Multilevel Monte Carlo path simulation

Michael B. Giles*

Abstract

We show that multigrid ideas can be used to reduce the computational complexity of estimating an expected value arising from a stochastic differential equation using Monte Carlo path simulations. In the simplest case of a Lipschitz payoff and an Euler discretisation, the computational cost to achieve an accuracy of $O(\epsilon)$ is reduced from $O(\epsilon^{-3})$ to $O(\epsilon^{-2}(\log\epsilon)^2)$. The analysis is supported by numerical results showing significant computational savings.

1 Introduction

In Monte Carlo path simulations which are used extensively in computational finance, one is interested in the expected value of a quantity which is a functional of the solution to a stochastic differential equation. To be specific, suppose we have a multi-dimensional SDE with general drift and volatility terms,

$$\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W(t), \quad 0 < t < T, \tag{1}$$

and given initial data S_0 we want to compute the expected value of $f(S(T))$ where $f(S)$ is a scalar function with a uniform Lipschitz bound, i.e. there exists a constant c such that

$$|f(U) - f(V)| \ \le c\,\|U - V\|\,, \quad \forall\ U, V. \tag{2}$$

A simple Euler discretisation of this SDE with timestep h is

$$\widehat{S}_{n+1} = \widehat{S}_n + a(\widehat{S}_n, t_n)\,h + b(\widehat{S}_n, t_n)\,\Delta W_n,$$

and the simplest estimate for $E[f(S_T)]$ is the mean of the payoff values $f(\widehat{S}_{T/h})$, from N independent path simulations,

$$\widehat{Y} = N^{-1}\sum_{i=1}^{N} f(\widehat{S}^{(i)}_{T/h}).$$

*Professor of Scientific Computing, Oxford University Computing Laboratory

It is well established that, provided $a(S,t)$ and $b(S,t)$ satisfy certain conditions [1, 15, 20], the expected mean-square-error (MSE) in the estimate $\widehat{Y}$ is asymptotically of the form

$$\text{MSE} \approx c_1 N^{-1} + c_2 h^2,$$

where c_1, c_2 are positive constants. The first term corresponds to the variance in $\widehat{Y}$ due to the Monte Carlo sampling, and the second term is the square of the $O(h)$ bias introduced by the Euler discretisation.

To make the MSE $O(\epsilon^2)$, so that the r.m.s. error is $O(\epsilon)$, requires that $N = O(\epsilon^{-2})$ and $h = O(\epsilon)$, and hence the computational complexity (cost) is $O(\epsilon^{-3})$ [4]. The main theorem in this paper proves that the computational complexity for this simple case can be reduced to $O\left(\epsilon^{-2}(\log \epsilon)^2\right)$ through the use of a multilevel method which reduces the variance, leaving unchanged the bias due to the Euler discretisation. The multilevel method is very easy to implement and can be combined, in principle, with other variance reduction methods such as stratified sampling [7] and quasi Monte Carlo methods [16, 17, 19] to obtain even greater savings.

The method extends the recent work of Kebaier [14] who proved that the computational cost of the simple problem described above can be reduced to $O(\epsilon^{-2.5})$ through the appropriate combination of results obtained using two levels of timestep, h and $O(h^{1/2})$. This is closely related to a more generally applicable approach of quasi control variates analysed by Emsermann and Simon [5].

Our technique generalises Kebaier's approach to multiple levels, using a geometric sequence of different timesteps $h_l = M^{-l} T$, $l = 0, 1, \ldots, L$, for integer $M \geq 2$, with the smallest timestep h_L corresponding to the original h which determines the size of the Euler discretisation bias. This idea of using a geometric sequence of timesteps comes from the multigrid method for the iterative solution of linear systems of equations arising from the discretisation of elliptic partial differential equations [2, 21]. The multigrid method uses a geometric sequence of grids, each typically twice as fine in each direction as its predecessor. If one were to use only the finest grid, the discretisation error would be very small, but the computational cost of a Jacobi or Gauss-Seidel iteration would be very large. On a much coarser grid, the accuracy is much less, but the cost is also much less. Multigrid solves the equations on the finest grid, by computing corrections using all of the grids, thereby achieving the fine grid accuracy at a much lower cost. This is a very simplified explanation of multigrid, but it is the same essential idea which will be used here, retaining the accuracy/bias associated with the smallest timestep, but using calculations with larger timesteps to reduce the variance in a way that minimises the overall computational complexity.

A similar multilevel Monte Carlo idea has been used by Heinrich [9] for parametric integration, in which one is interested in evaluating a quantity $I(\lambda)$ which is defined as a multi-dimensional integral of a function which has a parametric dependence on λ. Although the details of the method are quite different from Monte Carlo path simulation, the analysis of the computational complexity is quite similar.

The paper begins with the introduction of the new multilevel method and an outline of its asymptotic accuracy and computational complexity for the simple problem described above. The main theorem and its proof are then presented. This establishes the computational complexity for a broad category of applications and numerical discretisations with certain properties. The applicability of the theorem to the Euler discretisation is a consequence of its well-established weak and strong convergence properties. The paper then discusses some refinements to the method and its implementation, and the effects of different payoff functions and numerical discretisations. Finally, numerical results are presented to provide support for the theoretical analysis, and directions for further research are outlined.

2 Multilevel Monte Carlo method

Consider Monte Carlo path simulations with different timesteps $h_l = M^{-l}T$, $l = 0, 1, \ldots, L$. For a given Brownian path $W(t)$, let P denote the payoff $f(S(T))$, and let $\widehat{S}_{l,M^l}$ and $\widehat{P}_l$ denote the approximations to $S(T)$ and P using a numerical discretisation with timestep h_l.

It is clearly true that

$$E[\widehat{P}_L] = E[\widehat{P}_0] + \sum_{l=1}^{L} E[\widehat{P}_l - \widehat{P}_{l-1}].$$

The multilevel method independently estimates each of the expectations on the right-hand side in a way which minimises the computational complexity.

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 that one might use 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. This

is easily implemented by first constructing the Brownian increments for the simulation of the discrete path leading to the evaluation of $\widehat{P}_l^{(i)}$, and then summing them in groups of size M to give the discrete Brownian increments for the evaluation of $\widehat{P}_{l-1}^{(i)}$. 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. The same inverse dependence on N_l would apply in the case of a more sophisticated estimator using stratified sampling or a zero-mean control variate to reduce the variance.

The variance of the combined estimator $\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l$ is

$$V[\widehat{Y}] = \sum_{l=0}^{L} N_l^{-1} V_l.$$

The computational cost, if one ignores the asymptotically negligible cost of the final payoff evaluation, 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 $\sqrt{V_l \, h_l}$. This calculation of an optimal number of samples N_l is similar to the approach used in optimal stratified sampling [7], except that in this case we also include the effect of the different computational cost of the samples on different levels.

