Report no. 06/03

Multi-level Monte Carlo path simulation

Michael B. Giles
giles@comlab.ox.ac.uk

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.

Key words and phrases: computational finance, Monte Carlo, multi-level

Oxford University Computing Laboratory
Numerical Analysis Group
Wolfson Building
Parks Road
Oxford, England OX1 3QD

May, 2006

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 vector SDE with general drift and volatility terms,

$$\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W(t), \quad 0 < t < T, \tag{1.1}$$

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

$$|f(U) - f(V)| \ \leq c\,\|U - V\|\,, \quad \forall\, U, V. \tag{1.2}$$

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,

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

It is well established that, provided $a(S,t)$ and $b(S,t)$ satisfy certain conditions [13, 9, 1], the expected mean-square-error (MSE) in the estimate $\widehat{Y}$ is asymptotically of the form

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

where c_1, c_2 are positive constants. The first term comes from the variance in $\widehat{Y}$ due to the Monte Carlo sampling, and the second term comes from 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 multi-level method which reduces the variance, leaving unchanged the bias due to the Euler discretisation. The multi-level method is very easy to implement and can be combined, in principle, with other variance reduction methods such as stratified sampling [5] and quasi Monte Carlo methods [11, 12, 10] to obtain even greater savings.

The method extends the recent work of Kebaier [8] 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 $O(h^{1/2})$. Our technique generalises this approach to multiple levels, using a geometric

sequence of different timesteps $h_l = M^{-l}, l = 0, 1, \ldots, L$, 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 [14, 2]. 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.

The paper begins with the introduction of the new multi-level 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 Multi-level 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 multi-level 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

$$\widehat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left(\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)} \right). \tag{2.1}$$

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 $\sqrt{V_l \, h_l}$.

In the particular case of the Euler discretisation and the Lipschitz payoff function, provided $a(S,t)$ and $b(S,t)$ satisfy certain conditions [13, 9, 1], there is $O(h)$ weak convergence, so that

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

and $O(h^{1/2})$ strong convergence, so that

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

From the latter, together with the Lipschitz property (1.2), it follows that

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

and therefore $V[\widehat{P}_l - P] = O(h_l)$. Furthermore,

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

Hence for the simple estimator (2.1), 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

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

as $\epsilon \to 0$, then $h_l = M^{-L} = O(\epsilon)$, and so the bias error $E[\widehat{P}_l - P]$ is $O(\epsilon)$. 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 multi-level method conditional on certain features of the underlying numerical approximation and the multi-level estimators. This approach is similar to that used by Duffie and Glynn [4].

Theorem 1 *Let P denote a functional of the solution of stochastic differential equation (1.1) 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 multi-level 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 $\frac{1}{\sqrt{2}}\,M^{-\alpha}\epsilon < c_1\,h_l^\alpha \leq \frac{1}{\sqrt{2}}\,\epsilon$ and hence, because of properties i) and ii),

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

We now need to consider the different possible values for β. If $\beta{=}1$, we set

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

so that

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

Together with the $\frac{1}{2}\epsilon^2$ bound on the square of the bias error, this establishes the ϵ^2 upper bound on the estimator MSE. To bound the computational complexity C we begin with an upper bound on L given by

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

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

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

where

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

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

$$\begin{aligned} C &\leq c_3 \sum_{l=0}^{L} N_l \, h_l^{-1} \\ &\leq c_3 \left(2\,\epsilon^{-2}(L+1)^2\, c_2 + \sum_{l=0}^{L} h_l^{-1} \right) \\ &\leq c_3 \left(2\,\epsilon^{-2}(L+1)^2\, c_2 + \frac{M^2}{M-1} \left(\frac{\epsilon}{2c_1} \right)^{-1/\alpha} \right), \end{aligned}$$

since

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

Using the upper bound for $L+1$ and noting that $\epsilon^{-1/\alpha} \leq \epsilon^{-2}(\log \epsilon)^2$ it follows that $C \leq c_4\, \epsilon^{-2}(\log \epsilon)^2$ where

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

For $\beta > 1$, setting

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

it is easily checked that the MSE is again less than ϵ^2. The corresponding complexity bound is a decreasing geometric series, resulting in the bound stated in the theorem.

