

Analysis of multilevel Monte Carlo using the Milstein discretisation

Michael B. Giles Kristian Debrabant
mike.giles@maths.ox.ac.uk debrabant@imada.sdu.dk

Andreas Rößler
roessler@math.uni-luebeck.de

September 1, 2018

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. In this paper we analyse its efficiency when using the Milstein discretisation; this has an improved order of strong convergence compared to the standard Euler-Maruyama method, and it is proved that this leads to an improved order of convergence of the variance of the multilevel estimator. Numerical results are also given for basket options to illustrate the relevance of the analysis.

1 Introduction

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

$$dS(t) = a(S(t), t) dt + b(S(t), t) dW(t), \quad 0 \leq t \leq T, \quad (1)$$

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

Using a standard Monte Carlo method with a numerical discretisation with first order weak convergence, to achieve an RMS error of ε would require $O(\varepsilon^{-2})$ independent paths, each with $O(\varepsilon^{-1})$ timesteps, giving a computational complexity which is $O(\varepsilon^{-3})$. However, the multilevel Monte Carlo (MLMC) approach of Giles [4, 3] reduces the cost to $O(\varepsilon^{-2})$ under certain circumstances.

The key identity underlying the method is

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{\ell=1}^L \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]. \quad (2)$$

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 ℓ and $\ell-1$. The quantity $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ can be estimated using N_ℓ independent samples by

$$\widehat{Y}_\ell = N_\ell^{-1} \sum_{i=1}^{N_\ell} \left(\widehat{P}_\ell^{(i)} - \widehat{P}_{\ell-1}^{(i)} \right). \quad (3)$$

Note that the difference $\widehat{P}_\ell^{(i)} - \widehat{P}_{\ell-1}^{(i)}$ comes from two discrete approximations with different timesteps but the same Brownian path; this difference is often small because of the strong convergence properties of the numerical discretisation scheme. The variance of this simple estimator is $\mathbb{V}[\widehat{Y}_\ell] = N_\ell^{-1} V_\ell$ where V_ℓ is the variance of a single sample. It is the convergence of V_ℓ as $\ell \rightarrow \infty$ which is the focus of this paper, because if $V_\ell = O(2^{-\beta\ell})$ with $\beta > 1$ then an overall RMS accuracy of ε can be achieved at a computational cost which is $O(\varepsilon^{-2})$ [4, 6] through selecting an optimal number of samples on each level.

For the MLMC method based on the simple Euler-Maruyama discretisation with a uniform timestep of size h , Giles, Higham and Mao [7] proved that $V_\ell = O(h)$ 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 $V_\ell = o(h^{1/2-\delta})$, for any $\delta > 0$, for barrier options (in which the payoff is zero if the underlying asset 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 in that case that $V_\ell = O(h^{1/2} \log h)$.

option	Euler-Maruyama		Milstein	
	numerical	analysis	numerical	analysis
Lipschitz	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
Asian	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
lookback	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2(\log h)^2)$
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: Orders of convergence for V_ℓ as observed numerically and proved analytically for both the Euler-Maruyama and Milstein discretisations; δ can be any strictly positive constant.

As summarised in Table 1, numerical simulations [4] suggest that all of these results are near-optimal.

For the MLMC method based on the Milstein discretisation, numerical simulations [3] suggest that $V_\ell = O(h^2)$ for European options with a uniform Lipschitz payoff and for Asian and lookback options, and $V_\ell = O(h^{3/2})$ 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.

The numerical analysis will be performed for scalar SDEs, but we begin the paper by presenting numerical results for basket options, in which each of the underlying assets has a drift and volatility which does not depend on the value of the other assets. Under these conditions, the Milstein discretisation can be applied to each asset individually, avoiding the need to simulate Lévy areas as required in general for multi-dimensional SDEs.

2 Numerical results for basket options

2.1 Milstein discretisation

The Milstein discretisation of equation (1) for a scalar SDE using a uniform timestep h is

$$\widehat{S}_{n+1} = \widehat{S}_n + a_n h + b_n \Delta W_n + \frac{1}{2} b'_n b_n ((\Delta W_n)^2 - h) \quad (4)$$

In the above equation, the subscript n is used to denote the timestep index with $t_n \equiv nh$ and a_n, b_n, b'_n correspond to $a, b, \partial b / \partial S$ evaluated at \widehat{S}_n, t_n . In

addition, ΔW_n is the Brownian path increment $W_{n+1} - W_n$, where $W_n \equiv W(t_n)$. Kloeden & Platen [10] proved that under certain conditions, which will be defined later, the Milstein scheme gives $O(h)$ strong convergence, and for a Lipschitz European payoff this immediately leads to the result that $V_\ell = O(h_\ell^2)$. This remains true for a put or call basket option based on the arithmetic average of several underlying assets, each of which is simulated using the Milstein discretisation.

Reference [3] developed and tested MLMC treatments for the tougher challenges of Asian, lookback, barrier and digital options based on a single underlying asset; it is the numerical analysis of those treatments which is the focus of this paper. The key to the Asian, lookback and barrier option constructions is a conditional piecewise Brownian interpolation. Within the time interval $[t_n, t_{n+1}]$ we approximate the drift and volatility as being constant and use a Brownian interpolation conditional on the two end values \widehat{S}_n and \widehat{S}_{n+1} , giving

$$\widehat{S}(t) = \widehat{S}_n + \lambda(t) (\widehat{S}_{n+1} - \widehat{S}_n) + b_n \left(W(t) - W_n - \lambda(t) (W_{n+1} - W_n) \right), \quad (5)$$

where $\lambda(t) = (t - t_n)/h$. Standard results for the distribution of the extrema and averages of Brownian motions [9] can then be used to construct suitable multilevel estimators [3]; more details are given later in the numerical analysis section.

These treatments for a single underlying asset extend very naturally to basket options. If the value is dependent on a weighted average of J underlying assets, $\overline{S}(t) = \sum_{j=1}^J \mu_j S_j(t)$, each of which satisfies an SDE of the form (1) driven by Brownian motions $W_j(t)$ with correlation matrix Σ , then the important observation is that the weighted average of the Brownian interpolations for the J underlying assets gives

$$\begin{aligned} \overline{S}(t) &= \overline{S}_n + \lambda(t) (\overline{S}_{n+1} - \overline{S}_n) \\ &\quad + \sum_{j=1}^J \mu_j b_{j,n} \left(W_j(t) - W_{j,n} - \lambda(t) (W_{j,n+1} - W_{j,n}) \right) \\ &= \overline{S}_n + \lambda(t) (\overline{S}_{n+1} - \overline{S}_n) + \overline{b}_n \left(W(t) - W_n - \lambda(t) (W_{n+1} - W_n) \right) \end{aligned}$$

where $W(t)$ is another scalar Brownian motion which is a weighted combination of the $W_j(t)$, and \overline{b}_n is defined by

$$\overline{b}_n^2 = \sum_{i,j} \mu_i b_{i,n} \Sigma_{i,j} \mu_j b_{j,n}.$$

Since the Brownian interpolation for the basket average has the same form as the scalar interpolation, the multilevel estimators can be constructed in exactly the same way as in [3], but using \bar{b}_n .

Similarly, for the digital option which has a discontinuous payoff based on a single underlying, one can use a constant coefficient Brownian extrapolation conditional on the value \widehat{S}_{N-1} , one timestep before the end. Following an approach used for payoff smoothing for pathwise sensitivity analysis [9], the conditional expectation for the payoff can be evaluated analytically and this is then used to construct the multilevel estimator [3]. This treatment also extends naturally to the basket case by considering the weighted average of the J conditional extrapolations.

2.2 Numerical results

The numerical results to be presented are for a basket of five assets, each modelled as a geometric Brownian motion:

$$dS_j(t) = r S_j(t) dt + \sigma_j S_j(t) dW_j(t), \quad 0 < t < T,$$

using a constant risk-free interest rate $r = 0.05$, and five volatilities $\sigma = 0.2, 0.25, 0.3, 0.35, 0.4$. The initial asset values are $S_j(0) = 100$, the simulation interval is taken to be $T = 1$, and the driving Brownian motions have a correlation of 0.25. In each case, the option price is based on a simple arithmetic average of the five assets.

