Multilevel Monte Carlo methods

Mike Giles

Mathematical Institute, University of Oxford

London Probability Seminar

April 24, 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.

Monte Carlo method

Given a function f of a random input ω, to estimate the value of $\mathbb{E}[f]$ we can use the Monte Carlo estimate

$$N^{-1}\sum_{n=1}^{N} f(\omega^{(n)}).$$

based on N independent samples $\omega^{(n)}$.

By the Central Limit Theorem, as $N \to \infty$, the error in this estimate becomes Normally distributed, with variance $N^{-1}\mathbb{V}[f]$.

The error lies within 3 s.d. with probability 99.7%, giving us a confidence interval.

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

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.

Reduced Basis PDE approximation

Vidal-Codina, Nguyen, G, Peraire (2014) take a fine FE discretisation:

$$A(\omega)\, u = f(\omega)$$

and use a reduced basis approximation

$$u \approx \sum_{k=1}^{K} v_k\, u_k$$

to obtain a low-dimensional reduced system

$$A_r(\omega)\, v = f_r(\omega)$$

- larger $K \Longrightarrow$ greater accuracy at greater cost
- in multilevel treatment, K_ℓ varies with level
- brute force optimisation determines the optimal number of levels, and reduced basis size on each level

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 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

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

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

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

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

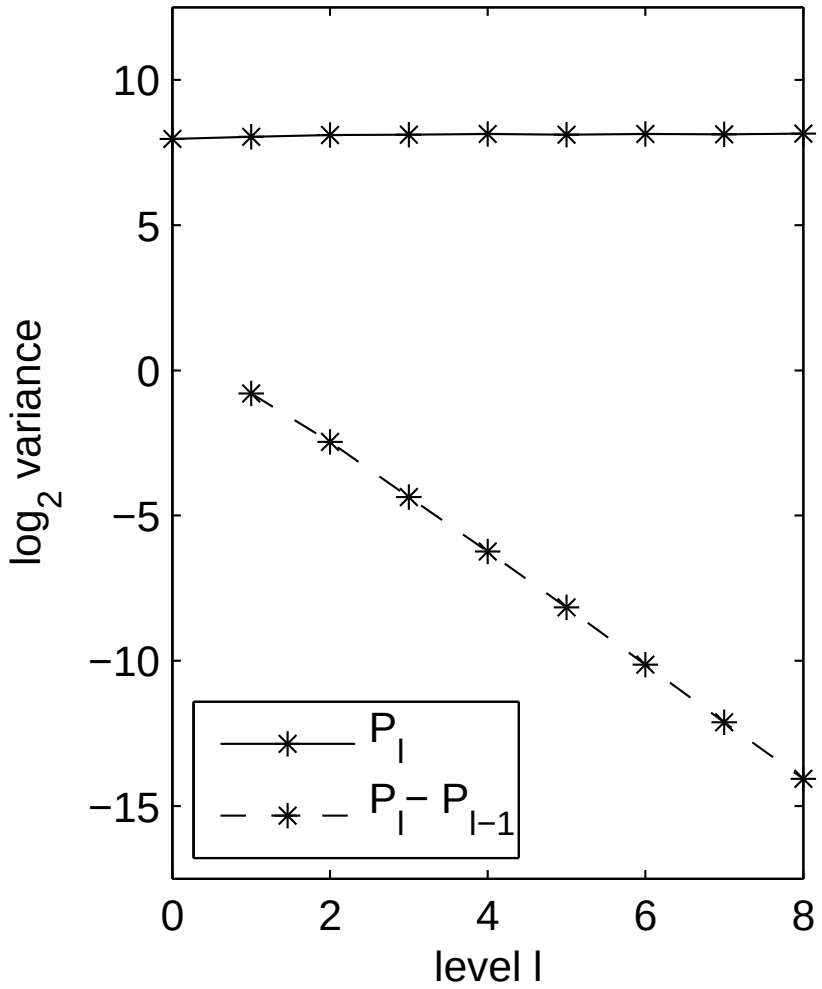
- Milstein numerical approximation

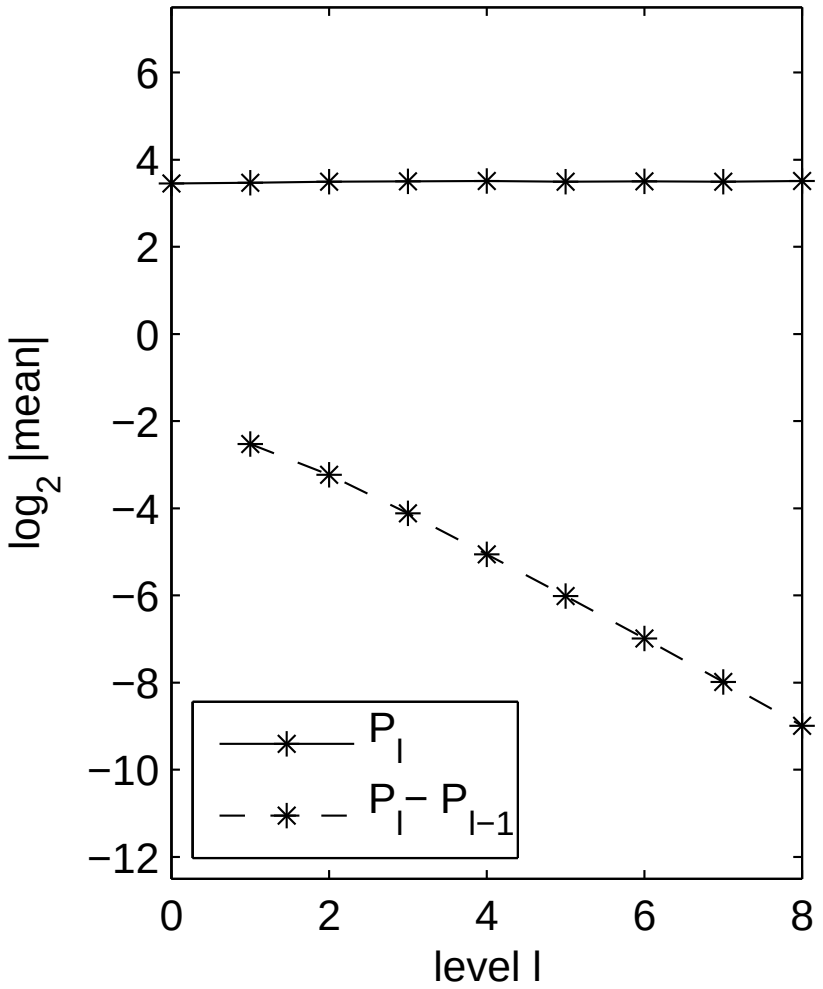
- standard call option is piecewise linear function of average at final time T

