

Multilevel quasi-Monte Carlo path simulation

Michael B. Giles and Ben J. Waterhouse

Abstract. This paper reviews the multilevel Monte Carlo path simulation method for estimating option prices in computational finance, and extends it by combining it with quasi-Monte Carlo integration using a randomised rank-1 lattice rule. Using the Milstein discretisation of the stochastic differential equation, it is demonstrated that the combination has much lower computational cost than either one on its own for evaluating European, Asian, lookback, barrier and digital options.

Key words. Multilevel, Monte Carlo, quasi-Monte Carlo, computational finance

AMS classification. 11K45, 60H10, 60H35, 65C05, 65C30, 68U20

1 Introduction

Giles [5, 6] has recently introduced a multilevel Monte Carlo path simulation method for the pricing of financial options. This improves the computational efficiency of Monte Carlo path simulation by combining results using different numbers of timesteps. This can be viewed as a generalisation of the two-level method of Kebaier [10] and is also similar in approach to Heinrich's multilevel method for parametric integration [9]. The first paper [6] (which was the second to appear in print due to a publication backlog) introduced the multilevel Monte Carlo method and proved that it can lower the computational complexity of path-dependent Monte Carlo evaluations. It also presented numerical results using the simplest Euler-Maruyama discretisation. The second paper [5] demonstrated that the computational cost can be further reduced by using the Milstein discretisation. This has the same weak order of convergence but an improved first order strong convergence, and it is the strong order of convergence which is central to the efficiency of the multilevel method.

In this paper we review the key ideas and introduce a new ingredient, the use of quasi-Monte Carlo (QMC) integration based on a randomised rank-1 lattice rule which further reduces the computational cost. To set the scene, we consider a scalar SDE with general drift and volatility terms,

$$dS(t) = a(S, t) dt + b(S, t) dW(t), \quad 0 < t < T, \quad (1.1)$$

First author: supported by Microsoft Corporation, the UK Engineering and Physical Sciences Research Council and the Oxford-Man Institute of Quantitative Finance.

Second author: supported by the Australian Research Council through a Linkage project between the University of New South Wales and Macquarie Bank.

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 r. m. s. error of ϵ would require $O(\epsilon^{-2})$ independent paths, each with $O(\epsilon^{-1})$ timesteps, giving a computational complexity which is $O(\epsilon^{-3})$. With the Euler-Maruyama discretisation the multilevel method reduces the cost to $O(\epsilon^{-2}(\log \epsilon)^2)$ for a European option with a payoff with a uniform Lipschitz bound [6], while the use of the Milstein discretisation further reduces the cost to $O(\epsilon^{-2})$ for a larger class of options, including Asian, lookback, barrier and digital options [5]. The paper begins by reviewing the multilevel approach, first with the Euler path discretisation and then with the superior Milstein discretisation. QMC methods based on rank-1 lattice rules are then introduced, with particular attention to Brownian Bridge construction and the use of randomisation to obtain confidence intervals. The combined multilevel QMC algorithm is presented and the following section provides numerical results for a range of options.

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

$$\mathbb{E}[\hat{P}_L] = \mathbb{E}[\hat{P}_0] + \sum_{l=1}^L \mathbb{E}[\hat{P}_l - \hat{P}_{l-1}]. \quad (2.1)$$

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

$$\hat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left(\hat{P}_l^{(i)} - \hat{P}_{l-1}^{(i)} \right). \quad (2.2)$$

The key point here is that the quantity $\hat{P}_l^{(i)} - \hat{P}_{l-1}^{(i)}$ comes from two discrete approxima-

tions with different timesteps but the same Brownian path. The variance of this simple estimator is $\mathbb{V}[\widehat{Y}_l] = N_l^{-1}V_l$ 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 $\sum_{l=0}^L \widehat{Y}_l$ is $\sum_{l=0}^L N_l^{-1}V_l$, 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 [2, 11, 21] there is $O(h^{1/2})$ strong convergence. From this it follows that $\mathbb{V}[\widehat{P}_l - P] = O(h_l)$ for a European option with a Lipschitz continuous payoff. Hence for the simple estimator (2.2), 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 $\mathbb{E}[\widehat{P}_L - P]$ 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 [6]:

Theorem 2.1 *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) $|\mathbb{E}[\widehat{P}_l - P]| \leq c_1 h_l^\alpha$
- ii) $\mathbb{E}[\widehat{Y}_l] = \begin{cases} \mathbb{E}[\widehat{P}_0], & l = 0 \\ \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}], & l > 0 \end{cases}$
- iii) $\mathbb{V}[\widehat{Y}_l] \leq c_2 N_l^{-1} h_l^\beta$

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

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

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 [6]. □

3 Milstein discretisation

The theorem proves that the best order of complexity is achieved using discretisations with $\beta > 1$. To achieve this for a scalar SDE, we use the Milstein discretisation of equation (1.1) which is

$$\widehat{S}_{n+1} = \widehat{S}_n + a_n h + b_n \Delta W_n + \frac{1}{2} \frac{\partial b_n}{\partial S} b_n ((\Delta W_n)^2 - h). \quad (3.1)$$

In the above equation, the subscript n is used to denote the timestep index, and a_n , b_n and $\partial b_n / \partial S$ 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 $\sigma(S, t)$.