The above analysis holds for any value of L. We now assume that $L \gg 1$, and consider the behaviour of V_l as $l \to \infty$. In the particular case of the Euler discretisation and the Lipschitz payoff function, provided $a(S,t)$ and $b(S,t)$ satisfy certain conditions [1, 15, 20], there is $O(h)$ weak convergence and $O(h^{1/2})$ strong convergence. Hence, as $l \to \infty$,

$$E[\widehat{P}_l - P] = O(h_l), \tag{4}$$

and

$$E[\, \|\widehat{S}_{l,M^l} - S(T)\|^2 \,] = O(h_l). \tag{5}$$

From the Lipschitz property (2), it follows that

$$V[\widehat{P}_l - P] \;\le\; E[\,(\widehat{P}_l - P)^2\,] \;\le\; c^2 \, E[\, \|\widehat{S}_{l,M^l} - S(T)\|^2 \,].$$

Combining this with (5) gives $V[\widehat{P}_l - P] = O(h_l)$. Furthermore,

$$\begin{aligned} (\widehat{P}_l - \widehat{P}_{l-1}) \;&=\; (\widehat{P}_l - P) \;-\; (\widehat{P}_{l-1} - P) \\ \Longrightarrow \quad V[\widehat{P}_l - \widehat{P}_{l-1}] \;&\le\; \Big((V[\widehat{P}_l - P])^{1/2} + (V[\widehat{P}_{l-1} - P])^{1/2} \Big)^2 . \end{aligned}$$

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 now chosen such that

$$L = \frac{\log \epsilon^{-1}}{\log M} + O(1),$$

as $\epsilon \to 0$, then $h_L = M^{-L} = O(\epsilon)$, and so the bias error $E[\widehat{P}_L - P]$ is $O(\epsilon)$, due to (4). Consequently, we obtain a MSE which is $O(\epsilon^2)$, with a computational complexity which is $O(\epsilon^{-2} L^2) = O(\epsilon^{-2} (\log \epsilon)^2)$.

3 Complexity theorem

The main theorem is worded quite generally so that it can be applied to a variety of financial models with output functionals which are not necessarily Lipschitz functions of the terminal state but may instead be a discontinuous function of the terminal state, or even path-dependent as in the case of barrier and lookback options. The theorem also does not specify which numerical approximation is used. Instead, it proves a result concerning the computational complexity of the multilevel method conditional on certain features of the underlying numerical approximation and the multilevel estimators. This approach is similar to that used by Duffie and Glynn [4].

Theorem 3.1 *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 Using the notation $\lceil x \rceil$ to denote the unique integer n satisfying the inequalities $x \leq n < x+1$, we start by choosing L to be

$$L = \left\lceil \frac{\log(\sqrt{2}\,c_1\,T^\alpha\,\epsilon^{-1})}{\alpha\,\log M} \right\rceil,$$

so that

$$\tfrac{1}{\sqrt{2}}\,M^{-\alpha}\epsilon \;<\; c_1\,h_L^\alpha \;\leq\; \tfrac{1}{\sqrt{2}}\,\epsilon, \tag{6}$$

and hence, because of properties i) and ii),

$$\left(E[\widehat{Y}] - E[P]\right)^2 \leq \tfrac{1}{2}\,\epsilon^2.$$

This $\frac{1}{2}\epsilon^2$ upper bound on the square of the bias error, together with the $\frac{1}{2}\epsilon^2$ upper bound on the variance of the estimator to be proved later, gives an ϵ^2 upper bound on the estimator MSE.

Also,

$$\sum_{l=0}^{L} h_l^{-1} \;=\; h_L^{-1} \sum_{l=0}^{L} M^{-l} \;<\; \frac{M}{M-1}\; h_L^{-1}$$

using the standard result for a geometric series, and

$$h_L^{-1} < M\left(\frac{\epsilon}{\sqrt{2}\,c_1}\right)^{-1/\alpha}.$$

due to the first inequality in (6). These two inequalities, combined with the observation that $\epsilon^{-1/\alpha} \leq \epsilon^{-2}$ for $\alpha \geq \frac{1}{2}$ and $\epsilon < e^{-1}$, give the following result which will be used later,

$$\sum_{l=0}^{L} h_l^{-1} \;<\; \frac{M^2}{M-1}\;\left(\sqrt{2}\,c_1\right)^{1/\alpha}\epsilon^{-2}. \tag{7}$$

We now need to consider the different possible values for β.

a) If $\beta=1$, we set $N_l = \lceil 2\,\epsilon^{-2}\,(L+1)\,c_2\,h_l \rceil$ so that

$$V[\widehat{Y}] \;=\; \sum_{l=0}^{L} V[\widehat{Y}_l] \;\leq\; \sum_{l=0}^{L} c_2\,N_l^{-1} h_l \;\leq\; \tfrac{1}{2}\,\epsilon^2,$$

which is the required upper bound on the variance of the estimator.

To bound the computational complexity C we begin with an upper bound on L given by

$$L \leq \frac{\log \epsilon^{-1}}{\alpha\,\log M} + \frac{\log(\sqrt{2}\,c_1\,T^{\alpha})}{\alpha\,\log M} + 1.$$

Given that $1<\log \epsilon^{-1}$ for $\epsilon<e^{-1}$, it follows that

$$L+1 \leq c_5\,\log \epsilon^{-1},$$

where

$$c_5 = \frac{1}{\alpha\,\log M} + \max\left(0, \frac{\log(\sqrt{2}\,c_1\,T^{\alpha})}{\alpha\,\log M}\right) + 2.$$

Upper bounds for N_l are given by

$$N_l \leq 2\epsilon^{-2}\,(L+1)\,c_2\,h_l + 1.$$

Hence the computational complexity is bounded by

$$C \;\leq\; c_3 \sum_{l=0}^{L} N_l\,h_l^{-1} \;\leq\; c_3\left(2\,\epsilon^{-2}(L+1)^2\,c_2 + \sum_{l=0}^{L} h_l^{-1}\right)$$

Using the upper bound for $L+1$ and inequality (7), and the fact that $1<\log \epsilon^{-1}$ for $\epsilon<e^{-1}$, it follows that $C \leq c_4\,\epsilon^{-2}(\log \epsilon)^2$ where

$$c_4 = 2\,c_3\,c_5^2\,c_2 + c_3\,\frac{M^2}{M-1}\;\left(\sqrt{2}\,c_1\right)^{1/\alpha}.$$

b) For $\beta>1$, setting

$$N_l = \left\lceil 2\,\epsilon^{-2}\,c_2\,T^{(\beta-1)/2}\;\left(1-M^{-(\beta-1)/2}\right)^{-1} h_l^{(\beta+1)/2} \right\rceil,$$

then

$$\sum_{l=0}^{L} V[\widehat{Y}_l] \;\leq\; \tfrac{1}{2}\,\epsilon^2\,T^{-(\beta-1)/2}\left(1-M^{-(\beta-1)/2}\right)\;\sum_{l=0}^{L} h_l^{(\beta-1)/2}.$$

Using the standard result for a geometric series,

$$\begin{aligned}\sum_{l=0}^{L} h_l^{(\beta-1)/2} \;&=\; T^{(\beta-1)/2} \sum_{l=0}^{L}\left(M^{-(\beta-1)/2}\right)^l \\ &<\; T^{(\beta-1)/2}\left(1-M^{-(\beta-1)/2}\right)^{-1}, \qquad (8)\end{aligned}$$

and hence we obtain an $\frac{1}{2}\,\epsilon^2$ upper bound on the variance of the estimator.