- digital call option is discontinuous function of average

Financial application

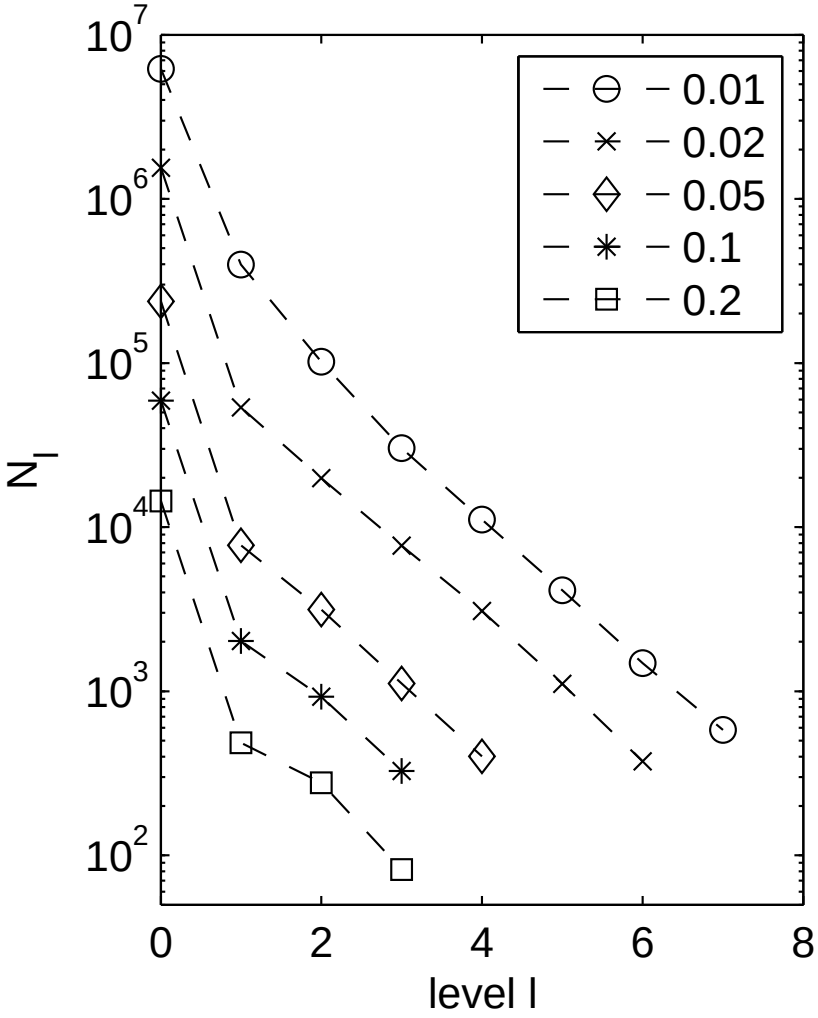
Standard call option:

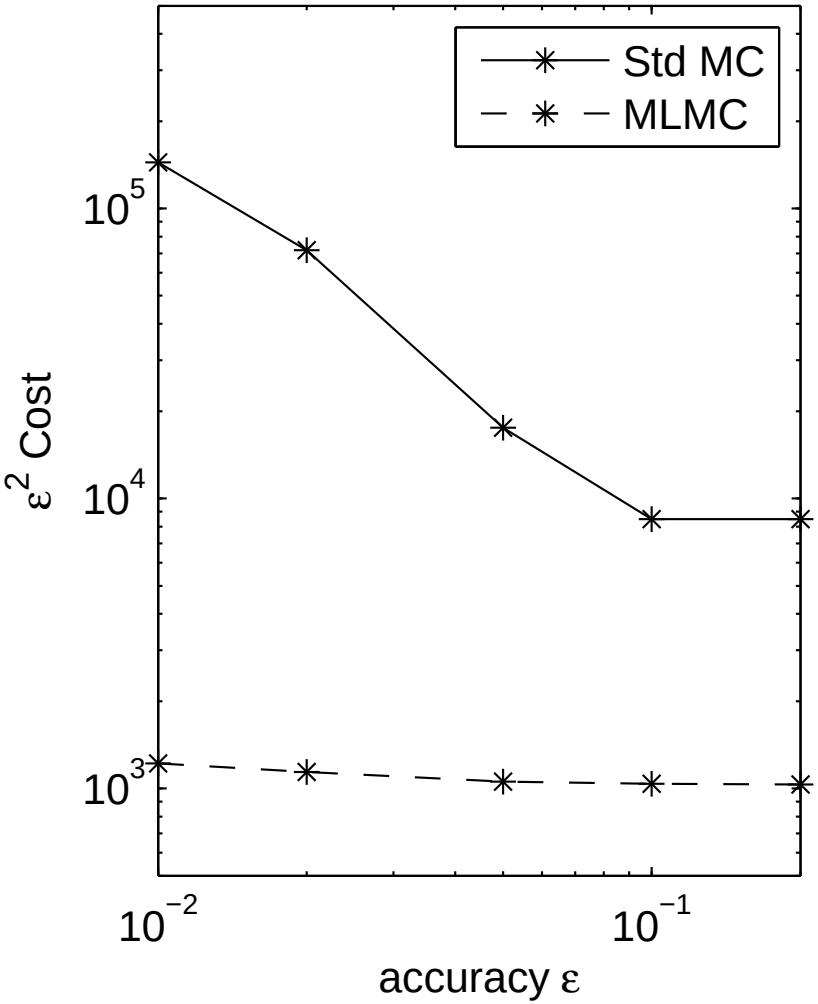




Financial application

Standard call option:





Lookback options

Payoff depends on the minimum attained by the path $S(t)$.

If the numerical approximation uses the minimum of the values at the discrete simulation times

$$\widehat{S}_{min} \equiv \min_j \widehat{S}_j$$

then we have two problems:

- $O(\sqrt{h})$ weak convergence
- $\widehat{S}_{\ell,min} - \widehat{S}_{\ell-1,min} = O(\sqrt{h_\ell})$ which leads to $V_\ell = O(h_\ell)$

Lookback options

To fix this, define a Brownian Bridge interpolation conditional on the endpoints for each timestep, with constant drift and volatility.

For the fine path, standard result for the sampling from the distribution of the minimum of a Brownian Bridge gives

$$\widehat{S}_{min} = \min_j \ \tfrac{1}{2}\left(\widehat{S}_j + \widehat{S}_{j-1} - \sqrt{(\widehat{S}_j - \widehat{S}_{j-1})^2 - 2\,h\,b_j^2 \log U_j}\right)$$

where the U_j are independent $U(0,1)$ random variables.

