Antithetic multilevel Monte Carlo estimation for multidimensional SDES

Michael B. Giles and Lukasz Szpruch

Abstract In this paper we develop antithetic multilevel Monte Carlo (MLMC) estimators for multidimensional SDEs driven by Brownian motion. Giles has previously shown that if we combine a numerical approximation with strong order of convergence $O(\Delta t)$ with MLMC we can reduce the computational complexity to estimate expected values of Lipschitz functionals of SDE solutions with a root-mean-square error of ε from $O(\varepsilon^{-3})$ to $O(\varepsilon^{-2})$. However, in general, to obtain a rate of strong convergence higher than $O(\Delta t^{1/2})$ requires simulation, or approximation, of Lévy areas. Recently, Giles and Szpruch [5] constructed an antithetic multilevel estimator that avoids the simulation of Lévy areas and still achieves an MLMC correction variance which is $O(\Delta t^2)$ for smooth payoffs and almost $O(\Delta t^{3/2})$ for piecewise smooth payoffs, even though there is only $O(\Delta t^{1/2})$ strong convergence. This results in an $O(\varepsilon^{-2})$ complexity for estimating the value of financial European and Asian put and call options. In this paper, we extend these results to more complex payoffs based on the path minimum. To achieve this, an approximation of the Lévy areas is needed, resulting in $O(\Delta t^{3/4})$ strong convergence. By modifying the antithetic MLMC estimator we are able to obtain $O(\varepsilon^{-2}\log(\varepsilon)^2)$ complexity for estimating financial barrier and lookback options.

1 Introduction

In his original MLMC paper [4], Giles showed that one could obtain a good MLMC variance for smooth payoffs by using a numerical approximation with good strong convergence properties. This is in contrast to the standard Monte Carlo approach to simulations of SDEs, where only a good weak order of convergence is required.

Michael B. Giles
Mathematical Institute, University of Oxford, e-mail: mike.giles@maths.ox.ac.uk

Lukasz Szpruch
Mathematical Institute, University of Oxford, e-mail: szpruch@maths.ox.ac.uk

For multidimensional SDEs, to obtain good strong convergence, simulation of the Lévy areas is required. Indeed, Clark & Cameron [1] proved for a particular SDE that it is impossible to achieve a better order of strong convergence than the Euler-Maruyama discretisation when using just the discrete increments of the underlying Brownian motion. The analysis was extended by Müller-Gronbach [8] to general SDEs. As a consequence, if we use the standard MLMC method with the Milstein scheme without simulating the Lévy areas the complexity will remain the same as for Euler-Maruyama. Recently, Giles and Szpruch [5] constructed an antithetic MLMC estimator, enabling one to neglect the Lévy areas and still obtain a multi-level correction estimator with a variance which decays at the same rate as the scalar Milstein estimator. They achieved an $O(\Delta t^2)$ MLMC variance for smooth payoffs and almost an $O(\Delta t^{3/2})$ variance for piecewise smooth payoffs, even though there is only $O(\Delta t^{1/2})$ strong convergence. This results in an $O(\varepsilon^{-2})$ complexity for estimating the value of European and Asian put and call options.

The question remains whether the approach can be extended to more complex payoffs such as those based on the minimum of the path over the simulation interval. For scalar SDEs with the Milstein discretisation, Giles [4] obtained $O(\varepsilon^{-2})$ complexity for such payoffs by combining MLMC with conditional Monte Carlo methods. In this paper, we extend these results to the multidimensional case. Unlike the previous multidimensional work, we find that a suitable approximation to the Lévy areas is required. By a suitable modification of the antithetic MLMC estimator we are able to obtain $O(\varepsilon^{-2}\log(\varepsilon)^2)$ complexity for payoffs corresponding to financial lookback and barrier options. We focus on simulations of Clark and Cameron's SDE since it captures the essence of simulations requiring Lévy area simulation to obtain higher that $O(\Delta t^{1/2})$ strong convergence property. Our results are supported by numerical experiments.

2 MLMC

Multilevel Monte Carlo simulation uses a number of levels of resolution, $\ell = 0, 1, \ldots, L$, with $\ell = 0$ being the coarsest, and $\ell = L$ being the finest. In the context of an SDE simulation, level 0 may have just one timestep for the whole time interval $[0, T]$, whereas level L might have 2^L uniform timesteps $\Delta t_L = 2^{-L}T$. If P denotes the payoff (or other output functional of interest), and P_ℓ denote its approximation on level ℓ, then the expected value $\mathbb{E}[P_L]$ on the finest level is equal to the expected value $\mathbb{E}[P_0]$ on the coarsest level plus a sum of corrections which give the difference in expectation between simulations on successive levels,

$$\mathbb{E}[P_L] = \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell - P_{\ell-1}]. \tag{1}$$

Let Y_0 be an estimator for $\mathbb{E}[P_0]$ using N_0 samples, and let Y_ℓ, $\ell > 0$, be an estimator for $\mathbb{E}[P_\ell - P_{\ell-1}]$ using N_ℓ samples. The simplest estimator is a mean of N_ℓ

independent samples, which for $\ell > 0$ is

$$Y_\ell = N_\ell^{-1} \sum_{i=1}^{N_\ell} (P_\ell^i - P_{\ell-1}^i). \tag{2}$$

The key point is that $P_\ell^i - P_{\ell-1}^i$ should come from two discrete approximations for the same underlying stochastic sample.

We recall the Theorem from [5]:

Theorem 1. *Let P denote a functional of the solution of a stochastic differential equation, and let P_ℓ denote the corresponding level ℓ numerical approximation. If there exist independent estimators Y_ℓ based on N_ℓ Monte Carlo samples, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2} \min(\beta, \gamma)$ and*

i) $|\mathbb{E}[P_\ell - P]| \leq c_1 2^{-\alpha \ell}$

ii) $\mathbb{E}[Y_\ell] = \begin{cases} \mathbb{E}[P_0], & \ell = 0 \\ \mathbb{E}[P_\ell - P_{\ell-1}], & \ell > 0 \end{cases}$

iii) $\mathbb{V}[Y_\ell] \leq c_2 N_\ell^{-1} 2^{-\beta \ell}$

iv) $C_\ell \leq c_3 N_\ell 2^{\gamma \ell}$, *where C_ℓ is the computational complexity of Y_ℓ*

then there exists a positive constant c_4 such that for any $\varepsilon < e^{-1}$ there are values L and N_ℓ for which the multilevel estimator $Y = \sum_{\ell=0}^{L} Y_\ell$, has a mean-square-error with bound $MSE \equiv \mathbb{E}\left[(Y - \mathbb{E}[P])^2 \right] < \varepsilon^2$ with a computational complexity C with bound

