
Improved Multilevel Monte Carlo Convergence using the Milstein Scheme

Mike Giles

Oxford University Computing Laboratory, Parks Road, Oxford, U.K.
Mike.Giles@comlab.ox.ac.uk

Summary. In this paper we show that the Milstein scheme can be used to improve the convergence of the multilevel Monte Carlo method for scalar stochastic differential equations. Numerical results for Asian, lookback, barrier and digital options demonstrate that the computational cost to achieve a root-mean-square error of ϵ is reduced to $O(\epsilon^{-2})$. This is achieved through a careful construction of the multilevel estimator which computes the difference in expected payoff when using different numbers of timesteps.

1 Introduction

In many financial engineering applications, one is interested in the expected value of a financial option whose payoff depends upon the solution of a stochastic differential equation. To be specific, we consider an SDE with general drift and volatility terms,

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

with given initial data S_0 . In the case of European and digital options, we are interested in the expected value of a function of the terminal state, $f(S(T))$, but in the case of Asian, lookback and barrier options the valuation depends on the entire path $S(t), 0 < t < T$.

Using a simple Monte Carlo method with a numerical discretisation with first order weak convergence, to achieve a root-mean-square error of $O(\epsilon)$ would require $O(\epsilon^{-2})$ independent paths, each with $O(\epsilon^{-1})$ timesteps, giving a computational complexity which is $O(\epsilon^{-3})$. We have recently introduced a new multilevel approach [Gil06] which reduces the cost to $O(\epsilon^{-2}(\log \epsilon)^2)$ when using an Euler path discretisation for a European option with a payoff with a uniform Lipschitz bound. This multilevel approach is related to the two-level method of Kebaier [Keb05], and is similar to the multi-level method proposed by Speight [Spe05] based on the quasi control variate method of Emsermann

and Simon [ES02]. There are also strong similarities to Heinrich's multilevel approach for parametric integration [Hei01].

In the previous work, it was also proved that the computational cost can be further reduced to $O(\epsilon^{-2})$ for numerical discretisations with certain multilevel convergence properties. The objective of this paper is to demonstrate that this improved complexity is attainable for scalar SDEs with a variety of exotic options through using the Milstein path discretisation. For European options with a Lipschitz continuous payoff, it can be proved that this is an immediate consequence of the improved strong order of convergence of the Milstein discretisation compared to the simpler Euler discretisation. However, for Asian, lookback, barrier and digital options, special numerical treatments have to be introduced, and that is the focus of the paper. Furthermore, no *a priori* convergence proofs have yet been constructed for these cases and so the paper relies on numerical demonstration of the effectiveness of the algorithms that have been developed.

The paper begins by reviewing the multilevel approach, and the theorem which describes its computational cost given certain properties of the numerical discretisation. The next section discusses the Milstein discretisation and the challenges of achieving higher order variance convergence within the multilevel method. Asian, lookback, barrier and digital options are all considered, and $O(\epsilon^{-2})$ computational cost is demonstrated for each through the use of Brownian interpolation to approximate the behaviour of paths within each timestep.

The final section indicates the direction of future research, including the need for *a priori* convergence analysis, the challenges of extending this work to multi-dimensional SDEs, and the use of quasi-Monte Carlo methods for further reduction of the computational complexity.

2 Multilevel Monte Carlo Method

Consider Monte Carlo path simulations with different timesteps $h_l = 2^{-l}T$, $l = 0, 1, \dots, L$. Thus on the coarsest level, $l = 0$, the simulations use just 1 timestep, while on the finest level, $l = L$, the simulations use 2^L timesteps. For a given Brownian path $W(t)$, let P denote the payoff, and let \hat{P}_l denote its approximation using a numerical discretisation with timestep h_l . Because of the linearity of the expectation operator, it is clearly true that

$$E[\hat{P}_L] = E[\hat{P}_0] + \sum_{l=1}^L E[\hat{P}_l - \hat{P}_{l-1}]. \quad (2)$$

This expresses the expectation on the finest level as being equal to the expectation on the coarsest level plus a sum of corrections which give the difference in expectation between simulations using different numbers of timesteps. The idea behind the multilevel method is to independently estimate each of the

expectations on the right-hand side in a way which minimises the overall variance for a given computational cost.

Let \widehat{Y}_0 be an estimator for $E[\widehat{P}_0]$ using N_0 samples, and let \widehat{Y}_l for $l > 0$ be an estimator for $E[\widehat{P}_l - \widehat{P}_{l-1}]$ using N_l paths. The simplest estimator is a mean of N_l independent samples, which for $l > 0$ is

$$\widehat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left(\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)} \right). \tag{3}$$

The key point here is that the quantity $\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)}$ comes from two discrete approximations with different timesteps but the same Brownian path. The variance of this simple estimator is $V[\widehat{Y}_l] = N_l^{-1} V_l$ where V_l is the variance of a single sample. Combining this with independent estimators for each of the other levels, the variance of the combined estimator $\widehat{Y} = \sum_{l=0}^L \widehat{Y}_l$ is $V[\widehat{Y}] = \sum_{l=0}^L N_l^{-1} V_l$, while its computational cost is proportional to $\sum_{l=0}^L N_l h_l^{-1}$. Treating the N_l as continuous variables, the variance is minimised for a fixed computational cost by choosing N_l to be proportional to $\sqrt{V_l h_l}$.