Using the N_l upper bound

$$N_l < 2\,\epsilon^{-2}\,c_2\,T^{(\beta-1)/2}\,\left(1-M^{-(\beta-1)/2}\right)^{-1}\,h_l^{(\beta+1)/2}+1,$$

the computational complexity is bounded by

$$C \le c_3\left(2\,\epsilon^{-2}\,c_2\,T^{(\beta-1)/2}\,\left(1-M^{-(\beta-1)/2}\right)^{-1}\sum_{l=0}^{L}h_l^{(\beta-1)/2}+\sum_{l=0}^{L}h_l^{-1}\right).$$

Using inequalities (7) and (8) gives $C \le c_4\,\epsilon^{-2}$ where

$$c_4 = 2\,c_3\,c_2\,T^{\,\beta-1}\,\left(1-M^{-(\beta-1)/2}\right)^{-2}+\;c_3\,\frac{M^2}{M-1}\,\left(\sqrt{2}\,c_1\right)^{1/\alpha}.$$

c) For $\beta<1$, setting

$$N_l = \left\lceil 2\,\epsilon^{-2}c_2\,h_L^{-(1-\beta)/2}\left(1-M^{-(1-\beta)/2}\right)^{-1}\,h_l^{(\beta+1)/2}\right\rceil,$$

then

$$\sum_{l=0}^{L}V[\widehat{Y}_l]\;<\;\tfrac{1}{2}\,\epsilon^2\,h_L^{(1-\beta)/2}\left(1-M^{-(1-\beta)/2}\right)\;\sum_{l=0}^{L}h_l^{-(1-\beta)/2}.$$

Since

$$\begin{aligned}\sum_{l=0}^{L}h_l^{-(1-\beta)/2} &= h_L^{-(1-\beta)/2}\sum_{l=0}^{L}\left(M^{-(1-\beta)/2}\right)^l\\ &< h_L^{-(1-\beta)/2}\,\left(1-M^{-(1-\beta)/2}\right)^{-1},\end{aligned}\qquad(9)$$

we again obtain an $\frac{1}{2}\,\epsilon^2$ upper bound on the variance of the estimator.

Using the N_l upper bound

$$N_l < 2\,\epsilon^{-2}\,c_2\,h_L^{-(1-\beta)/2}\,\left(1-M^{-(1-\beta)/2}\right)^{-1}\,h_l^{(\beta+1)/2}+1,$$

the computational complexity is bounded by

$$C \le c_3\left(2\,\epsilon^{-2}\,c_2\,h_L^{-(1-\beta)/2}\,\left(1-M^{-(1-\beta)/2}\right)^{-1}\sum_{l=0}^{L}h_l^{-(1-\beta)/2}+\sum_{l=0}^{L}h_l^{-1}\right).$$

Using inequality (9) gives

$$h_L^{-(1-\beta)/2}\,\left(1-M^{-(1-\beta)/2}\right)^{-1}\sum_{l=0}^{L}h_l^{-(1-\beta)/2} < h_L^{-(1-\beta)}\,\left(1-M^{-(1-\beta)/2}\right)^{-2}.$$

The first inequality in (6) gives

$$h_L^{-(1-\beta)} < \left(\sqrt{2}\,c_1\right)^{(1-\beta)/\alpha} M^{1-\beta}\,\epsilon^{-(1-\beta)/\alpha}.$$

Combining the above two inequalities, and also using inequality (7) and the fact that $\epsilon^{-2} < \epsilon^{-2-(1-\beta)/\alpha}$ for $\epsilon < e^{-1}$, gives $C \le c_4\,\epsilon^{-2-(1-\beta)/\alpha}$ where

$$c_4 = 2\,c_3\,c_2\,\left(\sqrt{2}\,c_1\right)^{(1-\beta)/\alpha} M^{1-\beta}\left(1-M^{-(1-\beta)/2}\right)^{-2} + \; c_3\,\frac{M^2}{M-1}\;\left(\sqrt{2}\,c_1\right)^{1/\alpha}.$$

□

The theorem and proof show the importance of the parameter β which defines the convergence of the variance V_l as $l \to \infty$. In this limit, the optimal N_l is proportional to $\sqrt{V_l\,h_l} = O(h_l^{(\beta+1)/2})$, and hence the computational effort $N_l h_l^{-1}$ is proportional to $O(h_l^{(\beta-1)/2})$. This shows that for $\beta > 1$ the computational effort is primarily expended on the coarsest levels, for $\beta < 1$ it is on the finest levels, and for $\beta = 1$ it is roughly evenly spread across all levels.

In applying the theorem in different contexts, there will often be existing literature on weak convergence which will establish the correct exponent α for condition i). Constructing estimators with properties ii) and iv) is also straightforward. The main challenge will be in determining and proving the appropriate exponent β for iii). An even bigger challenge might be to develop better estimators with a higher value for β.

In the case of the Euler discretisation with a Lipschitz payoff, there is existing literature on the conditions on $a(S,t)$ and $b(S,t)$ for $O(h)$ weak convergence and $O(h^{1/2})$ strong convergence [1, 15, 20], which in turn gives $\beta = 1$ as explained earlier.

The convergence is degraded if the payoff function $f(S(T))$ has a discontinuity. In this case, for a given timestep h_l, a fraction of the paths of size $O(h_l^{1/2})$ will have a final $\widehat{S}_{l,M^l}$ which is $O(h_l^{1/2})$ from the discontinuity. With the Euler discretisation, this fraction of the paths have an $O(1)$ probability of $\widehat{P}_l - \widehat{P}_{l-1}$ being $O(1)$, due to $\widehat{S}_{l,M^l}$ and $\widehat{S}_{l-1,M^{l-1}}$ being on opposite sides of the discontinuity, and therefore $V_l = O(h_l^{1/2})$ and $\beta = \frac{1}{2}$. Because the weak order of convergence is still $O(h_l)$ [1] so $\alpha = 1$, the overall complexity is $O(\epsilon^{-2.5})$, which is still better than the $O(\epsilon^{-3})$ complexity of the standard Monte Carlo method with an Euler discretisation. Further improvement may be possible through the use of adaptive sampling techniques which increase the sampling of those paths with large values for $\widehat{P}_l - \widehat{P}_{l-1}$ [8, 13, 18].

If the Euler discretisation is replaced by Milstein's method for a scalar SDE, its $O(h)$ strong convergence results in $V_l = O(h_l^2)$ for a Lipschitz payoff. Current research is investigating how to achieve a similar improvement in

the convergence rate for lookback, barrier and digital options, based on the appropriate use of Brownian interpolation [7], as well as the extension to multi-dimensional SDEs.

4 Extensions

4.1 Optimal M

The analysis so far has not specified the value of the integer M, which is the factor by which the timestep is refined at each level. In the multigrid method for the iterative solution of discretisations of elliptic PDEs, it is usually optimal to use $M=2$, but that is not necessarily the case with the multilevel Monte Carlo method introduced in this paper.

For the simple Euler discretisation with a Lipschitz payoff, $V[\widehat{P}_l - P] \approx c_0\, h_l$ asymptotically, for some positive constant c_0. This corresponds to the case $\beta=1$ in Theorem 3.1. From the identity

$$(\widehat{P}_l - \widehat{P}_{l-1}) = (\widehat{P}_l - P) \;-\; (\widehat{P}_{l-1} - P)$$

we obtain, asymptotically, the upper and lower bounds

$$\left(\sqrt{M}-1\right)^2 c_0\, h_l \le V[\widehat{P}_l - \widehat{P}_{l-1}] \le \left(\sqrt{M}+1\right)^2 c_0\, h_l,$$

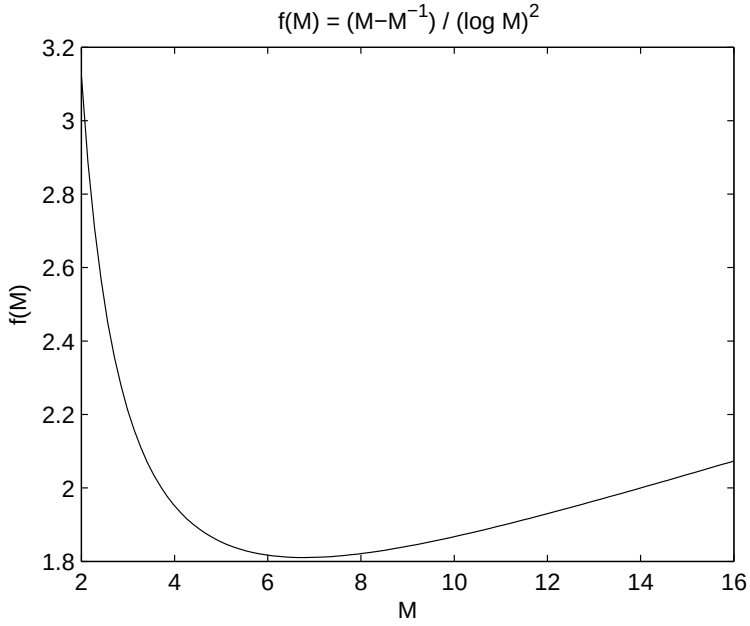

Figure 1: A plot of the function $(M-M^{-1})/(\log M)^2$

with the two extremes corresponding to perfect correlation and anti-correlation between $\widehat{P}_l - P$ and $\widehat{P}_{l-1} - P$.