This gives $O(h)$ weak convergence, but if we do something similar for the coarse path with a different set of $U's$ the variance will still be poor.

Lookback options

Instead, do the following:

- sample from the mid-point of the Brownian Bridge interpolation for the coarse timestep, using the Brownian path information from the fine path – this mid-point value is within $O(h_\ell)$ of the fine path simulation
- sample from the minima of each half of the coarse timestep using the same $U's$ as fine path
- take the minimum of the two minima, and then the minimum over all coarse timesteps.

This leads to an $O(h_\ell)$ difference in the computed minima for the coarse and fine paths, and is valid because the distribution for the coarse path minimum has not been altered.

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

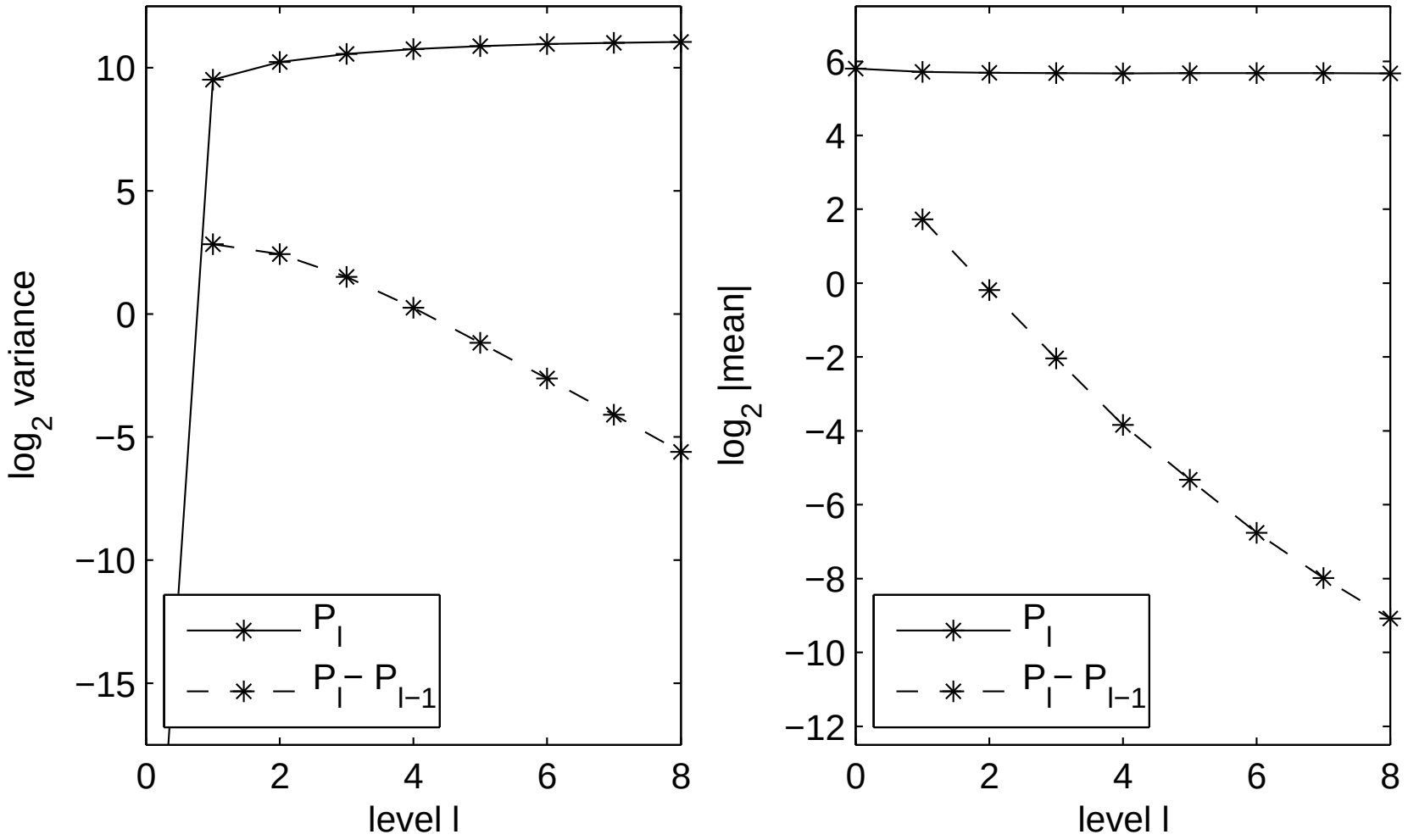
Three fixes:

- 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 $V_\ell = O(h_\ell^{3/2})$.