In the particular case of an Euler discretisation, provided $a(S, t)$ and $b(S, t)$ satisfy certain conditions [BT95, KP92, TT90] there is $O(h^{1/2})$ strong convergence. From this it follows that $V[\widehat{P}_l - P] = O(h_l)$ for a European option with a Lipschitz continuous payoff. Hence for the simple estimator (3), the single sample variance V_l is $O(h_l)$, and the optimal choice for N_l is asymptotically proportional to h_l . Setting $N_l = O(\epsilon^{-2} L h_l)$, the variance of the combined estimator \widehat{Y} is $O(\epsilon^2)$. If L is chosen such that $L = \log \epsilon^{-1} / \log 2 + O(1)$, as $\epsilon \rightarrow 0$, then $h_L = 2^{-L} = O(\epsilon)$, and so the bias error $E[\widehat{P}_L - P]$ is $O(\epsilon)$ due to standard results on weak convergence. Consequently, we obtain a Mean Square Error which is $O(\epsilon^2)$, with a computational complexity which is $O(\epsilon^{-2} L^2) = O(\epsilon^{-2} (\log \epsilon)^2)$.

This analysis is generalised in the following theorem:

Theorem 1. *Let P denote a functional of the solution of stochastic differential equation (1) for a given Brownian path $W(t)$, and let \widehat{P}_l denote the corresponding approximation using a numerical discretisation with timestep $h_l = M^{-l} T$.*

If there exist independent estimators \widehat{Y}_l based on N_l Monte Carlo samples, and positive constants $\alpha \geq \frac{1}{2}, \beta, c_1, c_2, c_3$ such that

- i) $E[\widehat{P}_l - P] \leq c_1 h_l^\alpha$*
- ii) $E[\widehat{Y}_l] = \begin{cases} E[\widehat{P}_0], & l = 0 \\ E[\widehat{P}_l - \widehat{P}_{l-1}], & l > 0 \end{cases}$*
- iii) $V[\widehat{Y}_l] \leq c_2 N_l^{-1} h_l^\beta$*

iv) C_l , the computational complexity of \widehat{Y}_l , is bounded by

$$C_l \leq c_3 N_l h_l^{-1},$$

then there exists a positive constant c_4 such that for any $\epsilon < e^{-1}$ there are values L and N_l for which the multilevel estimator

$$\widehat{Y} = \sum_{l=0}^L \widehat{Y}_l,$$

has a mean-square-error with bound

$$MSE \equiv E \left[\left(\widehat{Y} - E[P] \right)^2 \right] < \epsilon^2$$

with a computational complexity C with bound

$$C \leq \begin{cases} c_4 \epsilon^{-2}, & \beta > 1, \\ c_4 \epsilon^{-2} (\log \epsilon)^2, & \beta = 1, \\ c_4 \epsilon^{-2 - (1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

Proof. See [Gil06].

The remainder of this paper addresses the use of the Milstein scheme [Gla04, KP92] to construct estimators with variance convergence rates $\beta > 1$, resulting in an $O(\epsilon^{-2})$ complexity bound. Provided certain conditions are satisfied [KP92], the Milstein scheme gives $O(h)$ strong convergence. In the case of a Lipschitz continuous European payoff, this immediately leads to the result that $V_l = O(h_l^2)$, corresponding to $\beta=2$. Numerical results which are not presented here demonstrate this convergence rate, and the associated $O(\epsilon^{-2})$ complexity

This paper addresses the tougher challenges of Asian, lookback, barrier and digital options. These cases require some ingenuity to construct estimators for which $\beta > 1$. Unfortunately, there is no accompanying theoretical analysis as yet, and so the paper relies on numerical demonstration of their effectiveness.

3 Milstein Discretisation

For a scalar SDE, the Milstein discretisation of equation (1) is

$$\widehat{S}_{n+1} = \widehat{S}_n + a h + b \Delta W_n + \frac{1}{2} \frac{\partial b}{\partial S} b (\Delta W_n)^2. \quad (4)$$

In the above equation, the subscript n is used to denote the timestep index, and a , b and $\partial b/\partial S$ are evaluated at \widehat{S}_n, t_n .

All of the numerical results to be presented are for the case of geometric Brownian motion for which the SDE is

$$dS(t) = r S dt + \sigma S dW(t), \quad 0 < t < T.$$

By switching to the new variable $X = \log S$, it is possible to construct numerical approximations which are exact, but here we directly simulate the geometric Brownian motion using the Milstein method as an indication of the behaviour with more complicated models, for example those with a local volatility function $\sigma(S, t)$.

3.1 Estimator Construction

In all of the cases to be presented, we simulate the paths using the Milstein method. The refinement factor is $M = 2$, so each level has twice as many timesteps as the previous level. The difference between the applications is in how we use the computed discrete path data to estimate $E[\widehat{P}_l - \widehat{P}_{l-1}]$.