Suppose now that the value of $V[\widehat{P}_l - \widehat{P}_{l-1}]$ is given approximately by the geometric mean of the two bounds,

$$V[\widehat{P}_l - \widehat{P}_{l-1}] \approx (M-1)\, c_0\, h_l,$$

which corresponds to $c_2 = (M-1)c_0$ in Theorem 3.1. This results in

$$N_l \approx 2\epsilon^{-2}(L+1)\,(M-1)\,c_0\,h_l.$$

and so the computational cost of evaluating $\widehat{Y}_l$ is proportional to

$$N_l\,(h_l^{-1}+h_{l-1}^{-1}) = N_l\,h_l^{-1}\,(1+M^{-1}) \approx 2\,\epsilon^{-2}(L+1)\,(M-M^{-1})\,c_0.$$

Since $L = O(\log \epsilon^{-1}/\log M)$, summing the costs of all levels, we conclude that asymptotically, as $\epsilon \to 0$, the total computational cost is roughly proportional to

$$2\,\epsilon^{-2}(\log \epsilon)^2\ f(M),$$

where

$$f(M) = \frac{M-M^{-1}}{(\log M)^2}.$$

This function is illustrated in Figure 1. Its minimum near $M=7$ is about half the value at $M=2$, giving twice the computational efficiency. The numerical results presented later are all obtained using $M=4$. This gives most of the benefits of a larger value of M, but at the same time M is small enough to give a reasonable number of levels from which to estimate the bias, as explained in the next section.

4.2 Bias estimation and Richardson extrapolation

In the multilevel method, the estimates for the correction $E[\widehat{P}_l - \widehat{P}_{l-1}]$ at each level give information which can be used to estimate the remaining bias. In particular, for the Euler discretisation with a Lipschitz payoff, asymptotically, as $l \to \infty$

$$E[P - \widehat{P}_l] \approx c_1 h_l,$$

for some constant c_1 and hence

$$E[\widehat{P}_l - \widehat{P}_{l-1}] \ \approx\ (M-1)\,c_1 h_l \ \approx\ (M-1)\,E[P-\widehat{P}_l].$$

This information can be used in one of two ways. The first is to use it as an approximate bound on the remaining bias, so that to obtain a bias which has magnitude less than $\epsilon/\sqrt{2}$ one increases the value for L until

$$\left|\widehat{Y}_L\right| < \tfrac{1}{\sqrt{2}}\,(M-1)\,\epsilon.$$

Being more cautious, the condition which we use in the numerical results presented later is

$$\max\left\{M^{-1}\left|\widehat{Y}_{L-1}\right|,\left|\widehat{Y}_{L}\right|\right\}<\tfrac{1}{\sqrt{2}}\,(M-1)\,\epsilon. \tag{10}$$

This ensures that the remaining error based on an extrapolation from either of the two finest timesteps is within the desired range. This modification is designed to avoid possible problems due to a change in sign of the correction, $E[\widehat{P}_l-\widehat{P}_{l-1}]$, on successive refinement levels.

An alternative approach is to use Richardson extrapolation to eliminate the leading order bias. Since $E[P-\widehat{P}_L]\approx(M-1)^{-1}E[\widehat{P}_L-\widehat{P}_{L-1}]$, by changing the combined estimator to

$$\left(\sum_{l=0}^{L}\widehat{Y}_l\right)+(M-1)^{-1}\widehat{Y}_L=\frac{M}{M-1}\left\{\widehat{Y}_0+\sum_{l=1}^{L}\left(\widehat{Y}_l-M^{-1}\widehat{Y}_{l-1}\right)\right\},$$

the leading order bias is eliminated and the remaining bias is $o(h_L)$, usually either $O(h_L^{3/2})$ or $O(h_L^2)$. The advantage of re-writing the new combined estimator in the form shown above on the right-hand-side, is that one can monitor the convergence of the terms $\widehat{Y}_l-M^{-1}\widehat{Y}_{l-1}$ to decide when the remaining bias is sufficiently small, in exactly the same way as described previously for $\widehat{Y}_l$. Assuming the remaining bias is $O(h_L^2)$, the appropriate convergence test is

$$\left|\widehat{Y}_L-M^{-1}\widehat{Y}_{L-1}\right|<\tfrac{1}{\sqrt{2}}\,(M^2-1)\,\epsilon. \tag{11}$$

5 Numerical algorithm

Putting together the elements already discussed, the multilevel algorithm used for the numerical tests is as follows:

1. start with $L=0$
2. estimate V_L using an initial $N_L=10^4$ samples
3. define optimal $N_l,\ l=0,\ldots,L$ using Eqn. (12)
4. evaluate extra samples at each level as needed for new N_l
5. if $L\geq 2$, test for convergence using Eqn. (10) or Eqn. (11)
6. if $L<2$ or not converged, set $L:=L+1$ and go to 2.

The equation for the optimal N_l is

$$N_l = \left\lceil 2\,\epsilon^{-2}\,\sqrt{V_l\,h_l}\;\left(\sum_{l=0}^{L}\sqrt{V_l/h_l}\right)\right\rceil. \tag{12}$$

This makes the estimated variance of the combined multilevel estimator less than $\frac{1}{2}\,\epsilon^2$, while Equation (10) tries to ensure that the bias is less than $\frac{1}{\sqrt{2}}\,\epsilon$. Together, they should give a MSE which is less than ϵ^2, with ϵ being a user-specified r.m.s. accuracy.

In step 4, the optimal N_l from step 3 is compared to the number of samples already calculated at that level. If the optimal N_l is larger, then the appropriate number of additional samples are calculated. The estimate for V_l is then updated, and this improved estimate is used if step 3 is re-visited.

It is important to note that this algorithm is heuristic; it is not guaranteed to achieve a MSE error which is $O(\epsilon^2)$. The main theorem in Section 3 does provide a guarantee, but the conditions of the theorem assume *a priori* knowledge of the constants c_1 and c_2 governing the weak convergence and the variance convergence as $h \to 0$. These two constants are in effect being estimated in the numerical algorithm described above.

The accuracy of the variance estimate at each level depends on the size of the initial sample set. If this initial sample size were made proportional to ϵ^{-p} for some exponent $0 < p < 2 - 1/\alpha$, then as $\epsilon \to 0$ it could be proved that the variance estimate will converge to the true value with probability 1, without an increase in the order of the computational complexity.

