Pricing Exotic Options using Strong Convergence Properties*

Klaus Schmitz Abe and Michael Giles

Mathematical Institute, Oxford University, Oxford, U.K.
schmitz@maths.ox.ac.uk

Summary. In finance, the strong convergence properties of discretisations of stochastic differential equations (SDEs) are very important for the hedging and valuation of exotic options. In this paper we show how the use of the Milstein scheme can improve the convergence of the multi-level Monte Carlo method, so that the computational cost to achieve an accuracy of $O(e)$ is reduced to $O(\epsilon^{-2})$ for a Lipschitz payoff. The Milstein scheme gives first order strong convergence for all 1−dimensional systems (one Wiener process). However, for processes with two or more Wiener processes, such as correlated portfolios and stochastic volatility models, there is no exact solution for the iterated integrals of second order (Lévy area) and the Milstein scheme neglecting the Lévy area gives the same order of convergence as the Euler-Maruyama scheme. The purpose of this paper is to show that if certain conditions are satisfied, we can avoid the calculation of the Lévy area and obtain first convergence order by applying an orthogonal transformation. We demonstrate when the conditions of the 2−Dimensional problem permit this and give an exact solution for the orthogonal transformation. We present examples of pricing exotic options to demonstrate that the use of both the orthogonal Milstein scheme and the Multi-level Monte Carlo give a substantial reduction in the computation cost.

Key words: Stochastic volatility models, Milstein scheme, Lévy area, orthogonal transformation, pricing exotic options.

1 Introduction

We begin with a 2−Dimensional Itô stochastic differential equation (SDE) with a 2−Dimensional Wiener process:

$$\begin{aligned} dx &= \mu^{(x)}(x,y)\,dt + \sigma(x,y)\,d\widehat{W}_{1,t} \\ dy &= \mu^{(y)}(x,y)\,dt + \xi(x,y)\,d\widehat{W}_{2,t} \qquad\qquad \rho\,dt = \left\langle d\widehat{W}_{1,t}, d\widehat{W}_{2,t} \right\rangle \end{aligned} \tag{1}$$

* 14th European Conference on Mathematics for Industry (ECMI 2006), $10-14^{th}$ July 2006, Madrid, Spain.

Alternatively, in vector form:

$$dZ(t) = A_0\,(t, Z)\,dt + \sum_{k=1}^{2} A_k\,(t, Z)\,d\widehat{W}_{k,t} \qquad Z \in \mathbb{R}^2$$

This is in fact, only a symbolic representation for the stochastic integral equation:

$$Z(t) = Z(t_0) + \int_{t_0}^{t} A_0(s, Z)\,ds + \sum_{k=1}^{2} \int_{t_0}^{t} A_k(s, Z)\,d\widehat{W}_{k,s}$$

The first integral is a deterministic Riemann integral and the second is a stochastic integral.

Using the definition of correlation we can represent our system (1) in vector form with independent noise:

$$d\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mu^{(x)}(x,y) \\ \mu^{(y)}(x,y) \end{bmatrix} dt + \begin{bmatrix} \sigma(x,y) \\ \rho\,\xi(x,y) \end{bmatrix} dW_{1,t} + \begin{bmatrix} 0 \\ \widehat{\rho}\,\xi(x,y) \end{bmatrix} dW_{2,t} \qquad (2)$$

$$\langle dW_{1,t}, dW_{2,t} \rangle = 0 \qquad \widehat{\rho} = \sqrt{1-\rho^2}$$

The Milstein approximation is:

$$\begin{bmatrix} x_{t+\Delta t} \\ y_{t+\Delta t} \end{bmatrix} = \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} \mu^{(x_t)} \\ \mu^{(y_t)} \end{bmatrix} \Delta t + \begin{bmatrix} \sigma \\ \rho\xi \end{bmatrix} \Delta W_{1,t} + \begin{bmatrix} 0 \\ \widehat{\rho}\xi \end{bmatrix} \Delta W_{2,t} \qquad (3)$$

$$\frac{1}{2}\begin{bmatrix} \sigma\sigma_x + \rho\xi\sigma_y \\ \rho\sigma\xi_x + \rho^2\xi\xi_y \end{bmatrix} \left((\Delta W_{1,t})^2 - \Delta t \right) + \frac{1}{2}\begin{bmatrix} 0 \\ \widehat{\rho}^2\xi\xi_y \end{bmatrix} \left((\Delta W_{2,t})^2 - \Delta t \right)$$

$$+\frac{1}{2}\begin{bmatrix} \widehat{\rho}\xi\sigma_y \\ \widehat{\rho}\sigma\xi_x + 2\rho\widehat{\rho}\xi\xi_y \end{bmatrix} (\Delta W_{1,t}\Delta W_{2,t}) + \frac{1}{2}\left[A_1, A_2\right]\left[L_{(1,2)}\right]_t^{t+\Delta t}$$

where subscript x and y denote partial derivatives, $L_{(1,2)}$ is the Lévy area defined by:

$$\left[L_{(1,2)}\right]_t^{t+\Delta t} = \int_t^{t+\Delta t} \int_t^{S+\Delta t} dW_{1,U} dW_{2,S} - \int_t^{t+\Delta t} \int_t^{S+\Delta t} dW_{2,U} dW_{1,S}$$

and $[A_1, A_2]$ is the Lie bracket defined by (∂_{A_i} is the Jacobian matrix of A_i):

$$[A_1, A_2] = (\partial_{A_2} A_1 - \partial_{A_1} A_2) = \begin{bmatrix} -\widehat{\rho}\xi\sigma_y \\ \widehat{\rho}\sigma\xi_x \end{bmatrix}$$

The numerical difficulty is how to calculate the Lévy area $L_{(1,2)}$. The technique of Gaines and Lyons [6] can be used to sample the distribution for $L_{(1,2)}$ conditional on ΔW_1, ΔW_2. However there is no generalisation of this to higher dimensions apart from the approximation of [9], which has a significant computational cost.