In each case, the estimator for $E[\widehat{P}_l - \widehat{P}_{l-1}]$ is an average of values from N_l independent path simulations. For each Brownian input, the value which is computed is of the form $\widehat{P}_l^f - \widehat{P}_{l-1}^c$. Here \widehat{P}_l^f is a fine-path estimate using timestep $h = 2^{-l}T$, and \widehat{P}_{l-1}^c is the corresponding coarse-path estimate using timestep $h = 2^{-(l-1)}T$. To ensure that the identity (2) is correctly respected, to avoid the introduction of an undesired bias, we require that

$$E[\widehat{P}_l^f] = E[\widehat{P}_l^c]. \tag{5}$$

This means that the definitions of \widehat{P}_l when estimating $E[\widehat{P}_l - \widehat{P}_{l-1}]$ and $E[\widehat{P}_{l+1} - \widehat{P}_l]$ must have the same expectation.

In the simplest case of a European option, this can be achieved very simply by defining \widehat{P}_l^f and \widehat{P}_l^c to be the same; this is the approach which was used for all applications in the previous work using the Euler discretisation [Gil06]. However, for more challenging applications such as Asian, lookback, barrier and digital options, the definition of \widehat{P}_l^c will involve information from the discrete simulation of \widehat{P}_{l+1}^f , which is not available in computing \widehat{P}_l^f . The reason for doing this is to reduce the variance of the estimator, but it must be shown that equality (5) is satisfied. This will be achieved in each case through a construction based on a simple Brownian motion approximation.

3.2 Asian Option

The Asian option we consider has the discounted payoff

$$P = \exp(-rT) \max(0, \bar{S} - K),$$

where

$$\bar{S} = T^{-1} \int_0^T S(t) dt.$$

The simplest approximation of \bar{S} , which was used in previous work [Gil06], is

$$\bar{S} = T^{-1} \sum_0^{n_T-1} \frac{1}{2} h (\hat{S}_n + \hat{S}_{n+1}),$$

where $n_T = T/h$ is the number of timesteps. This corresponds to a piecewise linear approximation to $S(t)$ but improved accuracy can be achieved by approximating the behaviour within a timestep as simple Brownian motion, with constant drift and volatility, conditional on the computed values \hat{S}_n . Taking b_n to be the constant volatility within the time interval $[t_n, t_{n+1}]$, standard Brownian Bridge results (see section 3.1 in [Gla04]) give

$$\int_{t_n}^{t_{n+1}} S(t) dt = \frac{1}{2} h (S(t_n) + S(t_{n+1})) + b_n \Delta I_n,$$

where ΔI_n , defined as

$$\Delta I_n = \int_{t_n}^{t_{n+1}} (W(t) - W(t_n)) dt - \frac{1}{2} h \Delta W,$$

is a $N(0, h^3/12)$ Normal random variable, independent of ΔW . Using $b_n = b(\hat{S}_n, t_n)$, this gives the fine-path approximation

$$\bar{S} = T^{-1} \sum_0^{n_T-1} \left(\frac{1}{2} h (\hat{S}_n + \hat{S}_{n+1}) + b_n \Delta I_n \right).$$

The coarse path approximation is the same except that the values for ΔI_n are derived from the fine path values, noting that

$$\begin{aligned} & \int_{t_n}^{t_n+2h} (W(t) - W(t_n)) dt - h(W(t_n + 2h) - W(t_n)) \\ &= \int_{t_n}^{t_n+h} (W(t) - W(t_n)) dt - \frac{1}{2} h (W(t_n + h) - W(t_n)) \\ & \quad + \int_{t_n+h}^{t_n+2h} (W(t) - W(t_n + h)) dt - \frac{1}{2} h (W(t_n + 2h) - W(t_n + h)) \\ & \quad + \frac{1}{2} h (W(t_n + h) - W(t_n)) - \frac{1}{2} h (W(t_n + 2h) - W(t_n + h)), \end{aligned}$$

and hence

$$\Delta I^c = \Delta I^{f1} + \Delta I^{f2} + \frac{1}{2} h (\Delta W^{f1} - \Delta W^{f2}),$$

where ΔI^c is the value for the coarse timestep, and ΔI^{f1} and ΔW^{f1} are the values for the first fine timestep, and ΔI^{f2} and ΔW^{f2} are the values for the second fine timestep.

