# Numerical analysis of multilevel Milstein scheme without Lévy areas

Mike Giles

Lukas Szpruch

mike.giles,lukas.szpruch@maths.ox.ac.uk

Oxford University Mathematical Institute

Oxford-Man Institute of Quantitative Finance

Eighth IMACS Seminar on Monte Carlo Methods

Borovets, August 29 – September 2, 2011

# Outline

- Milstein discretisation and multilevel method
- Clark & Cameron model problem
- antithetic treatment and analysis
- generalisation

# Milstein discretisation

The Milstein discretisation of the SDE

$$\mathrm{d}S_i(t) = a_i(S)\,\mathrm{d}t + \sum_j b_{ij}(S)\,\mathrm{d}W_j(t), \quad 0<t<T$$

is

$$\begin{aligned}\widehat{S}_{i,n+1} &= \widehat{S}_{i,n} + a_i(\widehat{S}_n)\,\Delta t + \sum_j b_{ij}(\widehat{S}_n)\,\Delta W_{j,n} \\ &+ \sum_{j,k} c_{ijk}(\widehat{S}_n)\left(\Delta W_{j,n}\,\Delta W_{k,n} - \Omega_{jk}\,\Delta t - A_{jk,n}\right)\end{aligned}$$

where $\Omega_{jk}$ is the correlation, $c_{ijk} \equiv \frac{1}{2}\sum_l \dfrac{\partial b_{ij}}{\partial S_l}\, b_{lk}$, and

$$A_{jk,n} = \int_{t_n}^{t_{n+1}} (W_j(t) - W_j(t_n))\;\mathrm{d}W_k - (W_k(t) - W_k(t_n))\;\mathrm{d}W_j$$

# Standard Multilevel approach

To estimate $\mathbb{E}[P]$, where the payoff $P = f(S_T)$ can be approximated by $\widehat{P}_\ell$ using $2^\ell$ uniform timesteps, we use

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

$\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ is estimated using $N_\ell$ simulations with same $W(t)$ for both $\widehat{P}_\ell$ and $\widehat{P}_{\ell-1}$,

$$\widehat{Y}_\ell = N_\ell^{-1} \sum_{i=1}^{N_\ell} \left( \widehat{P}_\ell^{(i)} - \widehat{P}_{\ell-1}^{(i)} \right)$$

Because of strong convergence, on finer levels $\mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ is small and so few paths are required.

# Modified Multilevel approach

Sometimes better to use a different approximation for $\widehat{P}_\ell$ in $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ and $\mathbb{E}[\widehat{P}_{\ell+1} - \widehat{P}_\ell]$. The decomposition

$$\mathbb{E}[\widehat{P}_L^f] = \mathbb{E}[\widehat{P}_0^f] + \sum_{\ell=1}^{L} \mathbb{E}[\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c]$$

is still a valid telescoping sum provided $\mathbb{E}[\widehat{P}_\ell^f] = \mathbb{E}[\widehat{P}_\ell^c]$.

In this work, we use $\widehat{P}_\ell^c = f(\widehat{S}_\ell^c)$ and

$$\widehat{P}_\ell^f = \tfrac{1}{2}\left( f(\widehat{S}_\ell^{f1}) + f(\widehat{S}_\ell^{f2}) \right)$$

where $f1$ is the fine path, and $f2$ is an “antithetic twin”.

# Antithetic Multilevel estimator

**Lemma 0.1** *If* $f \in C^2(\mathbb{R}^d, \mathbb{R})$ *and there exist constants* $L_1, L_2$ *such that for all* $S \in \mathbb{R}^d$

$$\left\| \frac{\partial f}{\partial S} \right\| \leq L_1, \quad \left\| \frac{\partial^2 f}{\partial S^2} \right\| \leq L_2.$$

*then*

$$\mathbb{E}\left[ \left( \tfrac{1}{2}(f(\widehat{S}^{f1}) + f(\widehat{S}^{f2})) - f(\widehat{S}^{c}) \right)^2 \right]$$

$$\leq \; 2\, L_1^2 \; \mathbb{E}\left[ \left\| \tfrac{1}{2}(\widehat{S}^{f1} + \widehat{S}^{f2}) - \widehat{S}^{c} \right\|^2 \right] \; + \; \tfrac{1}{32}\, L_2^2 \; \mathbb{E}\left[ \left\| \widehat{S}^{f1} - \widehat{S}^{f2}) \right\|^4 \right].$$