These tests are based on ones performed previously in [5], but they have been updated to use the latest version of the high-level MLMC driver software [6] which estimates the rates of convergence of the weak error and the variance V_ℓ , and therefore determines (approximately) the optimal number of levels L and the number of samples N_ℓ on each level.¹

2.2.1 Asian Option

The Asian basket option has discounted payoff $P = \exp(-rT) \max\left(0, \overline{\bar{S}} - K\right)$, where $\overline{\bar{S}}$ is the time-average of the average of the underlying assets, and the strike is $K = 100$.

¹These numerical tests are included in the MATLAB software available online at <http://people.maths.ox.ac.uk/gilesm/mlmc/> and an archived version is kept on <http://people.maths.ox.ac.uk/gilesm/mlmc.html>

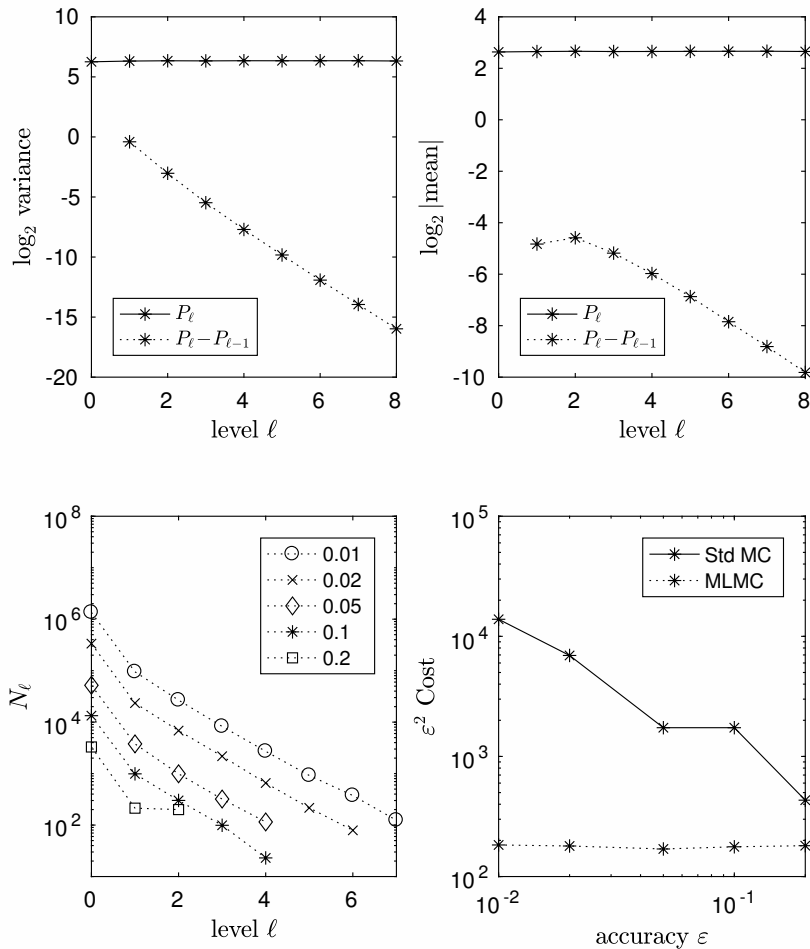


Figure 1: Asian option

The top left plot in Figure 1 shows the behaviour of the variance of both \widehat{P}_ℓ and $\widehat{P}_\ell - \widehat{P}_{\ell-1}$. The slope of the latter is approaching a value approximately equal to -2 , indicating that $V_\ell = O(h_\ell^2)$. On level $\ell = 2$, which has just 4 timesteps, V_ℓ is already almost 1000 times smaller than the variance $\mathbb{V}[\widehat{P}_\ell]$ of the standard Monte Carlo method with the same timestep. The top right plot shows that $|\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]|$ is approximately $O(h_\ell)$, corresponding to first order weak convergence. This is used to determine the number of levels that are required to reduce the bias to an acceptable level [3].

The bottom two plots have results from five multilevel calculations for different values of ε . Each line in the bottom left plot shows the values for

$N_\ell, \ell = 0, \dots, L$, with the values decreasing with ℓ because of the decrease in both V_ℓ and h_ℓ . It can also be seen that the value for L , the maximum level of timestep refinement, increases as the value for ε decreases, requiring a lower bias error [3]. The bottom right plot shows the variation with ε of $\varepsilon^2 C$ where the computational complexity C is defined as $C = \sum_{\ell} 2^\ell N_\ell$, which is the total number of fine grid timesteps on all levels. One line shows the results for the multilevel calculation and the other shows the corresponding cost of a standard Monte Carlo simulation of the same accuracy, i.e. the same bias error corresponding to the same value for L , and the same variance. It can be seen that $\varepsilon^2 C$ is almost constant for the multilevel method, as expected, whereas for the standard Monte Carlo method it increases with L . For the most accurate case, $\varepsilon = 0.01$, the multilevel method is approximately 100 times more efficient than the standard method.

2.2.2 Lookback Option

The basket lookback option we consider has the discounted payoff $P = \exp(-rT) \left(\overline{S}(T) - \min_{0 < t < T} \overline{S}(t) \right)$.

The top left plot in Figure 2 shows that the variance is $O(h_\ell^2)$, while the top right plot shows that the mean correction is $O(h_\ell)$. The bottom left plot shows that more levels are required to reduce the discretisation bias to the required level. Consequently, the savings relative to the standard Monte Carlo treatment are greater, up to a factor of approximately 200 for $\varepsilon = 0.01$. The computational cost of the multilevel method is almost perfectly proportional to ε^{-2} .

2.2.3 Barrier Option

The barrier option which is considered is a down-and-out call with payoff $P = \exp(-rT) (\overline{S}(T) - K)^+ \mathbf{1}_{\tau > T}$, where the notation $(x)^+$ denotes $\max(0, x)$, $\mathbf{1}_{\tau > T}$ is an indicator function taking value 1 if the argument is true, and zero otherwise, and the crossing time τ is defined as $\tau = \inf_{t > 0} \{ \overline{S}(t) < B \}$. The barrier value is taken to be $B = 85$, and the strike is again $K = 100$.

The top left plot in Figure 3 shows that the variance is approximately $O(h_\ell^{3/2})$. The reason for this is that an $O(h_\ell^{1/2})$ fraction of the paths have a minimum which lies within $O(h_\ell^{1/2})$ of the barrier. It will be proved that for