The Milstein discretisation defines the numerical approximation at the discrete times t_n . Within the time interval $[t_n, t_{n+1}]$ we use a constant coefficient Brownian interpolation conditional on the two end values,

$$\widehat{S}(t) = \widehat{S}_n + \lambda (\widehat{S}_{n+1} - \widehat{S}_n) + b_n \left(W(t) - W_n - \lambda (W_{n+1} - W_n) \right), \quad (3.2)$$

where

$$\lambda = \frac{t - t_n}{t_{n+1} - t_n}.$$

For the fine path, standard results on i) the expected average value, ii) the distribution of the minimum, and iii) the probability of crossing a certain value, will be used to obtain the value \widehat{P}_l for Asian, lookback and barrier options, respectively.

Exactly the same approach could also be used on the coarse path with half as many timesteps to obtain \widehat{P}_{l-1} . However, this would not give an estimator \widehat{Y}_l with variance convergence rate $\beta > 1$. To achieve the better convergence rate, we first use the value of the underlying Brownian motion $W(t)$ at the midpoint (which has already been sampled and used for the fine path calculation) to define an interpolated midpoint

$$\widehat{S}_{n+\frac{1}{2}} = \frac{1}{2}(\widehat{S}_{n+1} + \widehat{S}_n) + b_n \left(W_{n+\frac{1}{2}} - \frac{1}{2}(W_{n+1} + W_n) \right). \quad (3.3)$$

We can then use the Brownian interpolation (with volatility b_n) on each of the half-intervals $[t_n, t_{n+\frac{1}{2}}]$ and $[t_{n+\frac{1}{2}}, t_{n+1}]$ which each correspond to one of the timesteps on the fine path. A key point in this construction is that we have not altered the expected value for \widehat{P}_{l-1} , averaged over all underlying Brownian paths $W(t)$, compared to its evaluation on level $l-1$ on which it corresponds to the finer path; see [5] for further discussion of this important point.

4 Quasi-Monte Carlo method

QMC methods approximate an integral on a high-dimensional hypercube with an N -point equal-weight quadrature rule of the form

$$\int_{[0,1]^d} f(\mathbf{x}) \, d\mathbf{x} \approx \frac{1}{N} \sum_{i=0}^{N-1} f(\mathbf{x}_i).$$

This is the same form which is used in the Monte Carlo method. However, rather than choosing the d -dimensional points \mathbf{x}_i uniformly from the unit cube, as is the case with the Monte Carlo method, QMC methods choose the points in some deterministic manner.

Sobol' sequences [20] and digital nets [16] are two popular choices of QMC points, which have been previously used for financial applications [8, 14]. In this paper we use a rank-1 lattice rule [19] in which the points have the particularly simple construction

$$\mathbf{x}_i = \left\{ \frac{i}{N} \mathbf{z} \right\},$$

where \mathbf{z} is a d -dimensional vector with integer components and the notation $\{ \cdot \}$ denotes taking the fractional part of each component of the argument and disregarding the integer part so that \mathbf{x}_i lies within the half-open unit cube.

For Monte Carlo integration it is well known that the error is $O(N^{-1/2})$. In one dimension, the lattice rule is equivalent to a rectangle rule and can achieve $O(N^{-1})$ convergence of the error, for a sufficiently smooth integrand. For larger dimensions, it may be shown that for integrands with sufficient smoothness and dimensions which become progressively less important, there exist lattice rules for which the error decays

at $O(N^{-1+\varepsilon})$ for all $\varepsilon > 0$, see [12]. Unfortunately, many integrands in mathematical finance applications do not have the required smoothness and so we may not apply the theory to claim the $O(N^{-1+\varepsilon})$ convergence. However, experimentation suggests that this rate can in fact be achieved for many finance problems [13].

Two key aspects of the implementation of QMC methods are randomisation and the factorisation of the covariance matrix. If we neglect for the moment the discretisation errors which arise from finite timesteps, the standard Monte Carlo method has the attractive feature that it provides both an unbiased estimate of the desired value and a confidence interval for that estimate. The QMC method lacks this feature but it can be regained by re-defining the i^{th} point to be

$$\mathbf{x}_i = \left\{ \frac{i}{N} \mathbf{z} + \mathbf{\Delta} \right\}.$$

For a given offset vector $\mathbf{\Delta} \in [0, 1)^d$, this defines a set of N points, for which one can compute the average

$$\widehat{Y} = \frac{1}{N} \sum_{i=0}^{N-1} f(\mathbf{x}_i).$$

If we now treat $\mathbf{\Delta}$ as a random variable then the expected value of \widehat{Y} is equal to the desired integral, and therefore \widehat{Y} is an unbiased estimator. By choosing a number of different random offsets $\mathbf{\Delta}_1, \dots, \mathbf{\Delta}_q$ ($q = 32$ is used in this paper) and computing a separate \widehat{Y}_j for each, one can construct a confidence interval in the usual way.

