Multilevel Monte Carlo method

Mike Giles

mike.giles@maths.ox.ac.uk

Oxford-Man Institute of Quantitative Finance

Mathematical Institute, University of Oxford

Sylvestre Burgos, Lukas Szpruch, Yuan Xia (Oxford)

Des Higham, Xuerong Mao (Strathclyde)

Kristian Debrabant, Andreas Rößler (Darmstadt)

Rob Scheichl, Aretha Teckentrup (Bath), Andrew Cliffe (Nottingham)

Caltech Seminar, May 21, 2013

Outline

- standard Monte Carlo simulation
- multilevel Monte Carlo simulation
- financial SDE applications
- elliptic SPDE application
- conclusions

Monte Carlo simulation

In many applications want to estimate $\mathbb{E}[P(\omega)]$ where $\omega \in \Omega$ is an infinite-dimensional random variable.

- computational finance:
  - ω represents W_t, the Brownian motion in an SDE
  - P is the financial payoff function

- simulation of oil reservoirs & nuclear waste repositories:
  - ω represents $k(x)$, the permeability in an elliptic SPDE

$$-\nabla \cdot \big(k(x)\,\nabla p\big) = 0$$

  - P might be the flux of oil or contaminants across some boundary

Monte Carlo simulation

In MC simulation we estimate the expectation using

$$\widehat{Y} \;=\; N^{-1} \sum_{n=1}^{N} \widehat{P}(\omega^{(n)})$$

where $\omega^{(n)}$ are N independent samples

Note there are two sources of error here:

- *sampling error* due to the finite number of samples
- *bias* because $\widehat{P}(\omega)$ is an approximation to $P(\omega)$ due to
  - discretisation error (finite timesteps, finite grid size)
  - finite dimensional approximation to ω

Monte Carlo simulation

The mean square error is

$$\begin{aligned}
\mathbb{E}\left[\left(\widehat{Y}-\mathbb{E}[P]\right)^2\right] &= \mathbb{E}\left[\left(\widehat{Y}-\mathbb{E}[\widehat{Y}]\;+\;\mathbb{E}[\widehat{Y}]-\mathbb{E}[P]\right)^2\right] \\
&= \mathbb{E}\left[(\widehat{Y}-\mathbb{E}[\widehat{Y}])^2\right]+\left(\mathbb{E}[\widehat{Y}]-\mathbb{E}[P]\right)^2 \\
&= \mathbb{V}[\widehat{Y}]+\left(\mathbb{E}[\widehat{Y}]-\mathbb{E}[P]\right)^2 \\
&= N^{-1}\mathbb{V}[\widehat{P}]+\left(\mathbb{E}[\widehat{P}]-\mathbb{E}[P]\right)^2
\end{aligned}$$

- first term is due to sampling error
- second term is due to bias

Monte Carlo simulation

To achieve RMS accuracy of ε requires:

- $N = O(\varepsilon^{-2})$
- bias $= O(\varepsilon)$

The bias is due to the accuracy of the numerical approximation – using smaller timesteps or a finer grid reduces the bias, but increases the computational cost C.

If the cost per sample to achieve an $O(\varepsilon)$ bias is $O(\varepsilon^{-1/\alpha})$ then the total cost is $O(\varepsilon^{-2-1/\alpha})$.

The aim with multilevel is to reduce this to $O(\varepsilon^{-2})$, corresponding to an $O(1)$ cost per sample, on average.

Multilevel Monte Carlo

How can this be achieved?

Use the same philosophy as multigrid for iterative solution of large linear/nonlinear systems of equations:

- fine grid accuracy at coarse grid cost
- geometric sequence of grids

However, there's no iteration in Monte Carlo simulation, so in detail the method is quite different from traditional multigrid.

(Achi Brandt did some work in statistical physics which has some strong similarities to this work.)

Multilevel Monte Carlo

Consider Monte Carlo simulations with different levels of refinement, $\ell = 0, 1, \ldots, L$, with level L being the finest.

If $\widehat{P}_\ell$ is the approximation of P on level ℓ, then

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

Idea is to independently estimate each of the terms on the r.h.s., in a way which minimises the overall variance for a fixed computational cost.

Finest level is still the same, but will use very few samples at that level.