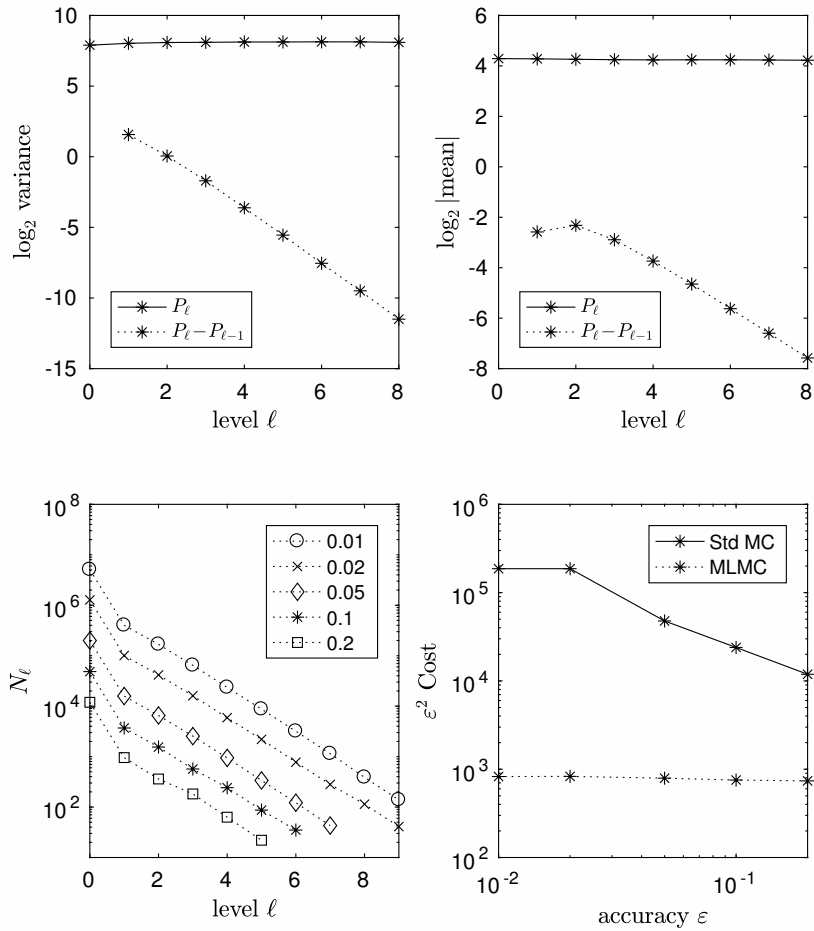


Figure 2: Lookback option

these paths the difference between the coarse and fine path payoff values is $O(h_\ell^{1/2})$, giving a contribution to the overall variance which is $O(h_\ell^{3/2})$.

The top right plot shows that the mean correction is $O(h_\ell)$, corresponding to first order weak convergence. The bottom right plot shows that the computational cost of the multilevel method is again almost perfectly proportional to ε^{-2} , and for $\varepsilon=0.01$ it is 200 times more efficient than the standard Monte Carlo method.

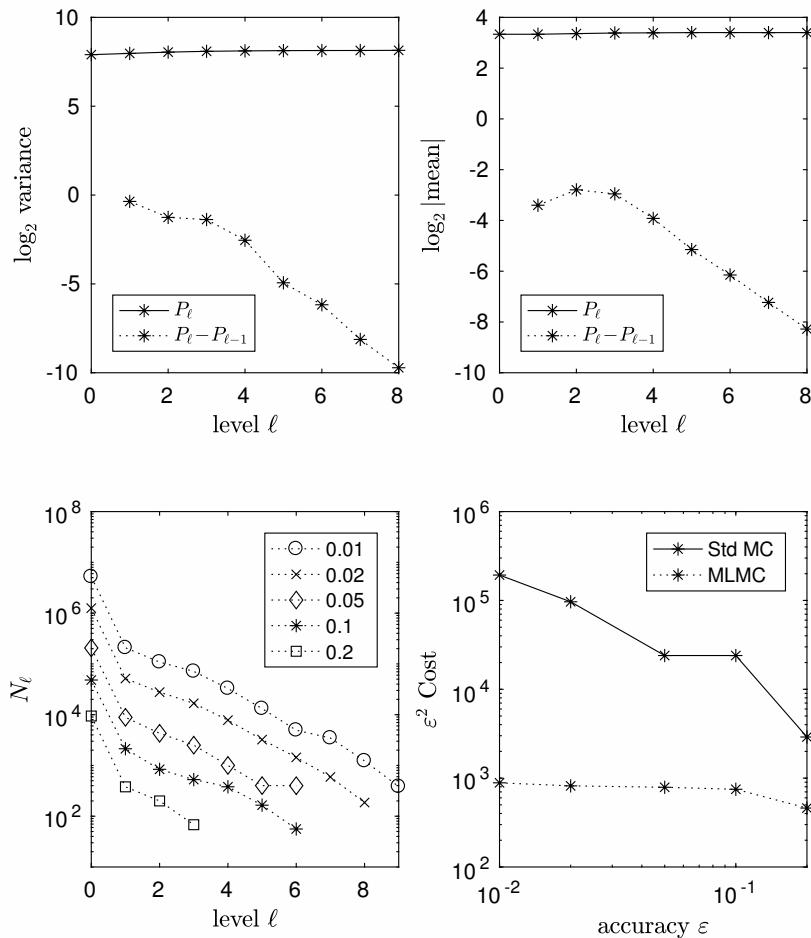


Figure 3: Barrier option

2.2.4 Digital Option

The digital option has the discounted payoff $P = \exp(-rT) K \mathbf{1}_{S(T) > K}$ with strike $K = 100$.

The top left plot in Figure 4 shows that the variance is approximately $O(h_\ell^{3/2})$. The reason for this is similar to the argument for the barrier option. $O(h_\ell^{1/2})$ of the paths have a minimum which lies within $O(h_\ell^{1/2})$ of the strike. The fine path and coarse path trajectories differ by $O(h_\ell)$, due to the first order strong convergence of the Milstein scheme and it will be proved that this results in an $O(h_\ell^{1/2})$ difference between the coarse and fine path evaluations.

One strikingly different feature is that the variance of the level 0 estimator

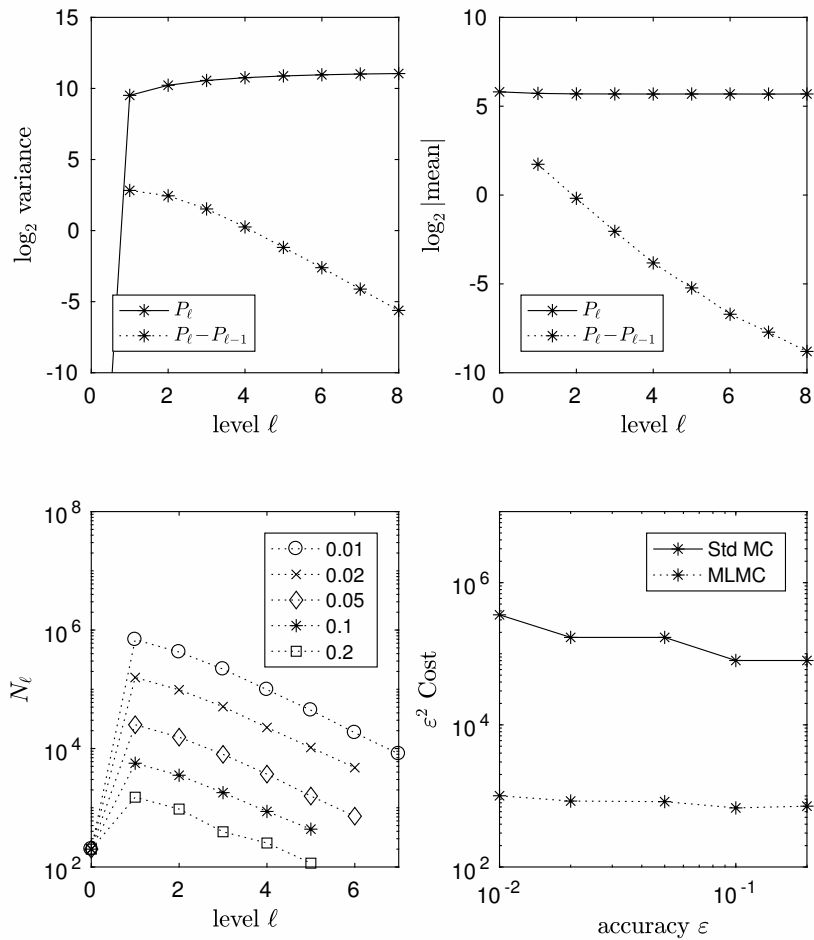


Figure 4: Digital option

is zero. This is because the multilevel treatment introduced in [3] uses a conditional expectation (based on a simple Brownian extrapolation for which the expectation is known analytically) evaluated one timestep before the end. At level $l = 0$ where there would usually be one timestep, there is no path simulation at all; one simply uses the analytic expression for the conditional expectation. This reduces the cost of the multilevel calculations even more than usual, giving more than a factor of 400 computational savings for $\epsilon = 0.01$.