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

In (2) we have used the same estimator for the payoff P_ℓ on every level ℓ, and therefore (1) is a trivial identity due to the telescoping summation. However, in [3] Giles explained that it can be better to use different estimators for the finer and coarser of the two levels being considered, P_ℓ^f when level ℓ is the finer level, and P_ℓ^c when level ℓ is the coarser level. In this case, we require that

$$\mathbb{E}[P_\ell^f] = \mathbb{E}[P_\ell^c] \quad \text{for } \ell = 0, \dots, L-1, \tag{3}$$

so that $\mathbb{E}[P_L^f] = \mathbb{E}[P_0^f] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell^f - P_{\ell-1}^c]$. The MLMC Theorem is still applicable to this modified estimator. The advantage is that it gives the flexibility to construct approximations for which $P_\ell^f - P_{\ell-1}^c$ is much smaller than the original $P_\ell - P_{\ell-1}$, giving a larger value for β, the rate of variance convergence in condition *iii)* in the theorem. In the next sections we demonstrate how suitable choice of P_ℓ^f and P_ℓ^c can dramatically increase the convergence of the variance of the MLMC estimator.

2.1 Milstein Scheme

Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete probability space with a filtration $\{\mathscr{F}_t\}_{t\geq 0}$ satisfying the usual conditions, and let $w(t)$ be a m-dimensional Brownian motion defined on the probability space. We consider the numerical approximation of SDEs of the form

$$\mathrm{d}x(t) = f(x(t))\,\mathrm{d}t + g(x(t))\,\mathrm{d}w(t), \tag{4}$$

where $x(t) \in \mathbb{R}^d$ for each $t \geq 0$, $f \in C^2(\mathbb{R}^d, \mathbb{R}^d)$, $g \in C^2(\mathbb{R}^d, \mathbb{R}^{d\times m})$ have bounded first and second derivatives, and for simplicity we assume a fixed initial value $x_0 \in \mathbb{R}^d$.

For Lipschitz continuous payoffs that depend on finite number of times $t_n^\ell = n\Delta t_\ell$, the MLMC variance can be estimated from the strong convergence of the numerical scheme, that is

$$\left(\mathbb{E}\big[\sup_{0\leq n\leq 2^\ell} \left\| x(t_n^\ell) - X_n^\ell \right\|^p \big] \right)^{1/p} = O(\Delta t_\ell^\xi) \quad \text{for} \quad p \geq 2.$$

For partition $\mathscr{P}_{\Delta t_\ell} := \{n\Delta t_\ell : n = 0, 1, 2, ..., 2^\ell = N\}$, where $\Delta t_\ell = T/N$, we consider the Milstein approximation X_n^ℓ with i^{th} component of the form

$$\begin{aligned}
X_{i,n+1}^\ell =\; & X_{i,n}^\ell + f_i(X_n^\ell)\,\Delta t_\ell + \sum_{j=1}^m g_{ij}(X_n^\ell)\,\Delta w_{j,n}^\ell \\
& + \sum_{j,k=1}^m h_{ijk}(X_n^\ell)\left(\Delta w_{j,n}^\ell \Delta w_{k,n}^\ell - \delta_{j,k}\,\Delta t_\ell - [A_{jk}^\ell]_{t_n}^{t_{n+1}} \right)
\end{aligned} \tag{5}$$

where $h_{ijk}(x) = \frac{1}{2}\sum_{l=1}^d g_{lk}(x)\,\frac{\partial g_{ij}}{\partial x_l}(x)$, $\delta_{j,k}$ is a Kronecker delta, $\Delta w_n^\ell = w((n+1)\Delta t_\ell) - w(n\Delta t_\ell)$ and $[A_{jk}^\ell]_{t_n}^{t_{n+1}}$ is the Lévy area defined as

$$[A_{jk}^\ell]_{t_n}^{t_{n+1}} = \int_{t_n^\ell}^{t_{n+1}^\ell} \left(w_j(t) - w_j(t_n^\ell) \right)\mathrm{d}w_k(t) - \int_{t_n^\ell}^{t_{n+1}^\ell} \left(w_k(t) - w_k(t_n^\ell) \right)\mathrm{d}w_j(t). \tag{6}$$

For the Milstein scheme $\xi = 1$ and therefore $\beta = 2$ for smooth payoffs, and hence MLMC has complexity $O(\varepsilon^{-2})$. However, there is no method for simulating Lévy areas with a cost per timestep similar to that of Brownian increments, apart from in dimension 2 [2, 9, 10]. Furthermore, within computational finance, options are often based on the continuously-monitored minimum (or maximum) or the path. The Milstein scheme gives an improved rate of convergence at the simulation times, but to maintain the strong order of convergence for such path-dependent options we use Brownian Bridge interpolation within each timestep $[t_n^\ell, t_{n+1}^\ell]$

$$\tilde{X}^\ell(t) = X_n^\ell + \lambda_\ell\,(X_{n+1}^\ell - X_n^\ell) + g(X_n^\ell)\left(w(t) - w(t_n^\ell) - \lambda\,\Delta w_n^\ell \right) \tag{7}$$

where $\lambda_\ell \equiv (t - t_n^\ell)/\Delta t_\ell$. Using this interpolant, we have the result [7]

$$\mathbb{E}\Big[\sup_{0\le t\le T}\big\|x(t)-\tilde{X}^\ell(t)\big\|^p\Big]=O(|\,\Delta t_\ell\log(\Delta t_\ell)\,|^p).$$

3 Antithetic MLMC estimator

The idea for the antithetic estimator is to exploit the flexibility of the more general MLMC estimator by defining $P^c_{\ell-1}$ to be the usual payoff $P(X^c)$ coming from a level $\ell-1$ coarse simulation X^c, and define P^f_ℓ to be the average of the payoffs $P(X^f),P(X^a)$ coming from an antithetic pair of level ℓ simulations, X^f and X^a.