2 Orthogonal Transformation

If we make an orthogonal transformation of the uncorrelated Wiener processes in (2), we do not change the distribution and we obtain:

$$\begin{aligned} d\widetilde{x} &= \mu^{(\widetilde{x})}(\widetilde{x},\widetilde{y})\,dt + \sigma(\widetilde{x},\widetilde{y})\,d\widetilde{\widehat{W}}_{1,t} \\ d\widetilde{y} &= \mu^{(\widetilde{y})}(\widetilde{x},\widetilde{y})\,dt + \xi(\widetilde{x},\widetilde{y})\,d\widetilde{\widehat{W}}_{2,t} \end{aligned} \tag{4}$$

where:

$$\begin{bmatrix} d\widetilde{\widehat{W}}_{1,t} \\ d\widetilde{\widehat{W}}_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \widehat{\rho} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix}$$

If we compute the Lie bracket for the new orthogonal process using independent Brownian paths $W_{1,t}$,$W_{2,t}$, we have:

$$[A_1, A_2] = \begin{bmatrix} -\widehat{\rho}\xi\sigma_{\widetilde{y}} - \sigma^2\theta_{\widetilde{x}} - \rho\sigma\xi\theta_{\widetilde{y}} \\ \widehat{\rho}\sigma\xi_{\widetilde{x}} - \rho\sigma\xi\theta_{\widetilde{x}} - \xi^2\theta_{\widetilde{y}} \end{bmatrix}$$

To avoid having to simulate the Lévy area, we need the Lie brackets to be identically zero [1], i.e., we need to impose the following conditions:

$$\begin{aligned} -\widehat{\rho}\xi\sigma_{\widetilde{y}} - \sigma^2\theta_{\widetilde{x}} - \rho\sigma\xi\theta_{\widetilde{y}} &= 0 \\ +\widehat{\rho}\sigma\xi_{\widetilde{x}} - \rho\sigma\xi\theta_{\widetilde{x}} - \xi^2\theta_{\widetilde{y}} &= 0 \end{aligned}$$

Simplifying we get:

$$\begin{aligned} \Phi &\doteq \frac{\partial\theta}{\partial\widetilde{x}} = \frac{-1}{\widehat{\rho}}\left(\frac{\xi\sigma_{\widetilde{y}}}{\sigma^2} + \frac{\rho\xi_{\widetilde{x}}}{\xi}\right) \\ \Psi &\doteq \frac{\partial\theta}{\partial\widetilde{y}} = \frac{1}{\widehat{\rho}}\left(\frac{\rho\sigma_{\widetilde{y}}}{\sigma} + \frac{\sigma\xi_{\widetilde{x}}}{\xi^2}\right) \end{aligned} \tag{5}$$

If we want to find a solution for θ, we must first determine when the system is consistent, or integrable. This requires that:

$$\frac{\partial\Phi}{\partial\widetilde{y}} = \frac{\partial^2\theta}{\partial\widetilde{x}\,\partial\widetilde{y}} = \frac{\partial\Psi}{\partial\widetilde{x}} \tag{6}$$

and the solution for θ is:

$$\theta(\widetilde{x},\widetilde{y}) = \int^{(\widetilde{x},\widetilde{y})} (\Phi\,d\widetilde{x} + \Psi\,d\widetilde{y}\,) \tag{7}$$

However, because not all SDEs satisfy condition (6), we also obtain the following SDE for θ:

$$d\theta = \frac{\partial\theta}{\partial\widetilde{x}}\,d\widetilde{x} + \frac{\partial\theta}{\partial\widetilde{y}}\,d\widetilde{y} = \left(\Phi\mu^{(\widetilde{x})} + \Psi\mu^{(\widetilde{y})}\right)dt + \sigma\Phi d\widetilde{\widehat{W}}_{1,t} + \xi\Psi d\widetilde{\widehat{W}}_{2,t}$$

If we choose to define θ in this way even when condition (6) is not satisfied then our system becomes a 3−Dimensional Itô process with two Wiener process inputs (θ−scheme):

$$\begin{bmatrix} d\widetilde{x} \\ d\widetilde{y} \\ d\theta \end{bmatrix} = \begin{bmatrix} \mu^{(\widetilde{x})} \\ \mu^{(\widetilde{y})} \\ \Phi\mu^{(\widetilde{x})} + \Psi\mu^{(\widetilde{y})} \end{bmatrix} dt + \begin{bmatrix} \sigma \\ 0 \\ \sigma\Phi \end{bmatrix} d\widetilde{\widehat{W}}_{1,t} + \begin{bmatrix} 0 \\ \xi \\ \xi\Psi \end{bmatrix} d\widetilde{\widehat{W}}_{2,t} \tag{8}$$

If we compute again the Lie brackets with independent noise, we obtain:

$$[A_1, A_2] = \begin{bmatrix} 0 \\ 0 \\ \widehat{\rho}\sigma\xi \left(\dfrac{\partial\ \Psi}{\partial\widetilde{x}} - \dfrac{\partial\Phi}{\partial\widetilde{y}} \right) \end{bmatrix} \tag{9}$$

Note that when condition (6) is satisfied this Lie bracket (9) is identically zero. In the remainder of the paper we investigate when particular applications satisfy condition (6), in which case one can discretise either (4) or (8) and when they do not, in which case one can only discretise (8) or the original untransformed SDE (1). Our objective is to try to achieve higher order strong convergence without the simulation of the Lévy areas.

When the Lie bracket is not equal to zero, the important question to be considered is how precisely does θ need to be calculated to obtain first strong order convergence in $\widetilde{x}$ and $\widetilde{y}$? For example, does neglecting the Lie bracket affect the accuracy of θ but not $\widetilde{x}$ and $\widetilde{y}$?