3 Numerical analysis

For clarity, the numerical analysis is now presented for financial options based on a single underlying asset, but for the reasons already discussed the numerical analysis extends immediately to basket options.

3.1 Preliminaries

The analysis builds on a large body of existing results which are included here for completeness.

3.1.1 Solution of the SDE and its Milstein discretisation

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

- 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

$$\begin{aligned} |a(x, t)| + |L_0 a(x, t)| + |L_1 a(x, t)| + |b(x, t)| + |L_0 b(x, t)| \\ + |L_1 b(x, t)| + |L_0 L_1 b(x, t)| + |L_1 L_1 b(x, t)| \leq K_2 (1 + |x|) \end{aligned}$$

- 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 [10].

Theorem 3.1. *Provided the assumptions A1-A3 are satisfied, then for all positive integers m , $\mathbb{E} \left[\sup_{0 < t < T} |S(t)|^m \right] < \infty$.*

Kloeden & Platen [10] define the following continuous time interpolant of the Milstein discretisation in (4) for $t_n \leq t \leq t_{n+1}$,

$$\begin{aligned} \widehat{S}_{KP}(t) &= \widehat{S}_n + a_n (t - t_n) + b_n (W(t) - W_n) \\ &\quad + \frac{1}{2} b'_n b_n ((W(t) - W_n)^2 - (t - t_n)), \end{aligned} \quad (6)$$

and prove the following result.

Theorem 3.2. *Provided the assumptions A1-A3 are satisfied, then for all positive integers m there exists a constant C_m such that*

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

3.1.2 Brownian bridge results

If the drift a and volatility b are constant, the scalar 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

$$S(t) = S_n + \lambda(t) (S_{n+1} - S_n) + b (W(t) - W_n - \lambda(t) (W_{n+1} - W_n)) \quad (7)$$

where again $\lambda(t) \equiv (t - t_n)/h$. 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 [9]).

Lemma 3.3. *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*

$$\int_{t_n}^{t_{n+1}} S(t) dt = \frac{1}{2} h (S_n + S_{n+1}) + b I_n \quad (8)$$

where

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

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

Lemma 3.4. *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*

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

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

Lemma 3.5. *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*

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

Corollary 3.6. *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 .

3.1.3 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 [2].

Lemma 3.7. *If $U_n, n = 1, \dots, N$ are independent samples from a uniform distribution on the unit interval $[0, 1]$, then for any positive integer m*

$$\mathbb{E} \left[\max_n |\log U_n|^m \right] = O((\log N)^m), \text{ as } N \rightarrow \infty. \quad (9)$$

Lemma 3.8. *If $Z_n, n = 1, \dots, N$ are independent samples from a standard Normal distribution, then for any positive integer m*

$$\mathbb{E} \left[\max_n |Z_n|^m \right] = O((\log N)^{m/2}), \text{ as } N \rightarrow \infty. \quad (10)$$

Corollary 3.9. *If $W_n(t), n = 1, \dots, N$ are independent Brownian paths on $[0, 1]$, conditional on $W_n(0) = W_n(1) = 0$, then for any positive integer m*

$$\mathbb{E} \left[\max_n \sup_{[0,1]} |W_n(t)|^m \right] = O((\log N)^{m/2}), \text{ as } N \rightarrow \infty. \quad (11)$$

Proof. From Corollary 3.6, for sufficiently large x the tail probability for $|W_n(t)|$ is less than that of a standard Normal random variable. \square

3.1.4 Extreme paths

Some of the proofs in [7] 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 3.10. *If X_ℓ is a scalar random variable defined on level ℓ of the multilevel analysis, and for each positive integer m , $\mathbb{E}[|X_\ell|^m]$ is uniformly bounded, then, for any $\delta > 0$,*

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

Proof. Follows immediately from Markov’s inequality

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

by choosing $m > p/\delta$. \square

Lemma 3.11. *If Y_ℓ is a scalar random variable on level ℓ , $\mathbb{E}[Y_\ell^2]$ is uniformly bounded, and for each $p > 0$, the indicator function $\mathbf{1}_{E_\ell}$ on level ℓ (which takes value 1 or 0 depending whether or not a path lies within some set E_ℓ) satisfies*

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

then for each $p > 0$,

$$\mathbb{E}[|Y_\ell| \mathbf{1}_{E_\ell}] = o(h_\ell^p).$$

Proof. Immediate consequence of Hölder inequality which gives

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

□

In the proofs in the main analysis, Lemma 3.10 will be used to establish the pre-conditions for Lemma 3.11, 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.

3.2 Analysis of the Milstein MLMC method

3.2.1 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, $\widehat{S}_n \equiv \widehat{S}(t_n)$ where $t_n = nh$. By approximating the drift and volatility as being constant within each timestep, we define the following Brownian interpolation based on equation (7),

$$\widehat{S}(t) = \widehat{S}_n + \lambda(t) (\widehat{S}_{n+1} - \widehat{S}_n) + b_n \left(W(t) - W_n - \lambda(t) (W_{n+1} - W_n) \right) \quad (12)$$

where again $\lambda(t) \equiv (t - t_n)/h$. The advantage of this interpolation compared to the standard Kloeden-Platen interpolant is that we can use Lemmas 3.3 – 3.5 in constructing the multilevel estimators. The accuracy of the interpolant relative to the Kloeden-Platen interpolant is given by the following theorem:

Theorem 3.12. *If $\widehat{S}(t)$ is the interpolant defined by (12) and $\widehat{S}_{KP}(t)$ is the Kloeden-Platen interpolant defined by (6) then for any positive integer m*

i)

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

ii)

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

iii)

$$\mathbb{E} \left[\left(\int_0^T (\widehat{S}(t) - \widehat{S}_{KP}(t)) dt \right)^2 \right] = O(h^3).$$

Proof. In each case we use the fact that $\mathbb{E}[\max_n |b'_n b_n|^m]$ is finite for all positive integers m due to Theorem 3.2 and Assumption A2. In addition, for $t \in [t_n, t_{n+1}]$, the difference between the two interpolants is

$$\widehat{S}(t) - \widehat{S}_{KP}(t) = \frac{1}{2} b'_n b_n Y(t),$$

where

$$\begin{aligned} Y(t) &= \lambda(t) (W_{n+1} - W_n)^2 - (W(t) - W_n)^2 \\ &= \lambda(t) (1 - \lambda(t)) (W_{n+1} - W_n)^2 - (W(t) - W_n - \lambda(t) (W_{n+1} - W_n))^2 \\ &\quad - 2 \lambda(t) (W_{n+1} - W_n) (W(t) - W_n - \lambda(t) (W_{n+1} - W_n)). \end{aligned}$$

i) Using Hölder's inequality, the assertion follows from

$$\mathbb{E} \left[\sup_{[0,T]} \left| \widehat{S}(t) - \widehat{S}_{KP}(t) \right|^m \right] \leq 2^{-m} \sqrt{\mathbb{E} \left[\max_n |b'_n b_n|^{2m} \right] \mathbb{E} \left[\sup_{[0,T]} |Y(t)|^{2m} \right]},$$

together with bounds on $\mathbb{E} \left[\sup_{[0,T]} |Y(t)|^{2m} \right]$ coming from Lemma 3.8 and Corollary 3.9.

ii) By setting $W(t) - W_n = \sqrt{\lambda(t)h} Z_1$ and $W_{n+1} - W(t) = \sqrt{(1-\lambda(t))h} Z_2$, with Z_1, Z_2 independent standard Normal random variables, one can prove that $|Y| \leq h \max(Z_1^2, Z_2^2)$ and hence $|Y|^m \leq h^m (Z_1^{2m} + Z_2^{2m})$. The assertion then follows from

$$\mathbb{E} \left[\left| \widehat{S}(t) - \widehat{S}_{KP}(t) \right|^m \right] = 2^{-m} \mathbb{E}[|b'_n b_n|^m] \mathbb{E}[|Y|^m],$$

and standard results for moments of Normal random variables.

iii) Defining $X_n := \int_{t_n}^{t_{n+1}} Y(t) dt$ we obtain

$$\mathbb{E} \left[\left(\int_0^T (\widehat{S}(t) - \widehat{S}_{KP}(t)) dt \right)^2 \right] = \frac{1}{4} \mathbb{E} \left[\left(\sum_{n=0}^{N-1} b'_n b_n X_n \right)^2 \right].$$

For $n > m$, $\mathbb{E}[b'_m b_m X_m b'_n b_n X_n] = 0$ since X_n is independent of $b'_m b_m X_m b'_n b_n$ and $\mathbb{E}[X_n] = 0$. In addition, the X_n are iid random variables, and therefore

$$\mathbb{E} \left[\left(\int_0^T (\widehat{S}(t) - \widehat{S}_{KP}(t)) dt \right)^2 \right] = \frac{1}{4} \mathbb{E}[X_0^2] \sum_{n=0}^{N-1} \mathbb{E}[(b'_n b_n)^2].$$

The proof is completed by noting that $\mathbb{E}[X_0^2] = O(h^4)$ due to standard results for moments of Brownian increments.