For a scalar SDE with n_T timesteps, the dimensionality of the problem is $d = n_T$, and the factorisation of the covariance matrix concerns the question of how best to map the different dimensions of the hypercube to the n_T Wiener increments in the Milstein discretisation. The expected value of a financial product whose value is determined by an asset whose dynamics are described by (1.1), discretised at times $t_n = nh$, is given by the integral

$$\int_{\mathbb{R}^d} p(\mathbf{x}) \frac{\exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right)}{(2\pi)^{d/2} \sqrt{\det \Sigma}} d\mathbf{x}.$$

Here $p(\mathbf{x})$ is the payoff function and the d -dimensional matrix $\Sigma_{i,j} = \min(t_i, t_j)$ is the covariance matrix for the elements of \mathbf{x} which are the underlying Wiener path values W_n . Taking a matrix A such that $A A^T = \Sigma$, and making the substitutions $\mathbf{x} = A \mathbf{y}$ and $\mathbf{y} = \Phi^{-1}(\mathbf{z})$ where Φ^{-1} is the inverse of the cumulative Normal distribution function taken componentwise, this can be reformulated as an integral over the unit cube

$$\int_{\mathbb{R}^d} p(A \mathbf{y}) \frac{\exp\left(-\frac{1}{2} \mathbf{y}^T \mathbf{y}\right)}{(2\pi)^{d/2}} d\mathbf{y} = \int_{[0,1]^d} p(A \Phi^{-1}(\mathbf{z})) d\mathbf{z}.$$

For Monte Carlo integration the choice of the matrix A makes no difference, but for QMC integration it is very important [3, 8, 14]. While any choice of A such that $A A^T = \Sigma$ is suitable, there are three established ways in which the matrix A may be

chosen. Firstly, A may be chosen to be the Cholesky factor of Σ . This is the simplest method and corresponds to taking the n^{th} component of \mathbf{x}_i to define ΔW_n through

$$\Delta W_n = \sqrt{h} \Phi^{-1}(x_{i,n}).$$

This would correctly map a uniform $[0, 1]$ distribution for x_n into a Normal distribution for ΔW_n with zero mean and variance h . This method is often referred to as the standard construction and is usually used for Monte Carlo integration due to the simplicity of its construction.

A second way in which A may be chosen is to use a Brownian Bridge construction [3, 8]. Under this method, the first component of \mathbf{x} is used to define $W(T)$, the second component defines $W(T/2)$ (conditional on the first), the third and fourth components define $W(T/4)$ and $W(3T/4)$ (conditional on the first two), and so on. Note that in the standard and Brownian Bridge constructions, the matrix A is not explicitly used, but rather implicitly used in the recursive construction.

The final way is known as the ‘‘Principal Components Analysis’’ (PCA) method. In this method A is chosen to be the matrix with n^{th} column equal to $\sqrt{\lambda_n} \mathbf{v}_n$ where λ_n is the n^{th} largest eigenvalue of Σ and \mathbf{v}_n is the corresponding eigenvector [8].

Several authors [3, 14, 8] have found the Brownian Bridge and PCA constructions to be much better for some problems, although it is known that there are problems from mathematical finance for which the standard construction performs much better than the Brownian Bridge, see [18]. In our numerical experiments we use the Brownian Bridge construction, since for our applications it consistently outperforms the standard construction.

The final implementation issue is the choice of the generating vector \mathbf{z} . We use a vector using the construction algorithm of Dick *et al* [4]. This particular type of lattice rule is said to be *extensible* since it can be used as a sequence with differing values of N . The construction algorithm is particularly efficient due to the fast FFT implementation technique of Nuyens and Cools [17].

5 Multilevel QMC algorithm

At level l in the multilevel formulation, N_l is defined to be the number of QMC points, and \widehat{Y}_l is the computed average of \widehat{P}_l (for $l = 0$) or $\widehat{P}_l - \widehat{P}_{l-1}$ (for $l > 0$) over the 32 sets of N_l QMC lattice points, each set having a different random offset. An unbiased estimate of its variance V_l is computed in the usual way from the differing values for the 32 averages.

On the assumption that there is first order weak convergence, the remaining bias at the finest level $\mathbb{E}[P - \widehat{P}_L]$ is approximately equal to \widehat{Y}_L . Being more cautious (to allow for the possibility that \widehat{Y}_l changes sign as l increases before settling into its first order

asymptotic convergence) we estimate the magnitude of the bias using

$$\max \left\{ \frac{1}{2} \left| \widehat{Y}_{L-1} \right|, \left| \widehat{Y}_L \right| \right\}.$$

The mean square error is the sum of the combined variance $\sum_{l=0}^L V_l$ plus the square of the bias $\mathbb{E}[P - \widehat{P}_L]$. We choose to make each of these smaller than $\epsilon^2/2$, so that overall we achieve a user-specified RMS accuracy of ϵ . The variance is reduced by increasing the number of lattice points on each level, while the bias is reduced by increasing the level of path refinement (i.e. increasing L).

Given this outline strategy, the multilevel QMC algorithm proceeds as follows:

1. start with $L=0$
2. get an initial estimate for V_L using 32 random offsets and $N_L = 1$
3. while $\sum_{l=0}^L V_l > \epsilon^2/2$, double N_l on the level with largest $V_l / (2^l N_l)$
4. if $L < 2$ or the bias estimate is greater than $\epsilon/\sqrt{2}$, set $L := L+1$ and go to step 2

Step 3 is based on the fact that doubling N_l will eliminate most of the variance V_l at a cost proportional to the product of the number of timesteps 2^l and the number of lattice points N_l . The choice of level l aims to maximise the reduction in variance per unit cost.

6 Numerical results

6.1 European call option

The European call option we consider has the discounted payoff

$$P = \exp(-rT) (S(T) - K)^+,$$

where the notation $(x)^+$ denotes $\max(0, x)$. Figure 6.1 shows the numerical results for parameters $S(0)=1$, $K=1$, $T=1$, $r=0.05$, $\sigma=0.2$.