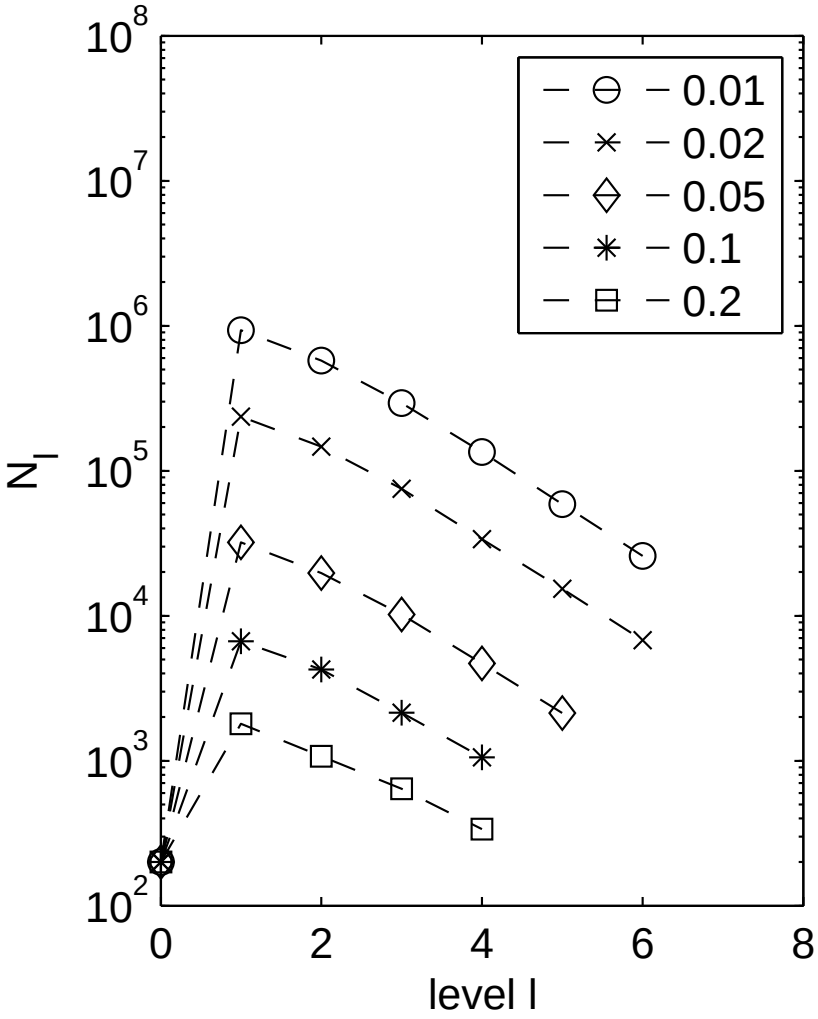
Financial application

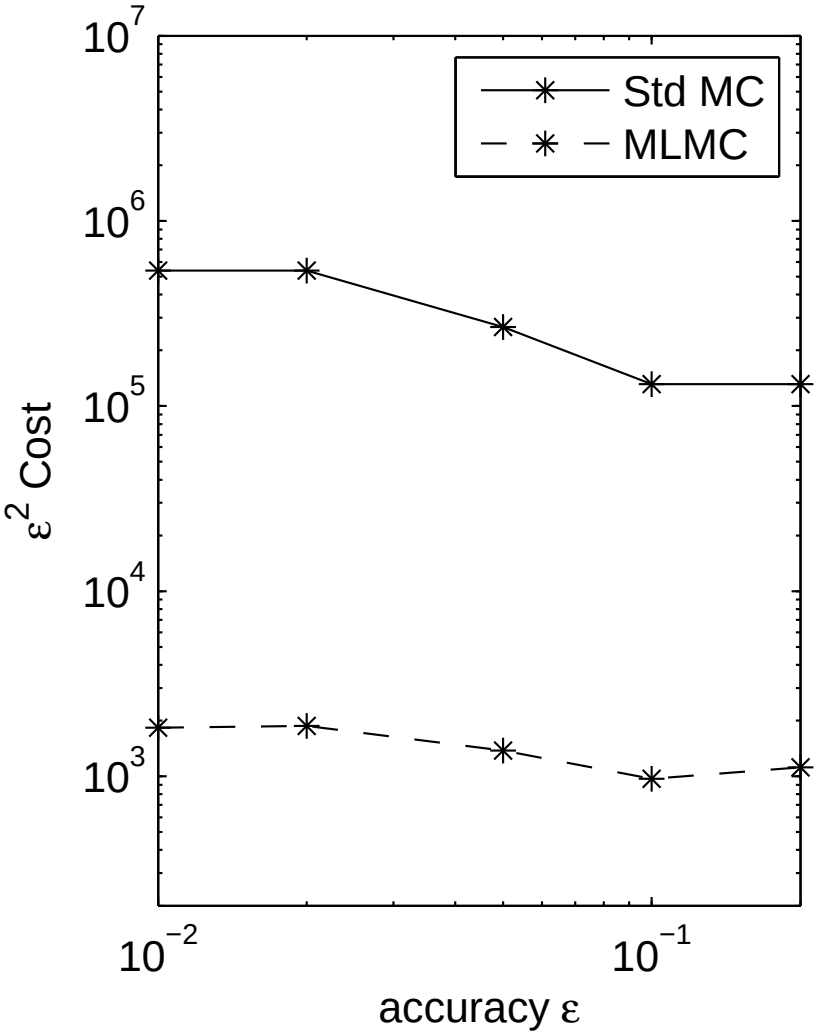
Digital call option:



Financial application

Digital call option:





Numerical Analysis

option	Euler		Milstein	
	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).

Greeks and jump diffusion

Greeks (Burgos, 2011)

- MLMC combines well with pathwise sensitivity analysis for Greeks
- Main concern is reduced regularity of "payoff"
- Techniques are similar to handling digital options

Finite activity rate Merton-style jump diffusion (Xia, 2011)

- if constant rate, no problem — use jump-adapted discretisation and coarse and fine paths jump at the same time
- if path-dependent rate, then it's trickier
  - use jump-adapted discretisation plus thinning (Glasserman & Merener)
  - could lead to fine and coarse paths jumping at different times $\Longrightarrow$ poor variance
  - instead use a change of measure to force jumps to be at the same time

Lévy processes

Infinite activity rate, general Lévy processes
(Dereich 2010; Marxen 2010; Dereich & Heidenreich 2011)

- on level ℓ, simulate jumps bigger than δ_ℓ ($\delta_\ell \to 0$ as $\ell \to \infty$)
- either neglect smaller jumps or use a Gaussian approximation
- multilevel problem: discrepancy in treatment of jumps which are bigger than δ_ℓ but smaller than $\delta_{\ell-1}$

Exact simulation (Cheng Zhu, Filippo Zinzani, Yuan Xia)

- with some popular exponential-Lévy models (variance-gamma, NIG) possible to directly simulate Lévy increments over fine timesteps
- sum them pairwise to get corresponding increments for coarse path
- very helpful for path-dependent options (Asian, lookback, barrier)

Heston stochastic volatility

Glasserman & Kim (2011) developed a series expansion for sampling from the integrated variance:

$$\left(\int_0^T V_s \, ds \;\middle|\; V_0 = v_0, V_t = v_t \right) \stackrel{d}{=} \sum_{n=1}^{\infty} x_n + \sum_{n=1}^{\infty} y_n + \sum_{n=1}^{\infty} z_n$$

where x_n, y_n, z_n are independent random variables.

Multilevel possibility:

- truncate series at $K_\ell \quad (K_\ell \to \infty \text{ as } \ell \to \infty)$
- should help for European options as well as path-dependent options

American options

Belomestny & Schoenmakers (2011) developed a multilevel implementation of upper bound dual pricing

- based on nested simulation algorithm of Andersen and Broadie (2004)
- requires sub-sampling at each timestep to estimate a conditional expectation (the continuation value)
- multilevel treatment uses a different number of sub-samples M_ℓ on each level $\quad (M_\ell \to \infty \text{ as } \ell \to \infty)$

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

$$\widehat{P}_\ell(\omega) - \widehat{P}(\omega) \approx c(\omega)\, 2^{-2\ell}$$

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

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

Elliptic SPDE

Elliptic PDE with random coefficient $k(\mathbf{x}, \omega)$:

$$-\nabla.\,(k(\mathbf{x}, \omega)\nabla p(\mathbf{x}, \omega)) = 0, \qquad \mathbf{x} \in D,$$

Model k as a lognormal random field, i.e. $\log k$ is a Gaussian field with mean 0 and covariance function

$$R(\mathbf{x}, \mathbf{y}) = \sigma^2 \exp\Big(- \|\mathbf{x}-\mathbf{y}\|_1/\lambda\Big)$$

Samples of $\log k$ are provided by a Karhunen-Loève expansion:

$$\log k(\mathbf{x}, \omega) = \sum_{n=0}^{\infty} \sqrt{\theta_n}\ \xi_n(\omega)\ f_n(\mathbf{x}),$$

where ξ_n are iid unit Normal random variables.

