and $b'_n$ are evaluated at $\hat{S}_n, t_n$. Kloeden and Platen [18] define a continuous time interpolant

\[ \hat{S}_{KP}(t) = \hat{S}_n + a_n (t-t_n) + b_n (W(t)-W_n) + \frac{1}{2} b'_n b_n \left( (W(t)-W_n)^2 - (t-t_n) \right), \quad (5) \]

for $t_n \leq t \leq t_{n+1}$ and prove the following result.

**Theorem 2.2** Provided the assumptions A1-A3 are satisfied, then for all positive integers $m$ there exists a constant $C_m$ such that

\[ \mathbb{E} \left[ \sup_{0 < t < T} |S(t) - \hat{S}_{KP}(t)|^m \right] < C_m h^m, \quad \mathbb{E} \left[ \sup_{0 < t < T} |\hat{S}_{KP}(t)|^m \right] < C_m. \]

### 2.2 Brownian bridge results

If the drift $a$ and volatility are constant, the SDE (1) has solution

\[ S(t) = S_0 + a t + b W(t) \]

and hence within the time interval $[t_n, t_{n+1}]$ of length $h$ we have

\[ S(t) = S_n + \lambda (S_{n+1} - S_n) + b (W(t) - W_n - \lambda (W_{n+1} - W_n)) \quad (6) \]

where $\lambda \equiv (t - t_n)/h$. This means that the deviation of $S(t)$ from a piecewise linear interpolation of the values $S_n \equiv S(t_n)$ is proportional to the deviation of $W(t)$ from its piecewise linear interpolation. It can be proved that the distribution of the latter is independent of the Brownian increment $W_{n+1} - W_n$, and furthermore we have the following results (see for example [14]).
Lemma 2.3  Conditional on $S_n$ and $S_{n+1}$, the distribution for the integral of $S(t)$ over the interval $[t_n, t_{n+1}]$ is given by

$$\int_{t_n}^{t_{n+1}} S(t) \, dt = \frac{1}{2} h (S_n + S_{n+1}) + b I_n$$  \hspace{1cm} (7)

where

$$I_n \equiv \int_{t_n}^{t_{n+1}} (W(t) - W_n - \lambda (W_{n+1} - W_n)) \, dt$$

is a $N(0, \frac{1}{12} h^3)$ Normal random variable, independent of $W_{n+1} - W_n$.

Lemma 2.4  Conditional on $S_n$ and $S_{n+1}$, the distributions for the minimum and maximum of $S(t)$ over the interval $[t_n, t_{n+1}]$ are given by

$$S_{n,\min} = \frac{1}{2} \left( S_n + S_{n+1} - \sqrt{(S_{n+1} - S_n)^2 - 2 b^2 h \log U_n} \right),$$

$$S_{n,\max} = \frac{1}{2} \left( S_n + S_{n+1} + \sqrt{(S_{n+1} - S_n)^2 - 2 b^2 h \log V_n} \right),$$

where $U_n$ and $V_n$ are each uniformly distributed on the unit interval $(0,1)$.

Lemma 2.5  Provided $b \neq 0$, conditional on $S_n$ and $S_{n+1}$, the probability that the minimum (or maximum) of $S(t)$ over the interval $[t_n, t_{n+1}]$ is less than (or greater than) some value $B$, is

$$\mathbb{P} \left( \inf_{[t_n, t_{n+1}]} S(t) < B \mid S_n, S_{n+1} \right) = \exp \left( \frac{-2 (S_n - B) + (S_{n+1} - B)^+}{b^2 h} \right),$$

$$\mathbb{P} \left( \sup_{[t_n, t_{n+1}]} S(t) > B \mid S_n, S_{n+1} \right) = \exp \left( \frac{-2 (B - S_n) + (B - S_{n+1})^+}{b^2 h} \right),$$

where the notation $(x)^+$ means $\max(x, 0)$.

Corollary 2.6  If $W(t)$ is a Brownian motion with $W(0) = W(1) = 0$, then for $x > 0$

$$\mathbb{P} \left( \sup_{[0,1]} W(t) > x \right) = \mathbb{P} \left( \inf_{[0,1]} W(t) < -x \right) = \exp(-2x^2),$$

and hence $\mathbb{E} \left[ \sup_{[0,1]} |W(t)|^m \right]$ is finite for all positive integers $m$.

2.3 Extreme values

The following results come from extreme value theory which determine the limiting distribution of the maximum of a large set of i.i.d. random variables [7].
**Lemma 2.7** If $U_n, n = 1, \ldots, N$ are independent samples from a uniform distribution on the unit interval $[0, 1]$, then for any positive integer $m$

$$
\mathbb{E} \left[ \max_n |\log U_n|^m \right] = \mathcal{O}((\log N)^m), \quad \text{as } N \to \infty. \quad (10)
$$

**Lemma 2.8** If $Z_n, n = 1, \ldots, N$ are independent samples from a standard Normal distribution, then for any positive integer $m$

$$
\mathbb{E} \left[ \max_n |Z_n|^m \right] = \mathcal{O}((\log N)^{m/2}), \quad \text{as } N \to \infty. \quad (11)
$$

**Corollary 2.9** If $W_n(t), n = 1, \ldots, N$ are independent Brownian paths on $[0, 1]$, conditional on $W_n(0) = W_n(1) = 0$, then for any positive integer $m$

$$
\mathbb{E} \left[ \max_n \sup_{[0,1]} |W_n(t)|^m \right] = \mathcal{O}((\log N)^{m/2}), \quad \text{as } N \to \infty. \quad (12)
$$

**Proof** From Corollary 2.6 for sufficiently large $x$ the tail probability for $|W_n(t)|$ is less than that of a standard Normal random variable.

### 2.4 Extreme paths

Some of the proofs in [11] use an argument that certain “extreme” paths make a negligible contribution to the overall expectation. This same argument will be employed in this paper but in a more compact form based on these two lemmas.

**Lemma 2.10** If $X_\ell$ is a scalar random variable defined on level $\ell$ of the multilevel analysis, and for each positive integer $m$, $\mathbb{E}[|X_\ell|^m]$ is uniformly bounded, then, for any $\delta > 0$,

$$
\mathbb{P} \left( |X_\ell| > h_\ell^{-\delta} \right) = o(h_\ell^p), \quad \forall p > 0.
$$

**Proof** Follows immediately from Markov’s inequality

$$
\mathbb{P} \left( |X_\ell| \geq h_\ell^{-\delta} \right) = \mathbb{P} \left( |X_\ell|^m \geq h_\ell^{-m\delta} \right) \leq h_\ell^{-m\delta} \mathbb{E}[|X_\ell|^m],
$$

by choosing $m > p/\delta$.

**Lemma 2.11** If $Y_\ell$ is a scalar random variable on level $\ell$, $\mathbb{E}[Y_\ell^2]$ is uniformly bounded, and for each $p > 0$, the indicator function $1_{E_\ell}$ on level $\ell$ (which takes value 1 or 0 depending whether or not a path lies within some set $E_\ell$) satisfies

$$
\mathbb{E}[1_{E_\ell}] = o(h_\ell^p),
$$

then for each $p > 0$,

$$
\mathbb{E}[|Y_\ell|1_{E_\ell}] = o(h_\ell^p).
$$
**Proof** Immediate consequence of Hölder inequality which gives

\[ E[|Y_\ell|1_{E_\ell}] \leq (E[Y_\ell^2])^{1/2} (E[1_{E_\ell}])^{1/2}. \]

In the proofs in the main analysis, Lemma 2.10 will be used to establish the preconditions for Lemma 2.11 from which it can be concluded, by choosing \( p \) sufficiently large, that the contribution of the extreme paths is negligible compared to the paths that are not extreme.