X^f will be defined in a way which corresponds naturally to the construction of X^c. Its antithetic "twin" X^a will be defined so that it has exactly the same distribution as X^f, conditional on X^c, which ensures that $\mathbb{E}[P(X^f)]=\mathbb{E}[P(X^a)]$ and hence (3) is satisfied, but at the same time $\big(X^f-X^c\big)\approx-(X^a-X^c)$ and therefore $\big(P(X^f)-P(X^c)\big)\approx-(P(X^a)-P(X^c))$, so that $\frac{1}{2}\big(P(X^f)+P(X^a)\big)\approx P(X^c)$. This leads to $\frac{1}{2}\big(P(X^f)+P(X^a)\big)-P(X^c)$ having a much smaller variance than the standard estimator $P(X^f)-P(X^c)$. It was proved in [5], that if $\big\|\frac{\partial P}{\partial x}\big\|\le L_1,\big\|\frac{\partial^2 P}{\partial x^2}\big\|\le L_2$. then for $p\ge 2$,

$$\mathbb{E}\left[\big(\tfrac{1}{2}(P(X^f)+P(X^a))-P(X^c)\big)^p\right]$$
$$\le\ 2^{p-1}L_1^p\,\mathbb{E}\left[\big\|\tfrac{1}{2}(X^f+X^a)-X^c\big\|^p\right]+2^{-p-1}L_2^p\,\mathbb{E}\left[\big\|X^f-X^a\big\|^{2p}\right].$$

In the multidimensional SDE we will show that the Milstein approximation with the Lévy areas set to zero, combined with the antithetic construction, leads to $X^f-X^a=O(\Delta t^{1/2})$ but $\frac{1}{2}(X^f+X^a)-X^c=O(\Delta t)$. Hence, the variance $\mathbb{V}[\frac{1}{2}(P^f_\ell+P^a_\ell)-P^c_{\ell-1}]$ is $O(\Delta t^2)$ for smooth payoffs, which is the same order obtained for scalar SDEs using the Milstein discretisation with its first order strong convergence.

4 Clark-Cameron example

The Clark and Cameron model problem [1] is

$$\mathrm{d}x_1(t)=\mathrm{d}w_1(t),\quad \mathrm{d}x_2(t)=x_1(t)\,\mathrm{d}w_2(t),\tag{8}$$

with $x_1(0)=x_2(0)=0$, and zero correlation between the two Brownian motions $w_1(t)$ and $w_2(t)$. These equations can be integrated exactly over a time interval $[t_n,t_{n+1}]$, where $t_n=n\Delta t$, to give

$$x_1(t_{n+1})=x_1(t_n)+\Delta w_{1,n}$$
$$x_2(t_{n+1})=x_2(t_n)+x_1(t_n)\Delta w_{2,n}+\tfrac{1}{2}\Delta w_{1,n}\Delta w_{2,n}+\tfrac{1}{2}[A_{12}]_{t_n}^{t_{n+1}}\tag{9}$$
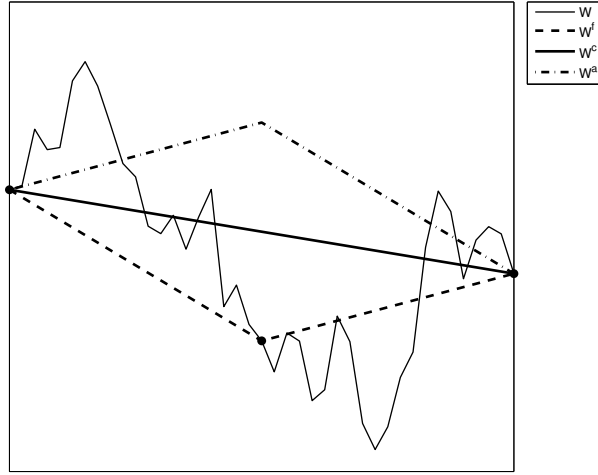
Fig. 1 Brownian path w, its piecewise linear interpolations w^c and w^f, and the antithetic w^a, for a single coarse timestep. The circles denote the points at which the Brownian path is sampled.

where $\Delta w_{i,n} \equiv w_i(t_{n+1}) - w_i(t_n)$, and $[A_{12}]_{t_n}^{t_{n+1}}$ is the Lévy area defined in (6). This corresponds exactly to the Milstein discretisation presented in (5), so for this simple model problem the Milstein discretisation is exact. The point of Clark and Cameron's paper is that for *any* numerical approximation $X(T)$ based solely on the set of discrete Brownian increments Δw, $\mathbb{E}[(x_2(T) - X_2(T))^2] \geq \frac{1}{4} T \Delta t$. Since in this section we use superscript f, a, c for fine X^f, antithetic X^a and coarse X^c approximations, respectively, we drop the superscript ℓ for the clarity of notation.

We define a coarse path approximation X^c with timestep Δt, and times $t_n \equiv n \Delta t$, by neglecting the Lévy area terms to give

$$X_{1,n+1}^c = X_{1,n}^c + \Delta w_{1,n}^{\ell-1}$$
$$X_{2,n+1}^c = X_{2,n}^c + X_{1,n}^c \Delta w_{2,n}^{\ell-1} + \tfrac{1}{2} \Delta w_{1,n}^{\ell-1} \Delta w_{2,n}^{\ell-1} \tag{10}$$

This is equivalent to replacing the true Brownian path by a piecewise linear approximation as illustrated in Figure 1. Similarly, we define the corresponding two half-timesteps of the first fine path approximation X^f. Using

$$\Delta w_{n+1}^{\ell-1} \equiv (w(t_{n+1}) - w(t_n))$$
$$= (w(t_{n+1}) - w(t_{n+1/2})) + (w(t_{n+1/2}) - w(t_n)) \equiv \Delta w_{n+1/2}^{\ell} + \Delta w_n^{\ell},$$

we can combine two half-timestep approximations to obtain an equation for the increment over the coarse timestep,

$$X^f_{1,n+1} = X^f_{1,n} + \Delta w^{\ell-1}_{1,n}$$
$$X^f_{2,n+1} = X^f_{2,n} + X^f_{1,n} \Delta w^{\ell-1}_{2,n} + \tfrac{1}{2} \Delta w^{\ell-1}_{1,n} \Delta w^{\ell-1}_{2,n} \tag{11}$$
$$+ \tfrac{1}{2} \left(\Delta w^\ell_{1,n} \Delta w^\ell_{2,n+1/2} - \Delta w^\ell_{2,n} \Delta w^\ell_{1,n+1/2} \right).$$

The antithetic approximation X^a_n is defined by exactly the same discretisation except that the Brownian increments $\Delta w^\ell_{n+1/2}$ and Δw^ℓ_{n+1} are swapped, as illustrated in Figure 1. This gives

$$X^a_{1,n+1} = X^a_{1,n} + \Delta w^{\ell-1}_{1,n},$$
$$X^a_{2,n+1} = X^a_{2,n} + X^a_{1,n} \Delta w^{\ell-1}_{2,n} + \tfrac{1}{2} \Delta w^{\ell-1}_{1,n} \Delta w^{\ell-1}_{2,n} \tag{12}$$
$$- \tfrac{1}{2} \left(\Delta w^\ell_{1,n} \Delta w^\ell_{2,n+1/2} - \Delta w^\ell_{2,n} \Delta w^\ell_{1,n+1/2} \right).$$

Swapping Δw^ℓ_n and $\Delta w^\ell_{n+1/2}$ does not change the distribution of the driving Brownian increments, and hence X^a has exactly the same distribution as X^f. Note also the change in sign in the last term in (11) compared to the corresponding term in (12). This is important because these two terms cancel when the two equations are averaged.