Elliptic SPDE

In multilevel treatment:

- different spatial grid resolution on each level
- truncate KL-expansion at different cutoffs K_ℓ

$$\log k_\ell(\mathbf{x},\omega) = \sum_{n=0}^{K_\ell} \sqrt{\theta_n}\ \xi_n(\omega)\ f_n(\mathbf{x}),$$

- (more efficient ways of generating $\log k_\ell$ use technique known as *circulant embedding*)

Iterative convergence

Most PDE solvers involve iterative solvers.

So far, have implicitly assumed we are converging the solution until the remaining error is negligible.

Alternatively, different levels in the multilevel formulation could use different numbers of iterations, or different convergence criteria.

No-one has tried this yet (as far as I know) – point here is that MLMC approach is very general and flexible, just need some hierarchy of approximations, with cost and accuracy increasing together

Other MLMC Applications

- 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 (Glynn, Rhee)
  - 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) 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 $\displaystyle\sum_{\ell=0}^{\infty} N_\ell = N, \qquad \mathbb{E}[N_\ell] = p_\ell N.$

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

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

MIMC

MLMC can be thought of as having a 1D set of levels, giving

$$\mathbb{E}[P] = \sum_{\ell \geq 0} \mathbb{E}[\Delta P_\ell].$$

where $\Delta P_\ell \equiv P_\ell - P_{\ell-1}$ with $P_{-1} \equiv 0$.

Combining MLMC with ideas from "sparse grids", MIMC extends this so that $\Delta_d P_{\boldsymbol{\ell}} \equiv P_{\boldsymbol{\ell}} - P_{\boldsymbol{\ell}-\mathbf{e}_d}$ where $\mathbf{e}_d$ is the unit vector in direction d. They then define the cross-difference

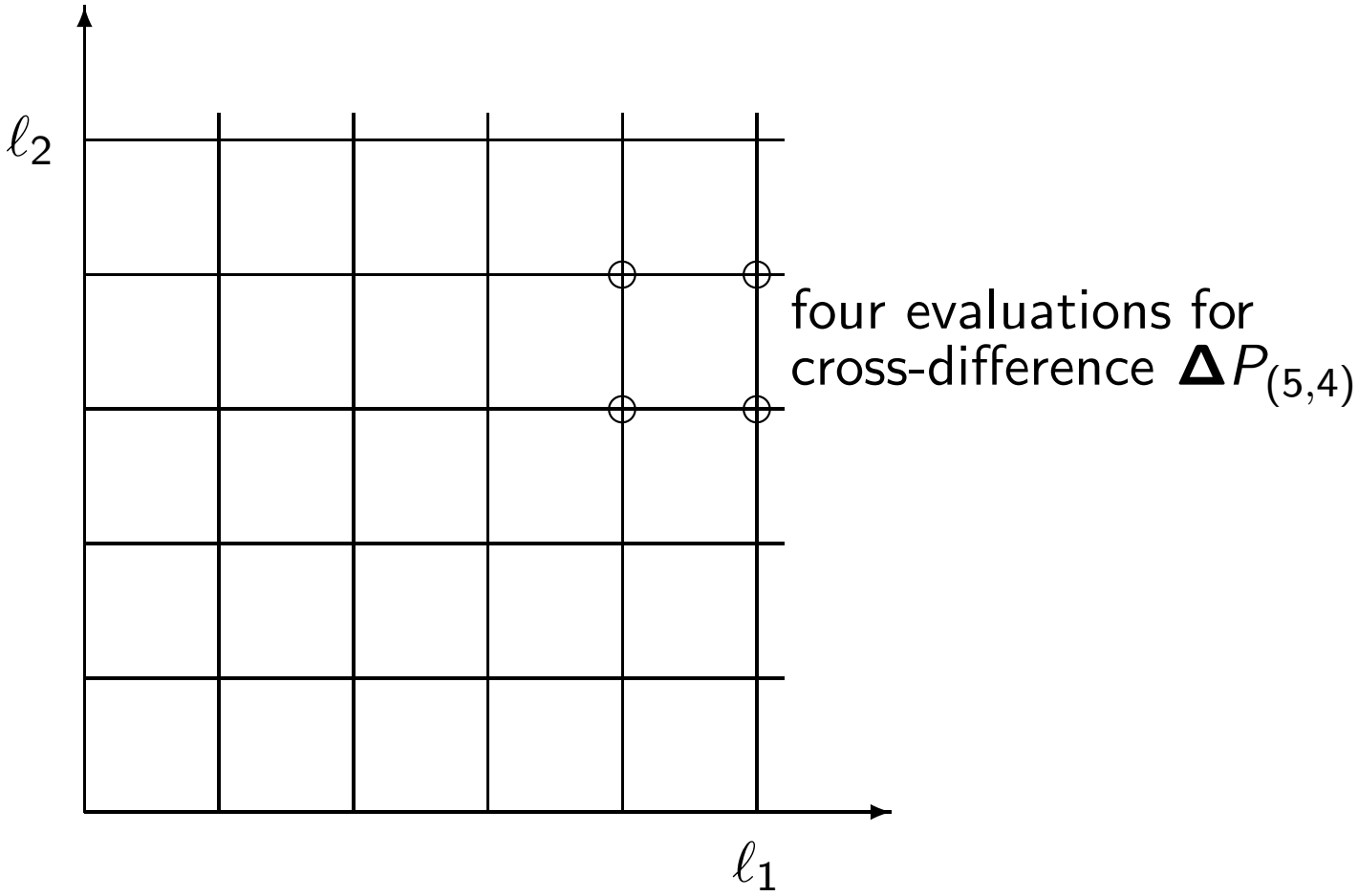
$$\boldsymbol{\Delta} P_{\boldsymbol{\ell}} \equiv \left(\prod_{d=1}^{D} \Delta_d \right) P_{\boldsymbol{\ell}}$$

and the telescoping sum becomes

$$\mathbb{E}[P] = \sum_{\boldsymbol{\ell} \geq 0} \mathbb{E}[\boldsymbol{\Delta} P_{\boldsymbol{\ell}}].$$