## 3 Analysis of the Milstein MLMC method

### 3.1 Brownian interpolation

In all of the cases to be analysed, the discrete paths are simulated using the Milstein method, with each level having twice as many timesteps as the previous level. This gives a set of values at discrete times, \( \hat{S}(t_n) \) where \( t_n = n \cdot h \). By approximating the drift and volatility as being constant within each timestep, we define the following Brownian interpolation based on equation (6),

\[
\hat{S}(t) = \hat{S}_n + \lambda (\hat{S}_{n+1} - \hat{S}_n) + b_n \left( W(t) - W_n - \lambda (W_{n+1} - W_n) \right)
\]

where \( \lambda \equiv (t - t_n)/h \). The advantage of this interpolation compared to the standard Kloeden-Platen interpolant is that we can use Lemmas 2.3 – 2.5 in constructing the multilevel estimators. The accuracy of the interpolant relative to the Kloeden-Platen interpolant is given by the following theorem:

**Theorem 3.1** If \( \hat{S}(t) \) is the interpolant defined by (13) and \( \hat{S}_{KP}(t) \) is the Kloeden-Platen interpolant defined by (5) then for any positive integer \( m \)

i)

\[
E \left[ \sup_{[0,T]} \left| \hat{S}(t) - \hat{S}_{KP}(t) \right|^m \right] = O((h \log h)^m),
\]

ii)

\[
\sup_{[0,T]} E \left[ \left| \hat{S}(t) - \hat{S}_{KP}(t) \right|^m \right] = O(h^m),
\]

iii)

\[
E \left[ \left( \int_0^T (\hat{S}(t) - \hat{S}_{KP}(t))^2 \, dt \right)^2 \right] = O(h^3).
\]
Proof. In each case we use the fact that $E[\max_n |b'_n b_n|^m]$ is finite due to Theorem 2.2 and Assumption A2. In addition, for $t \in [t_n, t_{n+1}]$, the difference between the two interpolants is
\[
\tilde{S}(t) - \tilde{S}_{KP}(t) = \frac{1}{2} b'_n b_n Y(t),
\]
where
\[
Y(t) = \lambda (W_{n+1} - W_n)^2 - (W(t) - W_n)^2
\]
\[
= \lambda (1 - \lambda) (W_{n+1} - W_n)^2 - (W(t) - W_n - \lambda (W_{n+1} - W_n))^2
\]
\[
- 2 \lambda (W_{n+1} - W_n) (W(t) - W_n - \lambda (W_{n+1} - W_n)).
\]
i) Using Hölder’s inequality, the assertion follows from
\[
E \left[ \sup_{[0,T]} |\tilde{S}(t) - \tilde{S}_{KP}(t)|^m \right] \leq 2^{-m} \sqrt{E \left[ \max_n |b'_n b_n|^{2m} \right] E \left[ \sup_{[0,T]} |Y(t)|^{2m} \right]},
\]
together with bounds on $E \left[ \sup_{[0,T]} |Y(t)|^{2m} \right]$ coming from Lemma 2.8 and Corollary 2.9.

ii) By setting $W(t) - W_n = \sqrt{\lambda h} Z_1$ and $W_{n+1} - W(t) = \sqrt{(1-\lambda) h} Z_2$, with $Z_1, Z_2$ independent standard Normal random variables, one can prove that
\[
|Y| \leq h \max(Z_1^2, Z_2^2) \implies |Y|^m \leq h^m \max(Z_1^{2m}, Z_2^{2m}) \leq h^m (Z_1^{2m} + Z_2^{2m})
\]
and hence the assertion follows from
\[
E \left[ |\tilde{S}(t) - \tilde{S}_{KP}(t)|^m \right] = 2^{-m} E[|b'_n b_n|^m] E[|Y|^m],
\]
and standard bounds for moments of Normal random variables.

iii) Defining $X_n := \int_{t_n}^{t_{n+1}} Y(t) \, dt$ we obtain
\[
E \left[ \left( \int_0^T (\tilde{S}(t) - \tilde{S}_{KP}(t)) \, dt \right)^2 \right] = \frac{1}{4} E \left[ \left( \sum_{n=0}^{N-1} b'_n b_n X_n \right)^2 \right].
\]
For $n > m$, $E[b'_n b_m X_m b'_n b_n] = 0$ since $X_n$ is independent of $b'_m b_m X_m b'_n b_n$ and $E[X_n] = 0$. In addition, the $X_n$ are iid random variables, and therefore
\[
E \left[ \left( \int_0^T (\tilde{S}(t) - \tilde{S}_{KP}(t)) \, dt \right)^2 \right] = \frac{1}{4} E[X_0^2] \sum_{n=0}^{N-1} E[(b'_n b_n)^2].
\]
The proof is completed by noting that $E[X_0^2] = \mathcal{O}(h^4)$ due to standard moment bounds for Brownian increments.
3.2 Estimator construction

For each Brownian input, the multilevel estimator (3) requires the calculation of the payoff difference \( \hat{P}_f^\ell - \hat{P}_c^{\ell-1} \). Here \( \hat{P}_f^\ell \) is a fine-path estimate using timestep \( h^\ell = 2^{-l}T \), and \( \hat{P}_c^{\ell-1} \) is the corresponding coarse-path estimate using timestep \( h = 2^{-(l-1)}T \). As explained in [9], to ensure that the identity (2) is correctly respected, it is required that

\[
E[\hat{P}_f^\ell] = E[\hat{P}_c^{\ell-1}].
\] (14)

In the simplest case of a European option, this can be achieved very simply by defining \( \hat{P}_f^\ell - 1 \) and \( \hat{P}_c^{\ell-1} \) to be the same. However, for the other applications the definition of \( \hat{P}_c^{\ell-1} \) involves information from the discrete simulation of \( \hat{P}_f^\ell \), which is not available in computing \( \hat{P}_f^\ell \). This is done to reduce the variance of the estimator, but it must be shown that equality (14) is satisfied. This will be achieved in each case through a construction based on the Brownian interpolant. In many cases this will involve evaluating the coarse path interpolant at the intermediate times \( t_n \) for odd values of \( n \), using the value for \( W_n \) which was used for the fine path.

The analysis of the variance of the multilevel estimator will often use the following decomposition of the difference between the Brownian interpolants for the fine and coarse paths,

\[
\tilde{S}_f(t) - \tilde{S}_c(t) = (\tilde{S}_f(t) - \tilde{S}_{KP}(t)) - (\tilde{S}_c(t) - \tilde{S}_{KP}(t)) + (\tilde{S}_{KP}(t) - S(t)) - (\tilde{S}_{KP}(t) - S(t))
\] (15)

with Theorem 3.1 bounding the error in the first two terms, and Theorem 2.2 bounding the error in the last two terms.

3.3 Lipschitz payoffs

Many European options, such as simple put and call options, have a payoff that is a Lipschitz function of the value of the underlying asset at maturity,

\[
P = f(S(T)).
\]

Discrete Asian options have a payoff which is a Lipschitz function of the value at maturity and the average of the underlying asset at a finite number of times \( T_m \),

\[
\bar{S} = M^{-1} \sum_{m=1}^{M} S(T_m).
\]

Both of these are special cases of a more general class of Lipschitz payoffs in which the payoff is a Lipschitz function of the values of the underlying asset at a finite number of times \( T_m \),

\[
P = f(S(T_1), S(T_2), \ldots, S(T_M)),
\]

with the Lipschitz bound

\[
\left| f(S_1^{(2)}, S_2^{(2)}, \ldots, S_M^{(2)}) - f(S_1^{(1)}, S_2^{(1)}, \ldots, S_M^{(1)}) \right| \leq L \sum_{m=1}^{M} \left| S_m^{(2)} - S_m^{(1)} \right|,
\]

\]
for some constant $L$. In the numerical discretisation the fine and coarse path payoffs are both defined by

$$
\hat{P} = f(\hat{S}(T_1), \hat{S}(T_2), \ldots, \hat{S}(T_M)),
$$

with $\hat{S}(t)$ given by the Brownian interpolation. Note that this will require the additional simulation of $W(T_m)$ if $T_m$ does not correspond to one of the existing timesteps.