The solid lines in the top left plot show the behaviour of the variance \widehat{P}_l , while the dashed lines show the variance of $\widehat{P}_l - \widehat{P}_{l-1}$. The four sets of calculations use different numbers of lattice points. The calculations with just one lattice point correspond to standard Monte Carlo. The calculations with 16, 256 and 4094 lattice points show the variance of the average over the set of lattice points multiplied by the number of lattice points; for standard Monte Carlo this quantity would be independent of the number of points, and therefore this is a fair basis of comparison which accounts for the cost of 4096 points being 4096 times greater than a single point. The solid line results show

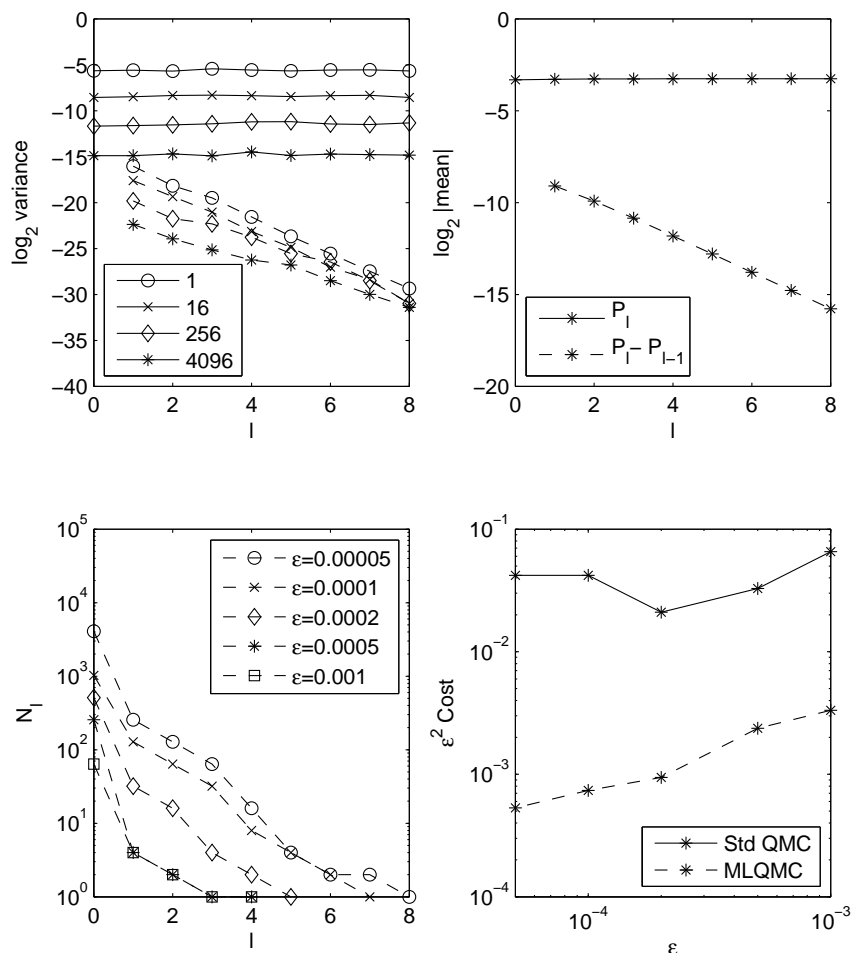


Figure 6.1 European call option

that the QMC method on its own is very effective in reducing the variance compared to the standard Monte Carlo method. The dashed line results show that in conjunction with the multilevel approach the QMC is effective at reducing the variance on the coarsest levels, but the benefits diminish on the finer levels. This is probably because the multilevel approach itself extracts much of the low-dimensional content in the integrand, so that on the finer levels the correction is predominantly high-dimensional and so the QMC approach is less effective. However, most of the computational cost of the multilevel method is on the coarsest levels, and so we will see that the combination does reduce the overall cost significantly.

The top right plot shows that $\mathbb{E}[\hat{P}_l - \hat{P}_{l-1}]$ is approximately $O(h_l)$, corresponding to the expected first order weak convergence. Each line in the bottom left plot shows the

values for $N_l, l = 0, \dots, L$, with the values decreasing with level l as expected. 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

$$C = 32 \sum_l 2^l N_l,$$

which is the total number of fine grid timesteps on all levels. One line shows the results for the multilevel QMC method and the other shows the corresponding cost of a standard QMC 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 roughly constant for the standard QMC method, and this is at a level which is comparable to that achieved previously using the multilevel method on its own. However, combining the multilevel method with QMC gives additional savings of factor 20-100, with the computational cost being approximately proportional to ϵ^{-1} . This is the best one could hope for using QMC since in the best cases its error is inversely proportional to the number of points, and hence, at best, inversely proportional to the computational cost.

6.2 Asian option

The Asian option we consider has the discounted payoff

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

where

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

On the fine path, integrating (3.2) and using standard Brownian Bridge results (see section 3.1 in [8]) gives

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

where

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

is a $N(0, h^3/12)$ Normal random variable, independent of ΔW . The coarse path approximation is similar except that the values for ΔI_n are derived from the fine path

