Multilevel Monte Carlo for multi-dimensional SDEs

Mike Giles

mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute

Oxford-Man Institute of Quantitative Finance

SIAM Conference on Financial Mathematics and Engineering

November 19-20, 2010

Multilevel approach

Given a scalar SDE driven by a Brownian diffusion

$$\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W(t),$$

to estimate $\mathbb{E}[P]$ where the path-dependent payoff P can be approximated by $\widehat{P}_l$ using 2^l uniform timesteps, we use

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

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

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

Multilevel approach

Using independent samples for each level, the variance of the combined estimator is

$$\mathbb{V}\left[\sum_{l=0}^{L} \widehat{Y}_l\right] = \sum_{l=0}^{L} N_l^{-1} V_l, \qquad V_l \equiv \begin{cases} \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}], & l > 0 \\ \mathbb{V}[\widehat{P}_0], & l = 0 \end{cases}$$

and the computational cost is proportional to $\displaystyle\sum_{l=0}^{L} N_l\, h_l^{-1}$

Hence, the variance is minimised for a fixed computational cost by choosing N_l to be proportional to $\sqrt{V_l\, h_l}$.

MLMC Theorem

Theorem: *Let P be a functional of the solution of a stochastic o.d.e., and $\widehat{P}_l$ the discrete approximation using a timestep $h_l = 2^{-l}\,T$.*

If *there exist independent estimators $\widehat{Y}_l$ based on N_l Monte Carlo samples, with computational complexity (cost) C_l, and positive constants $\alpha \geq \frac{1}{2}, \beta, c_1, c_2, c_3$ such that*

i) $\left| \mathbb{E}[\widehat{P}_l - P] \right| \leq c_1\, h_l^{\alpha}$

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) $C_l \leq c_3\, N_l\, h_l^{-1}$

MLMC Theorem

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*

$$\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l,$$

has Mean Square Error $MSE \equiv \mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[P] \right)^2 \right] < \varepsilon^2$

with a computational complexity C *with bound*

$$C \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > 1, \\ c_4\, \varepsilon^{-2} (\log \varepsilon)^2, & \beta = 1, \\ c_4\, \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

Previous Work

- First paper (*Operations Research, 2006 – 2008*) applied idea to SDE path simulation using Euler-Maruyama discretisation
- Second paper (*MCQMC 2006 – 2007*) used Milstein discretisation for scalar SDEs – improved strong convergence gives improved multilevel variance convergence

- Multilevel method is a generalisation of two-level control variate method of Kebaier (2005), and similar to ideas of Speight (2009) and also related to multilevel parametric integration by Heinrich (2001)

Numerical Analysis

If P is a Lipschitz function of $S(T)$, the value of the underlying at maturity, the strong convergence property

$$\left(\mathbb{E}\left[(\widehat{S}_N - S(T))^2\right]\right)^{1/2} = O(h^\gamma)$$

implies that $\mathbb{V}[\widehat{P}_l - P] = O(h_l^{2\gamma})$ and hence

$$V_l \;\equiv\; \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] = O(h_l^{2\gamma}).$$

Therefore $\beta = 1$ for Euler-Maruyama discretisation, and $\beta = 2$ for the Milstein discretisation.

However, in general, good strong convergence is neither necessary nor sufficient for good convergence for V_l.

Numerics and Analysis

	Euler		Milstein	
option	numerics	analysis	numerics	analysis
Lipschitz	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
Asian	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
lookback	$O(h)$	$O(h)$	$O(h^2)$	$o(h^{2-\delta})$
barrier	$O(h^{1/2})$	$o(h^{1/2-\delta})$	$O(h^{3/2})$	$o(h^{3/2-\delta})$
digital	$O(h^{1/2})$	$O(h^{1/2}\log h)$	$O(h^{3/2})$	$o(h^{3/2-\delta})$

Table: V_l convergence observed numerically (for GBM) and proved analytically (for more general SDEs)

Euler analysis due to G, Higham & Mao (*Finance & Stochastics, 2009*) and Avikainen (*Finance & Stochastics, 2009*). Milstein analysis due to G, Debrabant & Rößler