Figure 1 shows the numerical results for parameters $S(0) = 1, K = 1, T = 1, r = 0.05, \sigma = 0.2$. The top left plot shows the behaviour of the

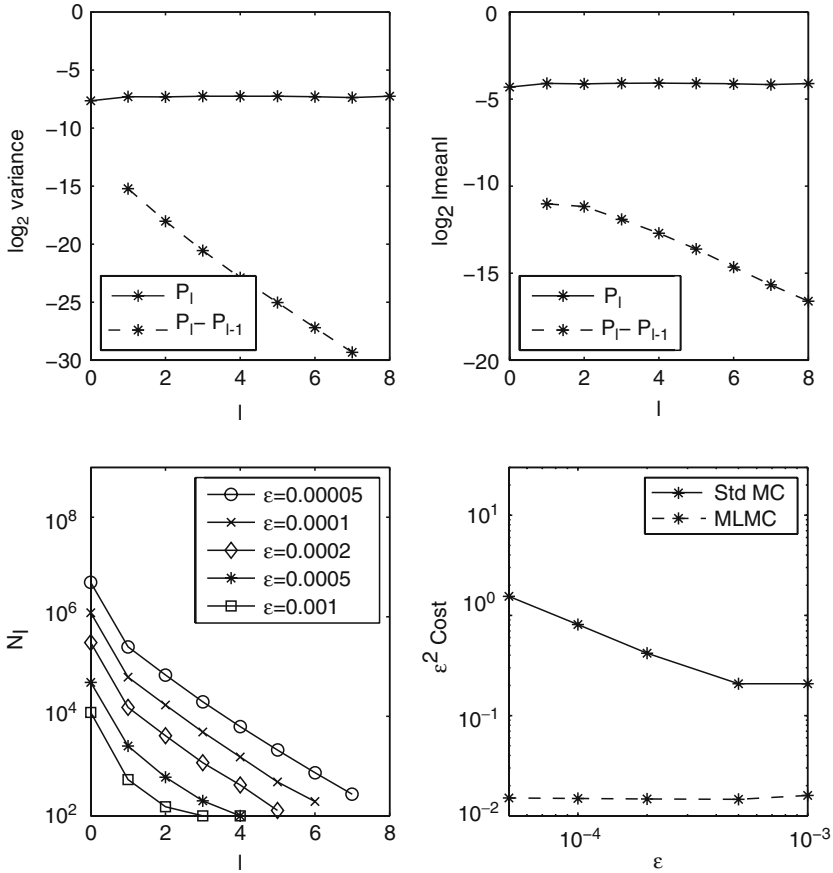


Fig. 1. Asian option

variance of both \widehat{P}_l and $\widehat{P}_l - \widehat{P}_{l-1}$. The slope of the latter is approaching a value approximately equal to -2 , indicating that $V_l = O(h_l^2)$, corresponding to $\beta = 2$. On level $l = 2$, which has just 4 timesteps, V_l is already more than 1000 times smaller than the variance $V[\widehat{P}_l]$ of the standard Monte Carlo method with the same timestep. The top right plot shows that $E[\widehat{P}_l - \widehat{P}_{l-1}]$ is approximately $O(h_l)$, corresponding to first order weak convergence, $\alpha = 1$. This is used to determine the number of levels that are required to reduce the bias to an acceptable level [Gil06].

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 $N_l, l = 0, \dots, L$, 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. The bottom right plot shows the variation with ϵ of $\epsilon^2 C$ where the

computational complexity C is defined as

$$C = \sum_l 2^l N_l,$$

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 $\epsilon^2 C$ is almost constant for the multilevel method, as expected, whereas for the standard Monte Carlo method it is approximately proportional to ϵ^{-1} . For the most accurate case, $\epsilon = 5 \times 10^{-5}$, the multilevel method is more than 100 times more efficient than the standard method.

3.3 Lookback Option

The lookback option we consider has the discounted payoff

$$P = \exp(-rT) \left(S(T) - \min_{0 < t < T} S(t) \right).$$

In previous work [Gil06], the minimum value of $S(t)$ over the path was approximated numerically by

$$\widehat{S}_{min} = \min_n \left(\widehat{S}_n - \beta^* b_n \sqrt{h} \right).$$

Here b_n is the volatility in the n^{th} timestep, and $\beta^* \approx 0.5826$ is a constant which corrects the $O(h^{1/2})$ leading order error due to the discrete sampling of the path, and thereby restores $O(h)$ weak convergence [BGK97]. However, using this approximation, the difference between the computed minimum values and fine and coarse paths is $O(h_i^{1/2})$, and hence the variance V_l is $O(h_i)$, corresponding to $\beta = 1$. In the previous work, this was acceptable because $\beta = 1$ is the best that can be achieved in general with the Euler path discretisation which was used, but in this work we aim to achieve an improved convergence rate using the Milstein scheme.

To achieve this, we again approximate the behaviour within a timestep as simple Brownian motion, with constant drift and volatility, conditional on the computed values \widehat{S}_n . For the time interval $[t_n, t_{n+1}]$, standard Brownian Interpolation results (see section 6.4 in [Gla04]) give the minimum of Brownian motion, conditional on the end values, as

$$\widehat{S}_{n,min} = \frac{1}{2} \left(\widehat{S}_n + \widehat{S}_{n+1} - \sqrt{\left(\widehat{S}_{n+1} - \widehat{S}_n \right)^2 - 2b_n^2 h \log U_n} \right), \quad (6)$$

where b_n is the constant volatility and U_n is a uniform random variable on $[0, 1]$.