For $\beta < 1$, the appropriate choice for N_l is

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

This again gives an MSE which is less than ϵ^2, but in this case the corresponding complexity bound is an increasing geometric series, resulting in the bound stated in the theorem. ■

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 [13, 9, 1], which in turn gives $\beta = 1$ as explained earlier.

The convergence is degraded if the payoff function $f(S(T))$ has a finite number of discontinuities. 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 a discontinuity. With the Euler discretisation, this fraction of the paths have an $O(1)$ probability of $\widehat{P}_l - \widehat{P}_{l-1}$ being $O(1)$,

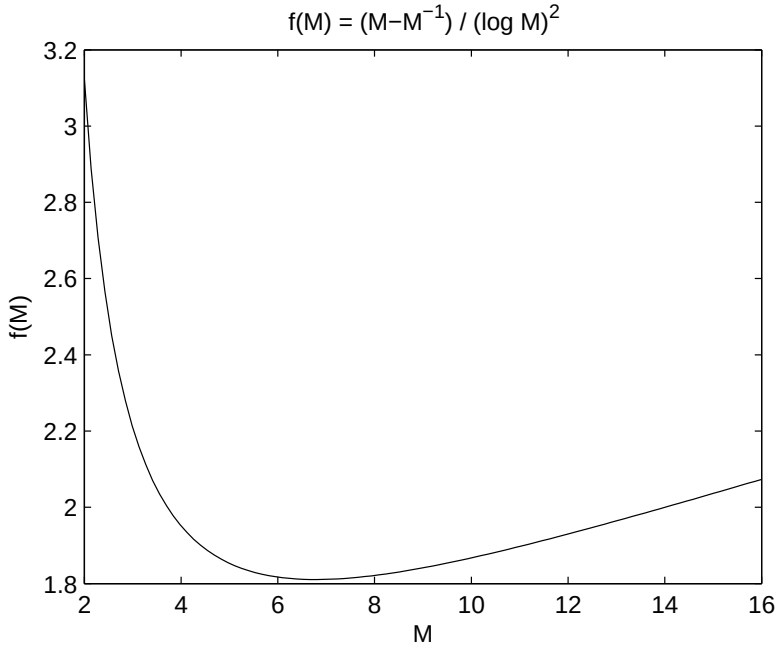

Figure 1: A plot of the function $(M-M^{-1})/(\log M)^2$

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 require the development of new methods with increased sampling of those paths with large values for $\widehat{P}_l - \widehat{P}_{l-1}$.

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. Future research will investigate the possibility of obtaining the same order of convergence for vector SDEs. This might be achievable by modifying the estimator to include a zero-mean control variate [5].

4 Extensions

4.1 Optimal M

In the analysis so far, the value of the integer M, which is the factor by which the timestep is refined at each level, has not been specified. 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 multi-level 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. 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

$$\left(\sqrt{M}-1\right)^2 c_1\, h_l \le V[\widehat{P}_l - \widehat{P}_{l-1}] \le \left(\sqrt{M}+1\right)^2 c_1\, h_l,$$

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 limits,

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

which results in

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

The computational cost of evaluating $\widehat{Y}_l$ is proportional to

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

Since $L = O(\log \epsilon^{-1} / \log M)$, we conclude that asymptotically the total computational cost for fixed ϵ is proportional to

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

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.

For $\beta<1$, it may be that even larger values for M will be optimal, whereas for $\beta>1$, values much closer to $M=2$ are likely to be best.

4.2 Bias estimation and Richardson extrapolation

In the multi-level 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, in the case of the Euler discretisation with a Lipschitz payoff, the correction from the final level $E[\widehat{P}_L - \widehat{P}_{L-1}]$ is asymptotically equal to $(M-1)\, E[P - \widehat{P}_L]$.

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 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

$$\max\left\{M^{-1}\left|\widehat{Y}_{L-1}\right|, \left|\widehat{Y}_L\right|\right\} < \tfrac{1}{\sqrt{2}}\,(M-1)\,\epsilon. \tag{4.1}$$

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, $E[\widehat{P}_l - \widehat{P}_{l-1}]$.