# Antithetic Multilevel estimator

**Proof** Defining $\overline{S}^f \equiv \frac{1}{2}(\widehat{S}^{f1}+\widehat{S}^{f2})$, Taylor expansion gives

$$\tfrac{1}{2}(f(\widehat{S}^{f1}) + f(\widehat{S}^{f2})) = f(\overline{S}^f) + \tfrac{1}{8}(\widehat{S}^{f1} - \widehat{S}^{f2})^T \frac{\partial^2 f}{\partial S^2}(\xi_1)\ (\widehat{S}^{f1} - \widehat{S}^{f2})$$

$$\begin{aligned}
\Longrightarrow \quad & \tfrac{1}{2}(f(\widehat{S}^{f1}) + f(\widehat{S}^{f2})) - f(\widehat{S}^c) \\
= \quad & \frac{\partial f}{\partial S}^T(\xi_2)\ (\overline{S}^f - \widehat{S}^c) + \tfrac{1}{8}(\widehat{S}^{f1} - \widehat{S}^{f2})^T \frac{\partial^2 f}{\partial S^2}(\xi_1)\ (\widehat{S}^{f1} - \widehat{S}^{f2}).
\end{aligned}$$

It follows that

$$\left| \tfrac{1}{2}(f(\widehat{S}^{f1}) + f(\widehat{S}^{f2})) - f(\widehat{S}^c) \right| \leq L_1 \left\| \overline{S}^f - \widehat{S}^c \right\| + \tfrac{1}{8} L_2 \left\| \widehat{S}^{f1} - \widehat{S}^{f2} \right\|^2$$

and squaring and taking the expectation gives the result. □

# Clark & Cameron problem

In their 1980 paper, Clark & Cameron considered the model problem:

$$\begin{aligned} \mathrm{d}X &= \mathrm{d}W_1 \\ \mathrm{d}Y &= X\,\mathrm{d}W_2 \end{aligned}$$

for independent Brownian paths $W_1, W_2$ and $X(0)\!=\!Y(0)\!=\!0$.

This can be integrated to give $X(t) = W_1(t)$ and

$$\begin{aligned} Y(t) &= \int_0^t W_1(s)\,\mathrm{d}W_2(s) \\ &= \tfrac{1}{2}\,W_1(t)\,W_2(t) + \tfrac{1}{2}\int_0^t W_1(s)\,\mathrm{d}W_2(s) - W_2(s)\,\mathrm{d}W_1(s) \end{aligned}$$

# Clark & Cameron problem

If we consider a set of times $t_n = n\,h$, then we get

$$Y(t_{n+1}) = Y(t_n) + X(t_n)\,\Delta W_{2,n} + \tfrac{1}{2}\Delta W_{1,n}\,\Delta W_{2,n} + \tfrac{1}{2}A_n,$$

where $\Delta W_{j,n} \equiv W_j(t_{n+1}) - W_j(t_n)$ and

$$A_n = \int_{t_n}^{t_{n+1}} W_1(s)\,\mathrm{d}W_2(s) - W_2(s)\,\mathrm{d}W_1(s).$$

This matches exactly the Milstein discretisation – i.e. the Milstein discretisation is exact for this problem

# Clark & Cameron problem

Summing over $n$ gives

$$Y(T) = \sum_n \left( X(t_n)\, \Delta W_{2,n} + \tfrac{1}{2} \Delta W_{1,n}\, \Delta W_{2,n} + \tfrac{1}{2} A_n \right)$$