□

3.2.2 Estimator construction

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

$$\mathbb{E}[\widehat{P}_{\ell-1}^f] = \mathbb{E}[\widehat{P}_{\ell-1}^c]. \quad (13)$$

In the simplest case of a European option, this can be achieved very simply by defining $\widehat{P}_{\ell-1}^f$ and $\widehat{P}_{\ell-1}^c$ to be the same. However, for the other applications the definition of $\widehat{P}_{\ell-1}^c$ involves information from the discrete simulation of \widehat{P}_ℓ^f , which is not available in computing $\widehat{P}_{\ell-1}^c$. This is done to reduce the variance of the estimator, but it must be shown that equality (13) 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,

$$\begin{aligned} \widehat{S}^f(t) - \widehat{S}^c(t) &= (\widehat{S}^f(t) - \widehat{S}_{KP}^f(t)) - (\widehat{S}^c(t) - \widehat{S}_{KP}^c(t)) \\ &\quad + (\widehat{S}_{KP}^f(t) - S(t)) - (\widehat{S}_{KP}^c(t) - S(t)) \end{aligned} \quad (14)$$

with Theorem 3.12 bounding the error in the first two terms, and Theorem 3.2 bounding the error in the last two terms.

3.2.3 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 , $P = f(S(T_1), S(T_2), \dots, S(T_M))$, with the Lipschitz bound

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

for some constant L . In the numerical discretisation the fine and coarse path payoffs are both defined by $\hat{P} = f(\hat{S}(T_1), \hat{S}(T_2), \dots, \hat{S}(T_M))$, 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 for the variance of the multilevel estimator:

Theorem 3.13. *For Lipschitz payoffs, $V_\ell = O(h_\ell^2)$.*

Proof. From the Lipschitz bound and Jensen's inequality we obtain

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

The decomposition (14) implies that

$$\begin{aligned} & \mathbb{E}[(\hat{S}^f(T_m) - \hat{S}^c(T_m))^2] \\ & \leq 4 \left(\mathbb{E}[(\hat{S}^f(T_m) - \hat{S}_{KP}^f(T_m))^2] + \mathbb{E}[(\hat{S}^c(T_m) - \hat{S}_{KP}^c(T_m))^2] \right. \\ & \quad \left. + \mathbb{E}[(\hat{S}_{KP}^f(T_m) - S(T_m))^2] + \mathbb{E}[(\hat{S}_{KP}^c(T_m) - S(T_m))^2] \right) \end{aligned}$$

and the proof is completed using Theorems 3.2 and 3.12. \square

3.2.4 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, $S(T)$.

The numerical approximation follows the approach used in [3], in which the fine and coarse path averages \widehat{S} are defined by integrating the interpolant (12). Because of Lemma 3.3, this gives

$$\int_0^T \widehat{S}^f(t) dt = \sum_{n=0}^{N-1} \frac{1}{2} h_\ell \left(\widehat{S}_n^f + \widehat{S}_{n+1}^f \right) + b_n I_n^f$$

where I_n^f 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

$$\begin{aligned} I_n^c &\equiv \int_{t_n}^{t_{n+2}} \left(W(t) - W_n - \frac{t - t_n}{2h_\ell} (W_{n+2} - W_n) \right) dt \\ &= I_n^f + I_{n+1}^f - \frac{1}{2} h_\ell (W_{n+2} - 2W_{n+1} + W_n). \end{aligned}$$

so I_n^c is obtained from the Brownian path data used for the fine path.

Theorem 3.14. *This approximation for continuous Asian payoffs has $V_\ell = O(h_\ell^2)$.*

Proof. Integrating (14) gives

$$\begin{aligned} \mathbb{E}[(\widehat{S}^f - \widehat{S}^c)^2] &\leq 4 \left(\mathbb{E}[(\widehat{S}^f - \widehat{S}_{KP}^f)^2] + \mathbb{E}[(\widehat{S}^c - \widehat{S}_{KP}^c)^2] \right. \\ &\quad \left. + \mathbb{E}[(\widehat{S}_{KP}^f - \bar{S})^2] + \mathbb{E}[(\widehat{S}_{KP}^c - \bar{S})^2] \right), \end{aligned}$$

and the proof is completed using the Lipschitz bound and Theorems 3.2 and 3.12. \square

3.2.5 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

$$\widehat{S}^f(t) = \widehat{S}_n^f + \lambda(t) (\widehat{S}_{n+1}^f - \widehat{S}_n^f) + b_n^f (W(t) - W_n - \lambda(t) (W_{n+1} - W_n))$$

where $\lambda(t) = (t - t_n)/h_\ell$ and $b_n^f \equiv b(\widehat{S}_n^f, t_n)$, and make use of Lemma 3.4 to simulate the minimum on the time interval as

$$\widehat{S}_{n,min}^f = \frac{1}{2} \left(\widehat{S}_n^f + \widehat{S}_{n+1}^f - \sqrt{\left(\widehat{S}_{n+1}^f - \widehat{S}_n^f \right)^2 - 2 (b_n^f)^2 h_\ell \log U_n} \right), \quad (15)$$

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 \widehat{P}_ℓ^f .

For the coarse path value $\widehat{P}_{\ell-1}^c$, we do something slightly different. Using the same conditional Brownian interpolation, for even n we again use equation (12) to define \widehat{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 $[t_{n+1}, t_{n+2}]$,

$$\begin{aligned} \widehat{S}_{n,min}^c &= \frac{1}{2} \left(\widehat{S}_n^c + \widehat{S}_{n+1}^c - \sqrt{\left(\widehat{S}_{n+1}^c - \widehat{S}_n^c \right)^2 - 2 (b_n^c)^2 h_\ell \log U_n} \right), \\ \widehat{S}_{n+1,min}^c &= \frac{1}{2} \left(\widehat{S}_{n+1}^c + \widehat{S}_{n+2}^c - \sqrt{\left(\widehat{S}_{n+2}^c - \widehat{S}_{n+1}^c \right)^2 - 2 (b_{n+1}^c)^2 h_\ell \log U_{n+1}} \right). \end{aligned} \quad (16)$$

Here $b_n^c = b_{n+1}^c \equiv b(\widehat{S}_n^c, t_n)$. Note the re-use of the same uniform random numbers U_n and U_{n+1} used to compute the fine path minimum. Also, $\min(\widehat{S}_{n,min}^c, \widehat{S}_{n+1,min}^c)$ for level ℓ has exactly the same distribution as $\widehat{S}_{n/2,min}^f$ for level $\ell-1$, since they are both based on the same approximate Brownian interpolation, and therefore equality (13) is satisfied.