3 Strong Convergence

If we apply any discrete approximation scheme to our system (1) and we want to numerically evaluate the strong convergence order of our approximations $\widehat{X}$, an exact solution is normally required. However, at present, there are no solutions available for many SDEs. Because we are only interested in the distribution of the solution, we can use the next theorems [8] to determine the order of convergence for our discrete time approximation without an exact solution.

Most models can be described through a SDE of the form:

$$dx(t) = \mu\left(x,t\right)dt + \sigma\left(x,t\right)dW(t) \qquad x(0) = x_0 \tag{10}$$

with W a M−Dimensional Brownian motion, μ mapping $\mathbb{R}^N \times [0,\infty)$ into $\mathbb{R}^N$, σ mapping $\mathbb{R}^N \times [0,\infty)$ into $\mathbb{R}^{N\times M}$ and x_0 a random N−Dimensional vector independent of W.

Theorem 1 (Existence and Uniqueness of Strong Solutions).
Suppose $E\left[||x_0||^2\right]$ *is finite and that there is a constant* K *for which for all* $t \in [T_0, T]$ *and all* $x, y \in \mathbb{R}^d$ *the following conditions are satisfied:*

$$\|\mu(x,t) - \mu(y,t)\| + \|\sigma(x,t) - \sigma(y,t)\| \leq K\|x-y\| \quad \textit{(Lipschitz condition)}$$

$$\|\mu(x,t)\| + \|\sigma(x,t)\| \leq K(1+\|x\|) \quad \textit{(Linear growth condition)}$$

Then the SDE (10) admits a strong solution x *and satisfies*

$$\left(E\left[||x(t)||^2\right] < \infty\right).$$

This solution is unique in the sense that if $\widehat{x}$ *is also a solution, then*

$$P\left(x(t) = \widehat{x}(t).\ \forall t \in [0,T]\right) = 1$$

Proof. Proofs and additional explanation can be found in [2] and [5].

Theorem 2 (Convergence Order without an Exact Solution).
A) If a discrete approximation $\widehat{x}$ *of (10) with time step* Δt *has strong convergence order* η*, i.e. there exist a constant* C_1 *such that:*

$$E\left[\ |x(T) - \widehat{x}(T, \triangle t)|\ \right] \leq C_1 \triangle t^{\eta} \tag{11}$$

Then there exist a positive constant, C_2*, such that:*

$$E\left[\ \left|\widehat{x}(T,\triangle t) - \widehat{x}\left(T, \frac{\triangle t}{2}\right)\right|\ \right] \leq C_2 \triangle t^{\eta} \tag{12}$$

B) Conversely, if it is known that the discretisation is strongly convergent and (12) holds for some positive constant C_2*, then the strong convergence order is* η*.*

Proof. A) If (11) is true for all Δt, then:

$$E\left[\ \left|x(T) - \widehat{x}\left(T, \frac{\triangle t}{2}\right)\right|\ \right] \leq C_1 \left(\frac{\triangle t}{2}\right)^{\eta} \tag{13}$$

Using the triangle law ($|A-B| \leq |A| + |B|$) and adding (11) and (13), we get:

$$E\left[\left|\widehat{x}(T,\triangle t) - \widehat{x}\left(T, \frac{\triangle t}{2}\right)\right|\right] \leq C_1\left(1 + \left(\frac{1}{2}\right)^{\eta}\right)\triangle t^{\eta}$$

B) Using the triangle law:

$$\begin{aligned} E\left[\ |x(T) - \widehat{x}(T,\triangle t)|\ \right] \ \leq\ & E\left[\ \left|x(T) - \widehat{x}\left(T, \left(\frac{1}{2}\right)^{M}\triangle t\right)\right|\ \right] \\ & + \sum_{m=0}^{M-1} E\left[\ \left|\widehat{x}\left(T, \left(\frac{1}{2}\right)^{m+1}\triangle t\right) - \widehat{x}\left(T, \left(\frac{1}{2}\right)^{m}\triangle t\right)\right|\ \right] \end{aligned}$$

Due to strong convergence:

$$\lim_{M\to\infty} E\left[\ \left| x(T) - \widehat{x}\left(T, \left(\frac{1}{2}\right)^M \triangle t\right)\right|\ \right] = 0$$

Hence, using (12):

$$E\left[\ |x(T) - \widehat{x}(T, \triangle t)|\ \right] \leq \sum_{m=0}^{\infty} C_2 \left(\frac{1}{2}\right)^{m\,\eta} \triangle t^{\eta} = \frac{C_2}{1-\left(\frac{1}{2}\right)^{\eta}}\ \triangle t^{\eta}$$

4 Stochastic volatility models

In this section we consider three different stochastic volatility models. All three have the following generic form:

$$dx = \mu^{(x)} dt + \alpha\, x^{\gamma_1} y^{\lambda_1} d\widehat{W}_{1,t} \qquad\qquad \rho dt = \left\langle d\widehat{W}_{1.t}, d\widehat{W}_{2.t} \right\rangle$$
$$dy = \mu^{(y)} dt + \beta\, x^{\gamma_2} y^{\lambda_2} d\widehat{W}_{2,t}$$

The integrability condition (6) becomes:

$$\lambda_C\, \lambda_1\, \beta^2\, y^{2\lambda_C} = -\gamma_C\, \gamma_2\, \alpha^2\, x^{2\gamma_C} \tag{14}$$
$$\gamma_C = \gamma_1 - \gamma_2 - 1; \qquad \lambda_C = \lambda_2 - \lambda_1 - 1$$

so then, for $\alpha, \beta, \gamma_2, \lambda_1 \neq 0$, we can conclude that θ is integrable if, and only if, $\lambda_C = \gamma_C = 0$, in which case the solution is:

$$\theta = \frac{1}{\widehat{\rho}}\left(\left(\frac{-\rho\gamma_2\alpha - \lambda_1\beta}{\alpha}\right)\log x + \left(\frac{\rho\lambda_1\beta + \gamma_2\alpha}{\beta}\right)\log y\right) \tag{15}$$

4.1 Quadratic volatility model (Case 1)

The first case we consider is

$$dx = x\,\overline{\mu}\, dt + x\, y\, d\widehat{W}_{1,t} \tag{16}$$
$$dy = k\,(\varpi - y)\, dt + \beta\, y^2 d\widehat{W}_{2,t}$$

with parameters $T=1$; $\rho=0.30$; $\overline{\mu}=0.05$; $k=1.8$; $\varpi=0.26$; $\beta=1$ and initial conditions $x(0)=1$; $y(0)=0.21$.