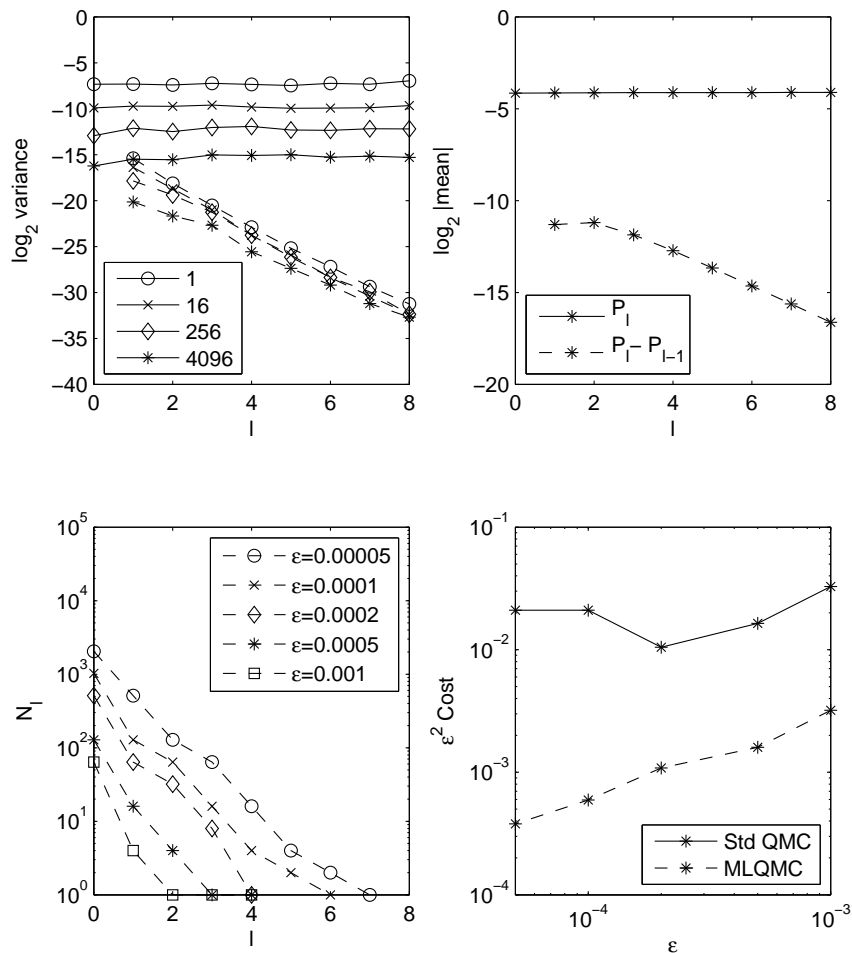


Figure 6.2 Asian option

values, noting that

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

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 6.2 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 variance of both \widehat{P}_l and $\widehat{P}_l - \widehat{P}_{l-1}$. The standard QMC method is effective at reducing the variance on all levels, but with the multilevel estimator its effectiveness diminishes at the finer levels. The bottom two plots again have results from five multilevel calculations for different values of ϵ . It can be seen that $\epsilon^2 C$ is very roughly constant for the standard QMC method (again at a level comparable to that achieved previously by the multilevel method on its own [5]), while $\epsilon^2 C$ decreases significantly with decreasing ϵ for the combined multilevel QMC method.

6.3 Lookback option

The lookback option we consider has the discounted payoff

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

For the fine path calculation on the time interval $[t_n, t_{n+1}]$, a standard Brownian interpolation result (see section 6.4 in [8]) gives the minimum value as

$$\widehat{S}_{n,min}^f = \frac{1}{2} \left(\widehat{S}_n^f + \widehat{S}_{n+1}^f - \sqrt{\left(\widehat{S}_{n+1}^f - \widehat{S}_n^f \right)^2 - 2 b_n^2 h \log U_n} \right), \quad (6.1)$$

where U_n is a uniform random variable on $[0, 1]$. Taking the minimum over all timesteps gives an approximation to $\min_{0 < t < T} S(t)$ from which \widehat{P}_l is calculated.

For the coarse path calculation, \widehat{P}_{l-1} is defined similarly, except that for each timestep the mid-point value is first constructed using (3.3), and then the minimum over the timestep is given by

$$\begin{aligned} \widehat{S}_{m,min}^c = \min \left\{ \frac{1}{2} \left(\widehat{S}_m^c + \widehat{S}_{m+\frac{1}{2}}^c - \sqrt{\left(\widehat{S}_{m+\frac{1}{2}}^c - \widehat{S}_m^c \right)^2 - b_m^2 h \log U_{2m-1}} \right), \right. \\ \left. \frac{1}{2} \left(\widehat{S}_{m+\frac{1}{2}}^c + \widehat{S}_{m+1}^c - \sqrt{\left(\widehat{S}_{m+1}^c - \widehat{S}_{m+\frac{1}{2}}^c \right)^2 - b_m^2 h \log U_{2m}} \right) \right\}. \end{aligned} \quad (6.2)$$

Note the re-use of the uniform random variables U_{2m-1} and U_{2m} from the two fine timesteps corresponding to this coarse timestep; it is this which ensures that the minimum from the coarse path is very close to the minimum from the fine path, resulting in a low variance for $\widehat{P}_l - \widehat{P}_{l-1}$.