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 stochastic sample $\omega^{(n)}$ for both levels

Variance is $N_\ell^{-1} V_\ell$ where $V_\ell = \mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$

Key point: V_ℓ gets progressively smaller as ℓ increases because $\widehat{P}_\ell, \widehat{P}_{\ell-1}$ both accurately approximate P for same ω

Multilevel Monte Carlo

If C_ℓ is cost of one sample on level ℓ, the variance of the combined estimator is $\sum_{\ell=0}^{L} N_\ell^{-1} V_\ell$ and its computational cost is $\sum_{\ell=0}^{L} N_\ell \, C_\ell$ so the variance is minimised for fixed cost by choosing $N_\ell \propto \sqrt{V_\ell / C_\ell}$, and then the cost on level ℓ is proportional to $N_\ell \, C_\ell \propto \sqrt{V_\ell \, C_\ell}$

To make RMS error ε

- choose constant of proportionality so variance is $\frac{1}{2}\,\varepsilon^2$
- choose L so that $\Big(\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P]\Big)^2 < \frac{1}{2}\,\varepsilon^2$

MLMC Theorem

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) $C_\ell \leq c_3\, 2^{\gamma\,\ell}$

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 a computational cost C with bound

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

Papers

Initial motivation was SDE applications in finance:

- first paper (*Operations Research, 2006 – 2008*) applied idea to SDE path simulation, and proved slightly less general form of the theorem
- second paper (*MCQMC 2006*) improved multilevel variance convergence using Milstein discretisation
- third paper with Higham & Mao (*Finance and Stochastics, 2009*) performed numerical analysis of discretisation in first paper
- new paper with Debrabant and Rößler analyses discretisation in second paper
- new paper with Szpruch uses antithetic treatment to avoid computing Lévy areas for Milstein discretisation

Other work

- Xia – jump-diffusion models
- Burgos – Greeks (sensitivities)
- Hoel, von Schwerin, Szepessy, Tempone – adaptive discretisations
- Dereich, Heidenreich – Lévy processes
- Hickernell, Müller-Gronbach, Niu, Ritter – complexity analysis
- Belomestny, Schoenmakers – American options

For more see:
`people.maths.ox.ac.uk/gilesm/mlmc_community.html`

SDEs

For the Milstein discretisation of the scalar SDE

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

we have

$$\mathbb{E}[(\widehat{S}_N - S(T))^2] = O(h_\ell^2)$$

and hence for a Lipschitz European payoff

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

The optimal N_ℓ is $O(\varepsilon^{-2}\, h_\ell^{3/2})$ and we obtain an $O(\varepsilon^2)$ MSE for an $O(\varepsilon^{-2})$ computational cost.

Call Option

Geometric Brownian motion:

$$\mathrm{d}S = r\,S\,\mathrm{d}t + \sigma\,S\,\mathrm{d}W, \qquad 0 < t < T,$$

$$T = 1, \quad S(0) = 100, \quad r = 0.05, \quad \sigma = 0.2$$

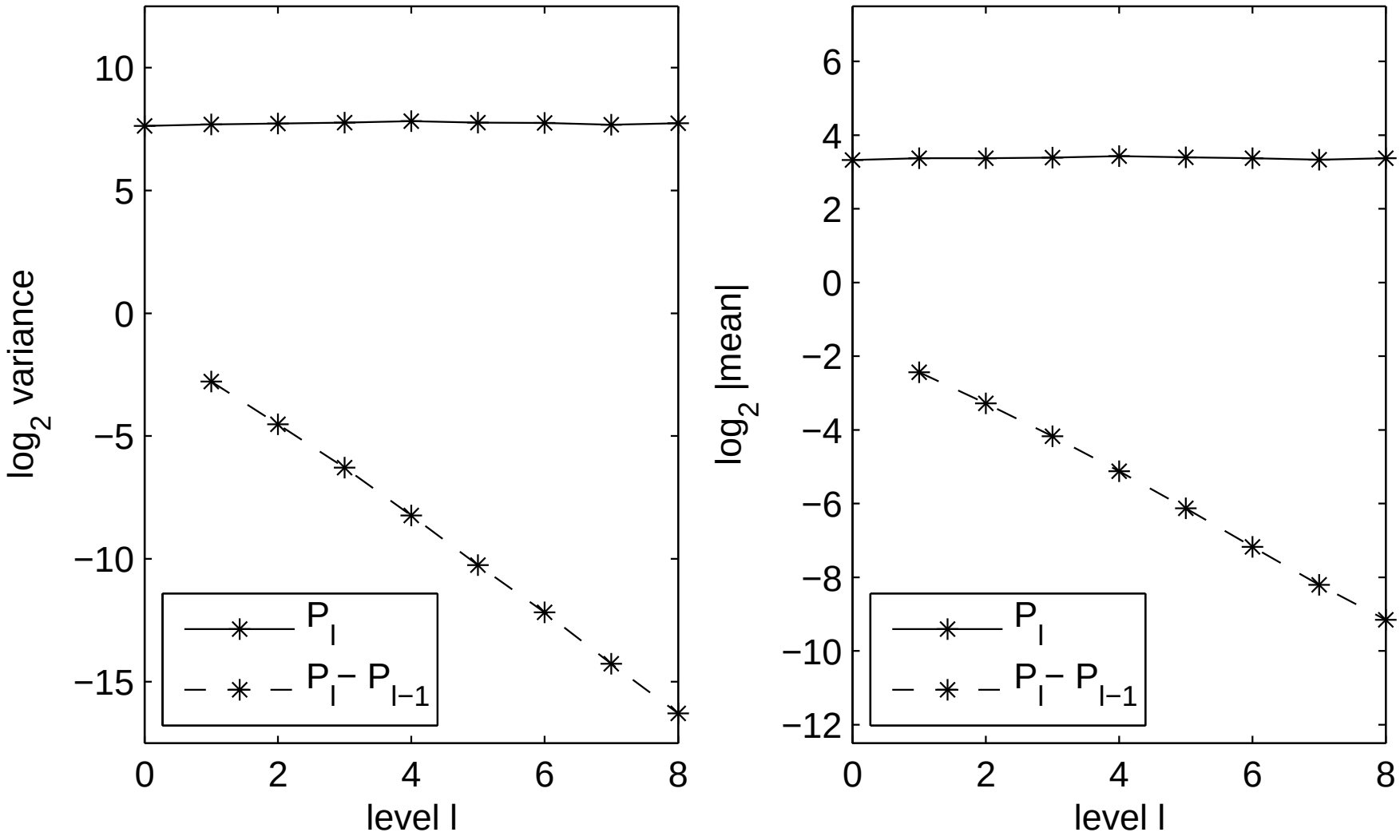
European call option with discounted payoff

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

with $K = 100$.