The fine-path value \widehat{P}_l^f is defined in this way using $b_n = b(\widehat{S}_n, t_n)$, and then taking the minimum over all timesteps to obtain the global minimum. However, for the coarse-path value \widehat{P}_{l-1}^c , we do something different. Again assuming simple Brownian motion conditional on the end-points, the value at the midpoint of the time interval $[t_n, t_{n+1}]$ is given by

$$\widehat{S}_{n+1/2} = \frac{1}{2} \left(\widehat{S}_n + \widehat{S}_{n+1} - b_n D_n \right), \tag{7}$$

where

$$D_n = W_{n+1} - 2W_{n+1/2} + W_n = (W_{n+1} - W_{n+1/2}) - (W_{n+1/2} - W_n),$$

is a $N(0, h)$ random variable which corresponds to a difference in the consecutive Brownian increments of a finer path with timestep $h/2$. Given this midpoint value, the minimum value over the full timestep is the smaller of the minima for each of the two half-timesteps,

$$\widehat{S}_{n,min} = \min \left\{ \frac{1}{2} \left(\widehat{S}_n + \widehat{S}_{n+1/2} - \sqrt{\left(\widehat{S}_{n+1/2} - \widehat{S}_n \right)^2 - b_n^2 h \log U_{1,n}} \right), \right. \\ \left. \frac{1}{2} \left(\widehat{S}_{n+1/2} + \widehat{S}_{n+1} - \sqrt{\left(\widehat{S}_{n+1} - \widehat{S}_{n+1/2} \right)^2 - b_n^2 h \log U_{2,n}} \right) \right\}. \tag{8}$$

In computing \widehat{P}_{l-1}^c , we use the values for $D_n, U_{1,n}$ and $U_{2,n}$ that come from the fine-path simulation for \widehat{P}_l^f . D_n is the difference of the Brownian increments for the two fine-path timesteps, and $U_{1,n}$ and $U_{2,n}$ are the uniform random variables used to compute the minima for the two fine-path timesteps. Since these all have the correct probability distribution, it follows that the expected values of (6) and (8) are identical, and therefore equality (5) is satisfied.

Figure 2 shows the numerical results for parameters $S(0) = 1, T = 1, r = 0.05, \sigma = 0.2$. The top left plot shows that the variance is $O(h_l^2)$, corresponding to $\beta = 2$, while the top right plot shows that the mean correction is $O(h_l)$, corresponding to first order weak convergence, $\alpha = 1$. 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 $\epsilon = 5 \times 10^{-5}$. The computational cost of the multilevel method is almost perfectly proportional to ϵ^{-2} .

3.4 Barrier Option

The barrier option which is considered is a down-and-out call for which the discounted payoff is

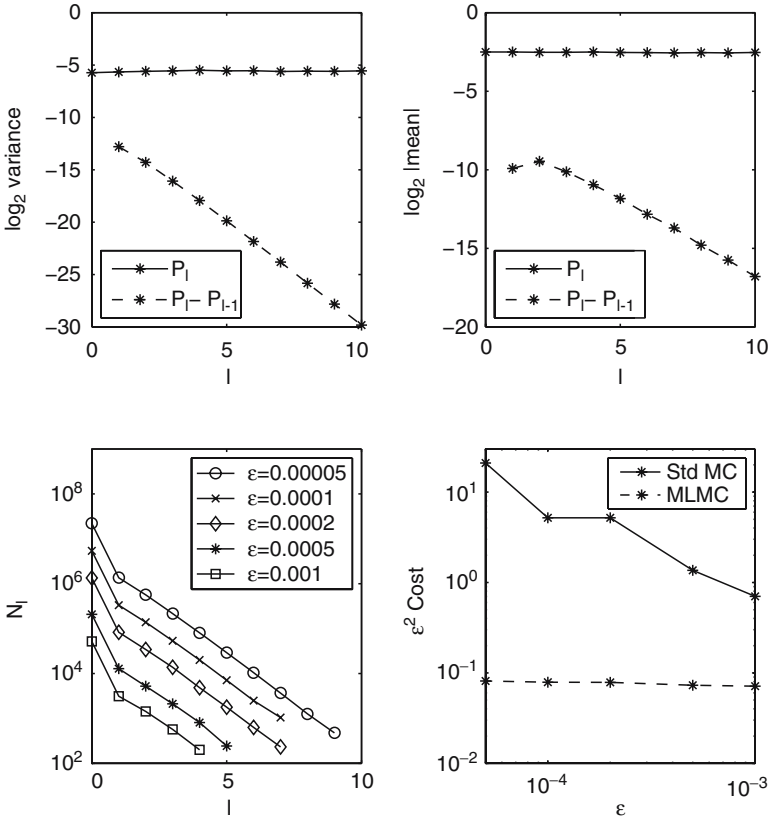


Fig. 2. Lookback option

$$P = \exp(-rT) (S(T) - K)^+ \mathbf{1}\{\tau > T\},$$

where the notation $(S(T) - K)^+$ denotes $\max(0, S(T) - K)$, $\mathbf{1}\{\tau > T\}$ is an indicator function taking value 1 if the argument is true, and zero otherwise, and the crossing time τ is defined as