Because $\lambda_C = \gamma_C = 0$, we can use either equation (4) together with (15), or the 3−Dimensional θ scheme (8). Because of the orthogonal transformation, neither requires the calculation of the Lévy area. Figure 1 and Table 1 show that, as expected, the Euler scheme and the Milstein scheme with zero Lévy areas (setting $L_{(1,2)}=0$ in (3)) give strong convergence order 0.5. On the other hand, the Milstein scheme (3) with a proper value for the distribution of the Lévy area (through simulating the Lévy area using N subintervals within each timestep) gives 1.0 order strong convergence, as do the three orthogonal θ−schemes.

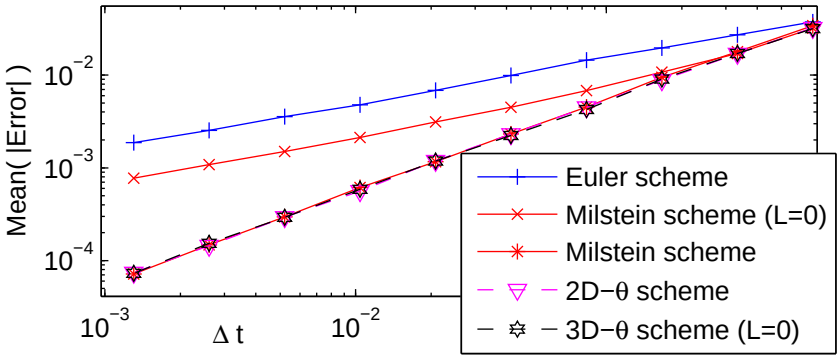

Fig. 1. Convergence test for Case 1.

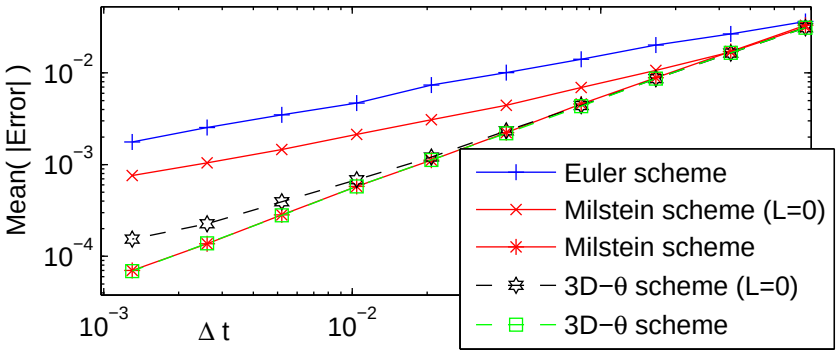

Fig. 2. Convergence test for Case 2.

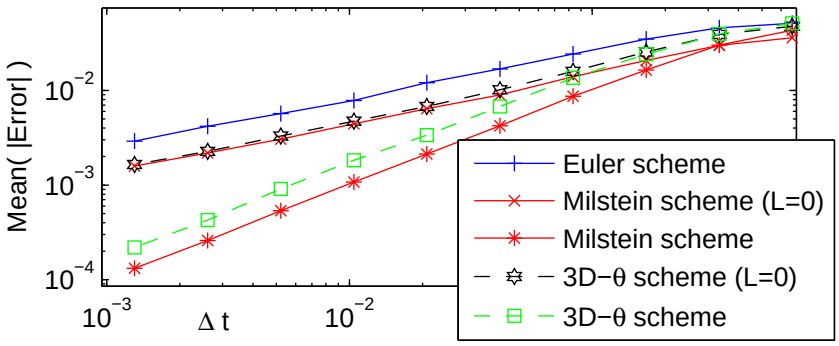

Fig. 3. Convergence test for Case 3.

Scheme	Description	C-1	C-2	C-3
Euler scheme	set $\triangle t{=}dt, \Delta W_i{=}dW_i$ in (2)	0.49	0.49	0.48
Milstein scheme (L=0)	Milstein–(3), set $L_{(1,2)}$ =0	0.56	0.56	0.51
Milstein scheme	Milstein–(3), simulate $L_{(1,2)}$	0.99	0.99	0.95
$2D-\theta$ scheme	Milstein–(4) with (15)	0.98	N/A	N/A
$3D-\theta$ scheme (L=0)	Milstein–(8), set $L_{(1,2)}$ =0	0.98	0.87	0.52
$3D-\theta$ scheme	Milstein–(8), simulate $L_{(1,2)}$	0.98	0.98	0.91

Table 1. Convergence orders η for all cases (N/A = not applicable)

Variance model (Case 2)

The second case we consider is the following stochastic variance model.

$$\begin{aligned} dx &= x\,\overline{\mu}\,dt + x\,\sqrt{y}\;d\widehat{W}_{1,t} \\ dy &= k\,(\varpi - y)\,dt + \beta\,y\;d\widehat{W}_{2,t} \end{aligned} \tag{17}$$

The parameters and initial conditions are the same except for $\beta = 0.5$; $\varpi = 0.26^2$; $y(0) = 0.21^2$, which are chosen so that x and y will have approximately the same relative volatility as in Case 1.