Other work

- Yuan Xia, G – jump-diffusion models
- Sylvestre Burgos, G – Greeks
- Hoel, von Schwerin, Szepessy, Tempone – adaptive discretisations
- Dereich, Heidenreich – Lévy processes
- Hickernell, Müller-Gronbach, Niu, Ritter – complexity analysis
- Müller-Gronbach, Ritter – parabolic SPDEs
- G, Reisinger – parabolic SPDEs
- Teckentrup, Scheichl, Cliffe, G – elliptic SPDEs
- Barth, Schwab, Zollinger – elliptic SPDEs

Multi-dimensional SDEs

The Milstein scheme for multi-dimensional SDEs is

$$\begin{aligned}\widehat{S}_{i,n+1} &= \widehat{S}_{i,n} + a_i\, h + \sum_j b_{ij}\, \Delta W_{j,n} \\ &\quad + \tfrac{1}{2} \sum_{j,k,l} \frac{\partial b_{ij}}{\partial S_l}\, b_{lk} \left(\Delta W_{j,n}\, \Delta W_{k,n} - \Omega_{jk}\, h - A_{jk,n} \right)\end{aligned}$$

where Lévy areas are defined as

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

- $O(h)$ strong convergence, but hard to simulate A_{jk}
- $O(h^{1/2})$ strong convergence in general if A_{jk} omitted

Discretisation error analysis

Suppose we ignore the Lévy area terms – what is the resulting difference between coarse and fine path approximations?

Let the coarse path approximation be

$$\widehat{S}^c_{n+1} = R(\widehat{S}^c_n)$$

and the fine path approximation be

$$\widehat{S}^f_{n+1} = R(\widehat{S}^f_n) + g_n$$

so to leading order the difference $\widehat{D}_n \equiv \widehat{S}^f_n - \widehat{S}^c_n$ satisfies

$$\widehat{D}_{n+1} = \frac{\partial R}{\partial S}\,\widehat{D}_n + g_n$$

Discretisation error analysis

Using a Brownian Bridge construction in which

$$W_{n+1/2} = \tfrac{1}{2}\left(W_n + W_{n+1} + Z\right)$$

where $Z \sim N(0, h_c)$, find that, to leading order,

$$g_{i,n} = \frac{1}{2} \sum_{j,k,l} \frac{\partial b_{ij}}{\partial S_l}\, b_{lk} \left(\Delta W_{j,n} Z_{k,n} - \Delta W_{k,n} Z_{j,n} \right)$$

Note: $g \equiv 0$ for scalar applications, and for vector applications satisfying the commutativity conditions

$$\sum_l \frac{b_{ij}}{\partial S_l}\, b_{lk} = \sum_l \frac{b_{ik}}{\partial S_l}\, b_{lj}, \quad \forall i, j, k$$

Discretisation error analysis

ΔW and Z are $O(\sqrt{h})$ and independent

$\Longrightarrow \; g_n = O(h)$ but $\mathbb{E}[g_n] = 0$ (to leading order)

$\Longrightarrow \; \widehat{D}_n = O(\sqrt{h})$ but $\mathbb{E}[\widehat{D}_n] = 0$ (to leading order)

Haven't achieved anything yet – really just shown $O(\sqrt{h})$ strong convergence when Lévy area is neglected.

(Best that can be achieved knowing just the discrete ΔW – Clark & Cameron, 1980)

Now comes the new idea – use antithetic variates in Brownian Bridge construction.

i.e. construct a second fine path using $-Z_n$ instead of Z_n

Antithetic treatment

Since g_n is linear in Z_n, this implies that, to leading order,

$$\widehat{D}_n^{(2)} = -\widehat{D}_n^{(1)}$$

Higher order terms in asymptotic error analysis give

$$\widehat{D}_n^{(1)} + \widehat{D}_n^{(2)} = O(h)$$

If the payoff function $f(S_T)$ is twice differentiable then

$$\begin{aligned}\tfrac{1}{2}\left(f(\widehat{S}^{f(1)})+f(\widehat{S}^{f(2)})\right) - f(\widehat{S}^c) \;&\approx\; \tfrac{1}{2}\left(\widehat{D}_n^{(1)} + \widehat{D}_n^{(2)}\right) f'(\widehat{S}^c) \\ &\quad + \tfrac{1}{4}\left((\widehat{D}_n^{(1)})^2 + (\widehat{D}_n^{(2)})^2\right) f''(\widehat{S}^c) \\ &= \; O(h)\end{aligned}$$