In [5] Giles and Szpruch proved the following result:

Lemma 1. *If X^f_n, X^a_n and X^c_n are as defined above, then*

$$X^f_{1,n} = X^a_{1,n} = X^c_{1,n}, \quad \tfrac{1}{2} \left(X^f_{2,n} + X^a_{2,n} \right) = X^c_{2,n}, \quad n = 1,2,\ldots,N \equiv 2^{\ell-1}.$$

and

$$\mathbb{E}\left[\left| X^f_{2,N} - X^a_{2,N} \right|^p \right] = O(\Delta t^{p/2}) \quad \text{for } p \geq 2.$$

This allows us to prove that for payoffs which are a smooth function of the final state the MLMC variance

$$\mathbb{V}\left[\tfrac{1}{2} \left(P(X^f_N) + P(X^a_N) \right) - P(X^c_N) \right]$$

has an $O(\Delta t^2)$ upper bound and therefore the complexity of the MLMC estimator is $O(\varepsilon^2)$. This matches the convergence rate and complexity for the multilevel method for scalar SDEs using the standard first order Milstein discretisation, and is much better than the $O(\Delta t)$ MLMC convergence obtained with the Euler-Maruyama discretisation. Very few financial payoff functions are twice differentiable on the entire domain, but Giles and Szpruch have proved that for piecewise smooth put and call options the variance converges with rate $O(\Delta t^{3/2})$, assuming local boundedness of the density of the SDE solution (4) near the strike [5].

To perform numerical experiments we closely follow the algorithm prescribed in [4, Section 5] with predefined root-mean-square errors $\varepsilon = [1,2,4,8,16] \times 10^{-4}$:
1. start with level $\ell = 0$
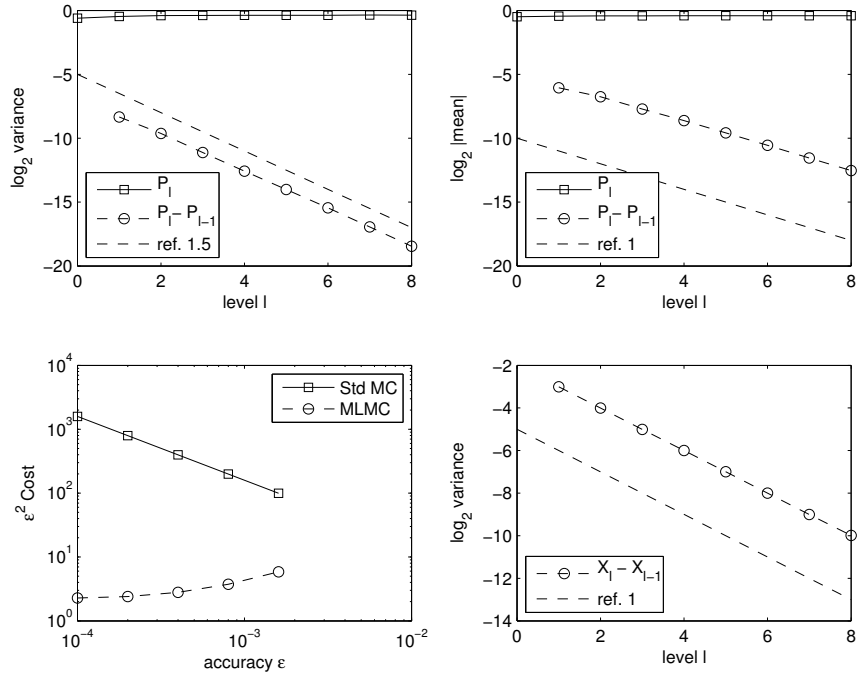2. estimate variance using initial 10^4 samples

Fig. 2 Clark/ Cameron model problem with payoff $\max(0, x_2(1) - 1)$.

3. evaluate optimal number of samples on each level as in [4, Section 5]
4. if $L \geq 2$, test for convergence [4, Section 5]
5. if $L < 2$ or not converged, set $\ell := \ell + 1$ and go to 2.

In addition, we used 10^6 samples to generate the plots where we estimate the rate of the strong and weak errors.

Figure 2 presents results for the payoff function $P = \max(0, x_2(1) - 1)$ applied to Clark and Camerson model problem with initial conditions $x_1(0) = x_2(0) = 1$. The top left plot with the superimposed reference slope with rate 1.5 shows that the variance $P_\ell - P_{\ell-1}$ is $O(\Delta t_\ell^{1.5})$. The top right plot shows that $\mathbb{E}[P_\ell - P_{\ell-1}] = O(\Delta t_\ell)$. The bottom left plot shows $\varepsilon^2 C$ where C is the computational complexity as defined in Theorem 1. The plot is versus ε, and the nearly horizontal line confirms that the MLMC complexity is $O(\varepsilon^{-2})$, whereas the standard Monte Carlo approach has complexity $O(\varepsilon^{-3})$. For accuracy $\varepsilon = 10^{-4}$, the antithetic MLMC is approximately 500 times more efficient than standard Monte Carlo. The bottom right plot shows that $\mathbb{V}[X_{2,N}^\ell - X_{2,N}^{\ell-1}] = O(\Delta t_\ell)$, corresponding to the standard strong convergence of order 0.5.

5 Subsampling of Levy areas

Consider now a Brownian path $w(t)$ on the interval $[0,T]$ with N sub-intervals of size $\Delta t = T/N$. We define $w_n \equiv w(n\Delta t)$ and $\Delta w_n \equiv w_{n+1} - w_n$.

Lemma 2. *The Lévy area for $w(t)$ can be expressed as*

$$[A_{jk}]_0^T = \sum_{n=0}^{N-1} \left((w_{j,n} - w_{j,0})\Delta w_{k,n} - (w_{k,n} - w_{k,0})\Delta w_{j,n} + [A_{jk}^s]_{n\Delta t}^{(n+1)\Delta t} \right)$$

where $[A_{jk}^s]_{n\Delta t}^{(n+1)\Delta t}$ is the Lévy area for the sub-interval $[n\Delta t, (n+1)\Delta t]$.