In this case $\lambda_C = 0.5$, and since the integrability condition is not satisfied it is not possible to use the 2D – θ scheme. Figure 2 and Table 1 show that the only schemes that achieved first order convergence are the Milstein schemes which simulate the Lévy area in the simulation. However, Figure 2 shows there is a remarkable difference between the original and the orthogonal scheme without the simulation of the Lévy area, not the improved order of convergence achieved in the first case (Table 1) but a much improved constant of proportionality.

Heston model (Case 3)

A particularly bad case for the orthogonal transformation is the Heston model [4]:

$$\begin{aligned} dx &= x\,\mu\,dt + x\,\sqrt{y}\;d\widehat{W}_{1,t} \\ dy &= k\,(\varpi - y)\,dt + \beta\,\sqrt{y}\;d\widehat{W}_{2,t} \end{aligned} \tag{18}$$

The parameters and initial conditions are the same as in Case 2 except for $\beta = 0.25$; this is again chosen so ensure that x and y will have approximately the same relative volatility as in the first two cases.

In this case, $\lambda_c = 1$. Figure 3 and Table 1 show that neither of the Milstein schemes in which the Lévy areas are set to zero performs very well. Both have order 0.5 strong convergence, and the constant of proportionality is not much better than for the Euler scheme. When the Lévy areas are simulated correctly, the two Milstein schemes do exhibit the expected first order strong convergence. This demonstrates the importance of the Lévy areas in this case.

5 Pricing Exotic Options using ML-MC

Usually, it is the weak convergence properties of numerical discretisations which are most important, because in financial applications one is mostly concerned with the accurate estimation of expected payoffs. However, in the recently developed Multilevel Monte Carlo path simulation method (ML-MC [3]), the strong convergence properties play a crucial role.

The key idea in the ML-MC approach is the use of a multilevel algorithm with different timesteps Δt on each level. Suppose level l uses 2^l timesteps of size $\Delta t_l = 2^{-l}\,T$, and define P_l to be the numerical approximation to the payoff on this level. Let L represent the finest level, with timesteps so small that the bias due to the numerical discretisation is smaller than the accuracy ϵ which is desired. Due to the linearity of the expectation operator, we can express the expectation on the finest grid as:

$$E\left[P_L\right] = E\left[P_0\right] + \sum_{l=1}^{L} E\left[P_l - P_{l-1}\right]$$

The quantity $E\left[P_l - P_{l-1}\right]$ represents the expected difference in the payoff approximation on levels l and $l-1$. This is estimated using a set of Brownian paths, with the same Brownian paths being used on both levels. This is where the strong convergence properties are crucial. The small difference between the terminal values for the paths computed on levels l and $l-1$ gives a small value for the payoff difference. Consequently, the variance

$$V_l = V[P_l - P_{l-1}]$$

decreases rapidly with level l. In particular, for a European option with a Lipschitz payoff, the order with which the variance converges to zero is double the strong order of convergence.

Using N_l independent paths to estimate $E\left[P_l - P_{l-1}\right]$, if we define the level 0 variance to be $V_0 = V[P_0]$ then the variance of the combined multilevel estimator is $\sum_{l=0}^{L} N_l^{-1} V_l$. The computational cost is proportional to the total number of timesteps: $\sum_{l=0}^{L} N_l \Delta t_l^{-1}$. Varying N_l to minimise the variance for a given computational cost gives a constrained optimisation problem whose solution is $N_l = C\sqrt{V_l \Delta t_l}$. The value for the constant of proportionality, C, is chosen to make the overall variance less than the ϵ^2, so that the r.m.s. error is less than ϵ.

The analysis in [3] shows that in the case of an Euler discretisation with a Lipschitz payoff, the computational cost of the ML-MC algorithm is $O(\epsilon^2 \log)$, which is significantly better than the $O(\epsilon^3)$ cost of the standard Monte Carlo method. Furthermore, the analysis shows that first order strong convergence should lead to an $O(\epsilon^2)$ cost for Lipschitz payoffs; this will be demonstrated in the results to come.

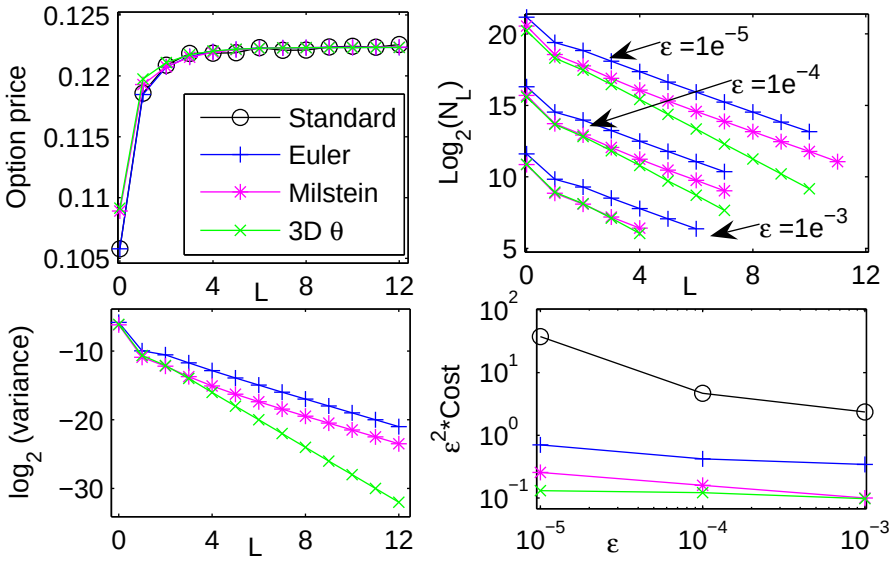

Fig. 4. European put option, Case 1. Top left: convergence in option value with grid level. Bottom left: convergence in the ML-MC variance with grid level. Top right: number of Monte Carlo paths N_l required on each level, depending on the desired accuracy. Bottom right: overall computational cost as a function of accuracy ϵ.