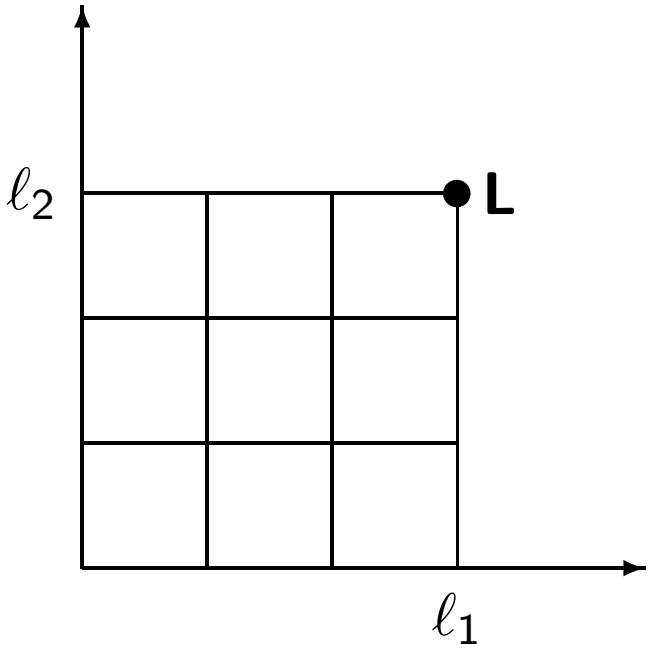
MIMC

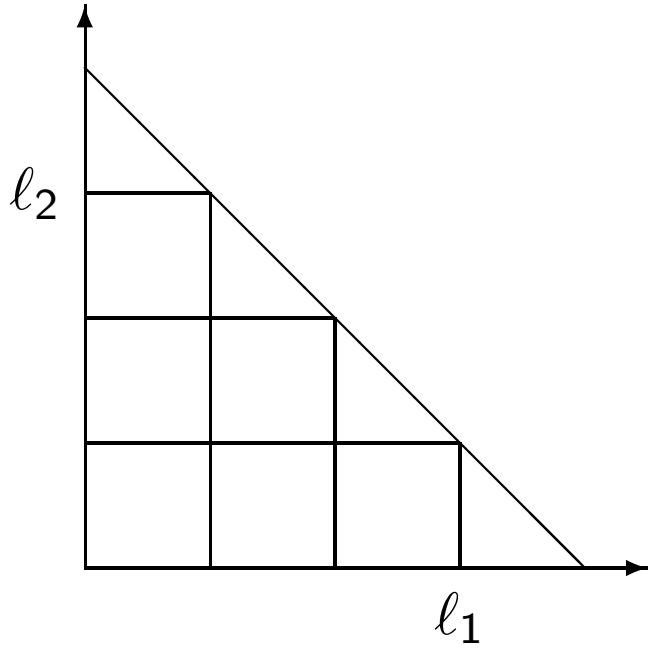
2D example with level indices (ℓ_1, ℓ_2):



MIMC

As with sparse grids, it is most efficient to retain a "triangular" set of differences.





This is all proved in theory which is similar to MLMC theory, but inevitably more complex.

MIMC

When does this help?

In d-dimensional PDE applications, as discussed earlier, can have

$$C_\ell \sim 2^{d\ell}, \quad V_\ell \sim 2^{-2\ell}$$

so $\beta < \gamma$ for $d > 2$, which does not give the optimal complexity.

The use of MIMC fixes this, so complexity of multi-dimensional PDE is similar to 1-dimensional PDE.

To do this, the grid refinement in each coordinate direction is one axis of the MIMC treatment – clear connection to sparse grid methods.

MIMC

Another example is nested simulation in which we want to estimate

$$\mathbb{E}\left[f\left(\mathbb{E}[u(W,Z)\,|\,Z\,]\right)\right]$$

where Z is a random variable ("scenario") and W represents a Brownian path and $u(W,Z)$ is a functional arising from an SDE simulation.

A simple Monte Carlo estimate for this has the form

$$N^{-1}\sum_{n-1}^{N} f\left(M^{-1}\sum_{m=1}^{M} u(W^{(m,n)},Z^{(n)})\right)$$

In this case:

- dim 1 corresponds to timestepping approximation: $h_\ell = 2^{-\ell_1}\,h_0$
- dim 2 corresponds to number of inner samples: $M_\ell = 2^{\ell_2}$

Initial analysis suggests this should work well.

MLQMC

To further improve the multilevel complexity, can use randomised QMC in place of MC.

G & Waterhouse (2008-9) used rank-1 lattice rules for scalar SDE applications

- far fewer samples required on coarsest levels
- almost no difference on finest levels
- overall, big savings when using Milstein discretisation
- in best case complexity was approximately $O(\varepsilon^{-1.5})$

Recently, Dick, Kuo, Schwab, Sloan have developed theory for MLQMC for PDEs.

MLQMC

Numerical algorithm (G, Waterhouse):

1. start with $L=0$
2. get an initial estimate for V_L using 32 random offsets and $N_L = 1$
3. while $\sum_{\ell=0}^{L} V_\ell > \varepsilon^2/2$, try to maximise variance reduction per unit cost by doubling N_ℓ on the level with largest $V_\ell \,/\, (C_\ell \; N_\ell)$
4. if $L<2$ or the bias estimate is greater than $\varepsilon/\sqrt{2}$, set $L := L+1$ and go back to step 2

New Feynman-Kac project (with Francisco Bernal)

Suppose that X_t satisfies the SDE

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

in bounded domain D, where W_t is an uncorrelated Brownian motion, and let

$$u(x_0, t_0) = \mathbb{E}\left[\int_{t_0}^{\tau} E(t,s)\, f(X_s, s)\,\mathrm{d}s + E(t_0, \tau)\, g(X_\tau, \tau) \mid X_{t_0} = x_0\right]$$

where τ is the first time at which X_t leaves D and

$$E(t_0, t_1) = \exp\left(-\int_{t_0}^{t_1} V(X_t, t)\,\mathrm{d}t\right).$$

Feynman-Kac theorem

If $f(x,t), g(x,t), V(x,t), a(x,t), b(x,t)$ are all Lipschitz continuous, then the Feynman-Kac theorem states that $u(x,t)$ satisfies the PDE

$$\frac{\partial u}{\partial t} + \sum_j a_j \frac{\partial u}{\partial x_j} + \tfrac{1}{2} \sum_{j,k,l} b_{j,k} b_{k,l} \frac{\partial^2 u}{\partial x_j \partial x_l} - V(x,t)\, u(x,t) + f(x,t) = 0$$

in domain D, subject to $u(x,t) = g(x,t)$, on the boundary ∂D.