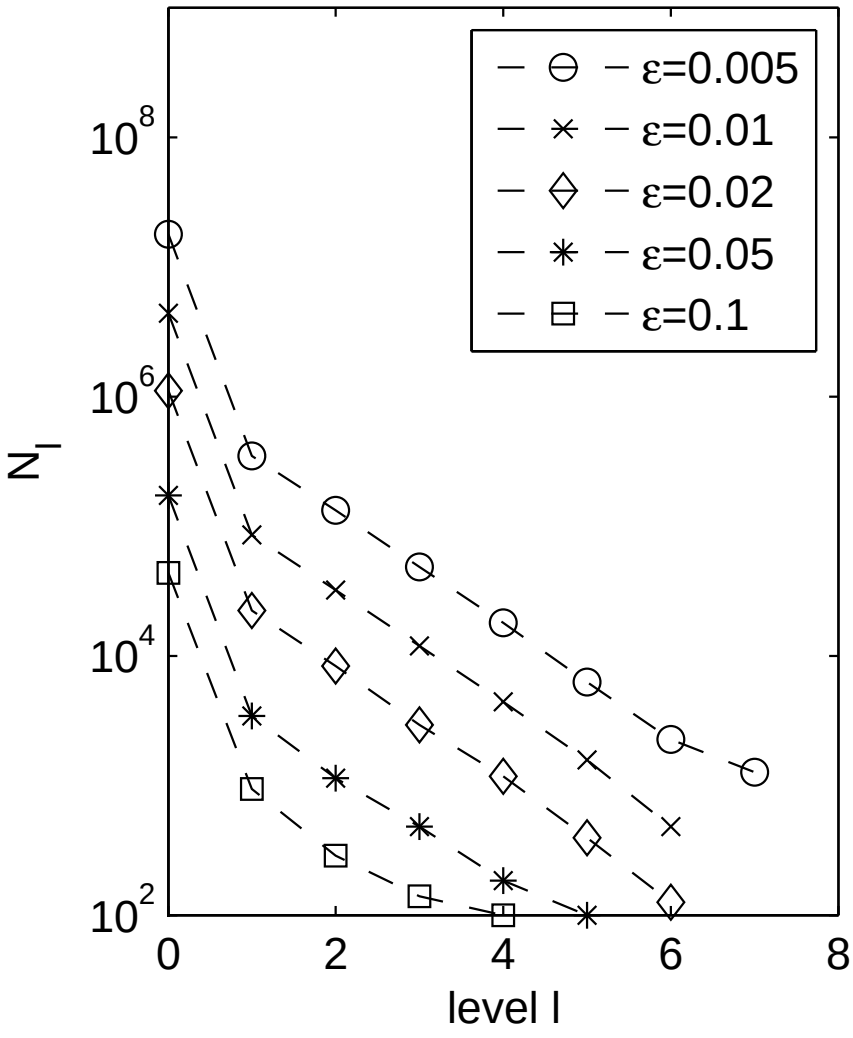
MLMC Results

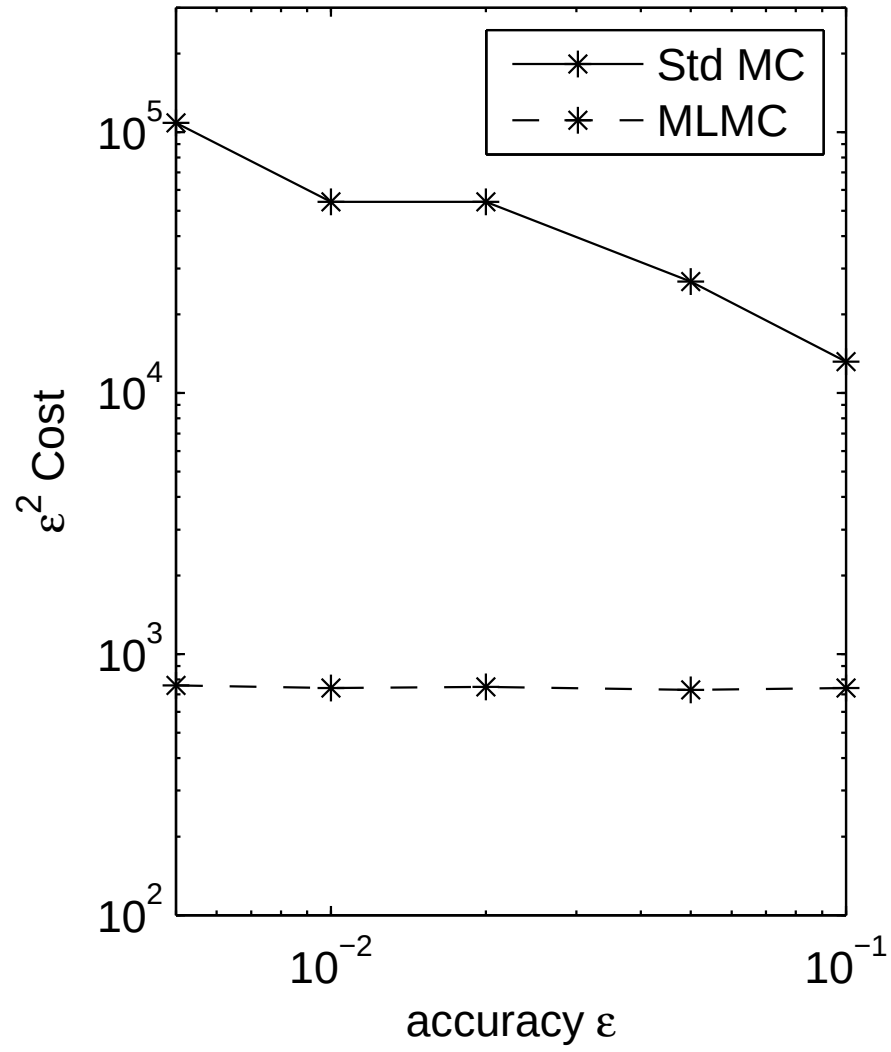
GBM: European call, $\exp(-rT)\ \max(S(T)-K,0)$



MLMC Results

GBM: European call, $\exp(-rT)\ \max(S(T)-K,0)$





Digital Option

What if we don't have the Lipschitz property?