5.1 European option

The first set of numerical results are for a European put option with strike K and maturity T, for which the payoff is given by:

$$P = \max\left(K - x(T), 0\right)$$

Using the Case 1 volatility model (16), with the same set of parameters as before and strike $K = 1.1$, we obtain the ML-MC results in Figure 4. The top left plot shows the weak convergence in the estimated value of the payoff as the finest grid level L is increased. All of the methods tend asymptotically to the same value. The bottom left plot shows the convergence of the quantity $V_l = V[P_l - P_{l-1}]$. The 3D – θ scheme exhibits second order convergence due to the first order strong convergence. The Milstein approximation (3) with the Lévy areas set equal to zero, and the Euler discretisation both give first order convergence, which is consistent with their 0.5 order strong convergence properties.

The top right plot shows three sets of results for different values of the desired r.m.s. accuracy ϵ. The ML-MC algorithm [3] uses the correction obtained at each level of timestep refinement to estimate the remaining bias due to the discretisation, and therefore determine the number of levels of refinement required. The results illustrate this, with the smaller values for ϵ

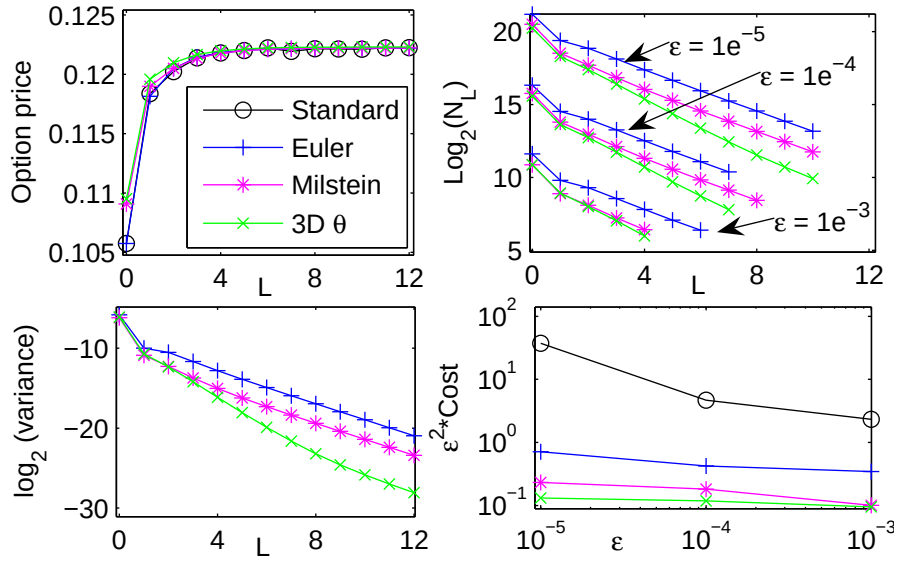

Fig. 5. European put option, Case 2. Top left: convergence in option value. Bottom left: convergence in ML-MC variance. Top right: number of Monte Carlo paths N_l required on each level. Bottom right: overall computational cost.

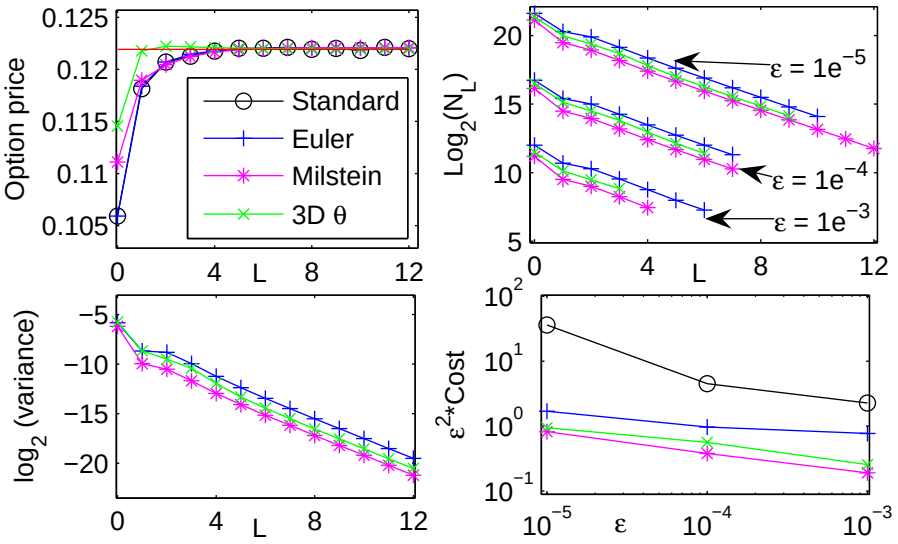

Fig. 6. European put option, Case 3. Top left: convergence in option value (red line is analytic value). Bottom left: convergence in ML-MC variance. Top right: number of Monte Carlo paths N_l required on each level. Bottom right: computational cost.

leading to more levels of refinement. To achieve the desired accuracy, it is also necessary to reduce the variance in the combined estimator to the required level, so any more paths (roughly proportional to ϵ^{-2}) are required for smaller values of ϵ. The final point to observe in this plot is how many fewer paths are required on the fine grid levels compared to the coarsest grid level for which there is just one timestep covering the entire time interval to maturity. This is a consequence of the variance convergence in the previous plot, together with the optimal choice for N_l described earlier.

The final bottom right plot shows the overall computational cost as a function of ϵ. The cost C_ϵ is defined as the total number of timesteps, summed over all paths and all grid levels. It is expected that C_ϵ will be $O(\epsilon^{-2})$ for the best ML-MC methods, and so the quantity which is plotted is $\epsilon^2 C_\epsilon$ versus ϵ. The results show that $\epsilon^2 C_\epsilon$ is almost perfectly independent of ϵ for the 3D – θ scheme, and varies only slightly with ϵ for the Milstein scheme. The Euler ML-MC scheme shows a bit more growth as $\epsilon \to 0$, which is consistent with the analysis in [3] which predicts that $C_\epsilon = O(\epsilon^{-2}(\log \epsilon)^{-2})$. The final comparison line is the standard Monte Carlo method using the Euler discretisation, for which $C_\epsilon = O(\epsilon^{-3})$.