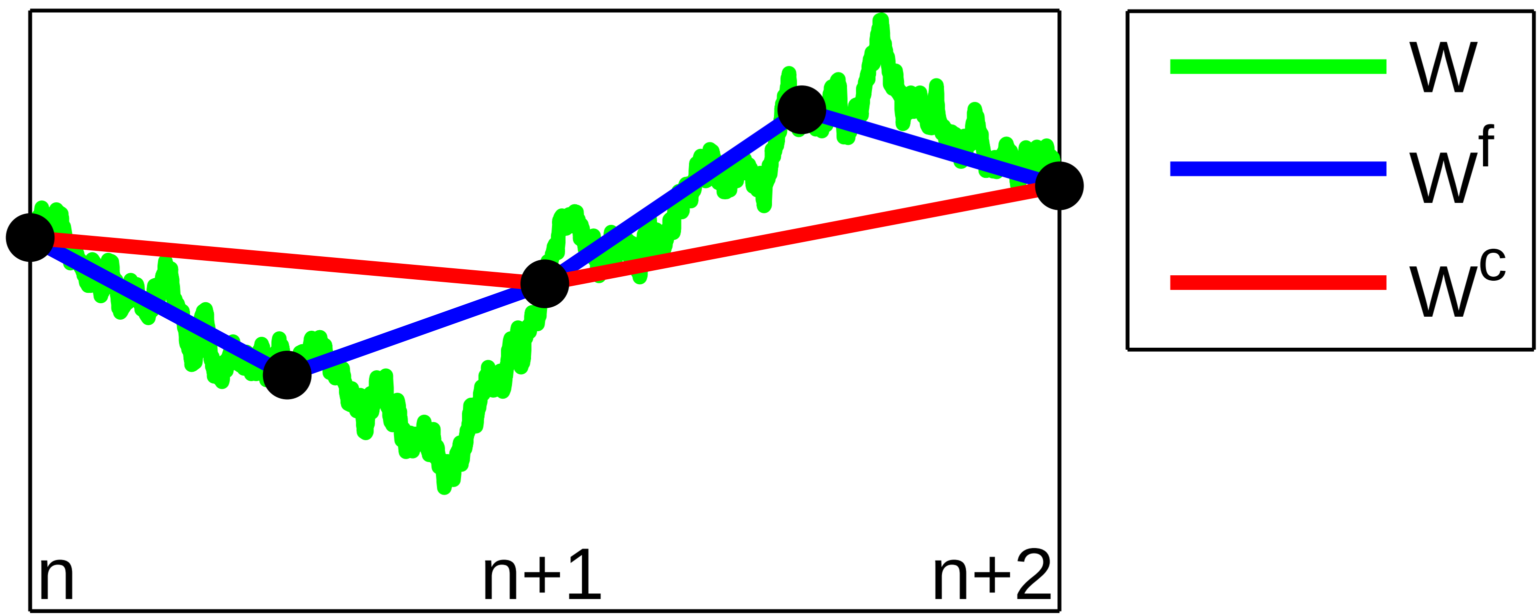
Key point of their paper: conditional on $\Delta W$ increments,

- $\mathbb{E}\left[ Y(T) \,|\, \Delta W \right] = \sum_n \left( X(t_n)\, \Delta W_{2,n} + \tfrac{1}{2} \Delta W_{1,n}\, \Delta W_{2,n} \right)$
- $\mathbb{V}\left[ Y(T) \,|\, \Delta W \right] = \tfrac{1}{4} \sum_n \mathbb{V}[A_n] = O(\Delta t)$

Hence, any numerical discretisation which uses only Brownian increments <u>cannot</u> in general achieve better than $O(\sqrt{\Delta t})$ strong convergence.

# Clark & Cameron problem

If $A_n$ is not known, best approximation sets it to zero,
– equivalent to a piecewise linear interpolation of the driving Brownian path.

