Multilevel Monte Carlo methods

Mike Giles

Mathematical Institute, University of Oxford

Applied Probability Frontiers:
Computational and Modeling Challenges

June 1-5, 2015

Objectives

In presenting the multilevel Monte Carlo method, I hope to emphasise:

- the simplicity of the idea
- its flexibility
- that it's not prescriptive, more an approach
- scope for improved performance through being creative
- lots of people working on a variety of applications

I will focus on ideas rather than lots of numerical results.

Control variate

Classic approach to variance reduction: approximate $\mathbb{E}[f]$ using

$$N^{-1} \sum_{n=1}^{N} \left\{ f(\omega^{(n)}) - \lambda \left(g(\omega^{(n)}) - \mathbb{E}[g] \right) \right\}$$

where

- control variate g has known expectation $\mathbb{E}[g]$
- g is well correlated with f, and optimal value for λ can be estimated by a few samples

For the optimal value of λ, the variance is reduced by factor $(1-\rho^2)$, where ρ is the correlation between f and g.

Two-level Monte Carlo

If we want to estimate $\mathbb{E}[f_1]$ but it is much cheaper to simulate $f_0 \approx f_1$, then since

$$\mathbb{E}[f_1] = \mathbb{E}[f_0] + \mathbb{E}[f_1 - f_0]$$

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} f_0^{(0,n)} \;+\; N_1^{-1} \sum_{n=1}^{N_1} \left(f_1^{(1,n)} - f_0^{(1,n)} \right)$$

Two differences from standard control variate method:

- $\mathbb{E}[f_0]$ is not known, so has to be estimated
- $\lambda = 1$

Benefit: if $f_1 - f_0$ is small, won't need many samples to accurately estimate $\mathbb{E}[f_1 - f_0]$, so cost will be reduced greatly.

Multilevel Monte Carlo

Natural generalisation: given a sequence $f_0, f_1, \ldots, f_L$

$$\mathbb{E}[f_L] = \mathbb{E}[f_0] + \sum_{\ell=1}^{L} \mathbb{E}[f_\ell - f_{\ell-1}]$$

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} f_0^{(0,n)} \; + \; \sum_{\ell=1}^{L} \left\{ N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(f_\ell^{(\ell,n)} - f_{\ell-1}^{(\ell,n)} \right) \right\}$$

with independent estimation for each level

Multilevel Monte Carlo

If we define

- C_0, V_0 to be cost and variance of f_0
- C_ℓ, V_ℓ to be cost and variance of $f_\ell - f_{\ell-1}$

then the total cost is $\displaystyle\sum_{\ell=0}^{L} N_\ell\, C_\ell$ and the variance is $\displaystyle\sum_{\ell=0}^{L} N_\ell^{-1} V_\ell$.

Using a Lagrange multiplier μ^2 to minimise the cost for a fixed variance

$$\frac{\partial}{\partial N_\ell} \sum_{k=0}^{L} \left(N_k\, C_k + \mu^2 N_k^{-1} V_k \right) = 0$$

gives

$$N_\ell = \mu \sqrt{V_\ell / C_\ell} \quad \Longrightarrow \quad N_\ell\, C_\ell = \mu \sqrt{V_\ell\, C_\ell}$$

Multilevel Monte Carlo

Setting the total variance equal to ε^2 gives

$$\mu = \varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell \, C_\ell} \right)$$

and hence, the total cost is

$$\sum_{\ell=0}^{L} N_\ell \; C_\ell \;=\; \varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell C_\ell} \right)^2$$

in contrast to the standard cost which is approximately $\varepsilon^{-2} \, V_0 \, C_L$.

The MLMC cost savings are therefore:

- V_L / V_0, if $\sqrt{V_\ell C_\ell}$ increases with level
- C_0 / C_L, if $\sqrt{V_\ell C_\ell}$ decreases with level

Multilevel Monte Carlo

This analysis treated the N_ℓ as real variables. Rounding them up to the nearest integer gives the following result:

Theorem: With V_ℓ and C_ℓ as defined previously, an estimate $\widehat{Y}$ with RMS accuracy ε,

$$\text{MSE} \;\equiv\; \mathbb{E}\left[(\widehat{Y} - \mathbb{E}[f_L])^2\right] \;\leq\; \varepsilon^2$$

can be obtained at computational cost

$$\varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell \, C_\ell} \right)^2 + \sum_{\ell=0}^{L} C_\ell$$

Note: this assumes perfect knowledge of V_ℓ and C_ℓ. In practice V_ℓ at least usually needs to be estimated.

Multilevel Path Simulation

Motivated by computational finance applications, in 2006 I introduced MLMC for SDEs (stochastic differential equations).

$$\mathrm{d}S_t = a(S_t, t)\,\mathrm{d}t + b(S_t, t)\,\mathrm{d}W_t$$

Level ℓ corresponds to approximation using 2^ℓ timesteps, giving approximate payoff $\widehat{P}_\ell$.

Choice of finest level L depends on weak error (bias).