The use of fewer Monte Carlo paths N_L is reflected directly in the computational cost of the process. For the most accurate case, $\epsilon = 0.00001$, the Euler, Milstein and 3D – θ version of the ML-MC scheme are roughly 50, 150 and 300 times more efficient that the standard Monte Carlo method using the Euler discretisation.

Figures 5 and 6 show the corresponding results for Cases 2 and 3, corresponding to the variance model (17) and the Heston model (18) respectively. For Case 2, the computational savings from using the ML-MC method are similar to Case 1, while for Case 3 the savings from the Euler, Milstein and 3D – θ versions of the ML-MC scheme are roughly 20, 40 and 40 in the most accurate case.

5.2 Digital option

The payoff for a digital option is given by:

$$P = H\left(x(T) - K\right)$$

where $H(x)$ is the Heaviside function ($H(x) = 1$ if $x > 0$, else $H(x) = 0$). Figure 7 shows the results using the Case 2 variance model (17) and strike $K = 1$. Because this payoff is not Lipschitz continuous, it shows the poorest benefits from the ML-MC approach. For the most accurate case, $\epsilon = 0.0001$, the Euler, Milstein and 3D – θ versions of the ML-MC scheme are roughly 3, 60 and 90 times more efficient than the standard method using the Euler scheme.

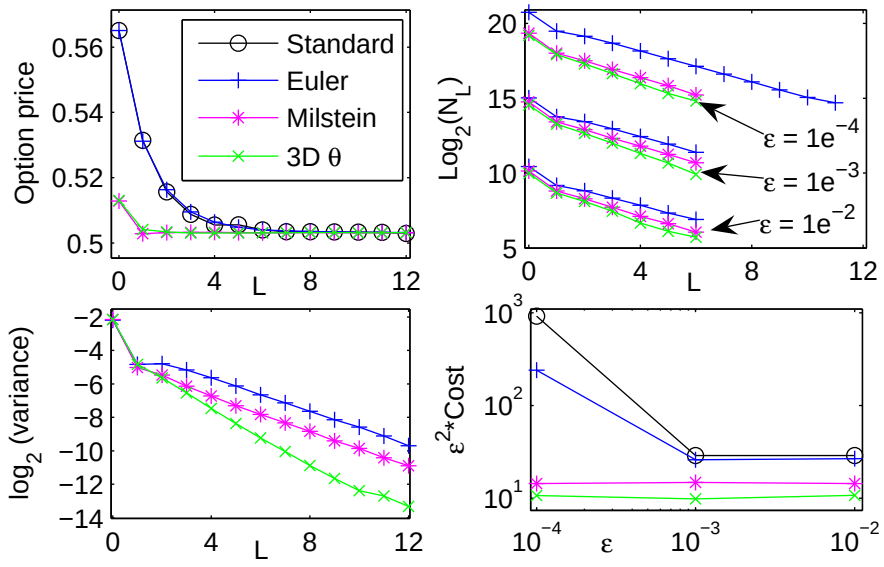

Fig. 7. Digital option, Case 2. Top left: convergence in option value. Bottom left: convergence in ML-MC variance. Top right: number of Monte Carlo paths N_l required on each level. Bottom right: overall computational cost.

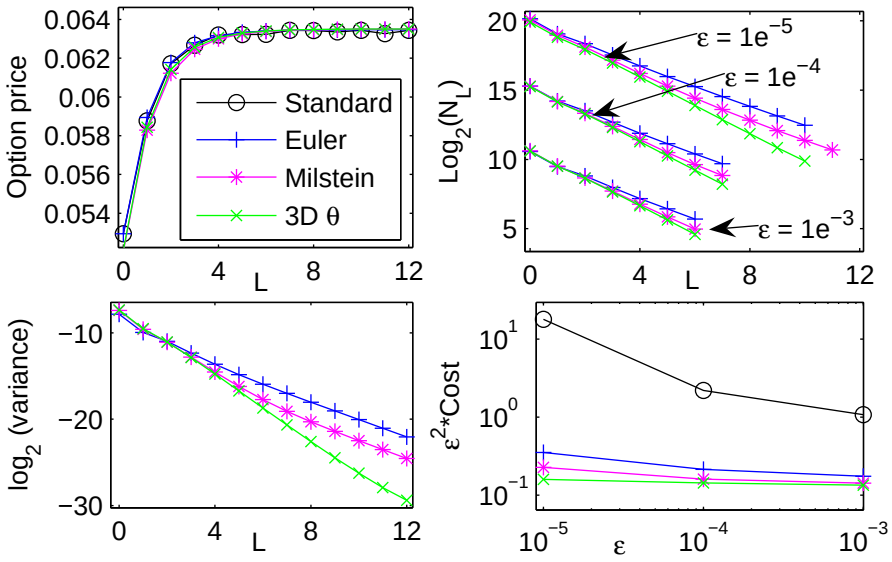

Fig. 8. Asian option, Case 2. Top left: convergence in option value. Bottom left: convergence in ML-MC variance. Top right: number of Monte Carlo paths N_l required on each level. Bottom right: overall computational cost.