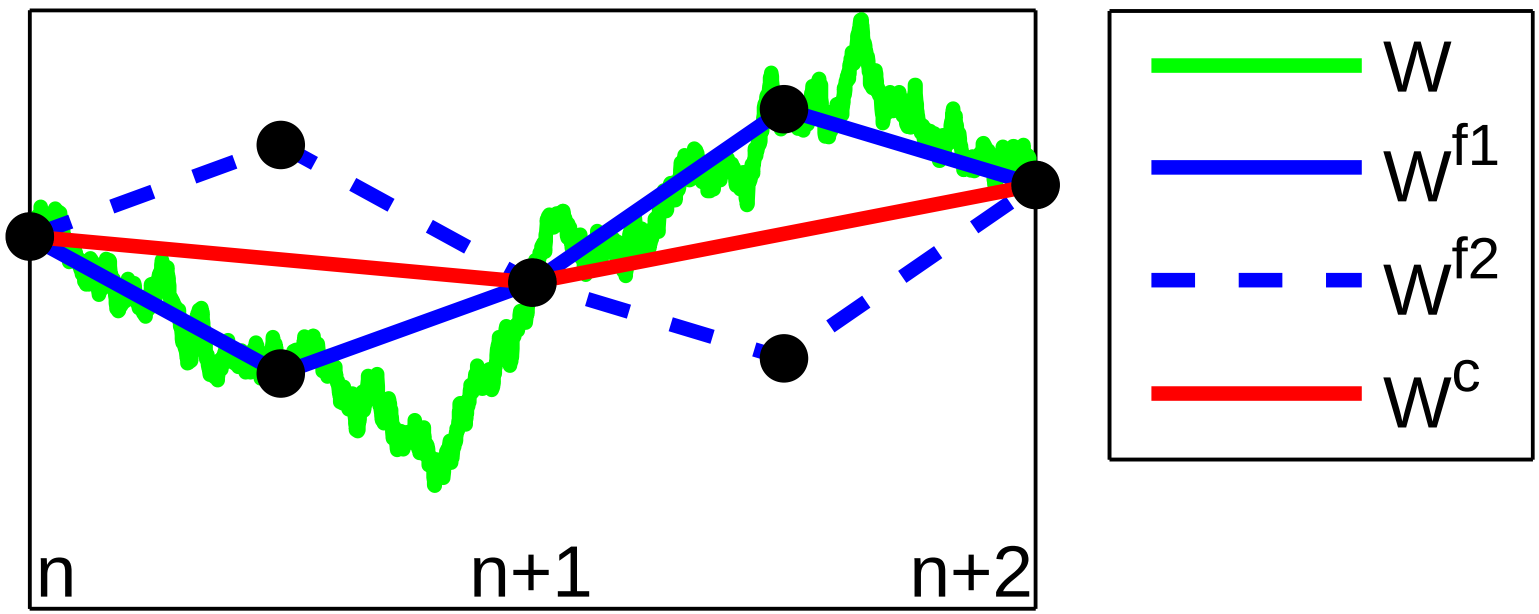



Coarse and fine paths use different interpolations

$$Y^f - Y^c = \sum_n A_n \Longrightarrow \mathbb{V}[Y^f - Y^c] = O(\Delta t)$$

# Clark & Cameron problem

Fine path "antithetic twin" swaps Brownian increments for odd and even timesteps – average of two piecewise linear Brownian paths matches coarse one




$$A_n^{f2} = -A_n^{f1} \quad \Longrightarrow \quad (Y^{f2} - Y^c) = -(Y^{f1} - Y^c)$$

Hence $\frac{1}{2}(Y^{f1} + Y^{f2}) = Y^c$

# Clark & Cameron problem

If the payoff function $f(X,Y)$ is twice-differentiable,

$$\begin{aligned} \tfrac{1}{2}\left( f(X,Y^{f1}) + f(X,Y^{f2}) \right) - f(X,Y^c) &= \tfrac{1}{2}\,\frac{\partial^2 f}{\partial Y^2}\,(Y^{f1}\!-\!Y^c)^2 \\ &= O(\Delta t) \end{aligned}$$

Hence, $\mathbb{V}[\widehat{P}_\ell \!-\! \widehat{P}_{\ell-1}] = O(\Delta t^2)$ – much better than before.