Proof. This follows from the definition of the Lévy area by expressing the integral over $[0,T]$ as the sum of integrals over each of the sub-intervals, and using the identity $w(t) - w(0) = (w_n - w_0) + (w(t) - w_n))$ to evaluate the integral on the nth sub-interval.

Ignoring the sub-interval Lévy areas $[A_{jk}^s]_{n\Delta t}^{(n+1)\Delta t}$, which corresponds to using the expected value of $[A_{jk}]_0^T$ conditional on $\{w(n\Delta t)\}_{0\leq n\leq N}$, gives the Lévy area approximation:

$$[L_{jk}]_0^T = \sum_{n=0}^{N-1} \left((w_{j,n} - w_{j,0})\Delta w_{k,n} - (w_{k,n} - w_{k,0})\Delta w_{j,n} \right).$$

We denote by $[L_{jk}^a]_0^T$ the corresponding antithetic quantity generated by reversing the order of the Brownian increments $\Delta w_N, \Delta w_{N-1}, \ldots, \Delta w_1$. The antithetic label is due to the following lemma:

Lemma 3.
$$[L_{jk}^a]_0^T = -[L_{jk}]_0^T.$$

Proof.

$$\begin{aligned}
[L_{jk}^a]_0^T &= \sum_{n=0}^{N-1} \sum_{m=0}^{n-1} \Delta w_{j,N-1-m}\Delta w_{k,N-1-n} - \Delta w_{k,N-1-m}\Delta w_{j,N-1-n} \\
&= \sum_{m'=0}^{N-1} \sum_{n'=m'+1}^{N-1} \Delta w_{j,n'}\Delta w_{k,m'} - \Delta w_{k,n'}\Delta w_{j,m'} \\
&= -\sum_{n'=0}^{N-1} \sum_{m'=0}^{n'-1} \Delta w_{j,m'}\Delta w_{k,n'} - \Delta w_{k,m'}\Delta w_{j,n'} \\
&= -[L_{jk}]_0^T
\end{aligned}$$

The second line in the proof uses the substitutions $m' = N-1-n$, $n' = N-1-m$, and the third line simply switches the order of summation.

5.1 Antithetic subsampling

In Section 4 we showed that by setting the Lévy area to zero and using a suitable antithetic treatment we obtained an MLMC variance with the same order as the Milstein scheme for scalar SDEs. However, to obtain similarly good results for payoffs which depend on the path minimum (or maximum) we are not able to completely neglect the Lévy areas. Instead, for reasons which would require a lengthy explanation and will be addressed in future work, we need to improve the rate of strong convergence from $1/2$ to $3/4$ by approximating the Lévy areas by sub-sampling the driving Brownian path. Let M_f denote the number of subsamples required to approximate the Lévy area on the fine timestep. The subsampling timestep is given by $\delta_\ell = 2^{-\ell} T / M_f$. Since we want to obtain

$$\mathbb{E} \left\| [L_{jk}]_{t_n}^{t_{n+1/2}} - [A_{jk}]_{t_n}^{t_{n+1/2}} \right\|^2 = O((2^{-\ell})^{3/2}),$$

we need to take $M_f \approx 2^{\ell/2}$ sub-samples in each fine timestep. By the same reasoning we take $M_c \approx 2^{(\ell-1)/2}$, with sub-sampling timestep $\delta_{\ell-1} = 2^{-(\ell-1)} T / M_c$. For implementation we round the exponents, using $M_f = 2^{\lceil \ell/2 \rceil}$ and $M_c = 2^{\lceil (\ell-1)/2 \rceil}$.
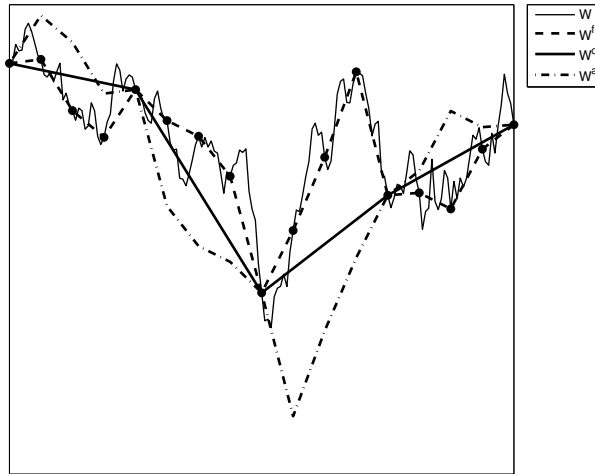

Fig. 3 Brownian path w, its piecewise linear interpolations w^c and w^f, and the antithetic w^a, for a single coarse timestep. The circles denote the points at which the Brownian path is sampled.

Figure 3 illustrates a case in which w^c has $M_c = 4$ sub-sampling intervals within each coarse timestep, and w^f has $M_f = 8$ sub-sampling intervals within each fine timestep (this corresponds to level $\ell = 5$). With sub-sampling, the piecewise linear antithetic fine path w^a is defined by a time-reversal of the Brownian increments within each of the coarse sub-sampling intervals. In the case illustrated, the first

coarse sub-sampling interval contains 4 fine sub-sampling intervals, so these 4 increments $\Delta w^{f,1}, \Delta w^{f,2}, \Delta w^{f,3}, \Delta w^{f,4}$ are re-ordered as $\Delta w^{f,4}, \Delta w^{f,3}, \Delta w^{f,2}, \Delta w^{f,1}$ to give the increments for w^a.

First we represent the Lévy area approximation on the coarse time interval as a sum of two approximations each with $M_c/2$ subsamples

$$[L_{jk}^c]_{t_n}^{t_{n+1}} = [L_{jk}^c]_{t_n}^{t_{n+1/2}} + [L_{jk}^c]_{t_{n+1/2}}^{t_{n+1}} + \left(\Delta w_{1,n+1/2}^\ell \Delta w_{2,n+1}^\ell - \Delta w_{2,n+1/2}^\ell \Delta w_{1,n+1}^\ell \right). \tag{13}$$

We can represent the Lévy area approximation for the first fine timestep within a coarse timestep as

$$[L_{jk}^f]_{t_n}^{t_{n+1/2}} = [L_{jk}^c]_{t_n}^{t_{n+1/2}} + \sum_{s=0}^{M_c/2-1} [L_{jk}^f]_{t_n+s\delta_{\ell-1}}^{t_n+(s+1)\delta_{\ell-1}}, \tag{14}$$

where $[L_{jk}^f]_{t_n+s\delta_{\ell-1}}^{t_n+(s+1)\delta_{\ell-1}}$ are Lévy area approximatiosn with $\frac{2M_f}{M_c}$ subsamples. Notice that $\frac{M_c}{2}\delta_{\ell-1} = 2^{-\ell}$. In the same way, we represent the Lévy area approximation for the antithetic path as

$$[L_{jk}^a]_{t_n}^{t_{n+1/2}} = [L_{jk}^c]_{t_n}^{t_{n+1/2}} + \sum_{s=0}^{M_c/2-1} [L_{jk}^a]_{t_n+s\delta_{\ell-1}}^{t_n+(s+1)\delta_{\ell-1}}. \tag{15}$$