We get the following result concerning the variance of the multilevel estimator:

**Theorem 3.2** This approximation for Lipschitz payoffs has $V_\ell = O(h_\ell^2)$.

**Proof** From the Lipschitz bound and Jensen’s inequality we obtain

$$
\forall [\hat{P}_\ell - \hat{P}_{\ell-1}] \leq E([\hat{P}_\ell - \hat{P}_{\ell-1}]^2) \leq L^2 M \sum_{m=1}^{M} E((\hat{S}(T_m) - \hat{S}(T_m))^2].
$$

The decomposition (15) implies that

$$
E((\hat{S}(T_m) - \hat{S}(T_m))^2] \leq 4 \left( E((\hat{S}(T_m) - \hat{S}(T_m))^2] + E((\hat{S}(T_m) - \hat{S}(T_m))^2] + E((\hat{S}(T_m) - \hat{S}(T_m))^2] + E((\hat{S}(T_m) - \hat{S}(T_m))^2] \right)
$$

and the proof is completed using the results from Theorems 2.2 and 3.1.

### 3.4 Asian options

Continuously monitored Asian options have a payoff that is a uniform Lipschitz function of two arguments, the average over the time interval

$$
\bar{S} \equiv T^{-1} \int_{0}^{T} S(t) \, dt,
$$

and the value at maturity, $S(T)$. We now consider two alternative numerical approximations.

#### 3.4.1 Treatment 1

The first treatment is that used in [9], in which the fine and coarse path averages $\bar{S}$ are defined by integrating the interpolant (13). Because of Lemma 2.3, this gives

$$
\int_{0}^{T} \hat{S}_{f}(t) \, dt = \sum_{n=0}^{N-1} \frac{1}{2} h_\ell \left( \hat{S}_n + \hat{S}_{n+1} \right) + b_n I^f_n
$$

where $I^f_n$ are independent $N(0, \frac{1}{12} h_\ell^3)$ variables. The payoff for the coarse path is defined similarly, but a straightforward calculation gives

$$
I^c_n \equiv \int_{t_n}^{t_{n+2}} \left( W(t) - \frac{t - t_n}{2h_\ell} (W_{n+2} - W_n) \right) dt
= I^f_n + I^f_{n+1} - \frac{1}{2} h_\ell (W_{n+2} - 2W_{n+1} + W_n),
$$

and so $I^c_n$ is obtained from the Brownian path information used for the fine path.
Theorem 3.3: This approximation for continuous Asian payoffs has $V_\ell = \mathcal{O}(h_\ell^2)$.

Proof: Integrating (15) gives

$$E[(\hat{S}f - \hat{S}c)^2] \leq 4 \left( E[(\hat{S}f - \hat{S}_{KP}f)^2] + E[(\hat{S}c - \hat{S}_{KP}c)^2] + E[(\hat{S}_{KP}f - \hat{S})^2] + E[(\hat{S}_{KP}c - \hat{S})^2] \right),$$

and the proof is completed using the Lipschitz bound and Theorems 2.2 and 3.1.

3.4.2 Treatment 2

The second treatment is the same as the first except that it omits the terms $I^f_n$ and $I^c_n$ and so the averages correspond to trapezoidal integration of the two interpolants, or alternatively they can be viewed as averages of the piecewise linear interpolants $\hat{S}_{PL}(t)$.

Theorem 3.4: This approximation for continuous Asian options has $V_\ell = \mathcal{O}(h_\ell^2)$.

Proof: The difference between the averages of the Brownian and piecewise linear interpolants is

$$T^{-1} \int_0^T \hat{S}(t) - \hat{S}_{PL}(t) \, dt = T^{-1} \sum_n b_n I_n.$$

Since the $I_n$ are iid $N(0, \frac{1}{12} h_\ell^2)$ variables, it follows that

$$E \left[ (\hat{S} - \hat{S}_{PL})^2 \right] = \frac{1}{12} T^{-2} h_\ell^3 \sum_n E[b_n^2],$$

and this is $\mathcal{O}(h_\ell^2)$ due to the finite bound for $E[\max_n b_n^2]$.

Since

$$\hat{S}_{PL}^f - \hat{S}_{PL}^c = (\hat{S}^f - \hat{S}^c) - (\hat{S}^f - \hat{S}_{PL}^f) + (\hat{S}^c - \hat{S}_{PL}^c),$$

it follows that

$$E[(\hat{S}_{PL}^f - \hat{S}_{PL}^c)^2] \leq 3 \left( E[(\hat{S}f - \hat{S}c)^2] + E[(\hat{S}^f - \hat{S}_{PL}^f)^2] + E[(\hat{S}^c - \hat{S}_{PL}^c)^2] \right).$$

The bounds on $E[(\hat{S} - \hat{S}_{PL})^2]$ together with the bound on $E[(\hat{S}^f - \hat{S}c)^2]$ from the proof of Theorem 3.3 prove that $E[(\hat{S}_{PL}^f - \hat{S}_{PL}^c)^2] = \mathcal{O}(h_\ell^2)$, and the result then follows from the assumed Lipschitz property of the payoff.

3.5 Lookback options

In lookback options the payoff is a uniform Lipschitz function of the value of the underlying at maturity $S(T)$, and either the minimum or the maximum of the underlying over the time interval. We will consider cases involving the minimum; the analysis for cases involving the maximum is very similar.
For the fine path simulation, we consider the conditional Brownian interpolation in the time interval \([t_n, t_{n+1}]\) defined by
\[
\hat{S}_n^f(t) = \hat{S}_n^f + \lambda (\hat{S}_{n+1}^f - \hat{S}_n^f) + b_n^f (W(t) - W_n - \lambda (W_{n+1} - W_n))
\]
where \(\lambda = (t - t_n)/h_\ell\) and \(b_n^f = b(\hat{S}_n^f, t_n)\), and make use of Lemma 2.4 to simulate the minimum on the time interval as
\[
\hat{S}_{n,\min}^f = \frac{1}{2} \left( \hat{S}_n^f + \hat{S}_{n+1}^f - \sqrt{\left( \hat{S}_{n+1}^f - \hat{S}_n^f \right)^2 - 2 \left( b_n^f \right)^2 h_\ell \log U_n} \right),
\]
where \(U_n\) is a uniform random variable on the unit interval. Taking the minimum over all timesteps gives the global minimum which is used to compute the fine path value \(\hat{P}_\ell^f\).

For the coarse path value \(\hat{P}_{\ell-1}^c\), we do something slightly different. Using the same conditional Brownian interpolation, for even \(n\) we again use equation (13) to define \(\hat{S}_{n+1}^c\). The minimum value over the interval \([t_n, t_{n+2}]\) can then be taken to be the smaller of the minima for the two intervals \([t_n, t_{n+1}]\) and \([t_{n+1}, t_{n+2}]\),
\[
\hat{S}_{n,\min}^c = \frac{1}{2} \left( \hat{S}_n^c + \hat{S}_{n+1}^c - \sqrt{\left( \hat{S}_{n+1}^c - \hat{S}_n^c \right)^2 - 2 \left( b_n^c \right)^2 h_\ell \log U_n} \right),
\]
\[
\hat{S}_{n+1,\min}^c = \frac{1}{2} \left( \hat{S}_{n+1}^c + \hat{S}_{n+2}^c - \sqrt{\left( \hat{S}_{n+2}^c - \hat{S}_{n+1}^c \right)^2 - 2 \left( b_{n+1}^c \right)^2 h_\ell \log U_{n+1}} \right),
\]
(17)

Here \(b_n^c = b_{n+1}^c = b(\hat{S}_n^c, t_n)\). Note the re-use of the same uniform random numbers \(U_n\) and \(U_{n+1}\) used to compute the fine path minimum. Also, \(\min(\hat{S}_{n,\min}^f, \hat{S}_{n,\min}^c)\) for level \(\ell\) has exactly the same distribution as \(\hat{S}_{n/2,\min}^f\) for level \(\ell-1\), since they are both based on the same approximate Brownian interpolation, and therefore equality (14) is satisfied.