The weakness in the heuristic algorithm lies in the bias estimation, and it does not appear to be easily resolved. Suppose the numerical algorithm determines that L levels are required. If $p(S)$ represents the probability density function for the final state $S(T)$ defined by the SDE, and $p_l(S)$, $l = 0, 1, \ldots L$ are the corresponding probability densities for the level l numerical approximations, then in general $p(S)$ and the $p_l(S)$ are likely to be linearly independent, and so

$$p(S) = g(S) + \sum_{l=0}^{L} a_l\,p_l(S),$$

for some set of coefficients a_l and a non-zero function $g(S)$ which is orthogonal to the $p_l(S)$. If we consider $g(S)$ to be an increment to the payoff function, then its numerical expectation on each level is zero, since

$$E_{p_l}[g] = \int g(S)\,p_l(S)\,\mathrm{d}S = 0,$$

while its true expectation is

$$E_p[g] = \int g(S)\, p(S)\, \mathrm{d}S = \int g^2(S)\, \mathrm{d}S > 0.$$

Hence, by adding an arbitrary amount of $g(S)$ to the payoff, we obtain an arbitrary perturbation of the true expected payoff, but the heuristic algorithm will on average terminate at the same level L with the same expected value.

This is a fundamental problem which also applies to the standard Monte Carlo algorithm. In practice, it may require additional *a priori* knowledge or experience to choose an appropriate minimum value for L to achieve a given accuracy. Being cautious, one is likely to use a value for L which is larger than required in most cases. In this case, the use of the multilevel method will yield significant additional benefits. For the standard Monte Carlo method, the computational cost is proportional to M^L, the number of timesteps on the finest level, whereas for the multilevel method with the Euler discretisation and a Lipschitz payoff the cost is proportional to L^2. Thus the computational cost of being cautious in the choice of L is much less severe for the multilevel algorithm than for the standard Monte Carlo.

Even better would be a multilevel application with a variance convergence rate $\beta > 1$; for this the computational cost is approximately independent of L, suggesting that one could use a value for L which is much larger than necessary. If there is a known value for L which is guaranteed to give a bias which is much less than ϵ, then it may be possible to define a numerical algorithm which will provably achieve a MSE error of ϵ^2 at a cost which is $O(\epsilon^{-2})$; this will be an area for future research.

In reporting the numerical results later, we define the computational cost as the total number of timesteps performed on all levels,

$$C = N_0 + \sum_{l=1}^{L} N_l\, (M^l + M^{l-1}).$$

The term $M^l + M^{l-1}$ reflects the fact that each sample at level $l > 0$ requires the computation of one fine path with M^l timesteps and one coarse path with M^{l-1} timesteps.

The computational costs are compared to those of the standard Monte Carlo method, which is calculated as

$$C^* = \sum_{l=0}^{L} N_l^* M^l,$$

where $N_l^* = 2\,\epsilon^{-2}\, V[P_l]$ so that the variance of the estimator is $\frac{1}{2}\,\epsilon^2$ as with the multilevel method. The summation over the grid levels corresponds to an

application of the standard Monte Carlo algorithm on each grid level to enable the estimation of the bias in order to apply the same heuristic termination criterion as the multilevel method.

Results are also shown for Richardson extrapolation in conjunction with both the multilevel and standard Monte Carlo methods. The costs for these are defined in the same way; the difference is in the choice of L, and the definition of the extrapolated estimator which has a slightly different variance.

6 Numerical results

6.1 Geometric Brownian motion

Figures 2-5 present results for a simple geometric Brownian motion,

$$\mathrm{d}S = r\,S\,\mathrm{d}t + \sigma\,S\,\mathrm{d}W, \quad 0 < t < 1,$$

with $S(0)=1$, $r=0.05$ and $\sigma=0.2$, and four different payoff options.

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

6.1.1 European option

The results in Figure 2 are for the European call option for which the discounted payoff function is

$$P = \exp(-r)\,\max(0, S(1)-1).$$

The top left plot shows the behaviour of the variance of both $\widehat{P}_l$ and $\widehat{P}_l - \widehat{P}_{l-1}$. The quantity which is plotted is the logarithm base M ($M=4$ for all numerical results in this paper) versus the grid level. The reason for this choice is that a slope of -1 corresponds to a variance which is exactly proportional to M^{-l}, which in turn is proportional to h_l. The slope of the line for $\widehat{P}_l - \widehat{P}_{l-1}$ is indeed approximately -1, indicating that $V_l = V[\widehat{P}_l - \widehat{P}_{l-1}] = O(h)$. For $l=4$, V_l is 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 the mean value and correction at each level. These two plots are both based on results from 4×10^6 paths. The slope of

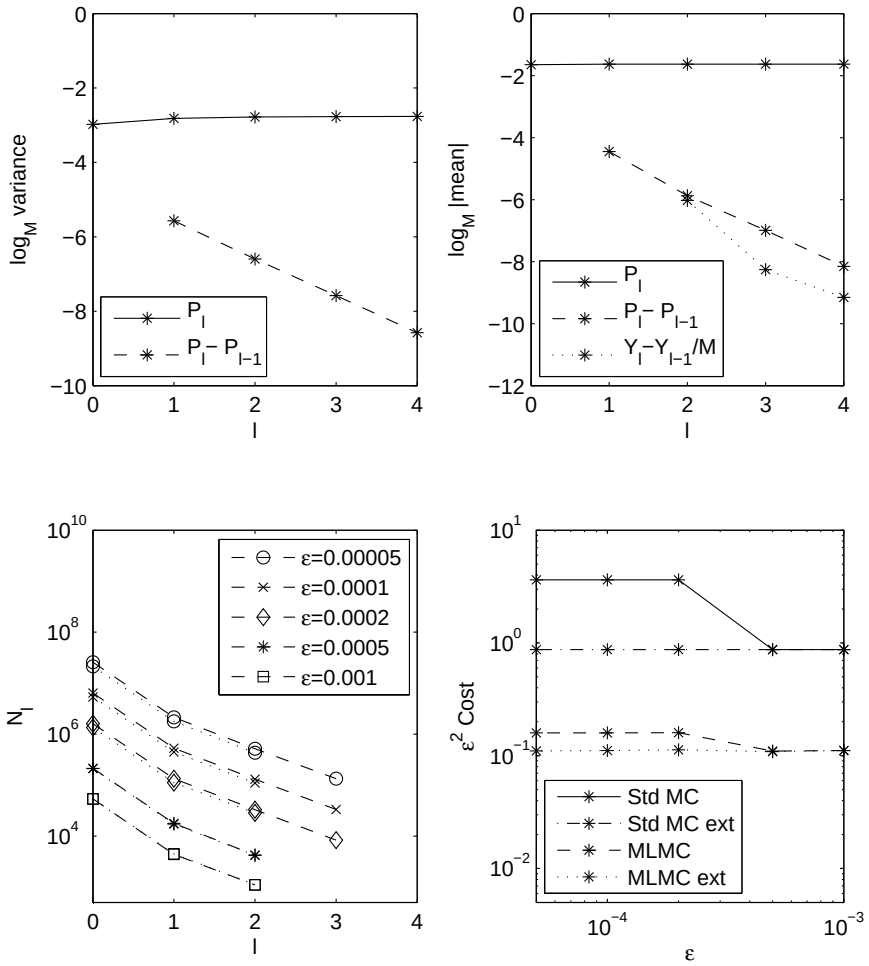

Figure 2: Geometric Brownian motion with European option (value ≈ 0.10).

approximately -1 again implies an $O(h)$ convergence of $E[\widehat{P}_l - \widehat{P}_{l-1}]$. Even at $l=3$, the relative error $E[P-\widehat{P}_l]/E[P]$ is less than 10^{-3}. Also plotted is a line for the multilevel method with Richardson extrapolation, showing significantly faster weak convergence.