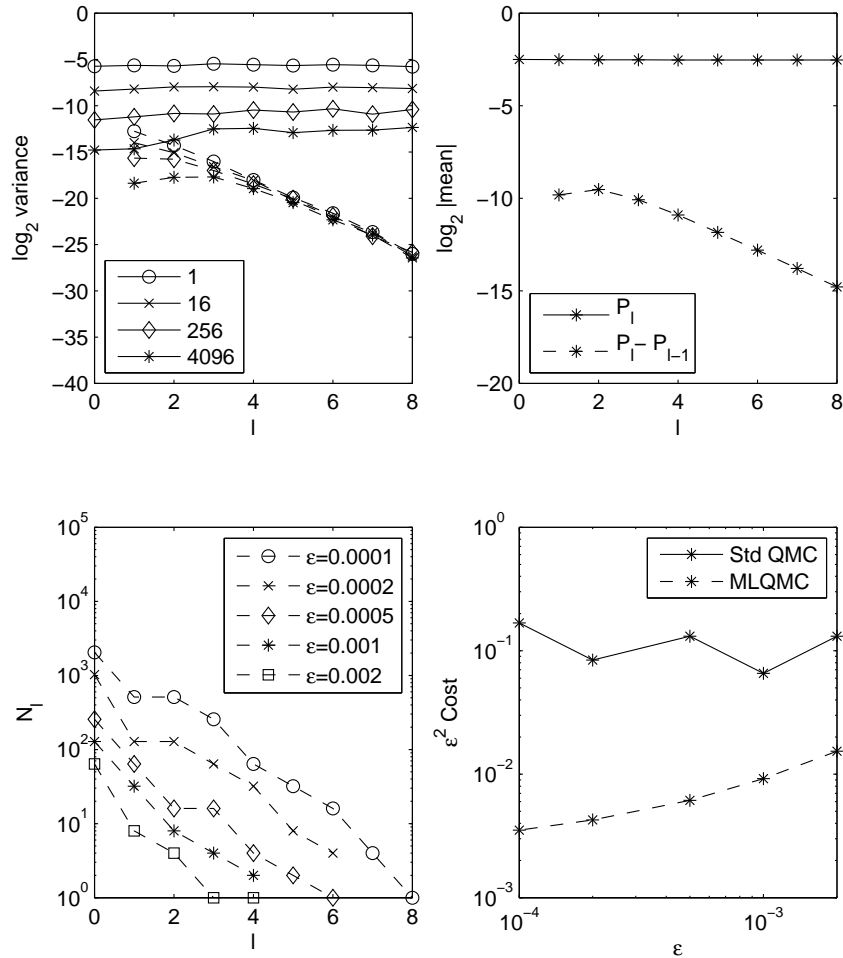


Figure 6.3 Lookback option

Figure 6.3 shows the results for parameters $S(0) = 1$, $T = 1$, $r = 0.05$, $\sigma = 0.2$. The results are qualitatively similar to the previous two cases. There is almost no improvement from using QMC on the finer levels, but nevertheless there is a big reduction in the overall cost compared to the multilevel method without QMC [5].

6.4 Barrier option

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

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

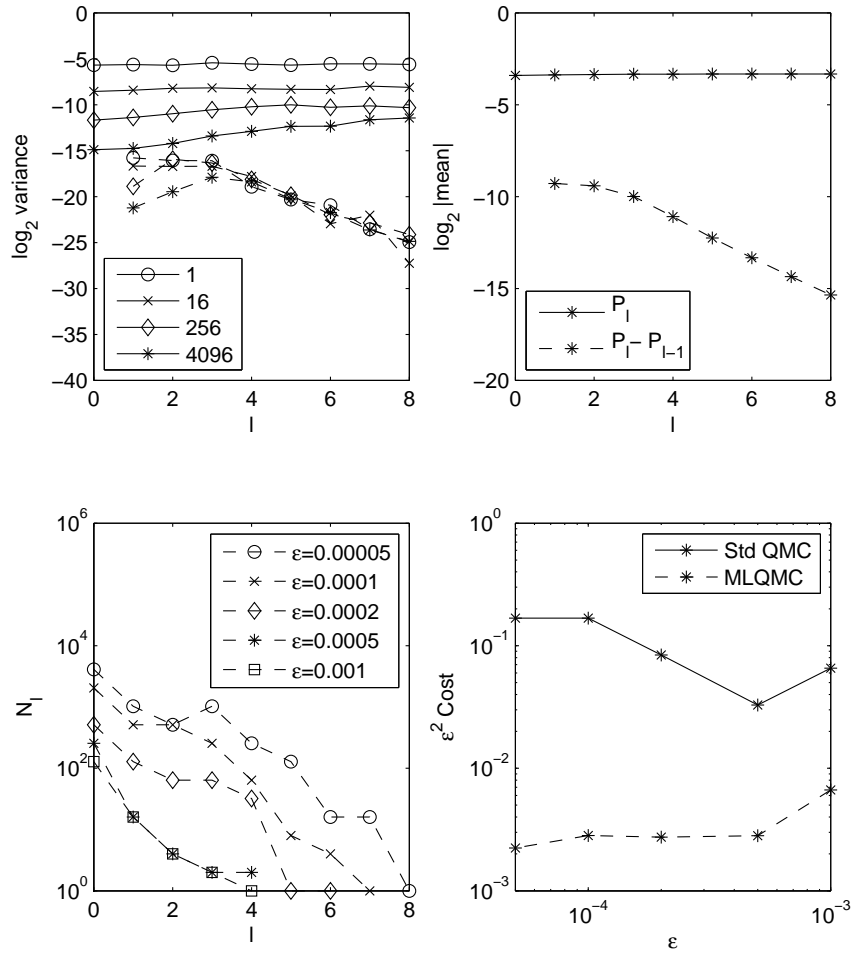


Figure 6.4 Barrier option

where $\mathbf{1}_{\tau > T}$ is an indicator function taking value 1 if the argument is true, and zero otherwise, and the barrier crossing time τ is $\tau = \inf_{t > 0} \{S(t) < B\}$.