**Theorem 3.5** The multilevel approximation for a lookback option which is a uniform Lipschitz function of \(S(T)\) and \(\inf_{[0,T]} S(t)\) has \(V_\ell = O(h_\ell^2 (\log h_\ell)^2)\).

**Proof** If \(\hat{S}_{n,\min}^f\) and \(\hat{S}_{n,\min}^c\) are the computed minima for the fine and coarse paths, then
\[
\left| \hat{S}_{n,\min}^f - \hat{S}_{n,\min}^c \right| \leq \max_n \left| \hat{S}_{n,\min}^f - \hat{S}_{n,\min}^c \right| \leq \max_n \left| \hat{S}_n^f - \hat{S}_n^c \right| + \max_n \left| \hat{D}_n^f - \hat{D}_n^c \right|,
\]
where
\[
\hat{D}_n^f = \frac{1}{2} \sqrt{\left( \hat{S}_{n+1}^f - \hat{S}_n^f \right)^2 - 2 \left( b_n^f \right)^2 h_\ell \log U_n}
\]
and \(\hat{D}_n^c\) is defined similarly. If \(\hat{D}_n^f\) and \(\hat{D}_n^c\) are both zero, then \(|\hat{D}_n^f - \hat{D}_n^c| = 0\). Otherwise, their sum is strictly positive and, using the inequality \(|x| - |y| \leq |x - y|\),
straightforward manipulations give

\[ \left| \hat{D}_n^f - \hat{D}_n^c \right| = \frac{(\hat{D}_n^f)^2 - (\hat{D}_n^c)^2}{\hat{D}_n^f + \hat{D}_n^c} \]

\[ \leq \frac{(S_{n+1}^f - S_n^f)^2 - (S_{n+1}^c - S_n^c)^2}{4(\hat{D}_n^f + \hat{D}_n^c)} + \frac{|(b_n^f)^2 - (b_n^c)^2|}{2(\hat{D}_n^f + \hat{D}_n^c)} h_\ell |\log U_n| \]

\[ \leq \frac{1}{2} \left| S_{n+1}^f - S_n^f \right| - \left| S_{n+1}^c - S_n^c \right| + \frac{1}{\sqrt{2}} \left| |b_n^f| - |b_n^c| \right| \sqrt{h_\ell |\log U_n|} \]

\[ \leq \frac{1}{2} \left( S_{n+1}^f - S_{n+1}^c + S_n^f - S_n^c \right) + \frac{1}{\sqrt{2}} \left| b_n^f - b_n^c \right| \sqrt{h_\ell |\log U_n|}, \]

and hence

\[ (S_{\min}^f - S_{\min}^c)^2 \leq 8 \max_n \left( S_n^f - S_n^c \right)^2 + h_\ell \left( \max_n (b_n^f - b_n^c)^2 \right) \left( \max_n |\log U_n| \right). \]

When \( n \) is even, assumption A1 gives

\[ (b_n^f - b_n^c)^2 \leq 2K_1^2 \left( S_n^f - S_n^c \right)^2, \]

while for odd \( n \) we have

\[ (b_n^f - b_n^c)^2 = \left( (b_n^f - b_{n-1}^f) + (b_{n-1}^f - b_{n-1}^c) \right)^2 \]

\[ \leq 2K_1^2 \left( S_n^f - S_{n-1}^f \right)^2 + 2K_1^2 \left( S_{n-1}^f - S_{n-1}^c \right)^2. \]

Now,

\[ \hat{S}_n^f - \hat{S}_{n-1}^f = a_{n-1} h_\ell + b_{n-1} \Delta W_{n-1} + \frac{b_n^c - b_{n-1}^c}{2}((\Delta W_{n-1})^2 - h_\ell). \]

Asymptotically, the dominant term on the right is \( b_{n-1} \Delta W_{n-1} \), and it can be proved using the Jensen and Hölder inequalities, the boundedness of \( \mathbb{E}[\max_n b_n^f] \) and Lemma 2.8 that

\[ \mathbb{E} \left[ \max_n (S_n^f - S_{n-1}^f)^2 \right] = O(h_\ell |\log h_\ell|), \]

from which it follows that

\[ \mathbb{E} \left[ \max_n (b_n^f - b_n^c)^2 \right] = O(h_\ell |\log h_\ell|). \]

From Lemma 2.7

\[ \mathbb{E} \left[ \max_n |\log U_n| \right] = O(|\log h_\ell|), \]

and hence,

\[ \mathbb{E} \left[ (S_{\min}^f - S_{\min}^c)^2 \right] = O(h_\ell^2 (\log h_\ell)^2), \]

and the final result then follows from the uniform Lipschitz property of the payoff function and the bound

\[ \max_n \mathbb{E} \left[ (\hat{S}_n^f - \hat{S}_n^c)^2 \right] = O(h_\ell^2). \]
3.6 Extreme paths

The analysis of the variance of the multilevel estimators for barrier and digital options will follow the extreme path approach used in \[11\]. We prepare for this with the following lemma in which we use the notation \( u < h^\alpha \) when \( u > 0 \) and there exists a constant \( c > 0 \) such that \( u < c h^\alpha \), for sufficiently small \( h \). Note that

\[
  u_1 < h^{\alpha_1}, \quad u_2 < h^{\alpha_2} \quad \Rightarrow \quad u_1 + u_2 < h^{\min(\alpha_1, \alpha_2)}, \quad u_1 u_2 < h^{\alpha_1 + \alpha_2}.
\]

Lemma 3.6 For any \( \gamma > 0 \), the probability that a Brownian path \( W(t) \), its increments \( \Delta W_n \equiv W((n+1)h) - W(nh) \), and the corresponding SDE solution \( S(t) \) and its fine (h) and coarse (2h) path approximations \( \hat{S}_f^n \) and \( \hat{S}_c^n \) satisfy any of the following extreme conditions

\[
  \max_n \left( \max(|S(nh)|, |\hat{S}_f^n|, |\hat{S}_c^n|) \right) > h^{-\gamma}
\]

\[
  \max_n \left( \max(|S(nh) - \hat{S}_c^n|, |S(nh) - \hat{S}_f^n|, |\hat{S}_f^n - \hat{S}_c^n|) \right) > h^{1-\gamma}
\]

\[
  \max_n |\Delta W_n| > h^{1/2-\gamma}
\]

\[
  \sup_{[0,T]} |\hat{S}_f(t) - S(t)| > h^{-1-\gamma},
\]

\[
  \sup_{[0,T]} |W(t) - \hat{W}(t)| > h^{1/2-\gamma},
\]

is \( o(h^p) \) for all \( p > 0 \). Here \( \hat{W}(t) \) is defined to be the piecewise linear interpolant of the discrete values \( W_n \).

Furthermore, if none of these extreme conditions is satisfied, and \( \gamma < \frac{1}{2} \), then

\[
  \max_n |\hat{S}_f^n - \hat{S}_c^n| < h^{1/2 - 2\gamma} \quad (18)
\]

\[
  \max_n |b_f^n - b_c^n| < h^{1/2 - 2\gamma} \quad (19)
\]

\[
  \max_n \max(|b_f^n|, |b_c^n|) < h^{-\gamma} \quad (20)
\]

\[
  \max_n |b_f^n - b_c^n| < h^{1/2 - 2\gamma} \quad (21)
\]

where \( b_c^n \) is defined to equal \( b_c^{n-1} \) if \( n \) is odd.

Proof The probability of the first two extreme conditions is \( o(h^p) \) for all \( p > 0 \) due to Theorems 2.1 and 2.2 and Lemma 2.10. Since

\[
  \mathbb{P} \left( \max_n |\Delta W_n| > h^{1/2-\gamma} \right) \leq \sum_n \mathbb{P} \left( |\Delta W_n| > h^{1/2-\gamma} \right),
\]

the probability of the third is \( o(h^p) \) for all \( p > 0 \) due to Lemma 2.10.