Multilevel decomposition gives

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

Multilevel Monte Carlo

Simplest estimator for $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ for $\ell > 0$ is

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

using same driving Brownian path for both levels

Standard analysis gives $\mathrm{MSE} = \left(\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P] \right)^2 + \sum_{\ell=0}^{L} N_\ell^{-1} V_\ell$

To make RMS error less than ε

- choose L so that $\left(\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P] \right)^2 < \frac{1}{2}\,\varepsilon^2$
- choose $N_\ell \propto \sqrt{V_\ell / C_\ell}$ so total variance is less than $\frac{1}{2}\,\varepsilon^2$

MLMC Theorem

(Slight generalisation of original version)

If there exist independent estimators $\widehat{Y}_\ell$ based on N_ℓ Monte Carlo samples, each costing C_ℓ, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2} \min(\beta, \gamma)$ and

i) $\left| \mathbb{E}[\widehat{P}_\ell - P] \right| \leq c_1 \, 2^{-\alpha \, \ell}$

ii) $\mathbb{E}[\widehat{Y}_\ell] = \begin{cases} \mathbb{E}[\widehat{P}_0], & \ell = 0 \\ \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}], & \ell > 0 \end{cases}$

iii) $\mathbb{V}[\widehat{Y}_\ell] \leq c_2 \, N_\ell^{-1} 2^{-\beta \, \ell}$

iv) $\mathbb{E}[C_\ell] \leq c_3 \, 2^{\gamma \, \ell}$

MLMC Theorem

then there exists a positive constant c_4 such that for any $\varepsilon < 1$ there exist L and N_ℓ for which the multilevel estimator

$$\widehat{Y} = \sum_{\ell=0}^{L} \widehat{Y}_\ell,$$

has a mean-square-error with bound $\mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[P]\right)^2\right] < \varepsilon^2$

with an expected computational cost C with bound

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

MLMC Theorem

Two observations of optimality:

- MC simulation needs $O(\varepsilon^{-2})$ samples to achieve RMS accuracy ε. When $\beta > \gamma$, the cost is optimal — $O(1)$ cost per sample on average.

  (Would need multilevel QMC to further reduce costs)

- When $\beta < \gamma$, another interesting case is when $\beta = 2\alpha$, which corresponds to $\mathbb{E}[\widehat{Y}_\ell]$ and $\sqrt{\mathbb{E}[\widehat{Y}_\ell^2]}$ being of the same order as $\ell \to \infty$. In this case, the total cost is $O(\varepsilon^{-\gamma/\alpha})$, which is the cost of a single sample on the finest level — again optimal.

MLMC generalisation

The theorem is for scalar outputs P, but it can be generalised to multi-dimensional (or infinite-dimensional) outputs with

i) $\left\| \mathbb{E}[\widehat{P}_\ell - P] \right\| \leq c_1 \, 2^{-\alpha\,\ell}$

ii) $\mathbb{E}[\widehat{Y}_\ell] = \begin{cases} \mathbb{E}[\widehat{P}_0], & \ell = 0 \\ \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}], & \ell > 0 \end{cases}$

iii) $\mathbb{V}[\widehat{Y}_\ell] \equiv \mathbb{E}\left[\left\| \widehat{Y}_\ell - \mathbb{E}[\widehat{Y}_\ell] \right\|^2 \right] \leq c_2 \, N_\ell^{-1} 2^{-\beta\,\ell}$

G, Nagapetyan & Ritter (2014) have used this for estimating cumulative distribution functions and probability density functions arising from SDEs.

Parametric Integration

This also connects to the first research on multilevel methods by Stefan Heinrich in 1999.

This was for parametric integration, in which x is a finite-dimensional random variable, and want to estimate

$$g(\lambda) \equiv \mathbb{E}[f(x,\lambda)]$$

for a range of values of the parameter λ.

In the simplest case, suppose

- λ is a scalar
- the parameter range is $0 \leq \lambda \leq 1$
- we use piecewise linear interpolation to approximate $g(\lambda)$

Parametric Integration

On the coarsest level, we only need to estimate $\mathbb{E}[f(x,0)]$ and $\mathbb{E}[f(x,1)]$.

On the next level we require $\mathbb{E}[f(x,\frac{1}{2})]$ but note that

$$\begin{aligned}\mathbb{E}[f(x,\tfrac{1}{2})] &= \tfrac{1}{2}\Big(\mathbb{E}[f(x,0)] + \mathbb{E}[f(x,1)]\Big) \\ &\quad + \mathbb{E}\left[f(x,\tfrac{1}{2}) - \tfrac{1}{2}(f(x,0) + f(x,1))\right]\end{aligned}$$

Provided $f(x,\lambda)$ is smooth with respect to λ then

$$\mathbb{V}\left[f(x,\tfrac{1}{2}) - \tfrac{1}{2}(f(x,0) + f(x,1))\right]$$

is smaller than $\mathbb{V}[f(x,\frac{1}{2})]$. This extends naturally to additional levels.