Antithetic treatment

Hence, for the multilevel estimator on level l we use

$$\widehat{Y}_l = N_l^{-1} \sum_{n=1}^{N_l} \tfrac{1}{2} \left(\widehat{P}_l^{(n1)} + \widehat{P}_l^{(n2)} \right) - \widehat{P}_{l-1}^{(n)}$$

and

$$\mathbb{V}[\widehat{Y}_l] = N_l^{-1} V_l$$

with

$$V_l = O(h^2).$$

This assumed the payoff function was twice differentiable. For a put or call option, more careful analysis near the strike gives $V_l = O(h^{3/2})$ – still enough to ensure the overall cost is $O(\varepsilon^{-2})$.

Numerical test

Heston stochastic volatility model:

$$\mathrm{d}S = r\, S\, \mathrm{d}t + \sqrt{v}\, S\, \mathrm{d}W_1, \qquad 0 < t < T,$$

$$\mathrm{d}v = \kappa(\theta - v) + \xi\, \sqrt{v}\, \mathrm{d}W_2, \qquad 0 < t < T,$$

with $T = 1$, $S(0) = 100$, $r = 0.05$, $\kappa = 1$, $\theta = 0.04$, $\xi = 0.25$ and correlation $\rho = -0.5$.

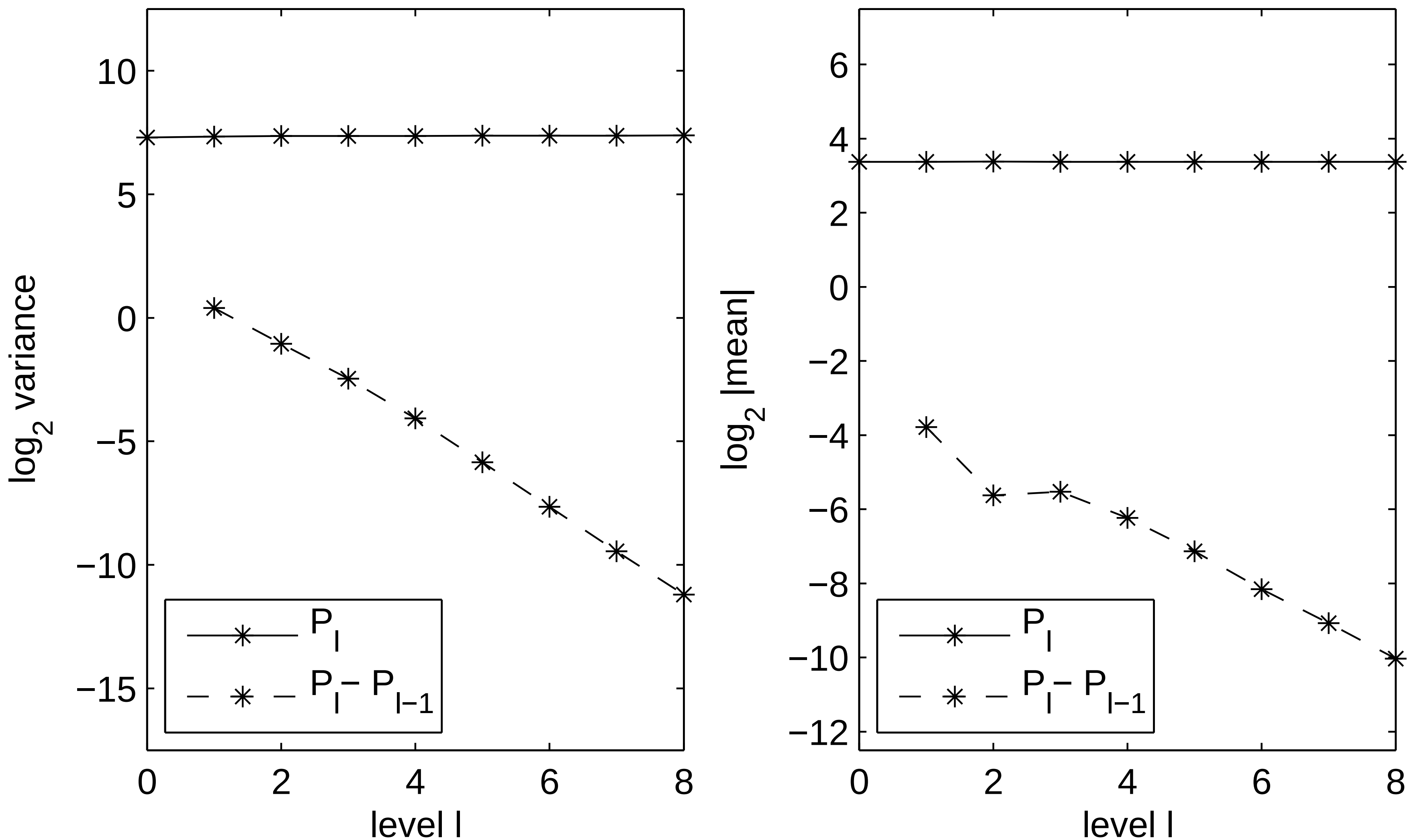
"Integrating factor" used for volatility discretisation to improve accuracy with large timesteps — Mark Broadie