Hence, can estimate $u(x,t)$ solution to a high-dimensional PDE at particular points (x,t), by Monte Carlo simulation of SDE.

This also extends to linear and nonlinear functionals of the PDE solution.

Numerical approximation

Let $\widehat{X}_t$ be the piecewise-constant Euler-Maruyama approximation and define

$$\widehat{E}(t_0, t_1) = \exp\left(-\int_{t_0}^{t_1} V(\widehat{X}_t, t)\,\mathrm{d}t\right),$$

and let

$$\widehat{u}(x_0, t_0) = \mathbb{E}\left[\int_{t_0}^{\widehat{\tau}} \widehat{E}(t, s)\, f(\widehat{X}_s, s)\,\mathrm{d}s + \widehat{E}(t_0, \widehat{\tau})\, g(\widehat{X}_{\widehat{\tau}}, \widehat{\tau}) \mid \widehat{X}_{t_0} = x_0\right].$$

with the Euler-Maruyama discretisation beginning at time t_0, with $\widehat{\tau}$ being the exit time.

Numerical approximation

The Euler-Maruyama method has strong accuracy $O(h^{1/2})$, and the natural definition of $\widehat{\tau}$ gives an $O(h^{1/2})$ weak error too.

For standard Monte Carlo method, ε RMS accuracy needs $O(\varepsilon^{-2})$ paths, each with $h = O(\varepsilon^2)$, so total cost is $O(\varepsilon^{-4})$

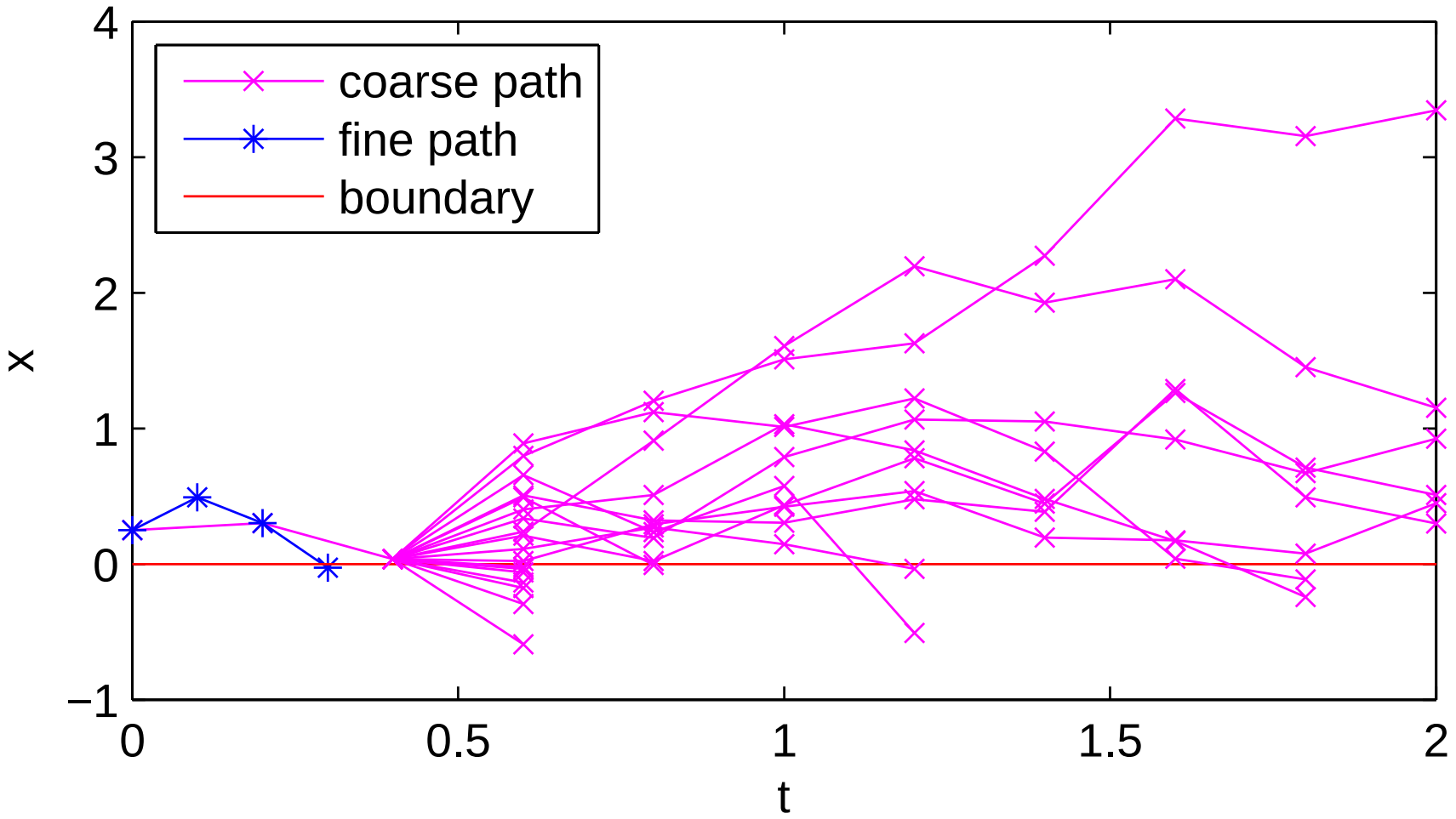
Gobet reduced this to $O(\varepsilon^{-3})$ by shifting the boundary by $O(h^{1/2})$ to improve the weak accuracy to $O(h)$.

Alternatively, Higham *et al* use MLMC to achieve $O(\varepsilon^{-3}|\log\varepsilon|^3)$ complexity without shifting the boundary.

Our aim is to improve it to $O(\varepsilon^{-2}|\log\varepsilon|^3)$.

MLMC challenge

When coarse or fine path exits the domain, the other is within $O(h^{1/2})$. However, there is a $O(h^{1/2})$ probability that it will not exit the domain until much later $\Longrightarrow V_\ell = O(h^{1/2})$.



MLMC challenge

How can we do better?

Similar to previous work on digital options, split second path into multiple copies, and average their outputs to approximate the conditional expectation.

$O(h^{1/2})$ expected time to exit for second path, so can afford to use $O(h^{-1/2})$ copies of second path.

This gives an approximation to the conditional expectation resulting in $\widehat{P}_\ell - \widehat{P}_{\ell-1} \approx O(h^{1/2})$, so $V_\ell \approx O(h)$.

Numerical results confirm this, and the numerical analysis is almost finished.