The bottom two plots have results from two sets of multilevel calculations, with and without Richardson extrapolation, for five different values of ϵ. Each line in the bottom left plot corresponds to one multilevel calculation and shows the values for $N_l, l = 0, \ldots, L$, 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.

The bottom right plot shows the variation of the computational complexity C (as defined in the previous section) with the desired accuracy ϵ. The plot is of $\epsilon^2 C$ versus ϵ, because we expect to see that $\epsilon^2 C$ is only very weakly

dependent on ϵ for the multilevel method. Indeed, it can be seen that without Richardson extrapolation $\epsilon^2 C$ is a very slowly increasing function of ϵ^{-1} for the multilevel methods, in agreement with the theory which predicts it to be asymptotically proportional to $(\log \epsilon)^2$. For the standard Monte Carlo method, theory predicts that $\epsilon^2 C$ should be proportional to the number of timesteps on the finest level, which in turn is roughly proportional to ϵ^{-1} due to the weak convergence property. This can be seen in the figure, with the "staircase" effect corresponding to the fact that $L=2$ for $\epsilon=0.001, 0.0005$ and $L=3$ for $\epsilon=0.0002, 0.0001, 0.0005$.

With Richardson extrapolation, *a priori* theoretical analysis predicts that $\epsilon^2 C$ for the standard Monte Carlo method should be approximately proportional to $\epsilon^{-1/2}$. However, with extrapolation the numerical results require no more than the minimum two levels of refinement to achieve the desired accuracy, and so $\epsilon^2 C$ is found to be independent of ϵ for the range of ϵ in the tests. Nevertheless, for the most accurate case with $\epsilon=5\times10^{-5}$, the multilevel method is still approximately 10 times more efficient than the standard Monte Carlo method when using extrapolation, and more than 60 times more efficient without extrapolation.

As a final check on the reliability of the heuristics in the multilevel numerical algorithm, ten sets of multilevel calculations have been performed for each value of ϵ, and the root-mean-square-error (RMSE) is computed and compared to the target accuracy of ϵ. For all cases, with and without Richardson extrapolation, the ratio RMSE/ϵ was found to be in the range 0.43–0.96, indicating that the algorithm is correctly achieving the desired accuracy.

6.1.2 Asian option

Figure 3 has results for the Asian option payoff, $P = \exp(-r)\, \max\left(0, \overline{S}-1\right)$, where

$$\overline{S} = \int_0^1 S(t)\, \mathrm{d}t,$$

which is approximated numerically by

$$\overline{S}_l = \sum_{n=1}^{N_l} \tfrac{1}{2}\left(\widehat{S}_n + \widehat{S}_{n-1}\right) h_l.$$

The $O(h_l)$ convergence of both V_l and $E[P_l - P_{l-1}]$ is similar to the European option case, but in this case the Richardson extrapolation does not seem to have improved the order of weak convergence. Hence, the reliability of the bias estimation and grid level termination must be questioned for the Richardson extrapolation. Without extrapolation, the multilevel method is up to 30 times more efficient than the standard Monte Carlo method.

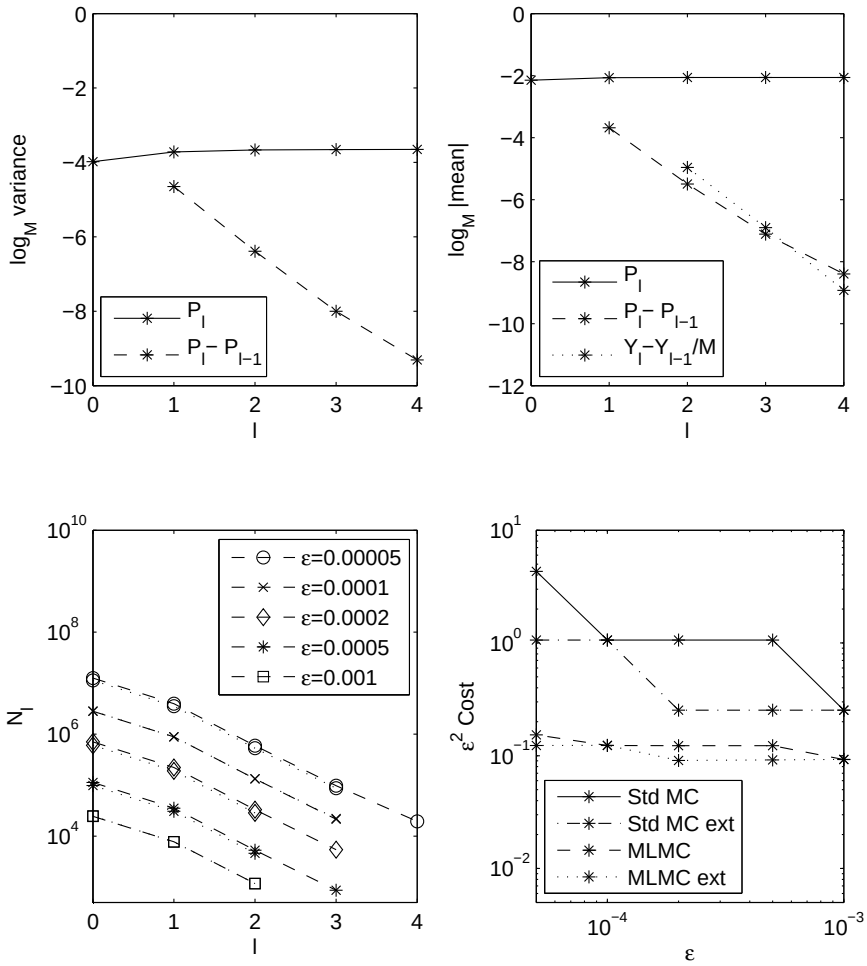

Figure 3: Geometric Brownian motion with Asian option (value ≈ 0.058).

6.1.3 Lookback option

The results in Figure 4 are for the lookback option

$$P = \exp(-r)\left(S(1) - \min_{0<t<1} S(t)\right).$$

The minimum value of $S(t)$ over the path is approximated numerically by

$$\widehat{S}_{min,l} = \left(\min_n \widehat{S}_n\right)\left(1 - \beta^* \sigma \sqrt{h_l}\right).$$

$\beta^* \approx 0.5826$ 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 [3]. Richardson extrapolation clearly works well in this case, improving the weak convergence to second order. This has a significant effect on the number