MLMC Theorem

MLMC Theorem allows a lot of freedom in constructing the multilevel estimator. I sometimes use different approximations on the coarse and fine levels:

$$\widehat{Y}_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(\widehat{P}_\ell^f(\omega^{(n)}) - \widehat{P}_{\ell-1}^c(\omega^{(n)}) \right)$$

The telescoping sum still works provided

$$\mathbb{E}\left[\widehat{P}_\ell^f\right] = \mathbb{E}\left[\widehat{P}_\ell^c\right].$$

Given this constraint, can be creative to reduce the variance

$$\mathbb{V}\left[\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c\right].$$

MLMC Theorem

Two examples:

- zero-mean control variate estimator: if

$$\widehat{P}_\ell(\omega^{(n)}) \approx \widehat{P}_{\ell-1}(\omega^{(n)}) + Z(\omega^{(n)})$$

where $\mathbb{E}[Z]=0$, then use

$$\widehat{P}^c_{\ell-1}(\omega^{(n)}) \equiv \widehat{P}_{\ell-1}(\omega^{(n)}), \qquad \widehat{P}^f_\ell(\omega^{(n)}) \equiv \widehat{P}_\ell(\omega^{(n)}) - Z(\omega^{(n)})$$

- antithetic estimator:

$$\widehat{P}^c_{\ell-1}(\omega^{(n)}) \equiv \widehat{P}_{\ell-1}(\omega^{(n)}), \qquad \widehat{P}^f_\ell(\omega^{(n)}) \equiv \tfrac{1}{2}\left(\widehat{P}_\ell(\omega^{(n)}) + \widehat{P}_\ell(\omega^{(n)}_{anti})\right)$$

where $\omega^{(n)}_{anti}$ is an antithetic "twin" with the same distribution as $\omega^{(n)}$.

MLMC Challenges

- not always obvious how to couple coarse and fine levels
  i.e. what does $\widehat{P}_\ell(\omega^{(n)}) - \widehat{P}_{\ell-1}(\omega^{(n)})$ mean?

- some creativity required to handle discontinuous functionals, where a small difference between the underlying coarse and fine simulations can produce an $O(1)$ difference in the output

- numerical analysis to determine the decay rate of V_ℓ can be tough

Brownian Diffusion SDEs

Brownian increments for coarse path obtained by summing increments for fine path – very simple and natural

I like the Milstein discretisation which gives first order strong convergence

$$\left(\mathbb{E} \left[\sup_{[0,T]} \| S_t - \widehat{S}_t \|^2 \right] \right)^{1/2} = O(h)$$

so for payoffs which are Lipschitz functions of the final state we get

$$\widehat{P}_\ell - \widehat{P}_{\ell-1} = O(h_\ell)$$

and hence $V_\ell = O(h_\ell^2)$.

However, not so easy for lookback, digital and barrier options. Also, in multiple dimensions sometimes requires Lévy areas, but can be avoided by an antithetic treatment, (G & Szpruch, 2013).