Theorem 3.15. *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 \widehat{S}_{min}^f and \widehat{S}_{min}^c are the computed minima for the fine and coarse paths, then

$$\left| \widehat{S}_{min}^f - \widehat{S}_{min}^c \right| \leq \max_n \left| \widehat{S}_{n,min}^f - \widehat{S}_{n,min}^c \right| \leq \max_n \left| \widehat{S}_n^f - \widehat{S}_n^c \right| + \max_n \left| \widehat{D}_n^f - \widehat{D}_n^c \right|,$$

where

$$\widehat{D}_n^f = \frac{1}{2} \sqrt{\left(\widehat{S}_{n+1}^f - \widehat{S}_n^f \right)^2 - 2(b_n^f)^2 h_\ell \log U_n}$$

and \widehat{D}_n^c is defined similarly. If \widehat{D}_n^f and \widehat{D}_n^c are both zero, then $|\widehat{D}_n^f - \widehat{D}_n^c| = 0$. Otherwise, their sum is strictly positive and, using the inequality $||x| - |y|| \leq |x - y|$, straightforward manipulations give

$$\begin{aligned} \left| \widehat{D}_n^f - \widehat{D}_n^c \right| &= \frac{\left| (\widehat{D}_n^f)^2 - (\widehat{D}_n^c)^2 \right|}{\widehat{D}_n^f + \widehat{D}_n^c} \\ &\leq \frac{\left| (\widehat{S}_{n+1}^f - \widehat{S}_n^f)^2 - (\widehat{S}_{n+1}^c - \widehat{S}_n^c)^2 \right|}{4(\widehat{D}_n^f + \widehat{D}_n^c)} + \frac{|(b_n^f)^2 - (b_n^c)^2| h_\ell |\log U_n|}{2(\widehat{D}_n^f + \widehat{D}_n^c)} \\ &\leq \frac{1}{2} \left| \left| \widehat{S}_{n+1}^f - \widehat{S}_n^f \right| - \left| \widehat{S}_{n+1}^c - \widehat{S}_n^c \right| \right| + \frac{1}{\sqrt{2}} \left| |b_n^f| - |b_n^c| \right| \sqrt{h_\ell |\log U_n|} \\ &\leq \frac{1}{2} \left(\left| \widehat{S}_{n+1}^f - \widehat{S}_{n+1}^c \right| + \left| \widehat{S}_n^f - \widehat{S}_n^c \right| \right) + \frac{1}{\sqrt{2}} |b_n^f - b_n^c| \sqrt{h_\ell |\log U_n|}, \end{aligned}$$

and hence

$$\left(\widehat{S}_{min}^f - \widehat{S}_{min}^c \right)^2 \leq 8 \max_n \left(\widehat{S}_n^f - \widehat{S}_n^c \right)^2 + h_\ell \max_n (b_n^f - b_n^c)^2 \max_n |\log U_n|.$$

When n is even, assumption A1 gives $(b_n^f - b_n^c)^2 \leq K_1^2 \left(\widehat{S}_n^f - \widehat{S}_n^c \right)^2$, while for odd n we have

$$\begin{aligned} (b_n^f - b_n^c)^2 &= \left((b_n^f - b_{n-1}^f) + (b_{n-1}^f - b_{n-1}^c) \right)^2 \\ &\leq 2K_1^2 \left(\widehat{S}_n^f - \widehat{S}_{n-1}^f \right)^2 + 2K_1^2 \left(\widehat{S}_{n-1}^f - \widehat{S}_{n-1}^c \right)^2. \end{aligned}$$

Now,

$$\widehat{S}_n^f - \widehat{S}_{n-1}^f = a_{n-1} h_\ell + b_{n-1} \Delta W_{n-1} + \frac{1}{2} b'_{n-1} b_{n-1} ((\Delta W_{n-1})^2 - h_\ell).$$

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 $\mathbb{E}[\max_n b_n^4]$ and Lemma 3.8 that

$$\mathbb{E} \left[\max_n (\widehat{S}_n^f - \widehat{S}_{n-1}^f)^2 \right] = O(h_\ell |\log h_\ell|),$$

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 3.7,

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

and hence,

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

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

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

□

3.2.6 Extreme paths

The analysis of the variance of the multilevel estimators for barrier and digital options will follow the extreme path approach used in [7]. We prepare for this with the following lemma in which we use the notation $u \prec h^\alpha$ when $u > 0$ and there exists a constant $c > 0$ such that $u < c h^\alpha$, for sufficiently small h . Note that

$$u_1 \prec h^{\alpha_1}, \quad u_2 \prec h^{\alpha_2} \quad \implies \quad u_1 + u_2 \prec h^{\min(\alpha_1, \alpha_2)}, \quad u_1 u_2 \prec h^{\alpha_1 + \alpha_2}.$$

Lemma 3.16. *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 \widehat{S}_n^f and*

\widehat{S}_n^c satisfy any of the following extreme conditions

$$\begin{aligned} \max_n \left(\max(|S(nh)|, |\widehat{S}_n^f|, |\widehat{S}_n^c|) \right) &> h^{-\gamma} \\ \max_n \left(\max(|S(nh) - \widehat{S}_n^c|, |S(nh) - \widehat{S}_n^f|, |\widehat{S}_n^f - \widehat{S}_n^c|) \right) &> h^{1-\gamma} \\ \max_n |\Delta W_n| &> h^{1/2-\gamma} \\ \sup_{[0,T]} \left| \widehat{S}^f(t) - S(t) \right| &> h^{1-\gamma}, \\ \sup_{[0,T]} |W(t) - \overline{W}(t)| &> h^{1/2-\gamma}, \end{aligned}$$

is $o(h^p)$ for all $p > 0$. Here $\overline{W}(t)$ 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

$$\max_n |\widehat{S}_n^f - \widehat{S}_{n-1}^f| \prec h^{1/2-2\gamma} \quad (17)$$

$$\max_n |b_n^f - b_{n-1}^f| \prec h^{1/2-2\gamma} \quad (18)$$

$$\max_n \max(|b_n^f|, |b_n^c|) \prec h^{-\gamma} \quad (19)$$

$$\max_n |b_n^f - b_n^c| \prec h^{1/2-2\gamma} \quad (20)$$

where b_n^c is defined to equal b_{n-1}^c if n is odd.

Proof. The probability of the first two extreme conditions is $o(h^p)$ for all $p > 0$ due to Theorems 3.1 and 3.2 and Lemma 3.10. Since

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

the probability of the third is $o(h^p)$ for all $p > 0$ due to Lemma 3.10.

The fourth extreme condition has a $o(h^p)$ probability because Theorems 3.2 and 3.12 together imply a uniform bound as $h \rightarrow 0$ for

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

for any $m > 0$. Similarly, the fifth is an extreme condition with $o(h^p)$ probability because of Corollary 3.6.

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

$$|\widehat{S}_{n+1}^f - \widehat{S}_n^f| < K_2 h(1+h^{-\gamma}) + K_2(1+h^{-\gamma})h^{1/2-\gamma} + \frac{1}{2}K_2(1+h^{-\gamma})(h^{1-2\gamma}+h)$$

and therefore (17) is satisfied provided $\gamma < \frac{1}{2}$ so that $h^{1/2-2\gamma}$ is the dominant term in the above inequality.