An alternative approach is to use Richardson extrapolation to eliminate the leading order bias. If $E[\widehat{P}_L - \widehat{P}_{L-1}] \approx c_1 h_L$, then $E[P - \widehat{P}_L] \approx c_1 (M-1)^{-1} h_L$ and so by changing the combined estimator to

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

the leading order bias is eliminated and the remaining bias is $o(h_L)$ (probably either $O(h^{3/2})$ or $O(h^2)$. 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 $M\widehat{Y}_l - \widehat{Y}_{l-1}$ to decide when the remaining bias is sufficiently small, in exactly the same way as described previously for $\widehat{Y}_l$.

5 Numerical results

Putting together the elements already discussed, the multi-level 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. (5.1)
4. evaluate extra samples at each level as needed for new N_l
5. if $L \geq 2$, test for convergence using Eqn. (4.1)
6. if $L < 2$ or not converged, set $L := L+1$ and go to 2.

The numerical results are all obtained using $M = 4$. This gives most of the benefits of a larger value of M, as analysed in the previous section, but at the same time M is small enough to give a reasonable number of data points to observe the asymptotic convergence.

The equation for the optimal N_l is

$$N_l = \left\lceil 2\, \epsilon^{-2} \sqrt{V_l\, h_l} \ \left(\sum_{l=0}^{L} \sqrt{V_l / h_l} \right) \right\rceil . \tag{5.1}$$

This makes the overall variance less than $\frac{1}{2}\epsilon^2$, while Equation (4.1) ensures that the bias is less than $\frac{1}{\sqrt{2}}\epsilon$. Together, they give a mean-square-error which is less than ϵ^2, with ϵ being a user-specified r.m.s. accuracy.

5.1 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 kinds of option payoff.

5.1.1 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 slope of the latter is approximately -1 showing 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 $O(h)$ 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}. These two plots are both based on results from 10^6 paths, for each value of the timestep.

The bottom two plots have results from five multi-level calculations for different values of ϵ. Each line in the bottom left plot 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 with ϵ of $\epsilon^2\,C$ where the computational complexity C is defined as

$$C = \sum_l N_l h_l^{-1}.$$

One line shows the results for the multi-level 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 a very slowly increasing function of ϵ^{-1} for the multi-level method, in agreement with the theory which predicts it to be proportional to $(\log\epsilon)^2$, 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 multi-level method is more than 60 times more efficient than the standard method.