Financial application

- basket of 5 underlying assets, modelled by Geometric Brownian Motion

$$\mathrm{d}S_i = r\, S_i\, \mathrm{d}t + \sigma_i\, S_i\, \mathrm{d}W_i$$

  with correlation between 5 driving Brownian motions

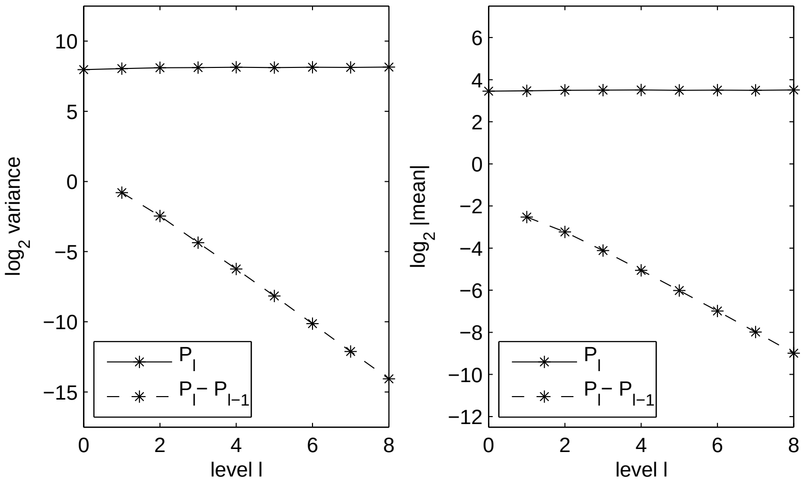
- Milstein numerical approximation

- standard call option is piecewise linear function of average at final time T

- digital call option is discontinuous function of average

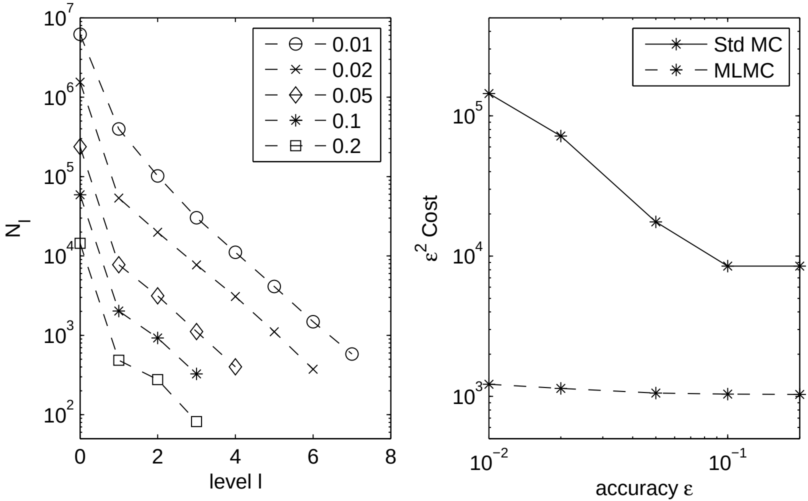
Financial application

Standard call option:



Financial application

Standard call option:



Digital options

In a digital option, the payoff is a discontinuous function of the final state.

Using the Milstein approximation, first order strong convergence means that $O(h_\ell)$ of the simulations have coarse and fine paths on opposite sides of a discontinuity.

Hence,

$$\widehat{P}_\ell - \widehat{P}_{\ell-1} = \begin{cases} O(1), & \text{with probability } O(h_\ell) \\ O(h_\ell), & \text{with probability } O(1) \end{cases}$$

so

$$\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] = O(h_\ell), \quad \mathbb{E}[(\widehat{P}_\ell - \widehat{P}_{\ell-1})^2] = O(h_\ell),$$

and hence $V_\ell = O(h_\ell)$, not $O(h_\ell^2)$

Digital options

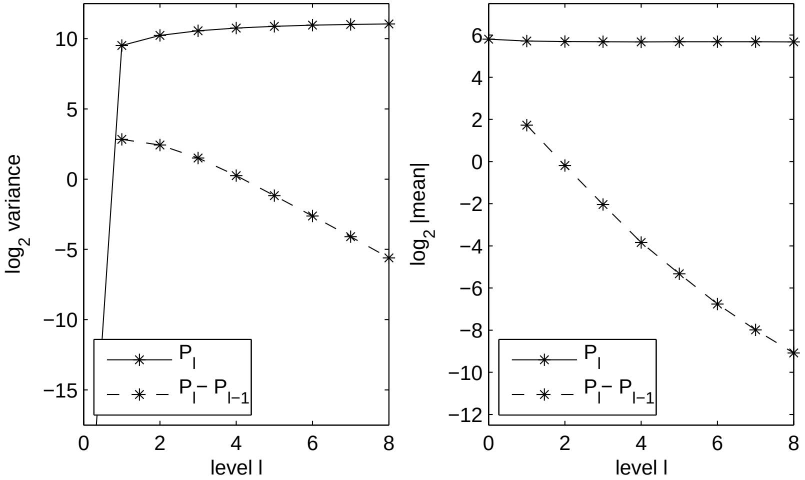
Three fixes:

- Conditional expectation: using the Euler discretisation instead of Milstein for the final timestep, conditional on all but the final Brownian increment, the final state has a Gaussian distribution, with a known analytic conditional expectation in simple cases
- Splitting: split each path simulation into M paths by trying M different values for the Brownian increment for the last fine path timestep
- Change of measure: when the expectation is not known, can use a change of measure so the coarse path takes the same final state as the fine path — difference in the "payoff" now comes from the Radon-Nikodym derivative

These all effectively smooth the payoff – end up with $V_\ell = O(h_\ell^{3/2})$.

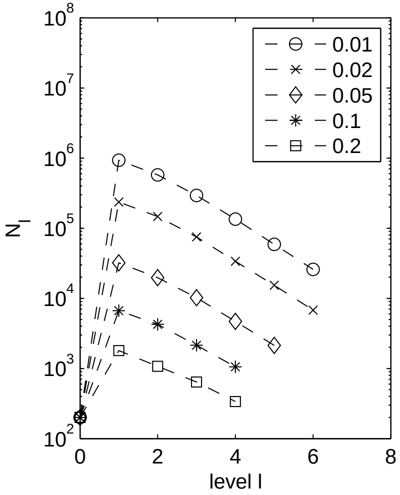
Financial application

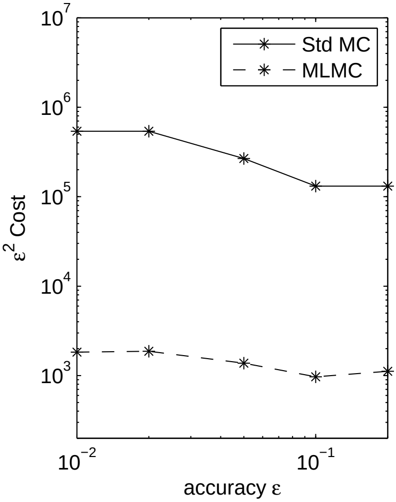
Digital call option:



Financial application

Digital call option:





Numerical 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 (2009) and Avikainen (2009).
Milstein analysis due to G, Debrabant & Rößler (2012).