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

When n is even, the bound in (20) follows from Assumption A1 and the bound on $|\widehat{S}_n^f - \widehat{S}_n^c|$, 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 (18) and the corresponding bound for $n-1$. \square

3.2.7 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 ,

$$P = f(S(T)) \mathbf{1}_{\tau > T},$$

with the crossing time τ 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 [3], instead computes for each timestep the probability that the minimum of the interpolant crosses the barrier, using the result from Lemma 3.5. This gives the conditional expectation for the payoff, conditional on the discrete Brownian increments of the fine path. For the fine path this gives

$$\widehat{P}_\ell^f = f(\widehat{S}_N^f) \prod_{n=0}^{N-1} (1 - \widehat{p}_n^f),$$

where

$$\widehat{P}_n^f = \exp \left(\frac{-2 (\widehat{S}_n^f - B)^+ (\widehat{S}_{n+1}^f - B)^+}{(b_n^f)^2 h_\ell} \right).$$

The payoff for the coarse path is similarly defined as

$$\widehat{P}_\ell^c = f(\widehat{S}_N^c) \prod_{n=0}^{N-1} (1 - \widehat{p}_n^c),$$

where

$$\widehat{p}_n^c = \exp \left(\frac{-2 (\widehat{S}_n^c - B)^+ (\widehat{S}_{n+1}^c - B)^+}{(b_n^c)^2 h_\ell} \right), \quad (21)$$

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

Equality (13) is satisfied in this case because

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

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

Theorem 3.17. *Provided $b_{min} \equiv \inf_{[0, T]} |b(B, t)| > 0$, and $\inf_{[0, T]} S(t)$ 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 [7], we start with

$$\begin{aligned} & \mathbb{V}[\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c] \\ & \leq \mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^2] \\ & = \mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^2 \mathbf{1}_{(i)}] + \mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^2 \mathbf{1}_{(ii)}] + \mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^2 \mathbf{1}_{(iii)}] \end{aligned}$$

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

(i) Paths are defined to be extreme if they satisfy any of the conditions of Lemma 3.16 for $0 < \gamma < \frac{1}{8}$. The Lipschitz bound for the payoff together with the bounds in Theorem 3.2 imply a uniform bound for $\mathbb{E}[(\widehat{P}_\ell^f)^4]$ and $\mathbb{E}[(\widehat{P}_{\ell-1}^c)^4]$ and therefore also for $\mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^4]$. Hence, by Theorem 3.11, $\mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^2 \mathbf{1}_{(i)}]$ 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

$$|\widehat{S}_n^f - S_{min}| \leq |\widehat{S}_n^f - \widehat{S}^f(\tau)| + |\widehat{S}^f(\tau) - S(\tau)|,$$

and noting that

$$\widehat{S}^f(\tau) - \widehat{S}_n^f = \frac{\tau - t_n}{h} \left(\widehat{S}_{n+1}^f - \widehat{S}_n^f \right) + b_n^f(W(\tau) - \overline{W}(\tau)),$$

we can conclude that $|\widehat{S}_n^f - S_{min}| \prec h_\ell^{1/2-2\gamma}$. Hence, for sufficiently small h_ℓ , $|\widehat{S}_n^f - S_{min}| < h_\ell^{1/2-4\gamma}$ and so \widehat{S}_n^f is guaranteed to be less than B . In addition, $\widehat{S}_n^f - \widehat{S}_n^c < h_\ell^{1-\gamma}$ and so, for sufficiently small h_ℓ , \widehat{S}_n^c is also guaranteed to be less than B and hence $\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c = 0$.

In the alternate case $S_{min} > B + h_\ell^{1/2-4\gamma}$, then

$$\min_n \min(\widehat{S}_n^f, \widehat{S}_n^c) > B + h_\ell^{1/2-4\gamma} - h_\ell^{1-\gamma}$$

and since $h_\ell^{1-\gamma} \prec h_\ell^{1/2-4\gamma}$ it follows that $\prod_n (1 - \widehat{p}_n^f)$ and $\prod_n (1 - \widehat{p}_n^c)$ are both equal to $1 - o(h_\ell^p)$ for all $p > 0$, and so $|\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c| \prec h_\ell^{1-\gamma}$ due to the Lipschitz condition and the bound on $\widehat{S}_N^f - \widehat{S}_N^c$. Hence, the contribution to $\mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^2]$ is at most $O(h_\ell^{2-2\gamma})$.

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

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

and

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

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

Assume $n \in R$. We have $\widehat{S}_n^f - \widehat{S}_n^c < h_\ell^{1-\gamma}$ and $\widehat{S}_{n+1}^f - \widehat{S}_{n+1}^c < h_\ell^{1-\gamma}$ due to the definition of extreme paths, and $b_n^f - b_n^c \prec h_\ell^{1/2-2\gamma}$, due to Lemma 3.16. If we now define

$$\begin{aligned} X_n^f &\equiv \frac{2(\widehat{S}_n^f - B)^+(\widehat{S}_{n+1}^f - B)^+}{(b_n^f)^2 h_\ell}, \\ X_n^c &\equiv \frac{2(\widehat{S}_n^c - B)^+(\widehat{S}_{n+1}^c - B)^+}{(b_n^c)^2 h_\ell}, \end{aligned}$$

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_ℓ , that $\min(|b_n^c|, |b_n^f|) > \frac{1}{2} b_{\min}$, and hence through repeated use of the following identity,

$$f_1 g_1 - f_2 g_2 = \frac{1}{2}(f_1 - f_2)(g_1 + g_2) + \frac{1}{2}(f_1 + f_2)(g_1 - g_2), \quad (22)$$

and the fact that $n \in R$ to bound terms such as $\widehat{S}_n^f - B$, we obtain $|X_n^f - X_n^c| \prec h_\ell^{1/2-4\gamma}$, and hence $|X_n^f - X_n^c| < h_\ell^{1/2-5\gamma}$, for sufficiently small h_ℓ . 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

$$\begin{aligned} 1 - \widehat{p}_n^c &= (1 - \widehat{p}_n^f) + (\widehat{p}_n^f - \widehat{p}_n^c) \\ &= (1 - \widehat{p}_n^f) + \widehat{p}_n^f (1 - \exp(X_n^f - X_n^c)) \\ &\leq (1 - \widehat{p}_n^f) + \widehat{p}_n^f \Delta_\ell. \end{aligned}$$

Since $g(\Delta) \equiv \prod_{n \in R} ((1 - p_n^f) + p_n^f \Delta) - \prod_{n \in R} (1 - p_n^f) - \Delta$ 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,

$$\prod_{n \in R} (1 - \widehat{p}_n^c) \leq \prod_{n \in R} (1 - \widehat{p}_n^f) + \Delta_\ell.$$

Similarly, $1 - \widehat{p}_n^f \leq (1 - \widehat{p}_n^c) + \widehat{p}_n^c \Delta_\ell$, which leads to

$$\prod_{n \in R} (1 - \widehat{p}_n^f) \leq \prod_{n \in R} (1 - \widehat{p}_n^c) + \Delta_\ell,$$

and therefore

$$\left| \prod_{n \in R} (1 - \widehat{p}_n^f) - \prod_{n \in R} (1 - \widehat{p}_n^c) \right| \leq \Delta_\ell.$$

Returning to the original products over all n ,

$$\left| \prod_n (1 - \widehat{p}_n^f) - \prod_n (1 - \widehat{p}_n^c) \right| \prec h_\ell^{1/2-5\gamma}.$$

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

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

3.2.8 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

$$P = \mathbf{1}_{S(T) > K},$$

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 3.2.3 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 $\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c = \pm 1$, resulting in $V_\ell = O(h_\ell)$. To improve the variance to $O(h_\ell^{3/2-\delta})$ for all $\delta > 0$ we follow the approach which was tested numerically in [3], using the technique of conditional expectation (see section 7.2.3 in [9]).

If \widehat{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 $a_{N-1}^f \equiv a(\widehat{S}_{N-1}^f, T - h_\ell)$ and volatility

$b_{N-1}^f \equiv b(\widehat{S}_{N-1}^f, T-h_\ell)$, the conditional expectation for the payoff is the probability that $\widehat{S}_N^f > K$ after one further timestep, which is

$$\widehat{P}_\ell^f = \Phi \left(\frac{\widehat{S}_{N-1}^f + a_{N-1}^f h_\ell - K}{|b_{N-1}^f| \sqrt{h_\ell}} \right), \quad (23)$$

where Φ is the cumulative Normal distribution.

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

$$\widehat{P}_{\ell-1}^c = \Phi \left(\frac{\widehat{S}_{N-2}^c + 2a_{N-2}^c h_\ell + b_{N-2}^c \Delta W_{N-2} - K}{|b_{N-2}^c| \sqrt{h_\ell}} \right). \quad (24)$$

The conditional expectation of (24) is equal to the conditional expectation of $\widehat{P}_{\ell-1}^f$ defined by (23) on level $\ell-1$, and so equality (13) is satisfied.