Numerical results

The test case comes from Gobet & Menozzi (2009)

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$$

in the domain $\|x\| \le 2,\ 0 \le t \le 1$ with

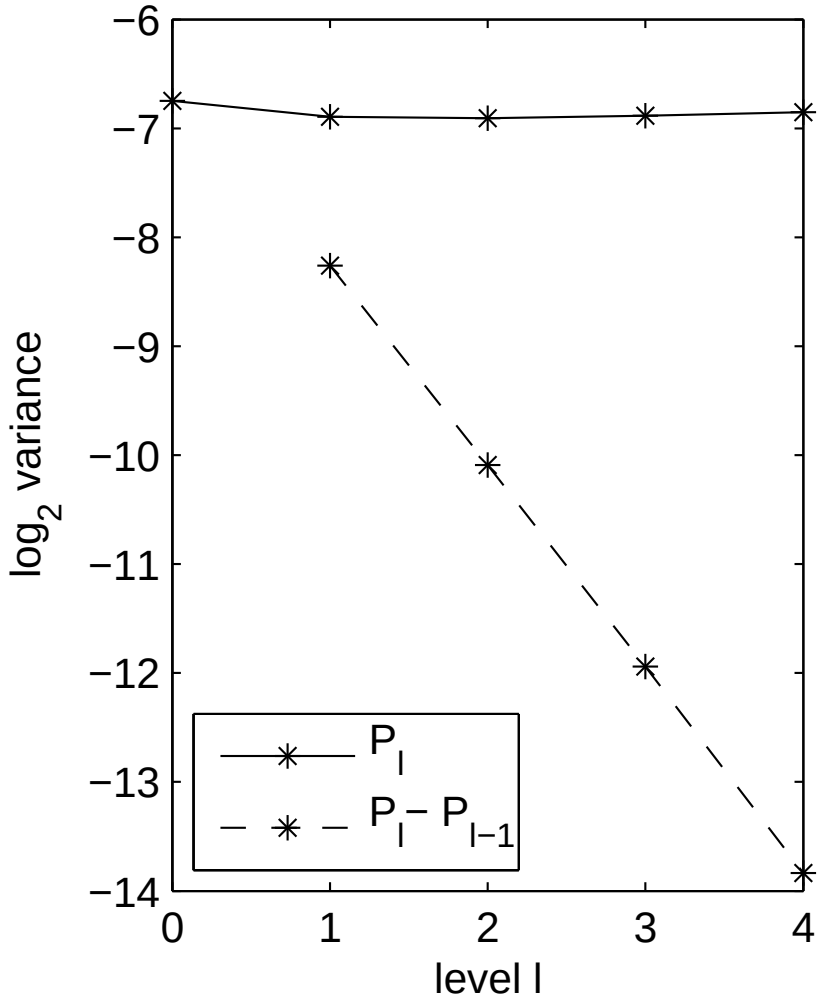
$$b(x) = \begin{pmatrix} x_2 \\ x_3 \\ x_1 \end{pmatrix}, \quad \sigma = \begin{pmatrix} (1+|x_3|)^{\frac{1}{2}} & 0 & 0 \\ \frac{1}{2}(1+|x_1|)^{\frac{1}{2}} & (\frac{3}{4})^{\frac{1}{2}}(1+|x_1|)^{\frac{1}{2}} & 0 \\ 0 & \frac{1}{2}(1+|x_2|)^{\frac{1}{2}} & (\frac{3}{4})^{\frac{1}{2}}(1+|x_2|)^{\frac{1}{2}} \end{pmatrix}$$

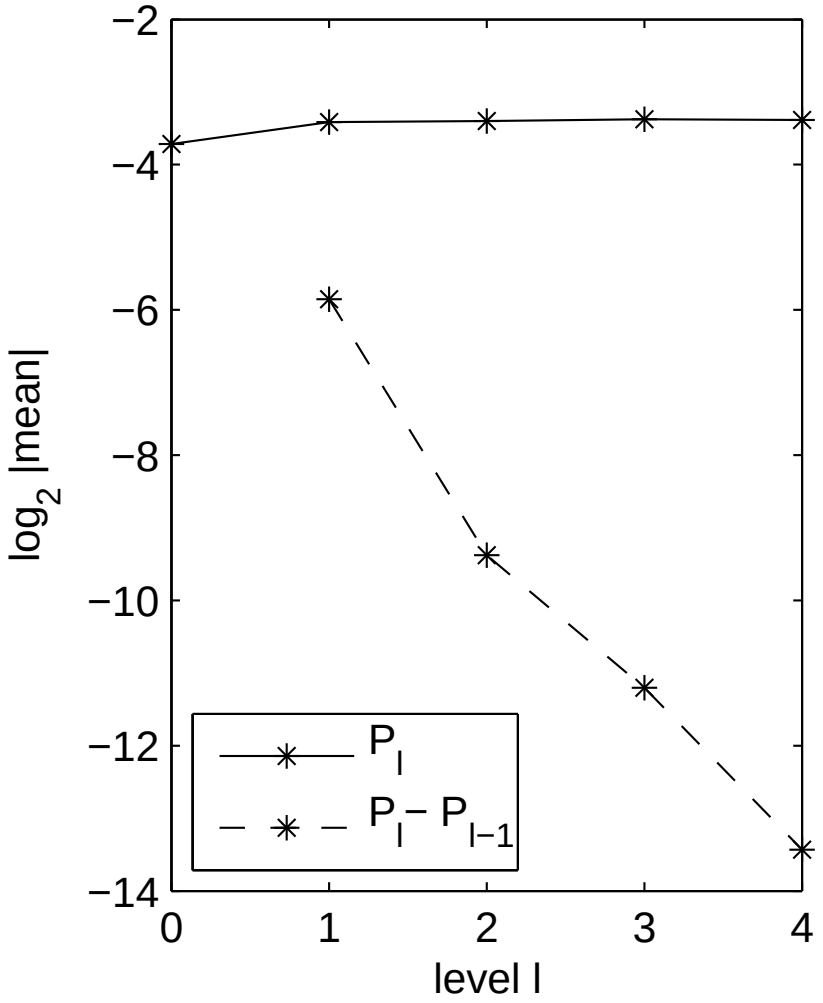
$V(x,t) \equiv 0$, and $f(x,t), g(x,t)$ are chosen so that the PDE solution is $u(x,t) = x_1 x_2 x_3$.

$X_0 = (0.56, 0.52, 0.33)^T$, so we are estimating $u(X_0, 0)$.

Timestep comes down by factor 4 on each level – better than factor 2 when $V_\ell = O(h_\ell)$. Gobet-Menozzi boundary shift used on each level.

Numerical results





$$V_\ell = O(h_\ell)$$

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

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