SPDEs

- quite natural application, with better cost savings than SDEs due to higher dimensionality
- range of applications
  - Graubner & Ritter (Darmstadt $\rightarrow$ Kaiserslautern) – parabolic
  - G, Reisinger (Oxford) – parabolic
  - Cliffe, G, Scheichl, Teckentrup (Bath/Nottingham) – elliptic
  - Barth, Jenny, Lang, Meyer, Mishra, Müller, Schwab, Sukys, Zollinger (ETH Zürich) – elliptic, parabolic, hyperbolic
  - Harbrecht, Peters (Basel) – elliptic
  - Efendiev (Texas A&M) – numerical homogenization
  - Vidal-Codina, G, Peraire (MIT) – reduced basis approximation
  - G, Hou, Zhang (Caltech) – numerical homogenization

Engineering Uncertainty Quantification

- consider 3D elliptic PDE, with uncertain boundary data
- use grid spacing proportional to $2^{-\ell}$ on level ℓ
- cost is $O(2^{+3\ell})$, if using an efficient multigrid solver
- 2nd order accuracy means that

$$
\begin{aligned}
& \widehat{P}_\ell(\omega) - \widehat{P}(\omega) \;\approx\; c(\omega)\, 2^{-2\ell} \\
\Longrightarrow \quad & \widehat{P}_{\ell-1}(\omega) - \widehat{P}_\ell(\omega) \;\approx\; 3\, c(\omega)\, 2^{-2\ell}
\end{aligned}
$$

- hence, $\alpha = 2,\ \beta = 4,\ \gamma = 3$
- cost is $O(\varepsilon^{-2})$ to obtain ε RMS accuracy

Stochastic chemical reactions

In stochastic simulations, each reaction is a Poisson process with a rate which depends on the current concentrations.

In the "tau-leaping" method (Euler-Maruyama method) the reaction rates are frozen at the start of the timestep, so for each reaction sample from a Poisson process $P(\lambda\,\Delta t)$ to determine the number of reactions in that timestep.

(As $\lambda\,\Delta t \to \infty$, the standard deviation becomes smaller relative to the mean, and it approaches the deterministic limit.)

Stochastic chemical reactions

Anderson & Higham (2011) developed a very efficient multilevel version of this algorithm – big savings because finest level usually has 1000's of timesteps.

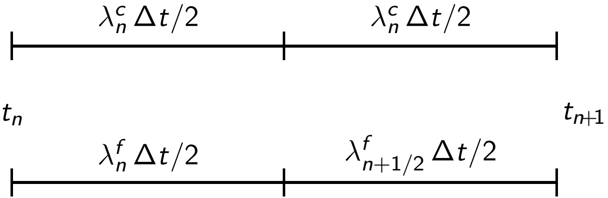
Key challenge: how to couple coarse and fine path simulations?

Crucial observation: $P(t_1) + P(t_2) \stackrel{d}{=} P(t_1 + t_2)$ for $t_1, t_2 \geq 0$

Stochastic chemical reactions

Solution:

- simulate the Poisson variable on the coarse timestep as the sum of two fine timestep Poisson variables
- couple the fine path and coarse path Poisson variables by using common variable based on smaller of two rates



If $\lambda_n^f < \lambda_n^c$, use $P(\lambda_n^c \Delta t/2) \sim P(\lambda_n^f \Delta t/2) + P((\lambda_n^c - \lambda_n^f)\Delta t/2)$

Other MLMC Applications

- multilevel QMC (Dick, G, Kuo, Scheichl, Schwab, Sloan)
- Lévy-driven SDEs (Dereich, Heidenreich)
- stochastic chemical reactions (Anderson & Higham, Tempone)
- mixed precision computation on FPGAs (Korn, Ritter, Wehn)
- MLMC for MCMC (Scheichl, Schwab, Stuart, Teckentrup)
- Coulomb collisions in plasma (Caflisch)
- nested simulation (Haji-Ali & Tempone, Hambly & Reisinger)

Recent MLMC Extensions

- unbiased estimation through randomisation of levels (Rhee & Glynn)
  - good for $\beta > \gamma$
- Richardson/Romberg extrapolation (Lemaire & Pagès)
  - good for $\beta < \gamma$
- Multi-Index Monte Carlo (Haji-Ali, Nobile & Tempone)
  - combines MLMC with sparse grid methods
  - potentially very important for SPDE applications

Unbiased MLMC

Rhee & Glynn (2014) (see also McLeish (2011)) use the estimator

$$Y = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{p_{\ell(n)}} (P^{(n)}_{\ell(n)} - P^{(n)}_{\ell(n)-1}),$$

where the level ℓ is selected randomly with probability p_ℓ.