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

Following a standard approach for continuously monitored barrier crossings (see section 6.4 in [Gla04]), for a particular Brownian path input sampled discretely at uniform intervals h , the conditional expectation of the payoff can be expressed as

$$\exp(-rT) (\widehat{S}_{n_T} - K)^+ \prod_{n=0}^{n_T-1} \widehat{p}_n,$$

where $n_T = T/h$ is again the number of timesteps, and \widehat{p}_n represents the probability that the path did not cross the barrier during the n^{th} timestep. If

we again approximate the motion within each timestep as simple Brownian motion conditional on the endpoint values, then

$$\hat{p}_n = 1 - \exp\left(\frac{-2(S_n - B)^+(S_{n+1} - B)^+}{b_n^2 h}\right). \tag{9}$$

This is the expression used to define the payoff \hat{P}_l^f for the fine-path calculation, with b_n set equal to $b(\hat{S}_n, t_n)$, as in the lookback calculation.

For the coarse path calculation, in which each timestep corresponds to two fine-path timesteps, we again use equation (7) to construct a midpoint value $\hat{S}_{n+1/2}$. Given this value, the probability that the simple Brownian path does not cross the barrier is

$$\begin{aligned} \hat{p}_n = & \left\{ 1 - \exp\left(\frac{-2(S_n - B)^+(S_{n+1/2} - B)^+}{b_n^2 h}\right) \right\} \\ & \times \left\{ 1 - \exp\left(\frac{-2(S_{n+1/2} - B)^+(S_{n+1} - B)^+}{b_n^2 h}\right) \right\}. \end{aligned} \tag{10}$$

The conditional expectation of (10) is equal to (9) and so equality (5) is satisfied.

Figure 3 shows the numerical results for parameters $S(0) = 1, K = 1, B = 0.85, T = 1, r = 0.05, \sigma = 0.2$. The top left plot shows that the variance is approximately $O(h_l^\beta)$ for a value of β slightly less than 2. An explanation for this is that a small $O(h_l^{1/2})$ fraction of the paths have a minimum which lies within $O(h_l^{1/2})$ of the barrier, for which the product $\prod \hat{p}_n$ is neither close to zero nor close to unity. The fine path and coarse path trajectories differ by $O(h_l)$, due to the first order strong convergence of the Milstein scheme. Since the \hat{p}_n have an $O(h_l^{-1/2})$ derivative, this results in the difference between $\prod \hat{p}_n$ for this small subset of coarse and fine paths being $O(h_l^{1/2})$, giving a contribution to the variance which is $O(h_l^{3/2})$.

The top right plot shows that the mean correction is $O(h_l)$, corresponding to first order weak convergence, $\alpha = 1$. The bottom right plot shows that the computational cost of the multilevel method is again almost perfectly proportional to ϵ^{-2} , and for $\epsilon = 5 \times 10^{-5}$ it is over 100 times more efficient than the standard Monte Carlo method.

3.5 Digital Option

The digital option which is considered has the discounted payoff

$$P = \exp(-rT) \mathbf{1}\{S(T) > K\}.$$

The standard numerical discretisation would be to simulate the path of $S(t)$ right up to the final time T . This is the approach adopted previously