The fourth extreme condition has a \( o(h^p) \) probability because Theorems 2.2 and 3.1 together imply a uniform bound as \( h \to 0 \) for

\[
  \mathbb{E} \left[ h^{-m+\gamma/2} \sup_{[0,T]} |\hat{S}_f(t) - S(t)|^m \right],
\]

where \( m \) is a positive integer.
for any \( m > 0 \). Similarly, the fifth is an extreme condition with \( o(h^p) \) probability because of Corollary 2.6.

If none of the extreme conditions is satisfied, then using Assumption A2 gives

\[
|\tilde{S}_{n+1}^f - \tilde{S}_{n}^f| < K_2 h (1 + h^{-\gamma}) + K_2 (1 + h^{-\gamma}) h^{1/2 - \gamma} + \frac{1}{2} K_2 (1 + h^{-\gamma}) h^{1 - 2\gamma} + h
\]

and therefore (18) is satisfied provided \( \gamma < \frac{1}{2} \) so that \( h^{1/2 - 2\gamma} \) is the dominant term in the above inequality.

(19) follows as a consequence because of Assumptions A1 and A3, and (20) is obtained similarly from Assumption A2 and the bound on \( |\tilde{S}_{n}^f| \) and \( |\tilde{S}_{n}^c| \).

When \( n \) is even, the bound in (21) follows from Assumption A1 and the bound on \( |\tilde{S}_{n}^f - \tilde{S}_{n}^c| \), while for odd \( n \) it requires the observation that

\[
|b_{n}^f - b_{n}^c| = |b_{n}^f - b_{n+1}^c| \leq |b_{n}^f - b_{n-1}^f| + |b_{n-1}^f - b_{n-1}^c|.
\]

and the bound then follows from (19) and the corresponding bound for \( n-1 \).

### 3.7 Barrier options

The barrier option which is considered is a down-and-out option for which the payoff is a Lipschitz function of the value of the underlying at maturity, provided the underlying has never dropped below a value \( B \),

\[
P = f(S(T)) \ 1_{T>\tau}.
\]

Here \( 1_{T>\tau} \) is an indicator function taking value 1 if the argument is true, and zero otherwise, and the crossing time \( \tau \) is defined as

\[
\tau = \inf_{t>0} \{ S(t) < B \}.
\]

One approach would be to follow the lookback approximation in computing the minimum of both the fine and coarse paths. However, the variance would be larger in this case because the payoff is a discontinuous function of the minimum. A better treatment, which is the one used in [9], instead computes for each timestep the probability that the minimum of the interpolant crosses the barrier, using the result from Lemma 2.5. This gives the conditional expectation for the payoff, conditional on the discrete Brownian increments of the fine path. For the fine path this gives

\[
\hat{P}_{\ell}^f = f(\tilde{S}_{N}^f) \prod_{n=0}^{N-1} (1 - \hat{p}_{n}^f),
\]

where

\[
\hat{p}_{n}^f = \exp \left(\frac{-2 (\tilde{S}_{n}^f - B)^+ (\tilde{S}_{n+1}^f - B)^+}{(b_{n}^f)^2 h_{\ell}}\right).
\]

The payoff for the coarse path is similarly defined as

\[
\hat{P}_{\ell}^c = f(\tilde{S}_{N}^c) \prod_{n=0}^{N-1} (1 - \hat{p}_{n}^c),
\]

15
where

\[ \hat{p}_n^c = \exp \left( -2 \frac{(\hat{S}_n^c - B)^+ (\hat{S}_{n+1}^c - B)^+}{(b_n^c)^2 h_\ell} \right), \]

and for odd values of \( n \), \( \hat{S}_n^c \) is defined by the usual interpolant and \( b_n^c \equiv b_{n-1}^c \).

Equality (14) is satisfied in this case because

\[ \mathbb{P} \left( \inf_{[t_n,t_{n+2}]} \hat{S}^c(t) > B \mid \hat{S}_n^c, \hat{S}_{n+1}^c \right) = \mathbb{E} \left[ \mathbb{P} \left( \inf_{[t_n,t_{n+2}]} \hat{S}^c(t) > B \mid \hat{S}_n^c, \hat{S}_{n+1}^c, \hat{S}_{n+2}^c \right) \right] \]

where the expectation on the r.h.s. is taken with respect to the distribution of the interpolated value \( \hat{S}_{n+1}^c \), conditional on \( \hat{S}_n^c, \hat{S}_{n+2}^c \).

**Theorem 3.7** Provided \( b_{min} \equiv \inf_{[0,T]} |b(B,t)| > 0 \), and \( \inf\limits_{[0,T]} S(t) \) has a bounded density in the neighbourhood of \( B \), then the multilevel estimator for a down-and-out barrier option has variance \( V_\ell = o(h_\ell^{3/2-\delta}) \) for any \( \delta > 0 \).

**Proof** The proof involves dividing the paths into the following three subsets:

(i) extreme paths;

(ii) paths which are not extreme and for which \( |S_{\min} - B| > h_\ell^{1/2-4\gamma} \) for \( 0 < \gamma < \frac{1}{8} \);

(iii) the rest.

Following the extreme path approach used in [11], we start with

\[ V[\hat{P}_\ell - \hat{P}_{\ell-1}] \leq \mathbb{E}[ (\hat{P}_\ell - \hat{P}_{\ell-1})^2 ] \]

\[ = \mathbb{E}[ (\hat{P}_\ell - \hat{P}_{\ell-1})^2 1_{(i)}] + \mathbb{E}[ (\hat{P}_\ell - \hat{P}_{\ell-1})^2 1_{(ii)}] + \mathbb{E}[ (\hat{P}_\ell - \hat{P}_{\ell-1})^2 1_{(iii)}] \]

where the indicator functions have unit value for paths within the respective subsets. Each of these is considered in turn, and their contributions to \( \mathbb{E}[ (\hat{P}_\ell - \hat{P}_{\ell-1})^2 ] \) are bounded.

(i) Paths are defined to be extreme if they satisfy any of the conditions of Lemma 3.6 for \( 0 < \gamma < \frac{1}{8} \). The Lipschitz bound for the payoff together with the bounds in Theorem 2.2 imply a uniform bound for \( \mathbb{E}[ (\hat{P}_\ell)^4 ] \) and \( \mathbb{E}[ (\hat{P}_{\ell-1})^4 ] \) and therefore also for \( \mathbb{E}[ (\hat{P}_\ell - \hat{P}_{\ell-1})^4 ] \). Hence, by Theorem 2.11 \( \mathbb{E}[ (\hat{P}_\ell - \hat{P}_{\ell-1})^2 1_{(i)}] \) is \( o(h_\ell^p) \) for all \( p > 0 \).

(ii) Suppose that \( S(t) \) attains its minimum at time \( \tau \in [t_n, t_{n+1}] \).

First we consider the case \( S_{\min} < B - h_\ell^{1/2-4\gamma} \). Starting with

\[ |\hat{S}_n^f - S_{\min}| \leq |\hat{S}_n^f - \hat{S}_n^f(\tau)| + |\hat{S}_n^f(\tau) - S(\tau)|, \]

and noting that

\[ \hat{S}_n^f(\tau) - \bar{S}_n^f = \frac{\tau - t_n}{h} (\hat{S}_{n+1}^f - \hat{S}_n^f) + b_n^f(\mathbb{W}(\tau) - \overline{\mathbb{W}(\tau)}), \]

16
we can conclude that \( |\hat{S}_n^f - S_{\text{min}}| < h_{1/2-2\gamma}^\ell \). Hence, for sufficiently small \( h_\ell \), \( |\hat{S}_n^f - S_{\text{min}}| < h_{1/2-2\gamma}^\ell \) and so \( \hat{S}_n^f \) is guaranteed to be less than \( B \). In addition, \( \hat{S}_n^f - \hat{S}_n^c < h_{1-\gamma}^\ell \) and so, for sufficiently small \( h_\ell \), \( \hat{S}_n^c \) is also guaranteed to be less than \( B \) and hence \( \hat{P}_\ell^f - \hat{P}_\ell^c_{-1} = 0 \).