It can also be expressed as

$$Y = \sum_{\ell=0}^{\infty} \left(\frac{1}{p_\ell\, N} \sum_{n=1}^{N_\ell} (P^{(n)}_\ell - P^{(n)}_{\ell-1}) \right).$$

where N_ℓ is random with $\displaystyle\sum_{\ell=0}^{\infty} N_\ell = N, \quad \mathbb{E}[N_\ell] = p_\ell N.$

Unbiased MLMC

It is unbiased because

$$
\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}\left[\frac{1}{p_{\ell'}}(P_{\ell'} - P_{\ell'-1})\right] \\
&= \sum_{\ell=0}^{\infty} p_\ell \, \mathbb{E}\left[\frac{1}{p_{\ell'}}(P_{\ell'} - P_{\ell'-1}) \mid \ell' = \ell\right] \\
&= \sum_{\ell=0}^{\infty} \mathbb{E}\left[P_\ell - P_{\ell-1}\right] \;=\; \mathbb{E}[P].
\end{aligned}
$$

Furthermore, defining $E_\ell = \mathbb{E}[P_\ell - P_{\ell-1}]$,

$$
\mathbb{V}[Y] \;=\; \sum_{\ell=0}^{\infty} \frac{1}{p_\ell}(V_\ell + E_\ell^2) - \left(\sum_{\ell=0}^{\infty} E_\ell\right)^2 \;\geq\; \sum_{\ell=0}^{\infty} \frac{1}{p_\ell}\, V_\ell,
$$

due to Jensen's inequality.

Unbiased MLMC

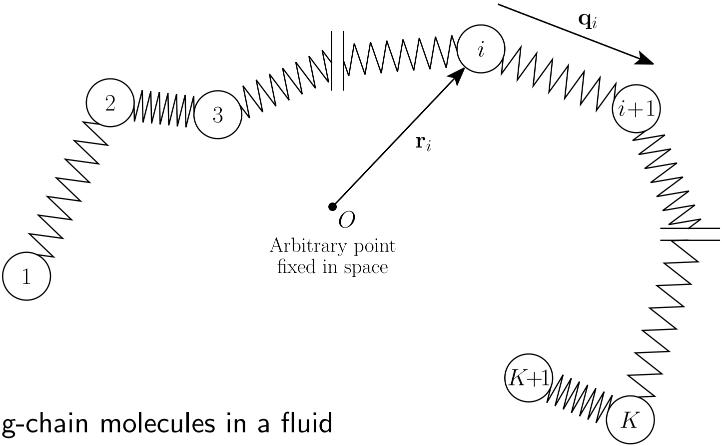
For both the variance and expected cost to be finite, need

$$\sum_{\ell=0}^{\infty} \frac{1}{p_\ell} V_\ell < \infty, \quad \sum_{\ell=0}^{\infty} p_\ell\, C_\ell < \infty.$$

Under the conditions of the MLMC Theorem, this is possible when $\beta > \gamma$ by choosing $p_\ell \propto 2^{-(\gamma+\beta)\ell/2}$, so that

$$\frac{1}{p_\ell} V_\ell \propto 2^{-(\beta-\gamma)\ell/2}, \quad p_\ell\, C_\ell \propto 2^{-(\beta-\gamma)\ell/2}.$$

FENE molecules in a fluid (Süli, Ye)



- long-chain molecules in a fluid
- modelled as ball-and-spring systems, subject to
  - force due to Finitely Extensible Nonlinear Elastic bond energy
  - force due to local rate-of-strain tensor $\partial v/\partial x$
  - random forcing due to fluid fluctuations

Modelling

- FENE (finitely extensible nonlinear elastic) model limits extension of molecular bonds
- motion of "balls" given by force balance:

$$\text{elastic force} + \text{random force} + \text{viscous drag} = 0$$

$$-\nabla V \quad + \quad R \quad - \quad k\ \left(\dot{r}_i - v(r_i)\right) \quad = \quad 0$$

  where V is the elastic potential, and v is the velocity of the fluid
- shifting to a moving frame of reference, a local Taylor series expansion gives

$$v(x) \approx \kappa\ x$$

  where κ is the local rate-of-strain tensor $\partial v / \partial x$

Modelling

This modelling leads to the following SDE for i^{th} "ball":

$$\mathrm{d}r_i = (\kappa\, r_i - \nabla_{r_i} V(r))\, \mathrm{d}t + \sqrt{2}\,\mathrm{d}W_i$$

where $\mathrm{d}W_i$ is the Brownian forcing, assumed to be independent of the forcing of the others, and

$$V(r) = \sum_{i=1}^{K} U_i(\|q_i\|^2/2)$$

with U_i being the elastic potential for the i^{th} bond.

Hence,

$$\mathrm{d}r_i = \left(\kappa\, r_i - \left(U'_{i-1}\left(\|q_{i-1}\|^2/2\right)\, q_{i-1} - U'_i\left(\|q_i\|^2/2\right)\, q_i\right)\right) \mathrm{d}t + \sqrt{2}\, \mathrm{d}W_i$$

if we define $q_0 \equiv q_{K+1} \equiv 0$ to account for non-existent bonds on either end.