Due to Proposition 3, $[L_{jk}^a]_{t_n+s\delta_{\ell-1}}^{t_n+(s+1)\delta_{\ell-1}} = -[L_{jk}^f]_{t_n+s\delta_{\ell-1}}^{t_n+(s+1)\delta_{\ell-1}}$. Hence

$$\left([L_{jk}^f]_{t_n}^{t_{n+1/2}} - [L_{jk}^c]_{t_n}^{t_{n+1/2}} \right) = -\left([L_{jk}^a]_{t_n}^{t_{n+1/2}} - [L_{jk}^c]_{t_n}^{t_{n+1/2}} \right),$$

which is the key antithetic property required for higher order MLMC variance convergence. We derive the analogous approximation for the second fine timestep within a coarse timestep. Returning to the Clark-Cameron example, where we focus only on the equation for x_2, with Lévy area approximation using M_c and M_f subsamples respectively, we obtain

$$X_{2,n+1}^c = X_{2,n}^c + X_{1,n}^c \Delta w_{2,n}^{\ell-1} + \tfrac{1}{2} \Delta w_{1,n}^{\ell-1} \Delta w_{2,n}^{\ell-1} + \tfrac{1}{2}[L_{jk}^c]_{t_n}^{t_{n+1}} \tag{16}$$

$$X_{2,n+1}^f = X_{2,n}^f + X_{1,n}^f \Delta w_{2,n}^{\ell-1} + \tfrac{1}{2} \Delta w_{1,n}^{\ell-1} \Delta w_{2,n}^{\ell-1} + \tfrac{1}{2}[L_{jk}^f]_{t_n}^{t_{n+1/2}} + \tfrac{1}{2}[L_{jk}^f]_{t_{n+1/2}}^{t_{n+1}}$$
$$+ \tfrac{1}{2}\left(\Delta w_{1,n}^\ell \Delta w_{2,n+1/2}^\ell - \Delta w_{2,n}^\ell \Delta w_{1,n+1/2}^\ell \right), \tag{17}$$

$$X_{2,n+1}^a = X_{2,n}^a + X_{1,n}^a \Delta w_{2,n}^{\ell-1} + \tfrac{1}{2} \Delta w_{1,n}^{\ell-1} \Delta w_{2,n}^{\ell-1} + \tfrac{1}{2}[L_{jk}^a]_{t_n}^{t_{n+1/2}} + \tfrac{1}{2}[L_{jk}^a]_{t_{n+1/2}}^{t_{n+1}}$$
$$+ \tfrac{1}{2}\left(\Delta w_{1,n}^\ell \Delta w_{2,n+1/2}^\ell - \Delta w_{2,n}^\ell \Delta w_{1,n+1/2}^\ell \right), \tag{18}$$

where we use Lévy areas approximations (13) and (14). We present a lemma that can be proved in a similar way to Lemma 3.1 in [5]:

Lemma 4. *If X_n^f, X_n^a and X_n^c are as defined above, and $N = 2^{\ell-1}$, then*

$$X_{1,n}^f = X_{1,n}^a = X_{1,n}^c, \quad \tfrac{1}{2}\left(X_{2,n}^f + X_{2,n}^a\right) = X_{2,n}^c, \quad n = 1, 2, \dots, N$$

and

$$\sup_{0 \le n \le N} \mathbb{E}\left[\left|X_{2,n}^f - X_{2,n}^a\right|^p\right] = O(\Delta t^{\frac{3}{4}p})$$

Our numerical experiments show that for lookback and barrier options the MLMC variance

$$\mathbb{V}\left[\tfrac{1}{2}\left(P(X^f) + P(X^a)\right) - P(X^c)\right]$$

has an $O(\Delta t^{3/2})$ upper bound. Since we use subsampling to approximate the Lévy areas, the computational cost corresponds to $\gamma = 3/2$ in Theorem 1, and as a consequence the complexity of the MLMC estimator is $O(\varepsilon^{-2}(\log\varepsilon)^2)$, whereas the standard Monte Carlo simulation complexity is $O(\varepsilon^{-3})$.

6 Lookback and barrier options

6.1 Lookback options

Lookback options are based on the minimum (or maximum) of the simulated path. As a specific example, we consider the payoff $P = x_2(T) - \min_{0 < t < T} x_2(t)$, based on the second component x_2 of the Clark and Cameron model problem.

To improve the convergence we use $\tilde{X}^\ell(t)$ defined in (7). We have

$$\min_{0 \le t < T} \tilde{X}_2^\ell(t) = \min_{0 \le n < 2^{\ell-1}} X_{2,n,min}^\ell,$$

where the minimum of the fine approximation over the fine timestep $[t_n^{\ell-1}, t_{n+1/2}^{\ell-1}]$ is given by [6]

$$X_{2,n,min}^\ell = \tfrac{1}{2}\left(X_{2,n}^\ell + X_{2,n+1/2}^\ell - \sqrt{\left(X_{2,n+1/2}^\ell - X_{2,n}^\ell\right)^2 - 2\,g_2(X_n^\ell)^2\,\Delta t_\ell \log U_n}\right),$$

$$\tag{19}$$

where U_n is a uniform random variable on the unit interval. The minima for the antithetic path are defined similarly, using the same uniform random numbers U_n.