European call option with discounted payoff

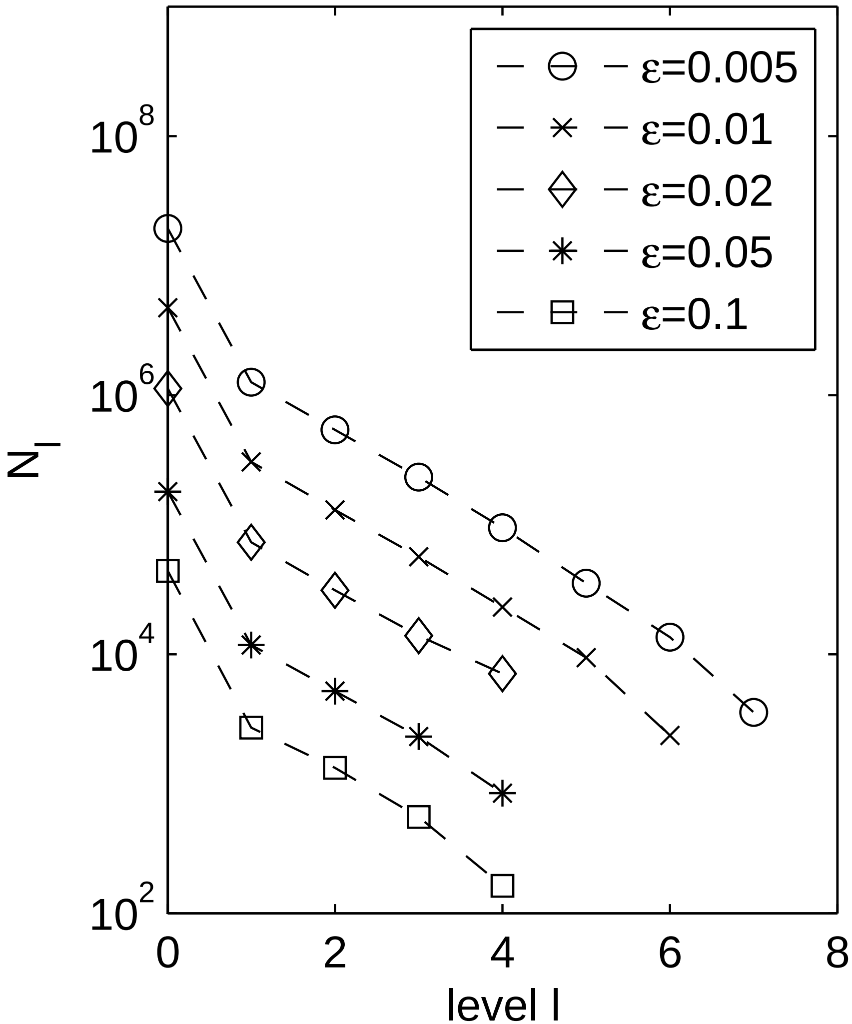
$$\exp(-rT)\ \max(S(T) - K, 0)$$

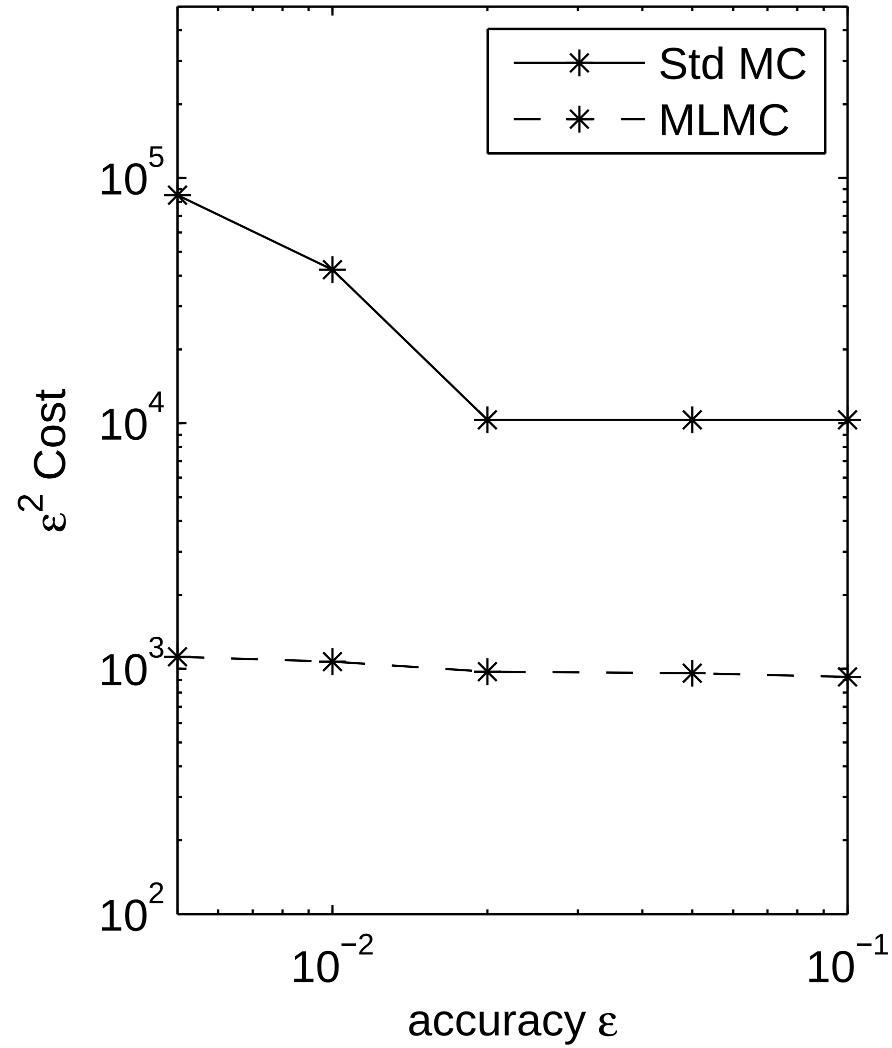
with strike $K = 100$.

Numerical test



Numerical test





Discontinuous payoffs

Antithetic treatment doesn't help with discontinuous payoffs:

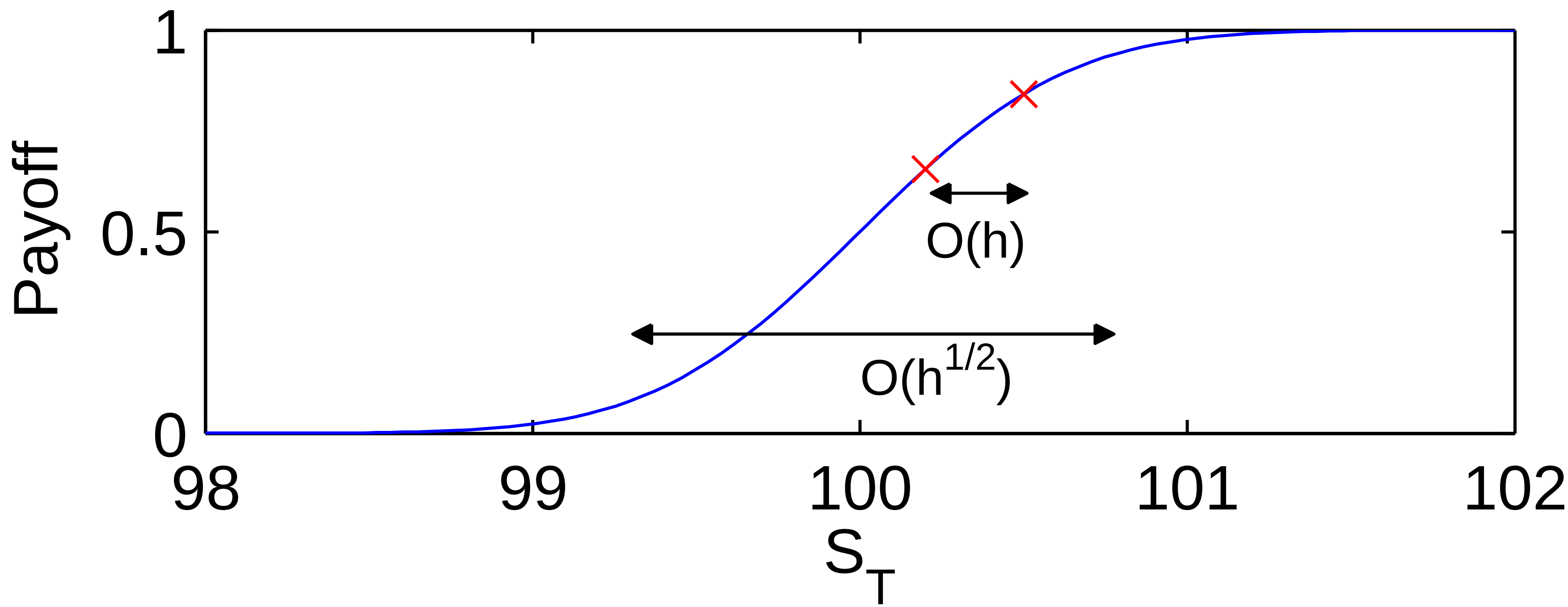
- $O(\sqrt{h})$ paths near enough to strike for fine and coarse paths to be on opposite sides
- these have $O(1)$ difference in payoffs, so

$$\mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] \approx \mathbb{E}[(\widehat{P}_l - \widehat{P}_{l-1})^2] = O(\sqrt{h})$$

For scalar SDEs, use conditional expectation one timestep before maturity:

- effectively smooths payoff over $O(\sqrt{h})$
- very helpful when $\widehat{S}^f - \widehat{S}^c = O(h)$
- minimal benefit when $\widehat{S}^f - \widehat{S}^c = O(\sqrt{h})$

Discontinuous payoffs



For paths in smoothed region, if $\widehat{S}_f - \widehat{S}_c = O(h)$ then

$$f'(S) = O(h^{-1/2}) \quad \Longrightarrow \quad \widehat{P}_l - \widehat{P}_{l-1} = O(h^{1/2})$$

and hence $\mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] = O(h^{3/2})$

Discontinuous payoffs

For multi-dimensional SDEs, approximate the Lévy areas by sub-sampling $W(t)$ within each timestep

Question: how many sub-samples to use?

- too few and there's no significant benefit
- too many and the computational cost is excessive
- what is optimal?

If each timestep is divided into M sub-intervals, error in each Lévy area approximation is $O(h\, M^{-1/2})$

Hence, strong convergence error and $\widehat{S}_f - \widehat{S}_c$ are both $O(h^{1/2} M^{-1/2})$, assuming $M \ll h^{-1}$

Discontinuous payoffs

Using antithetic treatment, for paths in smoothed region

$$\begin{aligned}\tfrac{1}{2}\left(f(\widehat{S}^{f(1)})+f(\widehat{S}^{f(2)})\right) - f(\widehat{S}^c) \;&\approx\; \tfrac{1}{2}\left(\widehat{D}_n^{(1)} + \widehat{D}_n^{(2)}\right) f'(\widehat{S}^c) \\ &\quad + \tfrac{1}{4}\left((\widehat{D}_n^{(1)})^2 + (\widehat{D}_n^{(2)})^2\right) f''(\widehat{S}^c) \\ &= \; O(h^{1/2} + M^{-1})\end{aligned}$$