In the alternate case \( S_{\text{min}} > B + h_{1/2-2\gamma}^\ell \), then

\[
\min_n \min(\hat{S}_n^f, \hat{S}_n^c) > B + h_{1/2-2\gamma}^\ell - h_{1-\gamma}^\ell
\]

and since \( h_{1-\gamma}^\ell < h_{1/2-2\gamma}^\ell \) it follows that \( \prod_n (1 - \hat{p}_n^f) \) and \( \prod_n (1 - \hat{p}_n^c) \) are both equal to \( 1 - o(h_\ell^p) \) for all \( p > 0 \), and so \( |\hat{P}_\ell^f - \hat{P}_\ell^c_{-1}| < h_{1-\gamma}^\ell \) due to the Lipschitz condition and the bound on \( \hat{S}_n^f - \hat{S}_n^c \). Hence, the contribution to \( \mathbb{E}[(\hat{P}_\ell^f - \hat{P}_\ell^c_{-1})^2] \) is at most \( O(h_{1/2-2\gamma}^2) \).

(iii) Our first step is to note that if any one of \( \hat{S}_n^f, \hat{S}_{n+1}^f, \hat{S}_n^c, \hat{S}_{n+1}^c \) is greater than \( B + h_{1/2-3\gamma}^\ell \), then the others will be greater than \( B + \frac{1}{2} h_{1/2-3\gamma}^\ell \), when \( h_\ell \) is sufficiently small, since \( |\hat{S}_n^f - \hat{S}_{n+1}^f| < h_{1/2-2\gamma}^\ell \) and \( \max(|\hat{S}_n^c - \hat{S}_{n+1}^c|, |\hat{S}_n^f - \hat{S}_{n+1}^f|) \leq h_{1-\gamma}^\ell \). In this case, \( \hat{p}_n^f \) and \( \hat{p}_n^c \) will both be \( o(h_\ell^p) \), and so

\[
\prod_n (1 - \hat{p}_n^f) = \prod_{n \in R} (1 - \hat{p}_n^f) + o(h_\ell^p),
\]

and

\[
\prod_n (1 - \hat{p}_n^c) = \prod_{n \in R} (1 - \hat{p}_n^c) + o(h_\ell^p),
\]

where \( R \) is the set of indices \( n \) for which none of \( \hat{S}_n^f, \hat{S}_{n+1}^f, \hat{S}_n^c, \hat{S}_{n+1}^c \) is greater than \( B + h_{1/2-3\gamma}^\ell \).

Assume \( n \in R \). We have \( \hat{S}_n^f - \hat{S}_n^c < h_{1-\gamma}^\ell \) and \( \hat{S}_{n+1}^f - \hat{S}_{n+1}^c < h_{1-\gamma}^\ell \) due to the definition of extreme paths, and \( b_n^f - b_n^c < h_{1/2-2\gamma}^\ell \), due to Lemma 3.6. If we now define

\[
X_n^f \equiv \frac{2(\hat{S}_n^f - B)^+(\hat{S}_{n+1}^f - B)^+}{(b_n^f)^2 h_\ell},
\]

\[
X_n^c \equiv \frac{2(\hat{S}_n^c - B)^+(\hat{S}_{n+1}^c - B)^+}{(b_n^c)^2 h_\ell},
\]

then when \( X_n^f \) and \( X_n^c \) are both strictly positive it follows, through the continuity of \( b(S, t) \) and for sufficiently small \( h_\ell \), that \( \min(|b_n^f|, |b_n^c|) > \frac{1}{2} b_{\text{min}} \), and hence through repeated use of the following identity,

\[
f_1 g_1 - f_2 g_2 = \frac{1}{2} (f_1 - f_2)(g_1 + g_2) + \frac{1}{2} (f_1 + f_2)(g_1 - g_2),
\]  

and the fact that \( n \in R \) to bound terms such as \( \hat{S}_n^f - B \), we obtain \( |X_n^f - X_n^c| < h_{1/2-4\gamma}^\ell \), and hence \( |X_n^f - X_n^c| < h_{1/2-5\gamma}^\ell \), for sufficiently small \( h_\ell \). The same bound can also be
achieved in the other cases in which at least one of $X_n^f$ and $X_n^c$ is equal to zero. If we define 
$$\Delta_\ell \equiv 1 - \exp(-h_\ell^{1/2-5\gamma}),$$
then we obtain
$$1 - \hat{p}_n^f = (1 - \hat{p}_n^f) + (\hat{p}_n^f - \hat{p}_n^c) = (1 - \hat{p}_n^f) + \hat{p}_n^f(1 - \exp(X_n^f - X_n^c)) \leq (1 - \hat{p}_n^f) + \hat{p}_n^f\Delta_\ell.$$ 

Since $g(\Delta) \equiv \prod_{n \in R} \left((1 - p_n^f) + p_n^f\Delta\right) - \prod_{n \in R} (1 - p_n^f) - \Delta$ is convex, $g(0) = 0$ and $g(1) = -\prod_{n \in R} (1 - p_n^f) \leq 0$, we conclude that $g(\Delta) \leq 0, \forall \Delta \in [0,1]$. Hence,
$$\prod_{n \in R} (1 - \hat{p}_n^c) \leq \prod_{n \in R} (1 - \hat{p}_n^f) + \Delta_\ell.$$ 

Similarly, $1 - \hat{p}_n^f \leq (1 - \hat{p}_n^f) + \hat{p}_n^c\Delta_\ell$, which leads to
$$\prod_{n \in R} (1 - \hat{p}_n^f) \leq \prod_{n \in R} (1 - \hat{p}_n^c) + \Delta_\ell,$$ 
and therefore
$$\left|\prod_{n \in R} (1 - \hat{p}_n^f) - \prod_{n \in R} (1 - \hat{p}_n^c)\right| \leq \Delta_\ell.$$ 

Returning to the original products over all $n$,
$$\left|\prod_n (1 - \hat{p}_n^f) - \prod_n (1 - \hat{p}_n^c)\right| \leq h_\ell^{1/2-5\gamma}.$$ 

This gives us $\tilde{P}_\ell^f - \tilde{P}_{\ell-1}^f \leq h_\ell^{1/2-6\gamma}$, because of the bound on $f(\tilde{S}_N^f)$ and $f(\tilde{S}_N^c)$, and so the contribution from set (iii) to $E[(\tilde{P}_\ell^f - \tilde{P}_{\ell-1}^c)^2]$ is at most $O(h_\ell^{3/2-16\gamma}).$

The proof is finally completed by choosing $\gamma < \min(\frac{1}{8}, \delta/16)$.

3.8 Digital options

A digital option has a payoff which is a discontinuous function of the value of the underlying asset at maturity, the simplest example being
$$P = 1_{S(T) > K},$$
which has a unit payoff iff $S(T)$ is greater than the strike $K$.

The difficulty with the digital option is that the approach used in section 3.3 will lead to an $O(h_\ell)$ fraction of the paths having coarse and fine path approximations to $S(T)$ on either side of the strike, producing $\tilde{P}_\ell^f - \tilde{P}_{\ell-1}^c = \pm 1$, resulting in $V_\ell = O(h_\ell)$. 

18
To improve the variance to $O(h_{\ell_1}^{3/2-\delta})$ for all $\delta > 0$ we follow the approach which was tested numerically in [9], using the technique of conditional expectation (see section 7.2.3 in [14]).

If $\hat{S}_{N-1}^f$ denotes the value of the fine path approximation one timestep before maturity, then if we approximate the motion thereafter as a simple Brownian motion with constant drift $a_{N-1}^f = a(\hat{S}_{N-1}^f, T - h_{\ell_1})$ and volatility $b_{N-1}^f = b(\hat{S}_{N-1}^f, T - h_{\ell_1})$, the conditional expectation for the payoff is the probability that $\hat{S}_N^f > K$ after one further timestep, which is

$$\hat{P}_{{\ell_1}} = \Phi\left(\frac{\hat{S}_{N-1}^f + a_{N-1}^f h_{\ell_1} - K}{|b_{N-1}^f| \sqrt{h_{\ell_1}}}\right),$$

(24)

where $\Phi$ is the cumulative Normal distribution.