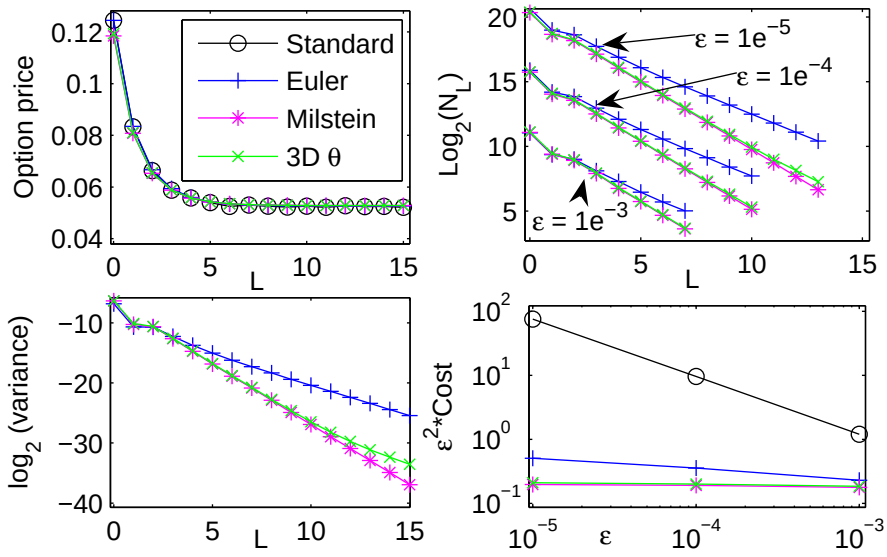

Fig. 9. Variance swap option, Case 2. Top left: convergence in option value. Bottom left: convergence in ML-MC variance. Top right: number of Monte Carlo paths N_l required on each level. Bottom right: overall computational cost.

5.3 Asian option

The payoff for an Asian call option is given by:

$$P = \max\left(\overline{x}(T) - K,\ 0\right)$$

where $\overline{x}$ is the arithmetic average which can be approximated numerically as:

$$\overline{x}(T) = \frac{1}{T}\int_0^T x(t)dt \approx \frac{\Delta t}{2\,T}\sum_{n=1}^{N_{\Delta t}}\left(\widehat{x}_n + \widehat{x}_{n-1}\right)$$

Using the Case 2 variance model (17), with strike $K = 1$, Figure 8 shows that for the most accurate case, $\epsilon = 0.00001$, the Euler, Milstein and 3D – θ versions of the ML-MC scheme are roughly 50, 80 and 110 times more efficient than the standard method using the Euler scheme.

5.4 Variance swap option

The payoff for a variance swap option is given by:

$$P = N\left(\overline{y}(T) - K_{var}\right)$$

where N is the nominal price and $\overline{y}(T)$ is the average of the variance in the time interval $[0, T]$ which can be approximated numerically in the same way as $\overline{y}(T)$ in the previous example.

Using the Case 2 variance model (17), $K_{var} = 0.26^2$ and $N = 10$, Figure 9 shows that for the most accurate case, $\epsilon = 0.00001$, the Euler, Milstein and 3D – θ versions of the ML-MC scheme are roughly 150, 380 and 360 times more efficient than the standard method using the Euler scheme. In this case, the Milstein method gives first order strong convergence for y, whereas the 3−Dimensional θ scheme gives similar accuracy initially but is tailing off towards order 0.5 strong convergence on the finest grids.

6 Conclusions

In finance, stochastic variance and volatility models are very important for the valuation of exotic options. We have shown that the use of the orthogonal θ scheme can achieve the first order strong convergence properties of the Milstein numerical discretisation without the expensive simulation of Lévy areas. In combination with the recently introduced Multilevel Monte Carlo method it can reduce substantially the computational cost in pricing exotic options, reducing the cost to achieve an r.m.s. error of size ϵ from $O(\epsilon^{-3})$ to $O(\epsilon^{-2})$).

The ML-MC works without any problems with all schemes and does not depend on the value of the parameters of the system. However, when a specific orthogonal transformation (θ scheme) is applied to a 2−Dimensional SDE it is only possible under certain conditions to avoid calculation of the Lévy area. The bias or error in the computation of the rotation angle θ that makes the Lie bracket equal to zero in the orthogonal scheme is crucial to obtain a better convergence order. When the conditions for integrability are satisfied, we can use the formula for θ to obtain the value of the rotation angle and obtain first order strong convergence. Otherwise, we have to use the 3−Dimensional transformation and check the magnitude of the Lie brackets to decide if it is likely to give computational savings in the solution of our system.

The numerical results demonstrate considerable computational savings when the orthogonal transformation is applied to either the quadratic volatility model (16) or the stochastic variance model (18). Unfortunately, similar savings are not achieved with the Heston model, and so the orthogonal transformation is not recommended in this case.

Acknowledgements

We are very grateful to Prof. T.J. Lyons for many very helpful discussions about this work. The first author would also like to thank CONACYT-MEXICO for their financial support.

References

1. A. B. Cruzeiro, P. Malliavin and A. Thalmaier: Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation. C. R. Acad. Sci. Paris, Ser. I, **338**, 481 – 486, (2004).
2. P. Glasserman: Monte Carlo Methods in Financial Engineering. Springer, (2004).
3. M. Giles: Multi-level Monte Carlo path simulation. Technical Report No. **NA06/03**, Oxford University Computing Laboratory, Parks Road, Oxford, U.K, (2006).
4. S. Heston: A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. The Review of Financial Studies, Vol. **6**, Issue 2, 327 – 343, (1993).
5. P. E. Kloeden and E. Platen: Numerical Solution of Stochastic Differential Equations. Springer, (1999).
6. J.G. Gaines and T.J. Lyons: Random Generation of Stochastic Area Integrals. SIAM Journal on Applied Mathematics, Vol. **54**, No. 4, 1132 – 1146, (1994).
7. P. Malliavin and A. Thalmaier: Stochastic Calculus of Variations in Mathematical Finance. Springer, (2005).
8. K. Schmitz-Abe and W. T. Shaw: Measure Order of Convergence without an Exact Solution, Euler vs Milstein Scheme. International Journal of Pure and Applied Mathematics, Vol. **24**, No 3, 365 – 381, (2005).
9. M. Wiktorsson: Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions. The Annals of Applied Probability, Vol. **11**, No 2, 470 – 487, (2001).