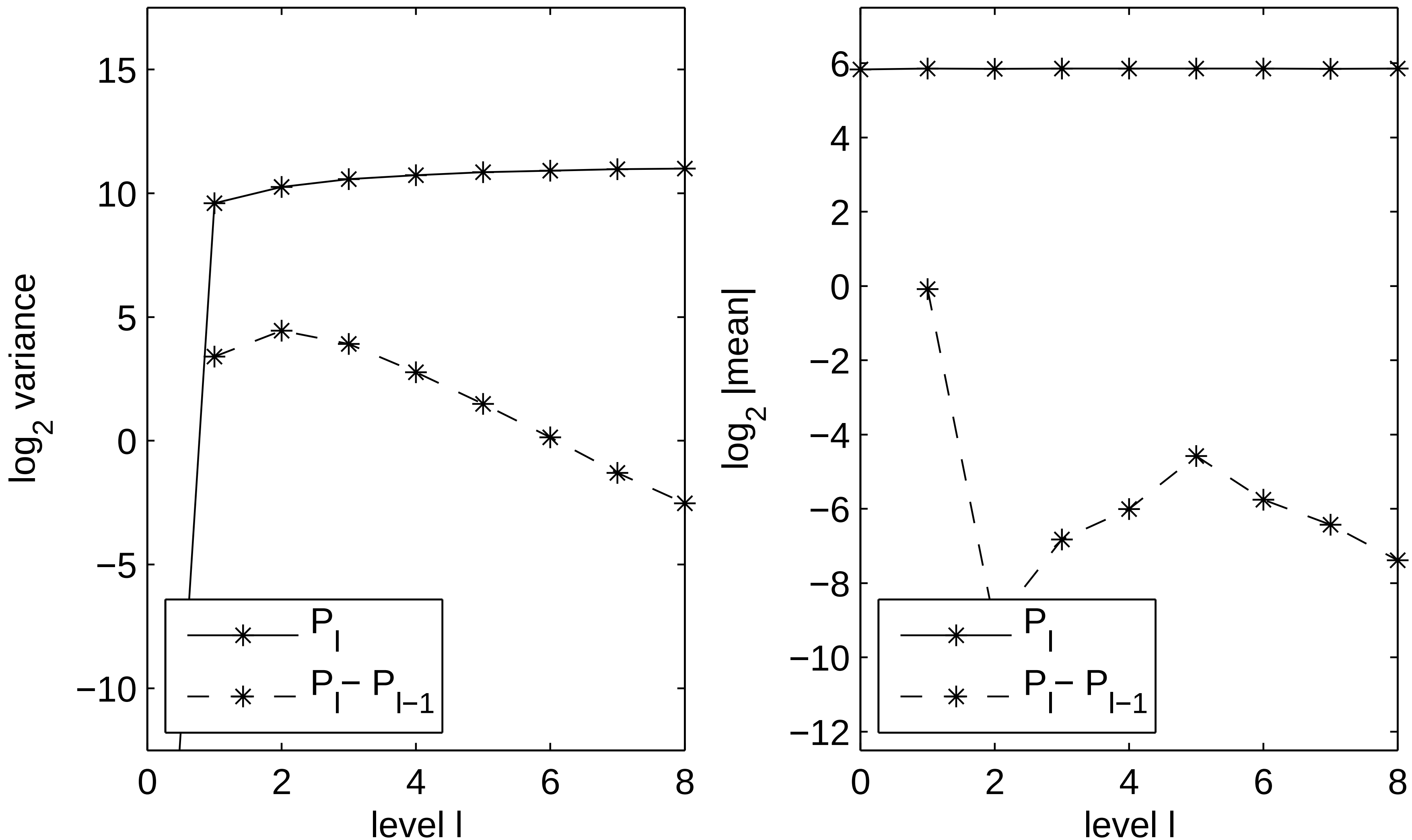
If $M^{-1} \gg h^{1/2}$, then doubling M doubles the cost per path, but reduces the variance by factor 4 — good!