5.1.2 Asian option

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

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

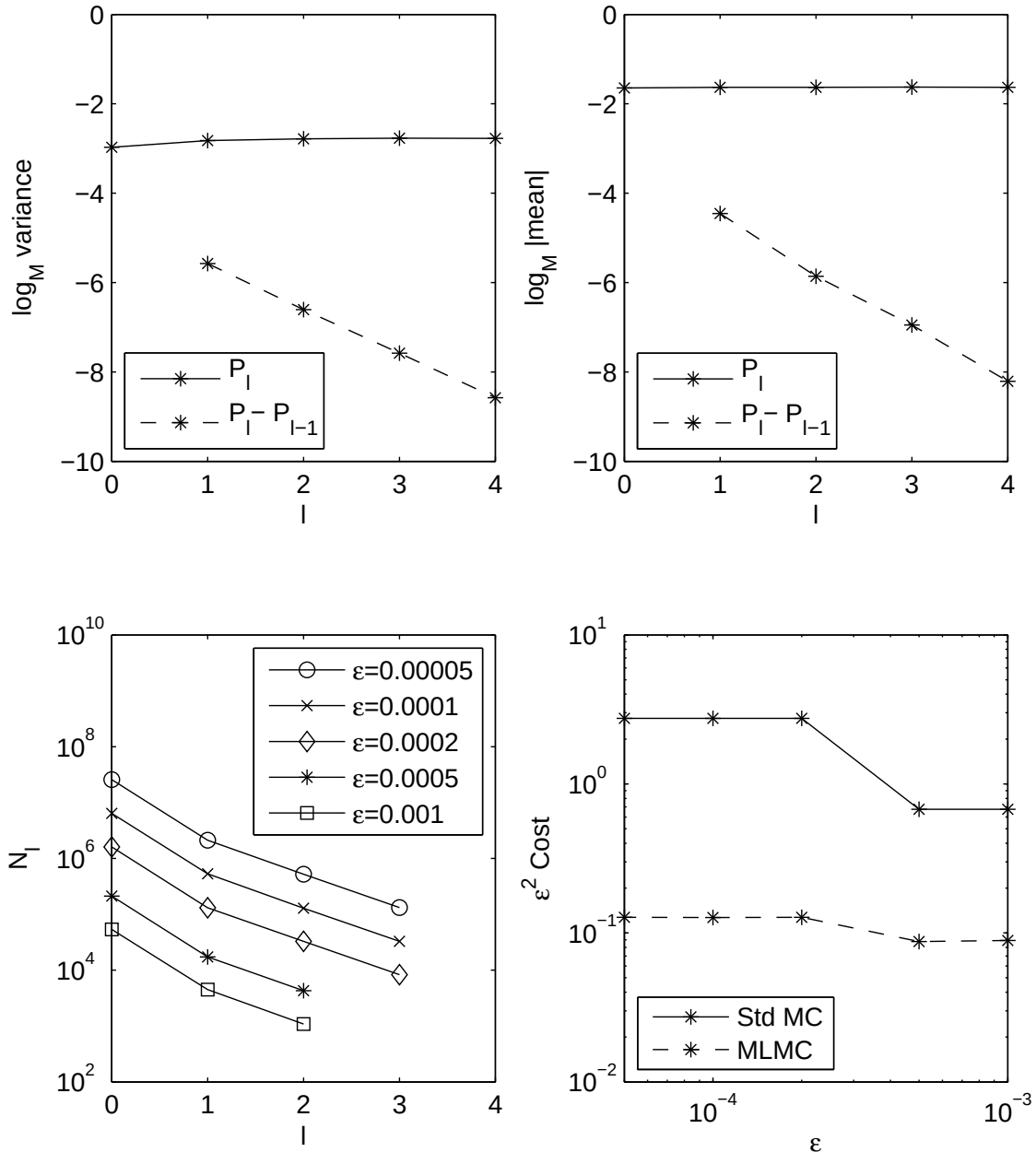

Figure 2: Geometric Brownian motion with European option

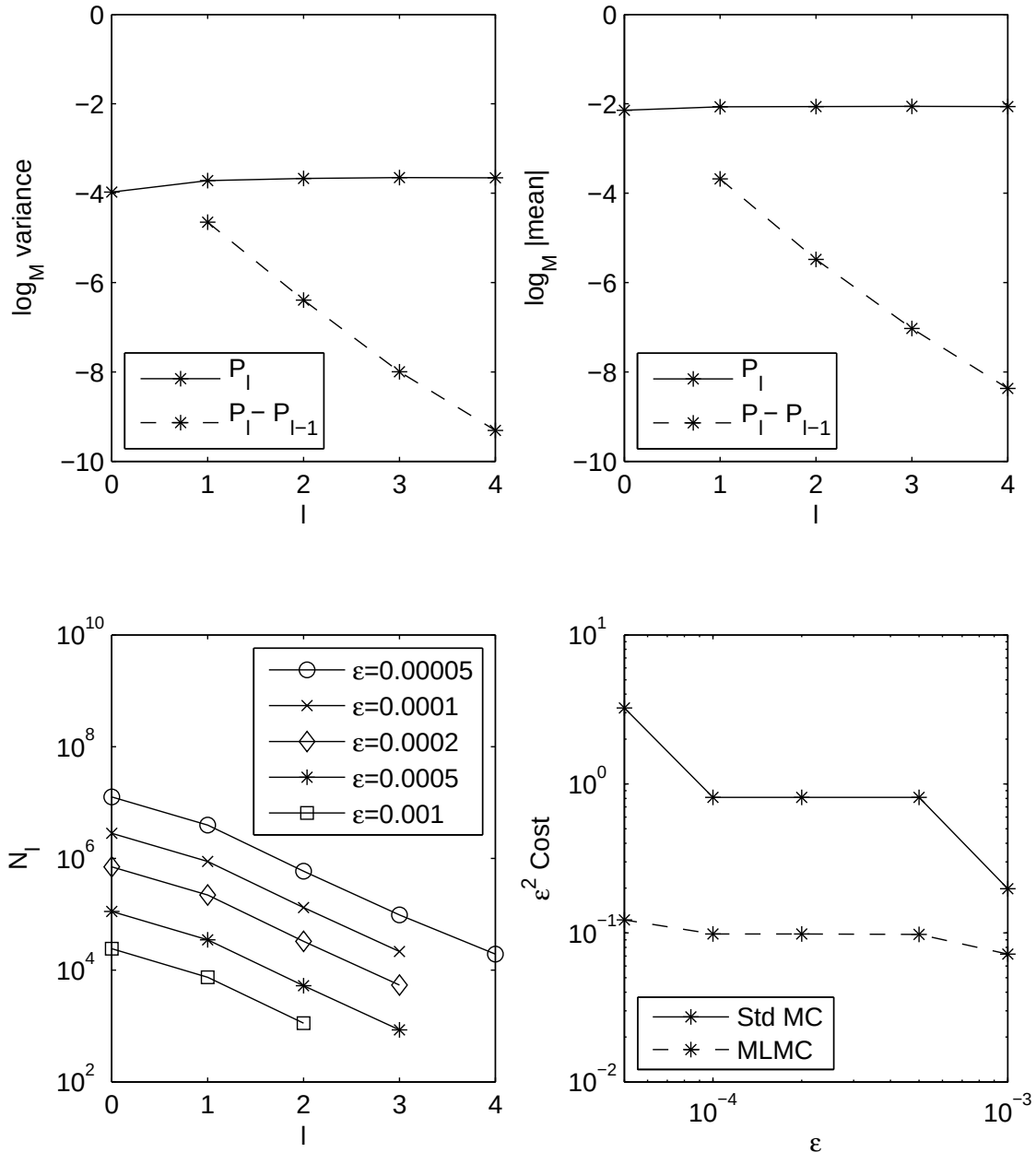

Figure 3: Geometric Brownian motion with Asian option

which is approximated numerically by

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

The results are qualitatively similar, with $O(h)$ convergence of both V_l and $E[P_l - P_{l-1}]$, and significant savings in computational complexity from using the multi-level method.

5.1.3 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]. Despite this, to get the desired accuracy requires larger values of L than for the other options, and hence the multi-level method gives greater computational savings than in the other cases.

5.1.4 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. This results in much larger values for N_l to achieve comparable accuracy, and the efficiency gains of the multi-level method are reduced accordingly.

Similar results have also been obtained for a barrier option.

5.2 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 [6],

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

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 $\mathrm{d}W_2$.