Modelling

The coupled system of SDEs can be written collectively as

$$\mathrm{d}q = (K\, q - D\, \nabla V)\, \mathrm{d}t + \sqrt{2}\; L\, \mathrm{d}W$$

where

- $V(q) \equiv \sum_i U_i(\|q_i\|^2/2)$ is the total bond energy, with $U_i(\|q_i\|^2/2) \to \infty$ as $\|q_i\|^2 \to 1$
- K is block diagonal, due to the fluid strain-rate tensor $\partial v/\partial x$
- L and D are of the form

$$L = \begin{pmatrix} -I & I & & \\ & -I & I & \\ & & -I & I \end{pmatrix}, \quad D = \begin{pmatrix} 2I & -I & \\ -I & 2I & -I \\ & -I & 2I \end{pmatrix} = L\, L^T.$$

Invariant distribution

The Fokker-Planck PDE for the probability density function $p(q,t)$ is

$$\frac{\partial p}{\partial t} = \nabla \cdot \Big((D\,\nabla V - K\,q)\,p + D\,\nabla p \Big),$$

so when $K\!=\!0$ the invariant (or equilibrium) distribution is

$$p_\infty(q) = C\,\exp(-V(q)).$$

In the particular case of the FENE model, we have

$$U_i(s) = -\beta\,\log(1-2s),$$

and therefore the invariant distribution for $K\!=\!0$ is

$$p_\infty(q) = C\,\prod_i (1-\|q_i\|^2)^\beta.$$

Contraction property

If κ is not too large, can prove the following result:

If two paths $q^{(1)}$, $q^{(2)}$ have

- <u>different</u> initial conditions
- <u>same</u> driving Brownian motion W_t

then

$$\left\| q^{(1)}(t) - q^{(2)}(t) \right\| \longrightarrow 0$$

exponentially as $t \to 0$.

Numerical approximation

The SDE is approximated as

$$q_{n+1} = q_n + (K\, q_n - D \, \nabla V(q_n))\, h_n + \sqrt{2}\; L\; \Delta W_n$$

using an adaptive timestep h_n.

No bond length should exceed 1 – try to ensure this through the restrictions:

$$\begin{aligned} h_n\, U_i'(\|q_{i,n}\|^2/2)\; \|q_{i,n}\| &\leq 1 - \|q_{i,n}\| \\ 5\sqrt{2\, h_n} &\leq 1 - \|q_{i,n}\| \end{aligned}$$

where $q_{i,n}$ is the i^{th} bond vector at timestep n (and then use clamping if this fails).

This sets an upper bound on the timestep – smaller values need to be chosen for accuracy.

Objective

Our objective is to numerically estimate $\lim_{T\to\infty} \mathbb{E}\left[P(T)\right]$ where

$$P(T) \;\equiv\; U'(\|q\|^2/2)\, q\, q^T \Big|_{t=T}$$

This corresponds to the stress exerted by the molecule on the fluid. In the future this will be applied to the fluid in a coupled simulation.

We start with a computation for a fixed, large T, then address the challenge of letting $T \to \infty$.

Multilevel Monte Carlo simulation

First challenge: how does MLMC work with adaptive time-stepping?

Actually, surprisingly easy — on level ℓ use

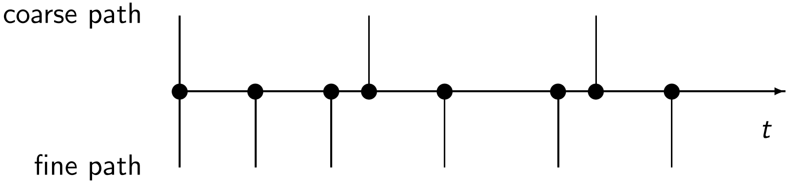
$$h_n = 2^{-\ell} \; \frac{\min_i (1 - \|q_{i,n}\|)^2}{\max(2\beta, 50)}$$

Coarse and fine paths each compute their own adaptive timesteps independently – this ensures the telescoping sum works correctly

But what is involved in coarse and fine paths using same driving Brownian motion?

Multilevel Monte Carlo simulation

As time proceeds, Brownian increments are generated as needed at discretisation times which are a union of coarse and fine path times:



The fact that the timesteps are not nested is not a problem – strong convergence still ensures a strong coupling between the coarse and fine paths, because both approximate the true path.

Multilevel Monte Carlo simulation

Second challenge: how to obtain expectation as $T \to \infty$?

Key idea here comes from other research by Glynn & Rhee (2014) on contracting Markov chains:

$$X_0 = x, \qquad X_{n+1} = \phi_n(X_n), \quad n \geq 0$$

where $\{\phi_n\}$ is a sequence of iid random functions such that

$$\sup_{x \neq y} \mathbb{E}\left[\left(\frac{d(\phi_n(x), \phi_n(y))}{d(x,y)} \right)^{2\gamma} \right] < 1$$

for some distance metric d, and some $\gamma \in (0,1)$.