For the coarse path, we do something slightly different. Using the same Brownian interpolation, we use equation (7) to define $\tilde{X}_{n+1/2}^{\ell-1} \equiv \tilde{X}^{\ell-1}((n+\tfrac{1}{2})\Delta t_{\ell-1})$. Given this interpolated value, the minimum value over the coarse interval can then be taken to be the smaller of the minima for the two fine intervals

$$X^{\ell-1}_{2,n,min} = \frac{1}{2} \left(X^{\ell-1}_{2,n} + \tilde{X}^{\ell-1}_{2,n+1/2} - \sqrt{\left(\tilde{X}^{\ell-1}_{2,n+1/2} - X^{\ell-1}_{2,n} \right)^2 - 2g_2(X^{\ell-1}_n)^2 \Delta t_\ell \log U_n} \right),$$

$$X^{\ell-1}_{2,n+1/2,min} = \frac{1}{2} \left(\tilde{X}^{\ell-1}_{2,n+1/2} + X^{\ell-1}_{2,n+1} - \sqrt{\left(X^{\ell-1}_{2,n+1} - \tilde{X}^{\ell-1}_{2,n+1/2} \right)^2 - 2g_2(X^{\ell-1}_n)^2 \Delta t_\ell \log U_{n+1/2}} \right)$$

$$(20)$$

Note that we use $g_2(X^{\ell-1}_n)$ for both fine timesteps, because we have used the Brownian Bridge with diffusion term $g_2(X^{\ell-1}_n)$ to derive both minima. If we changed $g_2(X^{\ell-1}_n)$ to $g_2(\tilde{X}^{\ell-1}_{n+1/2})$ in $X^{\ell-1}_{2,n+1/2,min}$, this would mean that we used different Brownian Bridge on the first and second half of the coarse timestep and as a consequence we would violate (3). Note also the re-use of the same uniform random numbers U_n and $U_{n+1/2}$ used to compute the fine path minima. To perform numerical experiments we closely follow the algorithm prescribed in [4]. The results in Figure 4 are for the Clark and Cameron model problem with this lookback payoff. The top left left plot shows the behaviour of the variance of both P_ℓ and $P_\ell - P_{\ell-1}$. The superimposed reference slope with rate 1.5 indicates that the variance $V_\ell = \mathbb{V}[P_\ell - P_{\ell-1}] = O(\Delta t_\ell^{1.5})$, corresponding to $O(\varepsilon^{-2}(\log \varepsilon)^2)$ computational complexity for the antithetic MLMC estimator. The top right plot shows that $\mathbb{E}[P_\ell - P_{\ell-1}] = O(\Delta t_\ell)$. The bottom left plot shows computational complexity C (as defined in Theorem 1) with desired accuracy ε. The plot is of $\varepsilon^2 C$ versus ε, because we expect to see that $\varepsilon^2 C$ is only weakly dependent on ε for MLMC. For standard Monte Carlo without subsampling of the Lévy areas, theory predicts that $\varepsilon^2 C$ 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 order. For accuracy $\varepsilon = 10^{-4}$, the antithetic MLMC is over 100 times more efficient than standard Monte Carlo. The bottom right plot shows that $\mathbb{V}[X^\ell_2 - X^{\ell-1}_2] = O(\Delta t_\ell^{3/2})$. This corresponds to the standard strong convergence of order 3/4.

6.2 Barrier options

The barrier option which is considered is a down-and-out option for which the payoff is a Lipschitz function of the value of the underlying at maturity, provided the underlying has never dropped below a value B, i.e. $P = f(x_2(T)) \mathbf{1}_{\{\tau > T\}}$, where the crossing time τ is defined as $\tau = \inf\{t : x_2(t) < B\}$. Using the Brownian Bridge interpolation, we can approximate $\mathbf{1}_{\{\tau > T\}}$ by $\prod_{n=0}^{2^{\ell-1}-1/2} \mathbf{1}_{\{X^\ell_{2,n,min} \geq B\}}$, where $X^\ell_{2,n,min}$ is defined in equation (19). This suggests following the lookback approximation in computing the minimum of both the fine and coarse paths. However, the variance would be larger in this case because the payoff is a discontinuous function of the minimum. A better treatment, which is the one used in [3], is to use the conditional Monte Carlo approach to further smooth the payoff. Since the process X^ℓ_n is Markovian, we have

$$\mathbb{E}\left[f(X_{2,N}^\ell)\prod_{n=0}^{N-1/2}\mathbf{1}_{\{X_{n,min}^\ell\geq B\}}\right] = \mathbb{E}\left[f(X_{2,N}^\ell)\,\mathbb{E}\left[\prod_{n=0}^{N-1/2}\mathbf{1}_{\{X_{2,n,min}^\ell\geq B\}}\mid X_0^\ell,\ldots,X_N^\ell\right]\right]$$

$$= \mathbb{E}\left[f(X_{2,N}^\ell)\prod_{n=0}^{N-1/2}\mathbb{E}\left[\mathbf{1}_{\{X_{2,n,min}^\ell\geq B\}}\mid X_n^\ell,X_{n+1/2}^\ell\right]\right]$$

$$= \mathbb{E}\left[f(X_{2,N}^\ell)\prod_{n=0}^{N-1/2}(1-p_n^\ell)\right],$$

where

$$p_n^\ell = \mathbb{P}\left(\inf_{t_n<t<t_{n+1/2}}\tilde{X}_2^\ell(t)<B\mid X_n^\ell,X_{n+1/2}^\ell\right) = \exp\left(\frac{-2\,(X_{2,n}^\ell-B)^+(X_{2,n+1/2}^\ell-B)^+}{g_2(X_n^\ell)^2\,\Delta t_\ell}\right)$$

The antithetic path is treated similarly. For the payoff for the coarse path we sub-sample $\tilde{X}_{n+1/2}^{\ell-1}$, as we did for the lookback option, to obtain
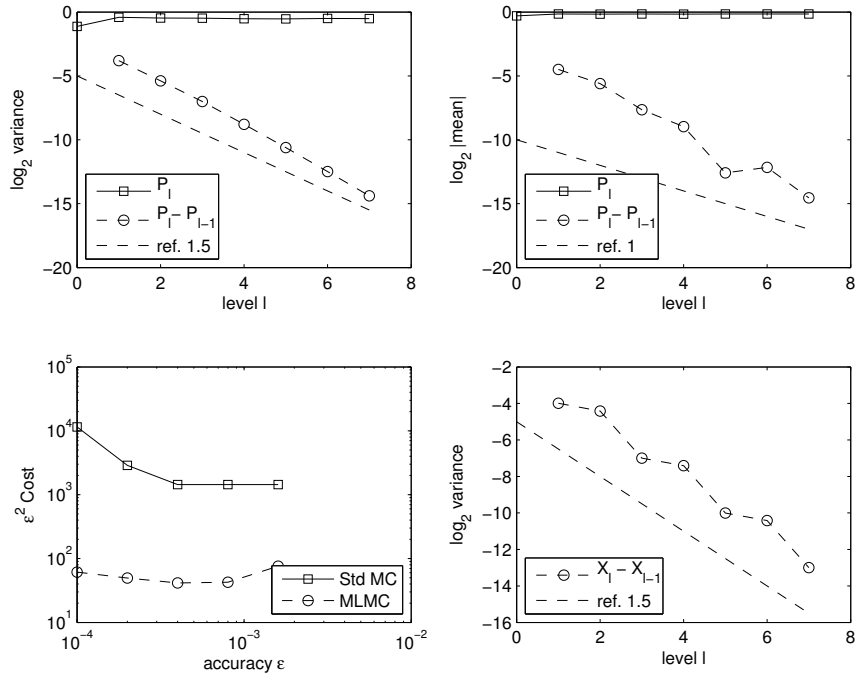