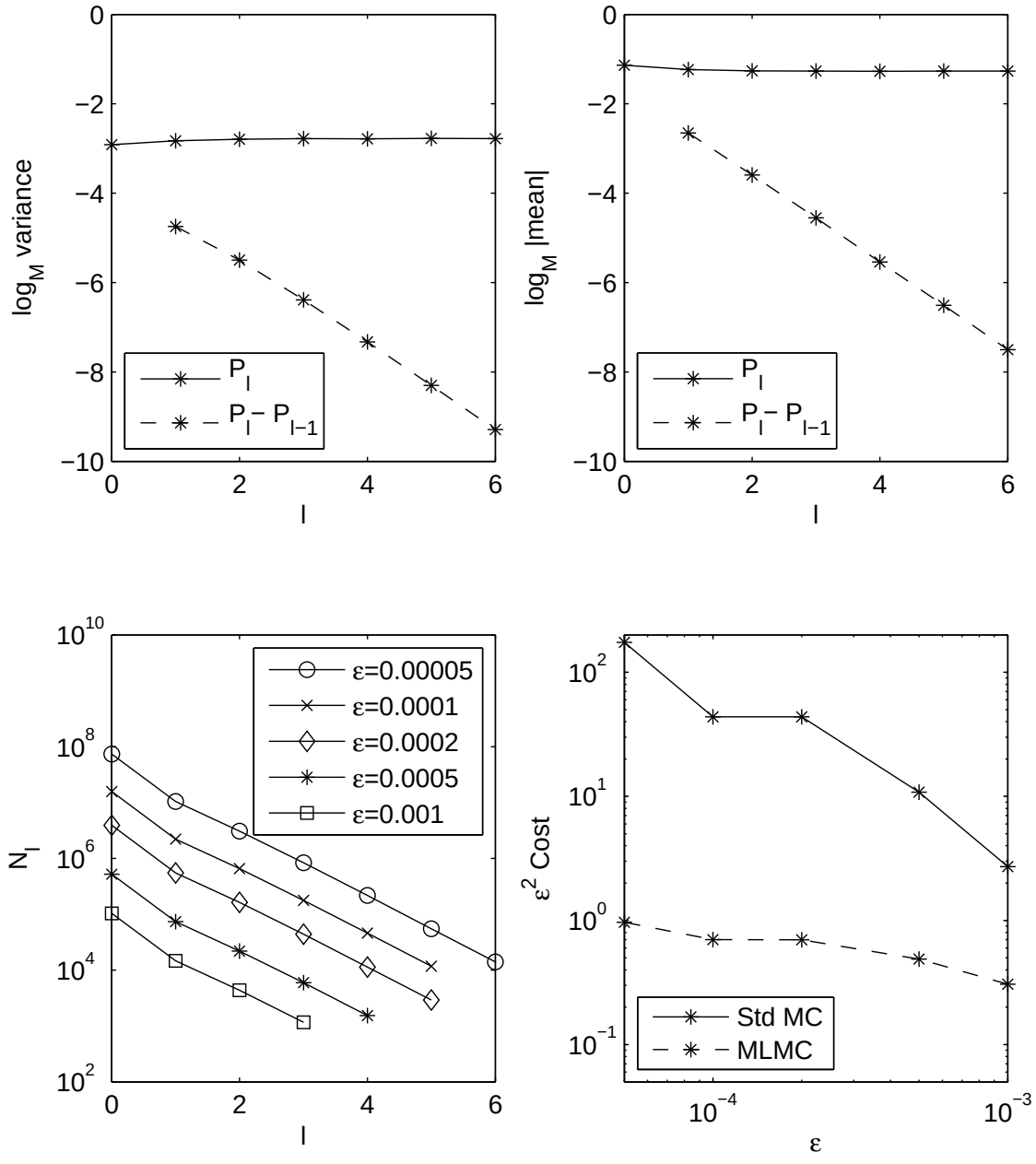

Figure 4: Geometric Brownian motion with lookback option

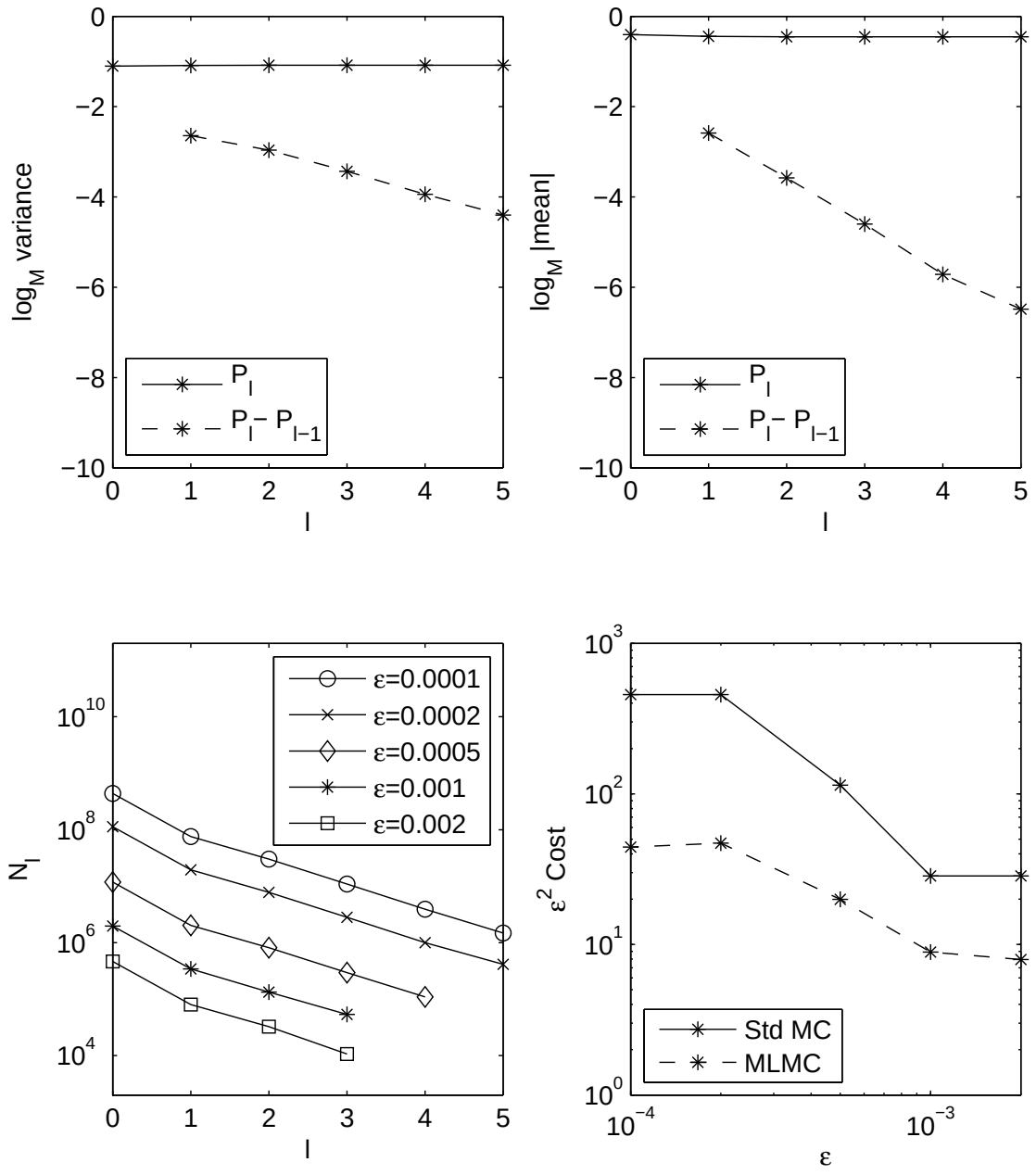

Figure 5: Geometric Brownian motion with digital option

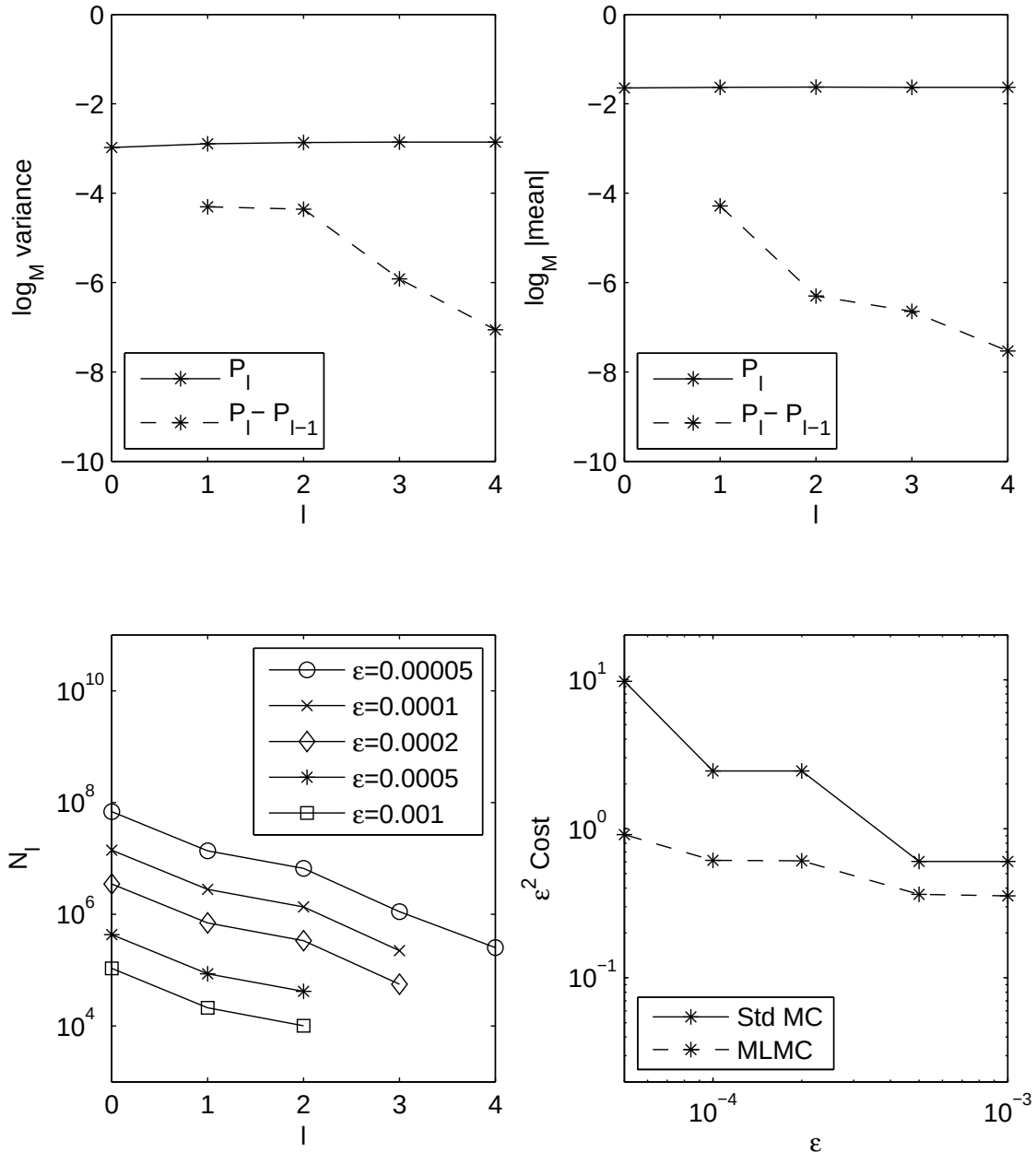

Figure 6: Heston model with European option

In the discretisation $\sqrt{V}$ is replaced by $\sqrt{\max(V,0)}$ 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 [7]. Consequently, the Euler discretisation has the same order of weak and strong convergence as for the geometric Brownian motion, and the multi-level method again yields computational savings which increase rapidly with the desired accuracy.

6 Concluding remarks

In this paper we have shown that a multi-level 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 multi-level 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. This interpretation is very much in the same spirit as the multigrid method for solving elliptic PDEs, in which the long wavelength corrections are computed on coarse grids, while the short wavelength corrections are computed on fine grids.