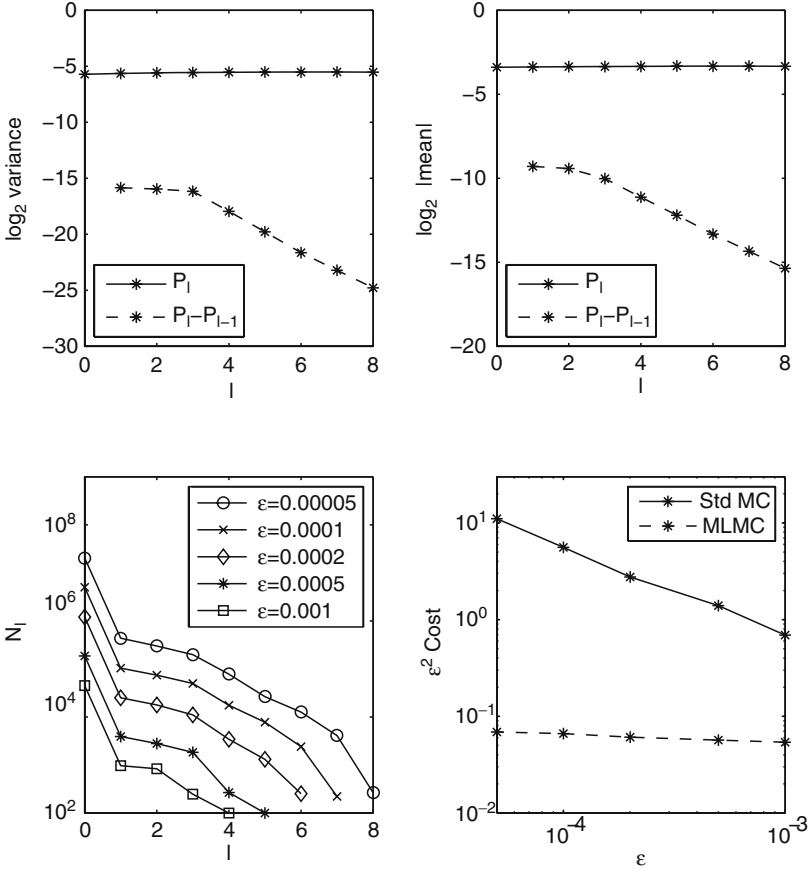


Fig. 3. Barrier option

for multilevel calculations using the Euler discretisation [Gil06]. In that case, the variance V_l was $O(h_l^{1/2})$, because $O(h_l^{1/2})$ of the paths terminate within $O(h_l^{1/2})$ of the strike K , and for these paths there is an $O(1)$ probability that the coarse and fine paths will terminate on opposite sides of the strike, giving an $O(1)$ value for $\hat{P}_l - \hat{P}_{l-1}$. Using the same approach with the Milstein method, there would be $O(h_l)$ of the paths terminating within $O(h_l)$ of the strike K , for which there would be an $O(1)$ probability that the coarse and fine paths would terminate on opposite sides of the strike. This would result in V_l being $O(h_l)$. This corresponds to $\beta = 1$ and would give a computational cost which is $O(\epsilon^{-2}(\log \epsilon)^2)$.

To achieve a better multilevel variance convergence rate, we instead smooth the payoff using the technique of conditional expectation (see section 7.2.3 in [Gla04]), terminating the path calculations one timestep before reaching the terminal time T . If \hat{S}_{n_T-1} denotes the value at this time, then if we

approximate the motion thereafter as a simple Brownian motion with constant drift a_{n_T-1} and volatility b_{n_T-1} , the probability that $\widehat{S}_{n_T} > K$ after one further timestep is

$$\widehat{p} = \Phi \left(\frac{\widehat{S}_{n_T-1} + a_{n_T-1}h - K}{b_{n_T-1}\sqrt{h}} \right), \tag{11}$$

where Φ is the cumulative Normal distribution.

For the fine-path payoff \widehat{P}_l^f we therefore use $\widehat{P}_l^f = \exp(-rT) \widehat{p}$, with $a_{n_T-1} = a(\widehat{S}_{n_T-1}, T - h)$ and $b_{n_T-1} = b(\widehat{S}_{n_T-1}, T - h)$. For the coarse-path payoff, we note that given the Brownian increment ΔW for the first half of the N^{th} timestep, then the probability that $\widehat{S}_{n_T} > K$ is

$$\widehat{p} = \Phi \left(\frac{\widehat{S}_{n_T-1} + a_{n_T-1}h + b_{n_T-1}\Delta W - K}{b_{n_T-1}\sqrt{h/2}} \right). \tag{12}$$

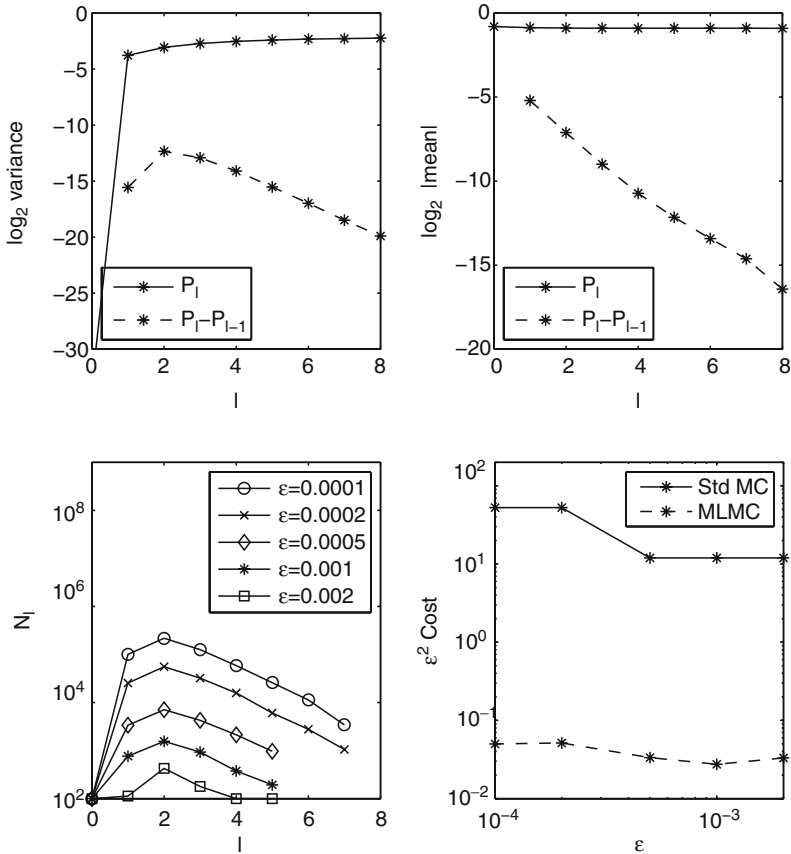


Fig. 4. Digital option

The value for ΔW is taken from the final timestep of the fine-path calculation, which corresponds to the first half of the N^{th} timestep in the coarse-path calculation. The conditional expectation of (12) is equal to (11), and so again equality (5) is satisfied.