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

Theorem 3.18. *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 $V_\ell = o(h_\ell^{3/2-\delta})$ for any $\delta > 0$.*

Proof. As in the proof of Theorem 3.17, we split the paths into three subsets:

- (i) extreme paths;
- (ii) paths which are not extreme and for which $|S_N - K| > h_\ell^{1/2-4\gamma}$;
- (iii) the rest

and we analyse the contributions to $\mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^2]$ from all three subsets.

(i) Paths are defined to be extreme if they satisfy any of the conditions of Lemma 3.16 for $0 < \gamma < \frac{1}{4}$. $\mathbb{E}[(\widehat{P}^f)^4]$ and $\mathbb{E}[(\widehat{P}^c)^4]$ are both finite, and hence the contribution of the extreme paths is $o(h_\ell^p)$, for all $p > 0$.

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

$$\begin{aligned} \frac{\widehat{S}_{N-1}^f + a_{N-1}^f h_\ell - K}{|b_{N-1}^f| \sqrt{h_\ell}} &= \frac{\widehat{S}_N^f - K}{|b_{N-1}^f| \sqrt{h_\ell}} \\ &- \frac{b_{N-1}^f}{|b_{N-1}^f| \sqrt{h_\ell}} \left(\Delta W_{N-1} + \frac{1}{2}(b')_{N-1}^f ((\Delta W_{N-1})^2 - h_\ell) \right), \end{aligned}$$

and similarly

$$\begin{aligned} \frac{\widehat{S}_{N-2}^c + 2a_{N-2}^c h_\ell + b_{N-2}^c \Delta W_{N-2} - K}{|b_{N-2}^c| \sqrt{h_\ell}} &= \frac{\widehat{S}_N^c - K}{|b_{N-2}^c| \sqrt{h_\ell}} \\ &- \frac{b_{N-2}^c}{|b_{N-2}^c| \sqrt{h_\ell}} \left(\Delta W_{N-1} + \frac{1}{2} (b')_{N-2}^c \left((\Delta W_{N-2} + \Delta W_{N-1})^2 - 2h_\ell \right) \right). \end{aligned}$$

Since the paths are not extreme, $|\Delta W_n| \leq h_\ell^{1/2-\gamma}$ and $|S(T) - \widehat{S}_N^f| \leq h_\ell^{1-\gamma}$, and due to Lemma 3.16 $|b_{N-1}^f| \prec h_\ell^{-\gamma}$. Consequently, if $S(T) > K + h_\ell^{1/2-3\gamma}$ then for sufficiently small h_ℓ it follows that

$$\frac{\widehat{S}_{N-1}^f + a_{N-1}^f h_\ell - K}{|b_{N-1}^f| \sqrt{h_\ell}} > C h_\ell^{-2\gamma},$$

for some suitably chosen constant C . A similar result follows for the corresponding coarse path, and hence for these paths $\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c = o(h_\ell^p)$, 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 $\mathbb{E}[(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c)^2 \mathbf{1}_{(ii)}]$ is $o(h_\ell^p)$ 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

$$b_{N-1}^f - b(K, T) = (b_{N-1}^f - b_N^f) + (b_N^f - b(K, T)),$$

using Assumption A1 and (18) with $\gamma < \frac{1}{4}$ we can conclude that for sufficiently small h_ℓ , $|b_{N-1}^f - b(K, T)| < \frac{1}{2} |b(K, T)|$ and in particular b_{N-1}^f is non-zero and of the same sign as $b(K, T)$. The same also applies to b_{N-2}^c and hence, exploiting the Lipschitz property $|\Phi(x_1) - \Phi(x_2)| \leq |x_1 - x_2|$,

$$\begin{aligned} & \left| \widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c \right| \\ & \leq \left| \frac{\widehat{S}_{N-1}^f + a_{N-1}^f h_\ell - K}{|b_{N-1}^f| \sqrt{h_\ell}} - \frac{\widehat{S}_{N-2}^c + 2a_{N-2}^c h_\ell + b_{N-2}^c \Delta W_{N-2} - K}{|b_{N-2}^c| \sqrt{h_\ell}} \right| \\ & \leq \left| \frac{\widehat{S}_N^f - K}{|b_{N-1}^f| \sqrt{h_\ell}} - \frac{\widehat{S}_N^c - K}{|b_{N-2}^c| \sqrt{h_\ell}} \right| \\ & \quad + \frac{1}{2} K_1 h_\ell^{-1/2} \{ (\Delta W_{N-1})^2 + (\Delta W_{N-2} + \Delta W_{N-1})^2 + 3h_\ell \}. \end{aligned}$$

Using the identity (22) we obtain

$$\begin{aligned} \frac{\widehat{S}_N^f - K}{|b_{N-1}^f| \sqrt{h_\ell}} - \frac{\widehat{S}_N^c - K}{|b_{N-2}^c| \sqrt{h_\ell}} &= \frac{1}{2\sqrt{h_\ell}} (\widehat{S}_N^f - \widehat{S}_N^c) \left(\frac{1}{|b_{N-1}^f|} + \frac{1}{|b_{N-2}^c|} \right) \\ &+ \frac{1}{2\sqrt{h_\ell}} (\widehat{S}_N^f + \widehat{S}_N^c - 2K) \left(\frac{|b_{N-2}^c| - |b_{N-1}^f|}{|b_{N-1}^f| |b_{N-2}^c|} \right). \end{aligned}$$

Using the bounds provided by Lemma 3.16, it follows that

$$\frac{\widehat{S}_N^f - K}{|b_{N-1}^f| \sqrt{h_\ell}} - \frac{\widehat{S}_N^c - K}{|b_{N-2}^c| \sqrt{h_\ell}} = O(h_\ell^{1/2-5\gamma}),$$

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

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 $O(h_\ell^2)$ for Lipschitz and Asian options, $O(h_\ell^2(\log h_\ell)^2)$ for lookback options, and $o(h_\ell^{3/2-\delta})$ for barrier and digital options, for any $\delta > 0$.

The MLMC theorem [4] also requires knowledge of the order of weak convergence. Theorems 3.2 and 3.12 together give $O(h)$ weak convergence for the Lipschitz and Asian options, and $O(h \log h)$ convergence for the lookback option. For the digital and barrier options, the analysis of the multilevel convergence can be modified to instead consider $\mathbb{E}[\widehat{P}_\ell - P]$, and hence it can be proved that the weak order of convergence is $o(h^{1-\delta})$ for any $\delta > 0$. From this, in all cases considered in this paper it can be concluded that the computational cost to achieve a RMS accuracy of ε is $O(\varepsilon^{-2})$.

The challenge for the future is to construct and analyse effective MLMC estimators for multi-dimensional SDEs which do not satisfy the commutativity condition, and therefore would require the simulation of Lévy areas to achieve first order strong convergence. Giles and Szpruch [8] have proved that if one ignores the Lévy area terms, it is nevertheless possible to construct an

efficient antithetic MLMC estimator, despite the strong convergence being of the same order as the Euler-Maruyama discretisation. They prove $V_\ell = O(h_\ell^2)$ for a European payoff which is twice differentiable, and $V_\ell = O(h_\ell^{3/2})$ for a payoff such as a put or call function which is continuous but not everywhere twice-differentiable. However, it is still an open problem to construct and analyse good estimators for lookback, barrier and digital options for general systems of SDEs.

References

- [1] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13(3):381–401, 2009.
- [2] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance*. Springer, 2008.
- [3] M.B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer, 2008.
- [4] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
- [5] M.B. Giles. Multilevel Monte Carlo for basket options. In M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference*, pages 1283–1290. IEEE, 2009.
- [6] M.B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.
- [7] M.B. Giles, D.J. Higham, and X. Mao. Analysing multilevel Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3):403–413, 2009.
- [8] M.B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Annals of Applied Probability*, 24(4):1585–1620, 2014.

- [9] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.
- [10] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.