There are three broad areas for further research arising from this work. The first is the determination of the exponent β which is critical to the application of the main theorem. This will require detailed analysis of each case, whether the output functional is a function of the final state $S(T)$, or depends on the entire path as in the case of barrier and lookback options.

A second direction for further study is an investigation into the use of standard variance reduction techniques in conjunction with the multi-level method. There should be no difficulty in using stratified sampling [5], for example, to reduce the computational cost, while leaving its order of complexity unchanged. It may also be possible to reduce the computational order of complexity by switching to quasi Monte Carlo methods [11, 12, 10].

In the examples in this paper, the estimators $\widehat{Y}_l$ are based on standard estimators for $\widehat{P}_l$ and $\widehat{P}_{l-1}$. The third area for future research is the development of new estimators specifically designed to estimate the quantity $E[\widehat{P}_l - \widehat{P}_{l-1}]$, with the goal of improving the exponent β. It is thought this might be achievable for Lipschitz payoffs through the use of a suitable zero-mean control variate based on the leading order term in an asymptotic expansion of the difference $\widehat{P}_l - \widehat{P}_{l-1}$. However, other approaches are likely to be required for discontinuous payoffs.

Acknowledgements

I am very grateful to Mark Broadie, Paul Glasserman and Terry Lyons for discussions on this work, and their very helpful comments on 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] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.

[6] 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.

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

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

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

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

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

[12] I.H. Sloan and S. Joe. *Lattice Methods for Multiple Integration*. Oxford University Press, 1994.

[13] 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.

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