Fig. 4 Clark / Cameron model with payoff $x_2(1)-\min x_2(t)$.

$$\mathbb{E}\left[f(X_{2,N}^{\ell-1})\prod_{n=0}^{N-1/2}\mathbf{1}_{\{X_{2,n,min}^{\ell-1}\geq B\}}\right] = \mathbb{E}\left[f(X_{2,N}^{\ell-1})\prod_{n=0}^{N-1/2}(1-p_n^{\ell-1})\right],$$

where, for integer n,

$$p_n^{\ell-1} = \exp\left(\frac{-2\,(X_{2,n}^{\ell-1}-B)^+(\tilde{X}_{n2,+1/2}^{\ell-1}-B)^+}{g_2(X_n^{\ell-1})^2\,\Delta t_\ell}\right),$$

$$p_{n+1/2}^{\ell-1} = \exp\left(\frac{-2\,(\tilde{X}_{2,n+1/2}^{\ell-1}-B)^+(X_{2,n+1}^{\ell-1}-B)^+}{g_2(X_n^{\ell-1})^2\,\Delta t_\ell}\right).$$

Note that the same $g_2(X_n^{\ell-1})$ is used to calculate both probabilities for the same reason as for the lookback option.

The results in Figure 5 are for barrier option with barrier $B = 0.1$. The top left left plot shows the behaviour of the variance of both P_ℓ and $P_\ell - P_{\ell-1}$. The superimpose reference slope with rate 1.5 indicates that the variance $V_\ell = \mathbb{V}[P_\ell - P_{\ell-1}] = O(\Delta t_\ell^{1.5})$. This corresponds to an $O(\varepsilon^2(\log\varepsilon)^2)$ computational complexity for the antithetic MLMC, due to the additional cost of the sub-sampling to approximate the Lévy areas. The top right plot shows that $\mathbb{E}[P_\ell - P_{\ell-1}] = O(\Delta t_\ell)$. The bottom left plot
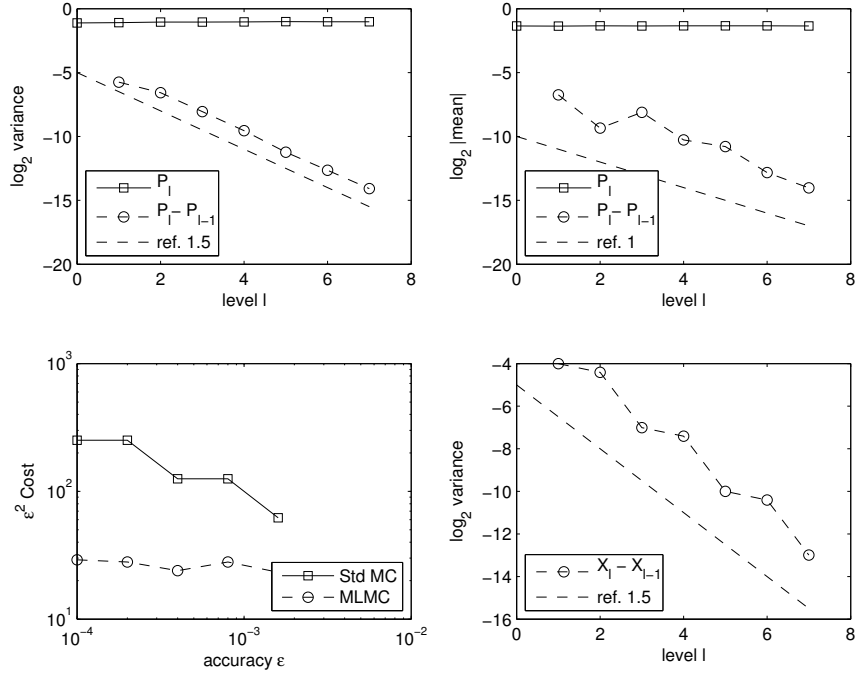

Fig. 5 Clark / Cameron model with payoff $\min(x_2(1)-1,0)\,\mathbf{1}_{\min x_2(t)>0.1}$.

shows the variation of the computational complexity C with desired accuracy ε. For standard Monte Carlo without subsampling of the Lévy areas, theory predicts that $\varepsilon^2 C$ 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 order. For accuracy $\varepsilon = 10^{-4}$, antithetic MLMC is almost 10 times more efficient than standard Monte Carlo. The bottom right plot shows that $\mathbb{V}[X_2^\ell - X_2^{\ell-1}] = O(\Delta t_\ell^{3/2})$. This corresponds to standard strong convergence of order 3/4.

7 Conclusions

In this paper we extended results from [3] and [5] to lookback and barrier options for multidimensional SDEs. By suitable modification of the antithetic MLMC estimator, using sub-sampling of the driving Brownian path to approximate the Lévy areas, we obtained $O(\varepsilon^{-2} \log(\varepsilon)^2)$ complexity for barrier and lookback options. Similar results have also been obtained for digital options which are a discontinuous function of the final state, but they have been omitted here due to lack of space.

References

1. J.M.C. Clark and R.J. Cameron. The maximum rate of convergence of discrete approximations for stochastic differential equations. In B. Grigelionis, editor, *Stochastic Differential Equations, No. 25 in Lecture Notes in Control and Information Sciences*. Springer-Verlag, 1980.
2. J.G. Gaines and T.J. Lyons. Random generation of stochastic integrals. *SIAM Journal of Applied Mathematics*, 54(4):1132–1146, 1994.
3. M.B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer-Verlag, 2008.
4. M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
5. M.B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multidimensional SDEs without Lévy area simulation. *Arxiv preprint arXiv:1202.6283*, 2012.
6. P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.
7. T. Müller-Gronbach. The optimal uniform approximation of systems of stochastic differential equations. *The Annals of Applied Probability*, 12(2):664–690, 2002.
8. T. Müller-Gronbach. *Strong approximation of systems of stochastic differential equations*. Habilitation thesis, TU Darmstadt, 2002.
9. T. Rydén and M. Wiktorsson. On the simulation of iterated Itô integrals. *Stochastic Processes and their Applications*, 91(1):151–168, 2001.
10. M. Wiktorsson. Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. *Annals of Applied Probability*, 11(2):470–487, 2001.