Multilevel Monte Carlo simulation

They are interested in

$$\lim_{M\to\infty} \left\{ \mathbb{E}[f(X_M)] \mid X_0 = x \right\}$$

which can be re-expressed as

$$\lim_{M\to\infty} \left\{ \mathbb{E}[f(X_0)] \mid X_{-M} = x \right\}$$

and they use multilevel with $M_\ell \to \infty$ as $\ell \to \infty$ and <u>same</u> random ϕ_n for *coarse* and *fine* paths for $-M_{\ell-1} \leq n < 0$.

This works because contraction property leads to effect of difference in values at $-M_{\ell-1}$ decaying exponentially, so

$$\left\| X_0^f - X_0^c \right\| \;\sim\; \exp(-c\, M_{\ell-1})$$

Multilevel Monte Carlo simulation

Back to our polymer application, instead of estimating

$$\lim_{T\to\infty} \left\{ \mathbb{E}[P(q(T))] \mid q(0) = q_0 \right\}$$

we use the same idea and estimate

$$\lim_{T\to\infty} \left\{ \mathbb{E}[P(q(0))] \mid q(-T) = q_0 \right\}$$

and use multilevel with $h_\ell \to 0$, $T_\ell \to \infty$ as $\ell \to \infty$ and the <u>same</u> Brownian motion $W(t)$ for coarse and fine paths for $-T_{\ell-1} < t < 0$.

This again works because of the contraction property which leads to effect of difference in q values at time $-T_{\ell-1}$ decaying exponentially, so

$$\left\| q^f(0) - q^c(0) \right\| \;=\; O(h_\ell) \;+\; O(\exp(-c\, T_{\ell-1}))$$

Conclusions

- multilevel idea is very simple
- challenge can be how to apply it in new situations
- discontinuous output functions can cause problems, but there is a lot of experience now in coping with this
- there are also "tricks" which can be used in situations with poor strong convergence
- being used for an increasingly wide range of applications; biggest computational savings when coarsest (helpful) approximation is much cheaper than finest
- currently, getting at least $100\times$ savings for SPDEs and stochastic chemical reaction simulations

References

Webpage for my research/papers:

`people.maths.ox.ac.uk/gilesm/mlmc.html`

Webpage for new 70-page *Acta Numerica* review and MATLAB test codes:

`people.maths.ox.ac.uk/gilesm/acta/`

– contains references to almost all MLMC research, including some very early related work by Achi Brandt

MLMC Community

Webpage: `people.maths.ox.ac.uk/gilesm/mlmc_community.html`

Abo Academi (Avikainen) – numerical analysis
Basel (Harbrecht) – elliptic SPDEs, sparse grids
Bath (Kyprianou, Scheichl, Shardlow, Yates) – elliptic SPDEs, MCMC, Lévy-driven SDEs, stochastic chemical modelling
Chalmers (Lang) – SPDEs
Duisburg (Belomestny) – Bermudan and American options
Edinburgh (Davie, Szpruch) – SDEs, numerical analysis
EPFL (Abdulle) – stiff SDEs and SPDEs
ETH Zürich (Jenny, Jentzen, Schwab) – SPDEs, multilevel QMC
Frankfurt (Gerstner, Kloeden) – numerical analysis, fractional Brownian motion
Fraunhofer ITWM (Iliev) – SPDEs in engineering
Hong Kong (Chen) – Brownian meanders, nested simulation in finance
IIT Chicago (Hickernell) – SDEs, infinite-dimensional integration, complexity analysis
Kaiserslautern (Heinrich, Korn, Ritter) – finance, SDEs, parametric integration, complexity analysis
KAUST (Tempone, von Schwerin) – adaptive time-stepping, stochastic chemical modelling
Kiel (Gnewuch) – randomized multilevel QMC
LPMA (Frikha, Lemaire, Pagès) – numerical analysis, multilevel extrapolation, finance applications
Mannheim (Neuenkirch) – numerical analysis, fractional Brownian motion
MIT (Peraire) – uncertainty quantification, SPDEs
Munich (Hutzenthaler) – numerical analysis
Oxford (Baker, Giles, Hambly, Reisinger) – SDEs, SPDEs, numerical analysis, finance applications, stochastic chemical modelling
Passau (Müller-Gronbach) – infinite-dimensional integration, complexity analysis
Stanford (Glynn) – numerical analysis, randomized multilevel
Strathclyde (Higham, Mao) – numerical analysis, exit times, stochastic chemical modelling
Stuttgart (Barth) – SPDEs
Texas A&M (Efendiev) – SPDEs in engineering
UCLA (Caflisch) – Coulomb collisions in physics
UNSW (Dick, Kuo, Sloan) – multilevel QMC
UTS (Baldeaux) – multilevel QMC
Warwick (Stuart, Teckentrup) – MCMC for SPDEs
WIAS (Friz, Schoenmakers) – rough paths, fractional Brownian motion, Bermudan options
Wisconsin (Anderson) – numerical analysis, stochastic chemical modelling