For the coarse path, we note that given the Brownian increment $\Delta W_{N-2}$ for the first half of the last coarse timestep (which comes from the fine path simulation), the probability that $\hat{S}_N^c > K$ is

$$\hat{P}_{{\ell_1}}^c = \Phi\left(\frac{\hat{S}_{N-2}^c + 2a_{N-2}^c h_{\ell_1} + b_{N-2}^c \Delta W_{N-2} - K}{|b_{N-2}^c| \sqrt{h_{\ell_1}}}\right).$$

(25)

The conditional expectation of (25) is equal to the conditional expectation of $\hat{P}_{\ell_1}^c$ defined by (24) on level $\ell_1 - 1$, and so equality (14) is satisfied.

A bound on the variance of the multilevel estimator is given by the following result:

**Theorem 3.8** Provided $b(K, T) \neq 0$, and $S(t)$ has a bounded density in the neighbourhood of $K$, then the multilevel estimator for a digital option has variance $V_{\ell} = o(h_{\ell_1}^{3/2-\delta})$ for any $\delta > 0$.

**Proof** As in the proof of Theorem 3.7, we split the paths into three subsets:

(i) extreme paths;

(ii) paths which are not extreme and for which $|S_N - K| > h_{\ell_1}^{1/2-\gamma}$;

(iii) the rest.

and we analyse the contributions to $\mathbb{E}[(\hat{P}_{{\ell_1}}^f - \hat{P}_{{\ell_1}}^c)^2]$ from all three subsets.

(i) Paths are defined to be extreme if they satisfy any of the conditions of Lemma 3.6 for $0 < \gamma < \frac{1}{2}$, $\mathbb{E}[(Pf)^4]$ and $\mathbb{E}[(Pc)^4]$ are both finite, and hence the contribution of the extreme paths is $o(h_{\ell_1}^p)$, for all $p > 0$.

(ii) If we define $\hat{S}_N^f$ and $\hat{S}_N^c$ to be the values which we would have obtained from the fine and coarse path simulations after the final timestep, then

$$\frac{\hat{S}_{N-1}^f + a_{N-1}^f h_{\ell_1} - K}{|b_{N-1}^f| \sqrt{h_{\ell_1}}} = \frac{\hat{S}_N^f - K}{|b_{N-1}^f| \sqrt{h_{\ell_1}}} - \frac{b_{N-1}^f}{|b_{N-1}^c| \sqrt{h_{\ell_1}}} \left(\Delta W_{N-1} + \frac{1}{2}(b')_{N-1}^f \left((\Delta W_{N-1})^2 - h_{\ell_1}\right)\right),$$

19
and similarly
\[
\frac{\tilde{S}^{f}_{N-2} + 2\tilde{a}^{c}_{N-2}h_{\ell} + b^{c}_{N-2}\Delta W_{N-2} - K}{|b^{c}_{N-2}| \sqrt{h_{\ell}}} = \frac{\tilde{S}^{c}_{N} - K}{|b^{c}_{N-2}| \sqrt{h_{\ell}}} - \frac{b^{c}_{N-2}}{|b^{c}_{N-2}| \sqrt{h_{\ell}}} \left( \Delta W_{N-1} + \frac{1}{2} (b')^{c}_{N-2} \left( (\Delta W_{N-2} + \Delta W_{N-1})^2 - 2h_{\ell} \right) \right).
\]

Since the paths are not extreme, \(|\Delta W_n| \leq h_{\ell}^{1/2-\gamma}\) and \(|S(T) - \tilde{S}^{f}_{N}| \leq h_{\ell}^{1-\gamma}\), and due to Lemma 3.6 \(|b^{f}_{N-1}| < h_{\ell}^{1-\gamma}\). Consequently, if \(S(T) > K + h_{\ell}^{1/2-3\gamma}\) then for sufficiently small \(h_{\ell}\) it follows that
\[
\frac{\tilde{S}^{f}_{N-1} + a^{f}_{N-1}h_{\ell} - K}{|b^{f}_{N-1}| \sqrt{h_{\ell}}} > C h_{\ell}^{-2\gamma},
\]
for some suitably chosen constant \(C\). A similar result follows for the corresponding coarse path, and hence for these paths \(\tilde{P}^{f}_{\ell} - \tilde{P}^{c}_{\ell-1} = o(h_{\ell}^{p})\), for all \(p > 0\). A similar argument applies to the other paths for which \(S(T) < K - h_{\ell}^{1/2-3\gamma}\), and hence \(\mathbb{E}[(\tilde{P}^{f}_{\ell} - \tilde{P}^{c}_{\ell-1})^2 1_{(iii)}] = o(h_{\ell}^{p})\) for all \(p > 0\) and so this contribution is also negligible.

(iii) This subset consists of non-extreme paths for which \(|S(T) - K| \leq h_{\ell}^{1/2-3\gamma}\).

Since
\[
b^{f}_{\ell-1} - b(K, T) = (b^{f}_{N-1} - b^{f}_{N}) + (b^{f}_{N} - b(K, T)),
\]
using Assumption A1 and (19) with \(\gamma < \frac{1}{2}\) we can conclude that for sufficiently small \(h_{\ell}\), \(|b^{f}_{N-1} - b(K, T)| < \frac{1}{2} |b(K, T)|\) and in particular \(b^{f}_{N-1}\) is non-zero and of the same sign as \(b(K, T)\). The same also applies to \(b^{c}_{N-2}\) and hence, exploiting the Lipschitz property \(|\Phi(x_1) - \Phi(x_2)| \leq |x_1 - x_2|\),
\[
\left| \tilde{P}^{f}_{\ell} - \tilde{P}^{c}_{\ell-1} \right| \leq \left| \frac{\tilde{S}^{f}_{N-1} + a^{f}_{N-1}h_{\ell} - K}{|b^{f}_{N-1}| \sqrt{h_{\ell}}} - \frac{\tilde{S}^{c}_{N-2} + 2a^{c}_{N-2}h_{\ell} + b^{c}_{N-2}\Delta W_{N-2} - K}{|b^{c}_{N-2}| \sqrt{h_{\ell}}} \right|
\]
\[
\leq \left| \frac{\tilde{S}^{f}_{N} - K}{|b^{f}_{N-1}| \sqrt{h_{\ell}}} - \frac{\tilde{S}^{c}_{N} - K}{|b^{c}_{N-2}| \sqrt{h_{\ell}}} \right| + \frac{1}{2} K_1 h_{\ell}^{-1/2} \left\{ (\Delta W_{N-1})^2 + (\Delta W_{N-2} + \Delta W_{N-1})^2 + 3 h_{\ell} \right\}.
\]

Using the identity (23) we obtain
\[
\frac{\tilde{S}^{f}_{N} - K}{|b^{f}_{N-1}| \sqrt{h_{\ell}}} - \frac{\tilde{S}^{c}_{N} - K}{|b^{c}_{N-2}| \sqrt{h_{\ell}}} = \frac{1}{2 \sqrt{h_{\ell}}} \left( \frac{1}{|b^{f}_{N-1}|} - \frac{1}{|b^{c}_{N-2}|} \right) \left( \tilde{S}^{f}_{N} - \tilde{S}^{c}_{N} \right) + \frac{1}{2 \sqrt{h_{\ell}}} \left( \tilde{S}^{f}_{N} + \tilde{S}^{c}_{N} - 2K \right) \left( \frac{|b^{c}_{N-2}| - |b^{f}_{N-1}|}{|b^{f}_{N-1}| |b^{c}_{N-2}|} \right)
\]

Using the bounds provided by Lemma 3.6 it follows that
\[
\frac{\tilde{S}^{f}_{N} - K}{|b^{f}_{N-1}| \sqrt{h_{\ell}}} - \frac{\tilde{S}^{c}_{N} - K}{|b^{c}_{N-2}| \sqrt{h_{\ell}}} = O(h_{\ell}^{1/2-5\gamma}),
\]

20
and hence $\hat{P}_\ell - \hat{P}_{\ell-1} = \mathcal{O}(h_\ell^{1/2-5\gamma})$. Since $\mathbb{E}[1_{(iii)}] = \mathcal{O}(h_\ell^{1/2-3\gamma})$ due to the bounded probability density for $S(T)$, it follows that $\mathbb{E}[(\hat{P}_\ell - \hat{P}_{\ell-1})^2 1_{(iii)}] = \mathcal{O}(h_\ell^{3/2-13\gamma})$. Choosing $\gamma < \min(\frac{1}{4}, \delta/13)$ completes the proof.