For the fine path simulation, following a standard approach for continuously monitored barrier crossings (see section 6.4 in [8]), the conditional expectation of the payoff can be expressed as

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

where \hat{p}_n , the probability the interpolated path did not cross the barrier during the n^{th}

timestep, is equal to

$$\hat{p}_n = 1 - \exp\left(\frac{-2(\hat{S}_n^f - B)^+(\hat{S}_{n+1}^f - B)^+}{b_n^2 h}\right). \quad (6.3)$$

For the coarse path calculation, we again use equation (3.3) to construct a midpoint value $\hat{S}_{m+1/2}^c$ for each timestep. Given this value, the probability that the Brownian interpolation path does not cross the barrier during the m^{th} coarse timestep is

$$\begin{aligned} \hat{p}_m^c &= \left\{ 1 - \exp\left(\frac{-2(\hat{S}_m^c - B)^+(\hat{S}_{m+1/2}^c - B)^+}{b_m^2 h}\right) \right\} \\ &\times \left\{ 1 - \exp\left(\frac{-2(\hat{S}_{m+1/2}^c - B)^+(\hat{S}_{m+1}^c - B)^+}{b_m^2 h}\right) \right\}. \end{aligned} \quad (6.4)$$

Figure 6.4 has the results for parameters $S(0) = 1$, $K = 1$, $B = 0.85$, $T = 1$, $r = 0.05$, $\sigma = 0.2$. The main features are similar, but the variance V_l decreases with level at a slightly lower rate in this case [5] and consequently $\epsilon^2 C$ for the combined multilevel QMC method does not decrease quite as much as ϵ is reduced compared to the previous examples.

6.5 Digital option

The digital option which is considered has the discounted payoff

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

To achieve a good multilevel variance convergence rate, we follow the same procedure used previously [5], smoothing the payoff using the technique of conditional expectation (see section 7.2.3 in [8]) in which we terminate the path calculations one timestep before reaching the terminal time T . If $\hat{S}_{n_T-1}^f$ denotes the fine path 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 $\hat{S}_{n_T}^f > K$ after one further timestep is

$$\hat{p}^f = \Phi\left(\frac{\hat{S}_{n_T-1}^f + a_{n_T-1}h - K}{b_{n_T-1}\sqrt{h}}\right), \quad (6.5)$$

where Φ is the cumulative Normal distribution. For the fine-path payoff \hat{P}_l^f we therefore use $\hat{P}_l^f = \exp(-rT) \hat{p}^f$.

For the coarse-path payoff, we note that given the Brownian increment ΔW for the first half of the last timestep, which is already known because it corresponds to the

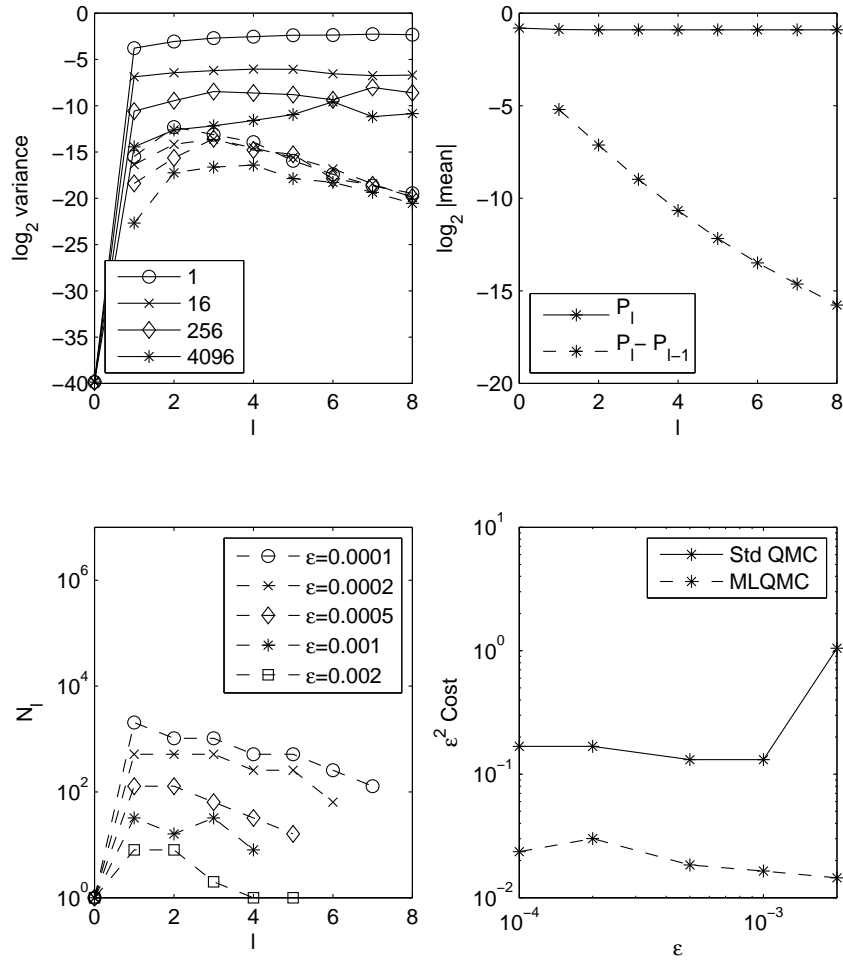


Figure 6.5 Digital option

last of the computed timesteps in the fine path calculation, then the probability that $\widehat{S}_{n_T/2}^c > K$ is

$$\widehat{p}^c = \Phi \left(\frac{\widehat{S}_{n_T/2-1}^c + a_{n_T-1}h + b_{n_T-1}\Delta W - K}{b_{n_T-1}\sqrt{h/2}} \right), \quad (6.6)$$

where $a_{n_T/2-1}$ and $b_{n_T/2-1}$ are the drift and volatility based on $\widehat{S}_{n_T/2-1}^c$.