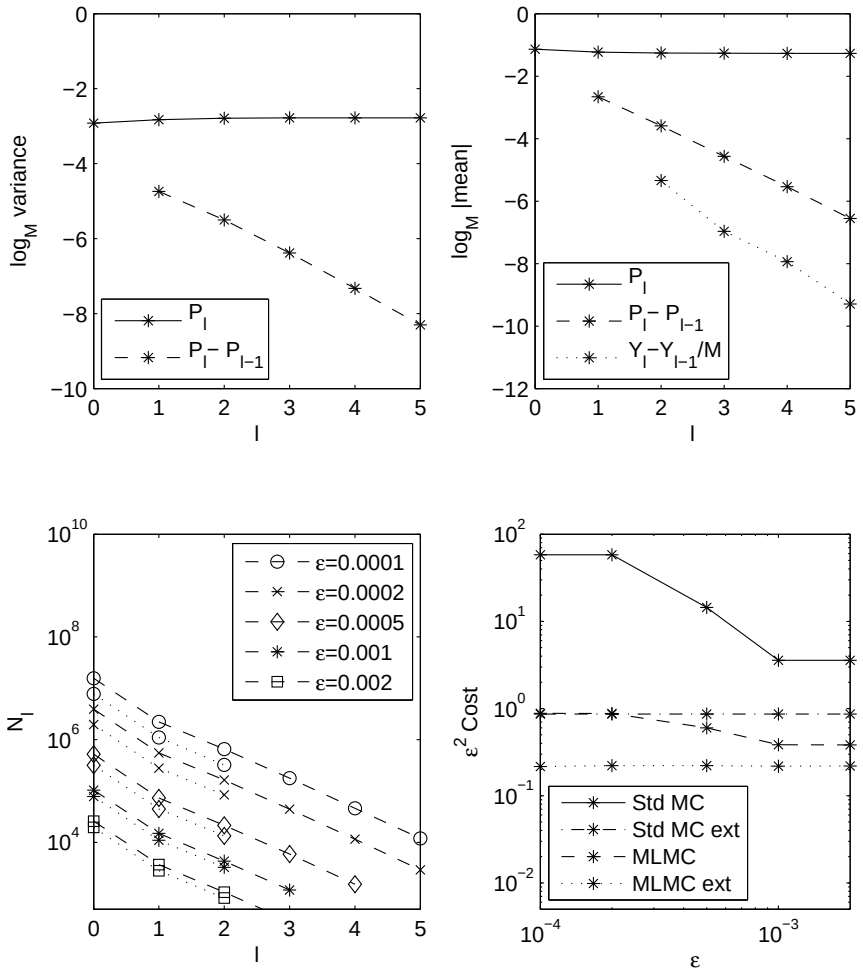

Figure 4: Geometric Brownian motion with lookback option (value ≈ 0.17).

of grid levels required, so that the multilevel method gives savings of up to factor 65 without extrapolation, but up to only 4 with extrapolation.

6.1.4 Digital option

The final payoff which is considered is a digital option, $P = \exp(-r)\, H(S(1)–1)$ where $H(x)$ is the Heaviside function. The results in Figure 5 show that $V_l = O(h_l^{1/2})$, instead of the $O(h_l)$ convergence of all of the previous options. Because of this, much larger values for N_l on the finer refinement levels are required to achieve comparable accuracy, and the efficiency gains of the multilevel method are reduced accordingly. Richardson extrapolation is extremely effective in this case, although the resulting order of weak convergence is unclear, but the multilevel method still offers some additional computational

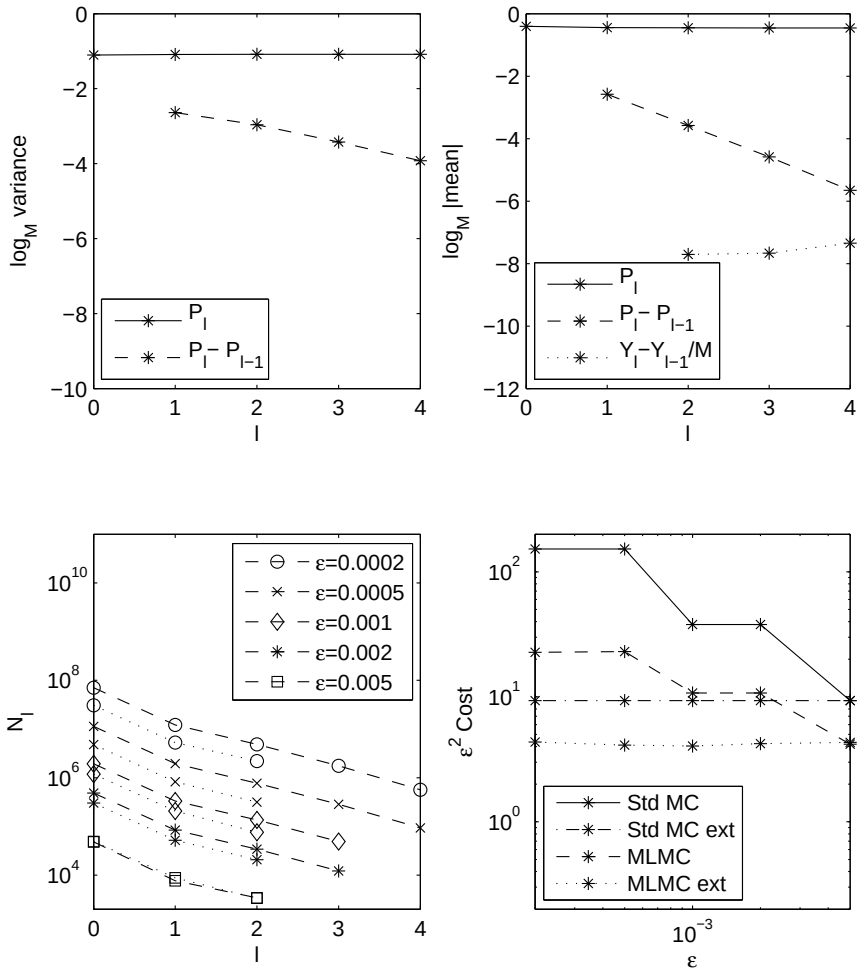

Figure 5: Geometric Brownian motion with digital option (value ≈ 0.53).

savings.

The accuracy of the heuristic algorithm is again tested by performing ten sets of multilevel calculations and comparing the RMSE error to the target accuracy ϵ. The ratio is in the range 0.55–1.0 for all cases, with and without extrapolation.

6.2 Heston stochastic volatility model

Figure 6 presents results for the same European call payoff considered previously, but this time based on the Heston stochastic volatility model [10],

$$\begin{aligned} \mathrm{d}S &= r\,S\,\mathrm{d}t + \sqrt{V}\,S\,\mathrm{d}W_1, \qquad\qquad 0 < t < 1 \\ \mathrm{d}V &= \lambda\,(\sigma^2 - V)\,\mathrm{d}t + \xi\,\sqrt{V}\;\mathrm{d}W_2, \end{aligned}$$

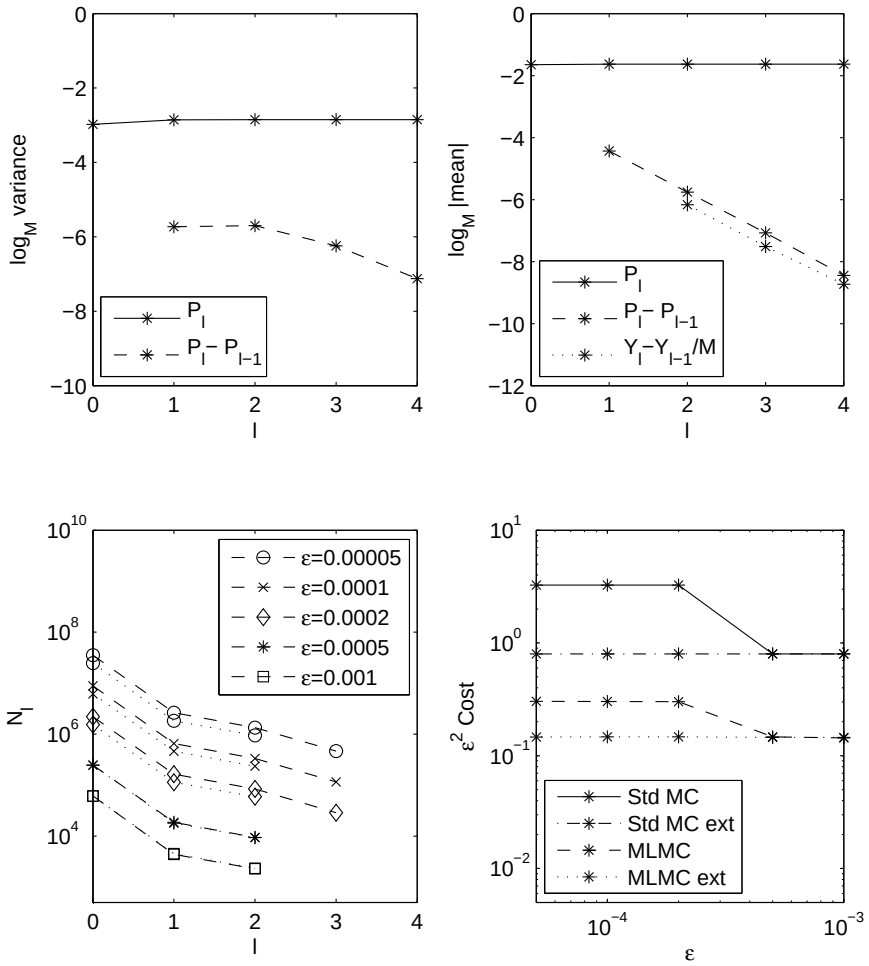

Figure 6: Heston model with European option (value ≈ 0.10).