4 Conclusions and future work

In this paper we have proved that when using the Milstein discretisation for a scalar SDE the variance of the multilevel estimator is $\mathcal{O}(h_\ell^2)$ for Lipschitz and Asian options, $\mathcal{O}(h_\ell^2(\log h_\ell)^2)$ for lookback options, and $\mathcal{o}(h_\ell^{3/2-\delta})$ for barrier and digital options, for any $\delta > 0$.

Condition i) of Theorem 1.1 requires knowledge of the order of weak convergence. Theorems 2.2 and 3.1 together give $\mathcal{O}(h)$ weak convergence for the Lipschitz and Asian options, and $\mathcal{O}(h \log h)$ convergence for the lookback option. For the digital and barrier options, the analysis of the multilevel convergence can be modified to instead consider $\mathbb{E}[\hat{P}_\ell - P]$, and hence it can be proved that the weak order of convergence is $\mathcal{o}(h^{1-\delta})$ for any $\delta > 0$. In all cases, the weak convergence rate satisfies the inequality $\alpha \geq \frac{1}{2}$ required by Theorem 1.1 and so it can be concluded that the computational cost to achieve a r.m.s. accuracy of $\varepsilon$ is $\mathcal{O}(\varepsilon^{-2})$ for all of the cases considered in this paper.

There are lots of directions for future research. One is the extension to jump-diffusion \cite{22} and exponential Lévy processes \cite{21}, to complement existing analyses for Lévy-driven SDEs \cite{19, 5, 6}. Others include the analysis of MLMC applied to the computation of sensitivities, usually referred to as “Greeks” \cite{3, 2}, and also stopping times \cite{20}. For path-dependent options there are possibilities of improving the order of convergence of the multilevel variance using adaptive algorithms \cite{16}, and the complexity can also be improved using quasi-Monte Carlo sampling \cite{13, 8}.

The most natural direction in which to extend the analysis in this paper is to multi-dimensional SDEs. In cases in which a certain commutativity condition is satisfied (see, for example, page 353 in \cite{14}) there is a simple generalisation of the Milstein discretisation which still requires only the increments of the driving Brownian motions. In this case, most of the analysis in this paper will extend quite naturally. The only problem is that if $W_1(t)$ and $W_2(t)$ are two correlated Brownian motions, then we do not have analytic expressions for the joint distribution of their minima and maxima over a given interval, conditional on the end values. This will cause problems in constructing the lookback and barrier estimators if the payoff depends on more than one minimum or maximum.

Clark & Cameron \cite{4} proved that in general it is not possible to achieve first order strong convergence for multi-dimensional SDEs using just the Brownian increments. If the commutativity condition is not satisfied, then the multi-dimensional Milstein discretisation requires the simulation of iterated Itô integrals known as Lévy areas to achieve $\mathcal{O}(h)$ strong convergence. If these terms are omitted, the strong convergence remains $\mathcal{O}(h^{1/2})$, but a new paper \cite{12} proves that it is still possible to achieve a good multilevel estimator through the use of antithetic variates. The variance convergence is $\mathcal{O}(h^2)$ for smooth European (and Asian) payoffs, and $\mathcal{O}(h^{3/2})$ for Lipschitz European (and Asian) payoffs such as put and call options which are smooth almost everywhere.
Extending this analysis to digital, lookback and barrier options will be a challenge for the future. The optimal treatment is likely to require sub-sampling of the Brownian paths within each timestep to approximate the Lévy areas, with the level of sub-sampling being a tradeoff between the cost and accuracy of the simulated Lévy areas.

Acknowledgements: The first author is grateful to Sylvestre Burgos, Lukasz Szpruch and Yuan Xia for their comments on the paper.

References

[1] Avikainen, R.: On irregular functionals of SDEs and the Euler scheme. Finance and Stochastics 13(3), 381–401 (2009)
[2] Burgos, S.: (in preparation). DPhil thesis, University of Oxford (2013)
[3] Burgos, S., Giles, M.: Computing Greeks using multilevel path simulation. In: L. Plaskota, H. Woźniakowski (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2010, pp. 281–296. Springer-Verlag (2012)
[4] Clark, J., Cameron, R.: The maximum rate of convergence of discrete approximations for stochastic differential equations. In: B. Grigelionis (ed.) Stochastic Differential Equations, no. 25 in Lecture Notes in Control and Information Sciences. Springer-Verlag (1980)
[5] Dereich, S.: Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. Annals of Applied Probability 21(1), 283–311 (2011)
[6] Dereich, S., Heidenreich, F.: A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. Stochastic Processes and their Applications 121(7), 1565–1587 (2011)
[7] Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events: for Insurance and Finance. Springer (2008)
[8] Gerstner, T., Noll, M.: Randomized multilevel quasi-Monte Carlo path simulation. World Scientific (2013)
[9] Giles, M.: Improved multilevel Monte Carlo convergence using the Milstein scheme. In: A. Keller, S. Heinrich, H. Niederreiter (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2006, pp. 343–358. Springer-Verlag (2008)
[10] Giles, M.: Multilevel Monte Carlo path simulation. Operations Research 56(3), 607–617 (2008)
[11] Giles, M., Higham, D., Mao, X.: Analysing multilevel Monte Carlo for options with non-globally Lipschitz payoff. Finance and Stochastics 13(3), 403–413 (2009)
[12] Giles, M., Szpruch, L.: Antithetic multilevel Monte Carlo estimation for multidimensional SDEs without Lévy area simulation. http://arxiv.org/pdf/1202.6283 (2012)
[13] Giles, M., Waterhouse, B.: Multilevel quasi-Monte Carlo path simulation. In: Advanced Financial Modelling, Radon Series on Computational and Applied Mathematics, pp. 165–181. de Gruyter (2009)
[14] Glasserman, P.: Monte Carlo Methods in Financial Engineering. Springer, New York (2004)

[15] Heinrich, S.: Multilevel Monte Carlo Methods, Lecture Notes in Computer Science, vol. 2179, pp. 58–67. Springer-Verlag (2001)

[16] Hoel, H., Schwerin, E.v., Szepessy, A., Tempone, R.: Adaptive multilevel Monte Carlo simulation. Numerical Analysis of Multiscale Computations 82, 217–234 (2012)

[17] Kebaier, A.: Statistical Romberg extrapolation: a new variance reduction method and applications to options pricing. Annals of Applied Probability 14(4), 2681–2705 (2005)

[18] Kloeden, P., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1992)

[19] Marxen, H.: The multilevel Monte Carlo method used on a Lévy driven SDE. Monte Carlo Methods and Applications 16(2), 167–190 (2010)

[20] Primozic, T.: Estimating expected first passage times using multilevel Monte Carlo algorithm. MSc thesis, University of Oxford (2011)

[21] Xia, Y.: (in preparation). DPhil thesis, University of Oxford (2013)

[22] Xia, Y., Giles, M.: Multilevel path simulation for jump-diffusion SDEs. In: L. Plaskota, H. Woźniakowski (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2010, pp. 695–708. Springer-Verlag (2012)