Figure 4 shows the numerical results for parameters $S(0) = 1$, $K = 1$, $T = 1$, $r = 0.05$, $\sigma = 0.2$. The top left plot shows that the variance is approximately $O(h_l^{3/2})$, corresponding to $\beta = 1.5$. The reason for this is similar to the argument for the barrier option. $O(h_l^{1/2})$ of the paths have a minimum which lies within $O(h_l^{1/2})$ of the strike, for which the \hat{p} is neither close to zero nor close to unity. The fine path and coarse path trajectories differ by $O(h_l)$, due to the first order strong convergence of the Milstein scheme. Since \hat{p} has an $O(h_l^{-1/2})$ derivative, this results in the difference between \hat{p} for the coarse and fine paths being $O(h_l^{1/2})$, and that results in the variance being $O(h_l^{3/2})$.

One strikingly different feature is that the variance of the level 0 estimator, V_0 , is zero. This is because at level $l = 0$ there would usually be only one timestep, and so here it is not simulated at all; one simply uses equation (11) to evaluate the payoff. This reduces the cost of the multilevel calculations even more than usual, leading to a factor 1000 computational savings for $\epsilon = 10^{-4}$.

4 Conclusions and Future Work

In this paper we have demonstrated numerically the ability of multilevel Monte Carlo path simulation using the Milstein discretisation to achieve an ϵ RMS error for a range of financial options at a computational cost which is $O(\epsilon^{-2})$. This requires the use of Brownian interpolation within each timestep for Asian, lookback and barrier options, and the use of conditional expectation to smooth the payoff of digital options.

There are three major directions for future research. The first is the theoretical analysis of the algorithms presented in this paper, to prove that they do indeed have variance convergence rates with $\beta > 1$. The analysis of earlier algorithms for lookback, barrier and digital options based on the Euler discretisation [Gil06] is currently being developed; it is hoped this can then be extended to the Milstein discretisation for scalar SDEs.

The second is the extension of the algorithms to multi-dimensional SDEs, for which the Milstein discretisation usually requires the simulation of Lévy areas [GL94, Gla04]. Current investigations indicate that this can be avoided for European options with a Lipschitz payoff through the use of antithetic variables. However, the extension to more difficult payoffs, such as the Asian, lookback, barrier and digital options considered in this paper, looks more challenging.

The third direction for future research is the use of quasi-Monte Carlo methods. The analysis in section 2 showed that the optimal number of samples

on level l is proportional to $\sqrt{V_l h_l}$. If $V_l = O(h_l^\beta)$, then this number is proportional to $h_l^{(\beta+1)/2}$. Since the cost of an individual sample is proportional to the number of timesteps, and hence inversely proportional to h_l , the computational cost on level l is proportional to $h_l^{(\beta-1)/2}$. For $\beta > 1$, this shows that the computational effort decreases geometrically as one moves to finer levels of discretisation. Thus, when using the Milstein discretisation most of the computational effort is expended on the coarsest levels of the multilevel computation. For these low dimensional levels it is reasonable to expect that quasi-Monte Carlo methods [KS05, Ecu04, Nie92] will be very much more effective than the standard Monte Carlo methods used in this paper.

Acknowledgements

This research was funded in part by a research grant from Microsoft Corporation, and in part by a Springboard Fellowship from the UK Engineering and Physical Sciences Research Council.

References

- [BGK97] M. Broadie, P. Glasserman, and S. Kou. A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4):325–348, 1997.
- [BT95] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations, I: convergence rate of the distribution function. *Probability Theory and Related Fields*, 104(1):43–60, 1995.
- [Ecu04] P. L’Ecuyer. Quasi-Monte Carlo methods in finance. In R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, editors, *Proceedings of the 2004 Winter Simulation Conference*, pages 1645–1655. IEEE Press, 2004.
- [ES02] M. Emsermann and B. Simon. Improving simulation efficiency with quasi control variates. *Stochastic Models*, 18(3):425–448, 2002.
- [Gil06] M.B. Giles. Multilevel Monte Carlo path simulation. Technical Report NA06/03, Oxford University Computing Laboratory, 2006 (to appear in *Operations Research*).
- [GL94] J.G. Gaines and T.J. Lyons. Random generation of stochastic integrals. *SIAM J. Appl. Math.*, 54(4):1132–1146, 1994.
- [Gla04] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.
- [Hei01] S. Heinrich. *Multilevel Monte Carlo Methods*, volume 2179 of *Lecture Notes in Computer Science*, pages 58–67. Springer-Verlag, 2001.
- [Keb05] A. Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to options pricing. *Annals of Applied Probability*, 14(4):2681–2705, 2005.
- [KP92] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, 1992.
- [KS05] F.Y. Kuo and I.H. Sloan. Lifting the curse of dimensionality. *Notices of the AMS*, 52(11):1320–1328, 2005.

- [Nie92] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, 1992.
- [Spe05] A. Speight. A multilevel approach to control variates. Working paper, Georgia State University, 2005.
- [TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8:483–509, 1990.