with $S(0)=1$, $V(0)=0.04$, $r=0.05$, $\sigma=0.2$, $\lambda=5$, $\xi=0.25$, and correlation $\rho=-0.5$ between $\mathrm{d}W_1$ and $\mathrm{d}W_2$.

The accuracy and variance are both improved by defining a new variable

$$W = e^{\lambda t}\,(V-\sigma^2),$$

and applying the Euler discretisation to the SDEs for W and S which results in the discrete equations

$$\begin{aligned}
\widehat{S}_{n+1} &= \widehat{S}_n + r\,\widehat{S}_n\,h + \sqrt{\widehat{V}_n^+}\,\widehat{S}_n\,\Delta W_{1,n} \\
\widehat{V}_{n+1} &= \sigma^2 + e^{-\lambda h}\left((\widehat{V}_n-\sigma^2) + \xi\,\sqrt{\widehat{V}_n^+}\,\Delta W_{2,n}\right),
\end{aligned}$$

Note that the $\sqrt{V}$ is replaced by $\sqrt{V^+} \equiv \sqrt{\max(V,0)}$ but as $h \to 0$ the

probability of the discrete approximation to the volatility becoming negative approaches zero, for the chosen values of λ, σ, ξ [12].

Because the volatility does not satisfy a global Lipschitz condition, there is no existing theory to predict the order of weak and strong convergence. The numerical results suggest the variance is decaying slightly slower than first order, while the weak convergence appears slightly faster than first order. The multilevel method without Richardson extrapolation gives savings of up to factor 10 compared to the standard Monte Carlo method. Using a reference value computed using the numerical method of Kahl and Jäckel [11], the ratio of the RMSE error to the target accuracy ϵ is found to be in the range 0.49–1.01.

The results with Richardson extrapolation are harder to interpret. The order of weak convergence does not appear to be improved. The computational cost is reduced, but this is due to the heuristic termination criterion which assumes the remaining error after extrapolation is second order, which it is not. Consequently, the ratio of the RMSE error to the target accuracy ϵ is in the range 0.66–1.23, demonstrating that the termination criterion is not reliable in combination with extrapolation for this application.

7 Concluding remarks

In this paper we have shown that a multilevel approach, using a geometric sequence of timesteps, can reduce the order of complexity of Monte Carlo path simulations. If we consider the generation of a discrete Brownian path through a recursive Brownian Bridge construction, starting with the end points W_0 and W_T at level 0, then computing the mid-point $W_{T/2}$ at level 1, then the interval mid-points $W_{T/4}, W_{3T/4}$ at level 2, and so on, then an interpretation of the multilevel method is that the level l correction, $E[\widehat{P}_l - \widehat{P}_{l-1}]$, corresponds to the effect on the expected payoff due to the extra detail that is brought into the Brownian Bridge construction at level l.

The numerical results for a range of model problems show that the multilevel algorithm is efficient and reliable in achieving the desired accuracy, whereas the use of Richardson extrapolation is more problematic; in some cases it works well but in other cases it fails to double the weak order of convergence and hence does not achieve the target accuracy.

There are a number of areas for further research arising from this work. One is the development of improved estimators giving a convergence order $\beta > 1$. For scalar SDEs, the Milstein discretisation gives $\beta = 2$ for Lipschitz payoffs, but more work is required to obtain improved convergence for look-

back, barrier and digital options. The extension to multi-dimensional SDEs is also challenging since, in most cases, the Milstein discretisation requires the simulation of Lévy areas [6, 7].

A second area for research concerns the heuristic nature of the multilevel numerical procedure. It would clearly be desirable to have a numerical procedure which is guaranteed to give a MSE which is less than ϵ^2. This may be achievable by using estimators with $\beta > 1$, so that one can use an excessively large value for L without significant computational penalty, thereby avoiding the problems with the bias estimation.

Thirdly, the multilevel method needs to be tested on much more complex applications, more representative of the challenges faced in the finance community. This includes payoffs which involve evaluations at multiple intermediate times in addition to the value at maturity, and basket options which involve high-dimensional SDE's.

Finally, it may be possible to further reduce the computational complexity by switching to quasi Monte Carlo methods such as Sobol sequences and lattice rules [16, 17]. This is likely to be particularly effective in conjunction with improved estimators with $\beta > 1$, because in this case the optimal N_l for the true Monte Carlo sampling leads to the majority of the computational effort being applied to extremely coarse paths. These are ideally suited to the use of quasi Monte Carlo techniques, which may be able to lower the computational cost towards $O(\epsilon^{-1})$ to achieve a MSE of ϵ^2.

Acknowledgements

I am very grateful to Mark Broadie, Paul Glasserman and Terry Lyons for discussions on this work, and their very helpful comments on drafts of the paper.

References

[1] 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.

[2] W.M. Briggs, V.E. Henson, and S.F. McCormick. *A Multigrid Tutorial (second edition)*. SIAM, 2000.

[3] M. Broadie, P. Glasserman, and S. Kou. A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4):325–348, 1997.

[4] D. Duffie and P. Glynn. Efficient Monte Carlo simulation of security prices. *Annals of Applied Probability*, 5(4):897–905, 1995.

[5] M. Emsermann and B. Simon. Improving simulation efficiency with quasi control variates. *Stochastic models*, 18(3):425–448, 2002.

[6] J.G. Gaines and T.J. Lyons. Random generation of stochastic integrals. *SIAM J. Appl. Math.*, 54(4):1132–1146, 1994.

[7] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.

[8] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. Multilevel splitting for estimating rare event probabilities. *Operations Research*, 47:585–600, 1999.

[9] S. Heinrich. *Multilevel Monte Carlo Methods*, volume 2179 of *Lecture Notes in Computer Science*, pages 58–67. Springer-Verlag, 2001.

[10] S.I. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[11] C. Kahl and P. Jäckel. Not-so-complex logarithms in the Heston model. *Wilmott*, pages 94–103, September 2005.

[12] C. Kahl and P. Jäckel. Fast strong approximation Monte-Carlo schemes for stochastic volatility models. Working Paper, ABN AMRO, 2006.

[13] H. Kahn. Use of different Monte Carlo sampling techniques. In H.A. Meyer, editor, *Symposium on Monte Carlo Methods*, pages 146–190. Wiley, 1956.

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

[15] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, 1992.

[16] F.Y. Kuo and I.H. Sloan. Lifting the curse of dimensionality. *Notices of the AMS*, 52(11):1320–1328, 2005.

[17] 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.

[18] J.S. Liu. *Monte Carlo strategies in scientific computing*. Springer, New York, 2001.

[19] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, 1992.

[20] 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.

[21] P. Wesseling. *An Introduction to Multigrid Methods*. John Wiley, 1992.