Optimum is when $M = O(h^{-1/2})$

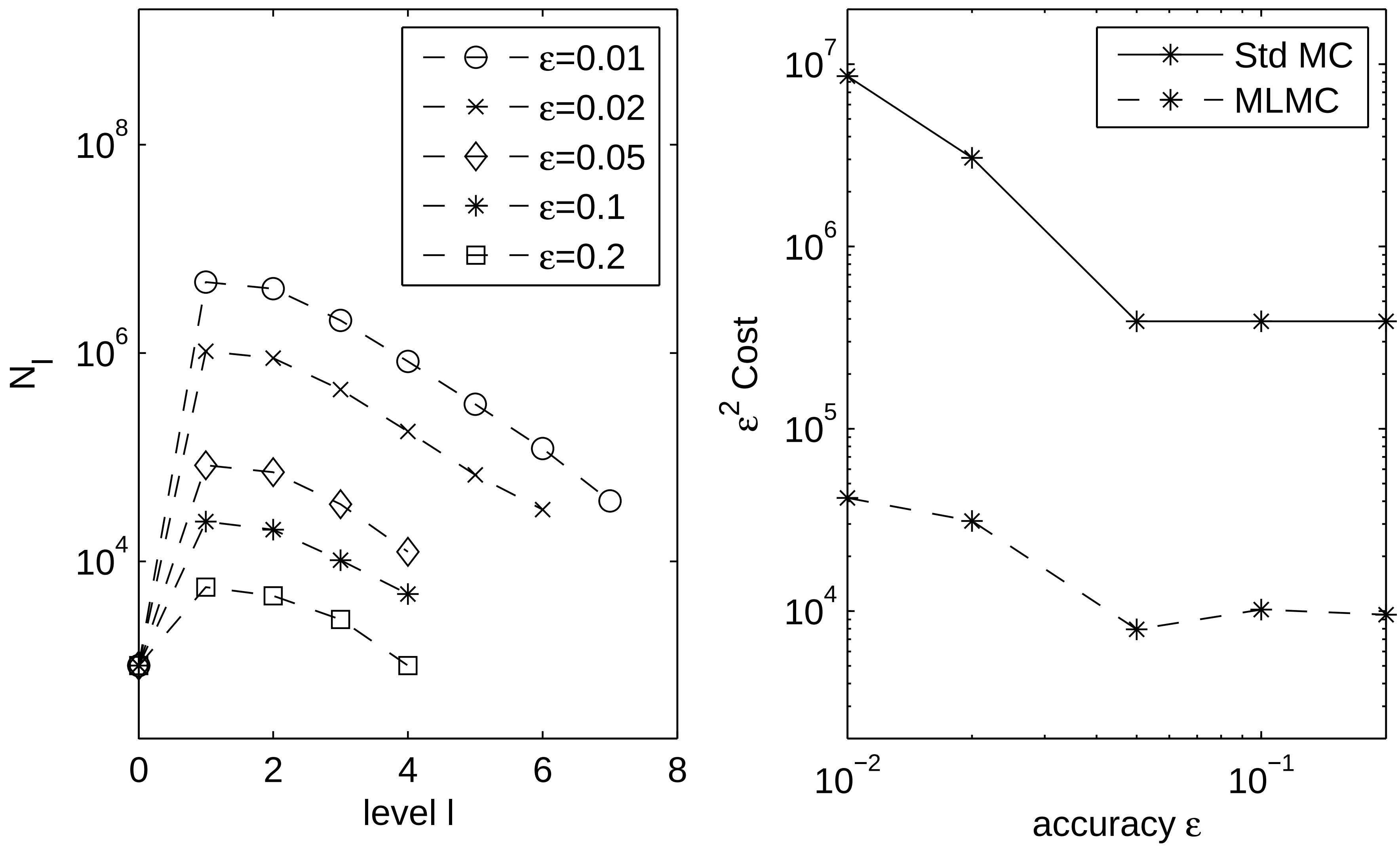
Multilevel variance is $O(h^{3/2})$ and cost is $O(h^{-1/2})$ per path; complexity analysis shows overall cost is $O(\varepsilon^{-2}(\log \varepsilon)^2)$.

Numerical test

Heston model for digital call $P = \exp(-rT)\, K\, \mathbf{1}_{S(T)>K}$



Numerical test



Conclusions

- multilevel method being adapted to increasingly more challenging applications
- for multi-dimensional SDEs with Lipschitz payoffs, neglecting the Lévy area terms in the Milstein scheme can still give good decay of the multilevel variance if antithetic variates are used
- for discontinuous payoffs, the Lévy areas need to be approximated but still get good decay of the variance

Papers are available from:
`www.maths.ox.ac.uk/gilesm/finance.html`