If $f(X,Y)$ is Lipschitz and twice-differentiable except on $K$, and $(X,Y^c)$ is within $O(\sqrt{\Delta t})$ of $K$ with probability $O(\sqrt{\Delta t})$, then a local analysis gives $\mathbb{V}[\widehat{P}_\ell \!-\! \widehat{P}_{\ell-1}] = O(\Delta t^{3/2})$

# Generalisation

For the general SDE

$$\mathrm{d}S_i(t) = a_i(S)\,\mathrm{d}t + \sum_j b_{ij}(S)\,\mathrm{d}W_j(t), \quad 0<t<T$$

we define the driving Brownian paths in the same way:

- fine path $W^{f1}(t)$ is piecewise linear interpolation with interval $\Delta t/2$
- fine path $W^{f2}(t)$ is "antithetic twin", swapping odd and even increments
- coarse path $W^{c}(t)$ is piecewise linear interpolation with interval $\Delta t$, and also average of the two fine paths

# Generalisation

**Assumptions:** $a(S)$ and $b(S)$ both twice differentiable with usual uniform Lipschitz bounds, and also uniformly bounded second derivatives.

**Lemma 0.2** *For all* $p \geq 1$*, there exists* $K_p$ *such that*

$$
\begin{aligned}
\mathbb{E}\left[\max_{0\leq n\leq N} \|\widehat{S}_n^c\|^p\right] &\leq K_p, \\
\mathbb{E}\left[\max_{0\leq n\leq N} \|\widehat{S}_n^{f1}\|^p\right] &\leq K_p, \\
\mathbb{E}\left[\max_{0\leq n\leq N} \|\widehat{S}_n^{f2}\|^p\right] &\leq K_p.
\end{aligned}
$$

Similar bounds hold for $a(S)$ and $b(S)$.

# Generalisation

**Lemma 0.3** *For all* $p \geq 1$*, there exists* $K_p$ *such that*

$$\mathbb{E}\left[\max_{0\leq n\leq N} \|\widehat{S}^c_n - S(t_n)\|^p\right] \leq K_p\, \Delta t^{p/2}$$

**Corollary 0.4** *For all* $p \geq 1$*, there exists* $K_p$ *such that*

$$\mathbb{E}\left[\max_{0\leq n\leq N} \|\widehat{S}^{f1}_n - \widehat{S}^c_n\|^p\right] \leq K_p\, \Delta t^{p/2}$$

$$\mathbb{E}\left[\max_{0\leq n\leq N} \|\widehat{S}^{f1}_n - \widehat{S}^{f2}_n\|^p\right] \leq K_p\, \Delta t^{p/2}$$

# Generalisation

**Lemma 0.5** *The equn for $\widehat{S}_n^{f1}$ over one coarse timestep is*

$$\begin{aligned}
\widehat{S}_{i,n+1}^{f1} &= \widehat{S}_{i,n}^{f1} + a_i(\widehat{S}_n^{f1})\,\Delta t + \sum_j b_{ij}(\widehat{S}_n^{f1})\,\Delta W_{j,n} \\
&+ \sum_{j,k} c_{ijk}(\widehat{S}_n^{f1})\left(\Delta W_{j,n}\Delta W_{k,n} - \Omega_{jk}\Delta t\right) \\
&- \sum_{j,k} c_{ijk}(\widehat{S}_n^{f1})\left(\delta W_{j,n}\delta W_{k,n+\frac{1}{2}} - \delta W_{k,n}\delta W_{j,n+\frac{1}{2}}\right) \\
&+ M_{i,n} \;+\; N_{i,n},
\end{aligned}$$

*where* $\mathbb{E}[M_n \,|\, \mathcal{F}_n] = 0$*, and for* $p \geq 1$ *there exists* $K_p$ *such that*

$$\mathbb{E}\left[\|M_n\|^p\right] \leq K_p\,\Delta t^{3p/2}, \quad \mathbb{E}\left[\|N_n\|^p\right] \leq K_p\,\Delta t^{2p}.$$