Figure 6.5 has the results for parameters $S(0) = 1$, $K = 1$, $T = 1$, $r = 0.05$, $\sigma = 0.2$. 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 (6.5) to evaluate the payoff. This essen-

tially eliminates the cost of the level 0 calculation, which is where the QMC method is usually most effective. Consequently, the cost of the combined multilevel QMC method remains approximately proportional to ϵ^{-2} , and is only slightly lower than the results obtained previously for the multilevel method without QMC [5]. However, we still get a factor 5-10 computational savings compared to the standard QMC on its own.

7 Conclusions and future work

In this paper we have demonstrated the benefits of combining rank-1 lattice rule quasi-Monte Carlo integration with multilevel Monte Carlo path simulation. Together, the computational cost is lower than using either one on its own.

There are two major directions for future research. The first is the extension of the algorithms to multi-dimensional SDEs, for which the Milstein discretisation usually requires the simulation of Lévy areas [8, 11]. Current investigations indicate that this can be avoided for European options with a Lipschitz payoff through the use of anti-thetic variables. However, the extension to more difficult payoffs, such as the Asian, lookback, barrier and digital options considered in this paper, looks more challenging and the direct simulation of the Lévy areas may be necessary.

The second direction for future research is the numerical analysis of multilevel methods. Müller-Gronbach and Ritter [15], Giles, Higham and Mao [7] and Avikainen [1] have obtained bounds on the convergence of the multilevel method using the Euler discretisation for different classes of output functional, but additional research is required for the Milstein discretisation.

Bibliography

- [1] R. Avikainen, *Convergence rates for approximations of functionals of SDEs*, Finance and Stochastics (to appear) (2009).
- [2] 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 (1995), pp. 43–60.
- [3] R.E. Caflisch, W.J. Morokoff, and A.B. Owen, *Valuation of Mortgage Backed Securities Using Brownian Bridges to Reduce Effective Dimension*, J. Comput. Finance 1 (1997), pp. 27–46.
- [4] J. Dick, F. Pillichshammer, and B.J. Waterhouse, *The construction of good extensible rank-1 lattices*, Math. Comp. 77 (2007), pp. 2345–2373.
- [5] M.B. Giles, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, Monte Carlo and Quasi-Monte Carlo Methods 2006 (A. Keller, S. Heinrich, and H. Niederreiter, eds.), Springer-Verlag, 2007, pp. 343–358.
- [6] ———, *Multilevel Monte Carlo path simulation*, Operations Research 56 (2008), pp. 981–986.

- [7] M.B. Giles, D. Higham, and X. Mao, *Analysing multilevel Monte Carlo for options with non-globally Lipschitz payoff*, Finance and Stochastics (to appear) (2009).
- [8] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York, 2004.
- [9] S. Heinrich, *Multilevel Monte Carlo Methods*, Lecture Notes in Computer Science, vol. 2179, pp. 58–67, Springer-Verlag, 2001.
- [10] A. Kebaier, *Statistical Romberg extrapolation: a new variance reduction method and applications to options pricing*, Annals of Applied Probability 14 (2005), pp. 2681–2705.
- [11] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1992.
- [12] F.Y. Kuo and I.H. Sloan, *Lifting the curse of dimensionality*, Notices of the AMS 52 (2005), pp. 1320–1328.
- [13] F.Y. Kuo, G.W. Wasilkowski, and B.J. Waterhouse, *Randomly shifted lattice rules for unbounded integrands*, J. Complexity 22 (2006), pp. 630–651.
- [14] P. L'Ecuyer, *Quasi-Monte Carlo methods in finance*, Proceedings of the 2004 Winter Simulation Conference (R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, eds.), IEEE Press, 2004, pp. 1645–1655.
- [15] T. Müller-Gronbach and K. Ritter, *Minimal errors for strong and weak approximation of stochastic differential equations*, Monte Carlo and Quasi-Monte Carlo Methods 2006 (A. Keller, S. Heinrich, and H. Niederreiter, eds.), Springer-Verlag, 2007, pp. 53–82.
- [16] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, 1992.
- [17] D. Nuyens and R. Cools, *Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces*, Math. Comp. 75 (2006), pp. 903–920 (electronic).
- [18] A. Papageorgiou, *The Brownian bridge does not offer a consistent advantage in Quasi-Monte Carlo integration*, J. Complex. 18 (2002), pp. 171–186.
- [19] I.H. Sloan and S. Joe, *Lattice Methods for Multiple Integration*, Oxford University Press, 1994.
- [20] I. M. Sobol', *Distribution of points in a cube and approximate evaluation of integrals*, Akademija Nauk SSSR. Žurnal Vyčislitel'noi Matematiki i Matematičeskoj Fiziki 7 (1967), pp. 784–802.
- [21] D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Analysis and Applications 8 (1990), pp. 483–509.

Author information

Michael B. Giles, Mathematical Institute and Oxford-Man Institute of Quantitative Finance, Oxford University, Oxford, United Kingdom.

Email: mike.giles@maths.ox.ac.uk

Ben J. Waterhouse, School of Mathematics and Statistics, University of New South Wales, Sydney, Australia.

Email: benjw@maths.unsw.edu.au