# Generalisation

**Lemma 0.6** *The equn for* $\widehat{S}^{f2}_n$ *over one coarse timestep is*

$$
\begin{aligned}
\widehat{S}^{f2}_{i,n+1} &= \widehat{S}^{f2}_{i,n} + a_i(\widehat{S}^{f2}_n)\,\Delta t + \sum_j b_{ij}(\widehat{S}^{f2}_n)\,\Delta W_{j,n} \\
&+ \sum_{j,k} c_{ijk}(\widehat{S}^{f2}_n)\left(\Delta W_{j,n}\Delta W_{k,n} - \Omega_{jk}\Delta t\right) \\
&+ \sum_{j,k} c_{ijk}(\widehat{S}^{f2}_n)\left(\delta W_{j,n}\delta W_{k,n+\frac{1}{2}} - \delta W_{k,n}\delta W_{j,n+\frac{1}{2}}\right) \\
&+ M_{i,n} + N_{i,n},
\end{aligned}
$$

*where* $\mathbb{E}[M_n \,|\, \mathcal{F}_n] = 0$*, and for* $p \geq 1$ *there exists* $K_p$ *such that*

$$
\mathbb{E}\left[\|M_n\|^p\right] \leq K_p\,\Delta t^{3p/2}, \quad \mathbb{E}\left[\|N_n\|^p\right] \leq K_p\,\Delta t^{2p}.
$$

# Generalisation

**Lemma 0.7** *The equn for* $\overline{S}^f_n \equiv \frac{1}{2}(\widehat{S}^{f1}_n + \widehat{S}^{f2}_n)$ *is*

$$
\begin{aligned}
\overline{S}^f_{i,n+1} &= \overline{S}^f_{i,n} + a_i(\overline{S}^f_n)\,\Delta t + \sum_j b_{ij}(\overline{S}^f_n)\,\Delta W_{j,n} \\
&+ \sum_{j,k} c_{ijk}(\overline{S}^f_n)\left(\Delta W_{j,n}\Delta W_{k,n} - \Omega_{jk}\Delta t\right) \\
&+ M_{i,n} + N_{i,n},
\end{aligned}
$$

*where* $\mathbb{E}[M_n \,|\, \mathcal{F}_n] = 0$*, and for* $p \geq 1$ *there exists* $K_p$ *such that*

$$
\mathbb{E}\left[\|M_n\|^p\right] \leq K_p\,\Delta t^{3p/2}, \quad \mathbb{E}\left[\|N_n\|^p\right] \leq K_p\,\Delta t^{2p}.
$$

# Generalisation

**Theorem 0.8** *For all* $p \geq 1$*, there exists* $K_p$ *such that*

$$\mathbb{E}\left[\max_{0\leq n\leq N} \|\overline{S}^f_n - \widehat{S}^c_n\|^p\right] \leq K_p\, \Delta t^p.$$

**Proof**

$$\begin{aligned}
\overline{S}^f_{i,n} - \widehat{S}^c_{i,n} \;&=\; \sum_{m<n} \Big(a_i(\overline{S}^f_{i,m}) - a_i(\widehat{S}^c_{i,m})\Big)\,\Delta t \\
&+\; \sum_{m<n}\sum_{j} \Big(b_{ij}(\overline{S}^f_{i,m}) - b_{ij}(\widehat{S}^c_{i,m})\Big)\,\Delta W_{j,m} \\
&+\; \sum_{m<n}\sum_{j,k} \Big(c_{ijk}(\overline{S}^f_{i,m}) - c_{ijk}(\widehat{S}^c_{i,m})\Big)\,(\Delta W_{j,n}\Delta W_{k,n} - \Omega_{jk}\Delta t) \\
&+\; \sum_{m<n} M_{i,m} + \sum_{m<n} N_{i,m}
\end{aligned}$$

# Generalisation

Using Burkholder-Davis-Gundy inequality, can prove that

$$Z_n \;\equiv\; \mathbb{E}\left[\max_{m<n} \| \overline{S}^f_m - \widehat{S}^c_m \|^p\right]$$

satisfies an inequality

$$Z_n \leq C_p \left( \Delta t^p + \sum_{m<n} Z_m \, \Delta t \right)$$

and desired result then comes from discrete Grönwall inequality.

# Conclusions

- MCQMC10 presentation gave numerical results showing effectiveness for Heston stochastic volatility model
- also gave an asymptotic analysis explanation
- new numerical analysis supports the observations and previous explanation
- further analysis treats case in which we approximate the Lévy areas by sub-sampling the Brownian path within each timestep – needed for discontinuous payoffs