A digital call payoff has the form

$$f(S(T)) = \begin{cases} 1, & S(T) > K \\ 0, & S(T) \leq K \end{cases}$$

Using the Milstein discretisation

- in most cases, fine and coarse paths are on same side of K, so $\widehat{P}_\ell - \widehat{P}_{\ell-1} = 0$
- for $O(h_\ell)$ of the paths, fine and coarse paths end up on different sides of K so $\widehat{P}_\ell - \widehat{P}_{\ell-1} = \pm 1$

Hence $\mathbb{E}[(\widehat{P}_\ell - \widehat{P}_{\ell-1})^2]$ and $\mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ are both $O(h_\ell)$.

Digital Option

Instead, can stop the simulation one timestep before the end and use a conditional expectation for the final value.

$$\widehat{P}_\ell = \mathbb{E}_Z[f(\widehat{S}_N) \,|\, \widehat{S}_{N-1}]$$

where, for the scalar SDE

$$\widehat{S}_N = \widehat{S}_{N-1} + a_{N-1}\, h + b_{N-1}\sqrt{h}\; Z$$

The key is that we know that

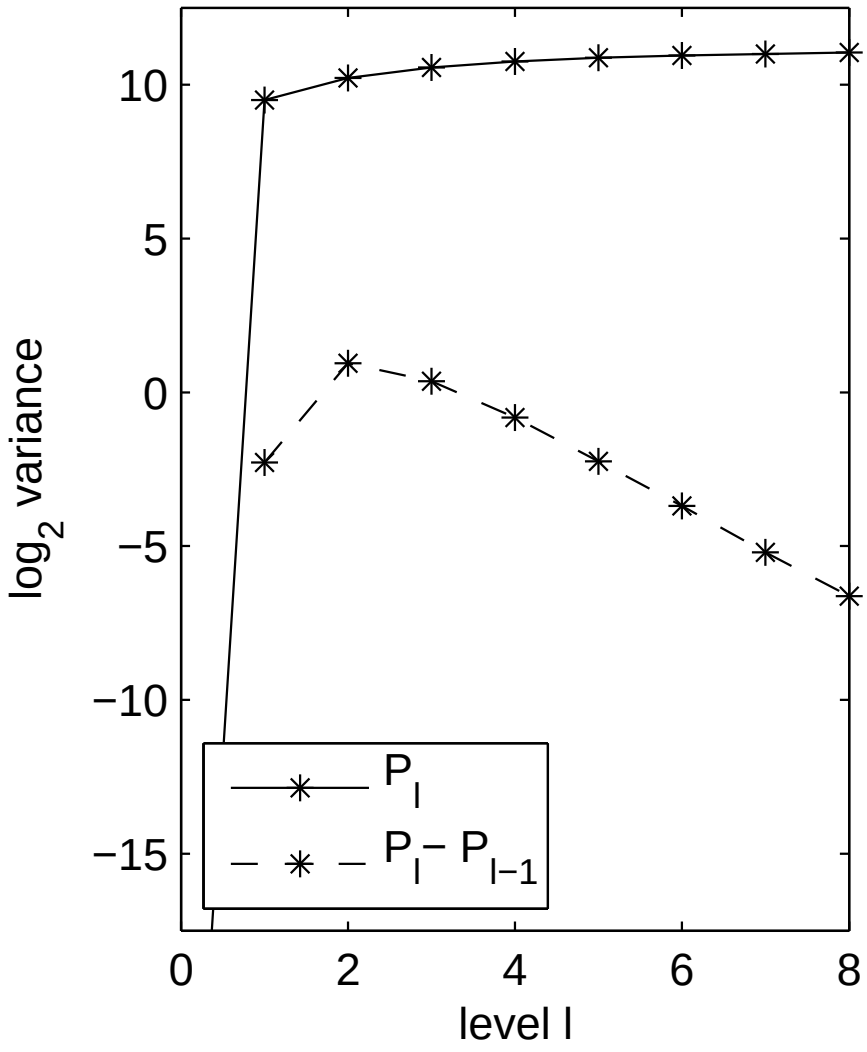
$$\mathbb{E}_Z[f(\widehat{S}_N)|\widehat{S}_{N-1}] = \Phi\left(\frac{\widehat{S}_{N-1} + a_{N-1}h - K}{b_{N-1}\sqrt{h}}\right)$$

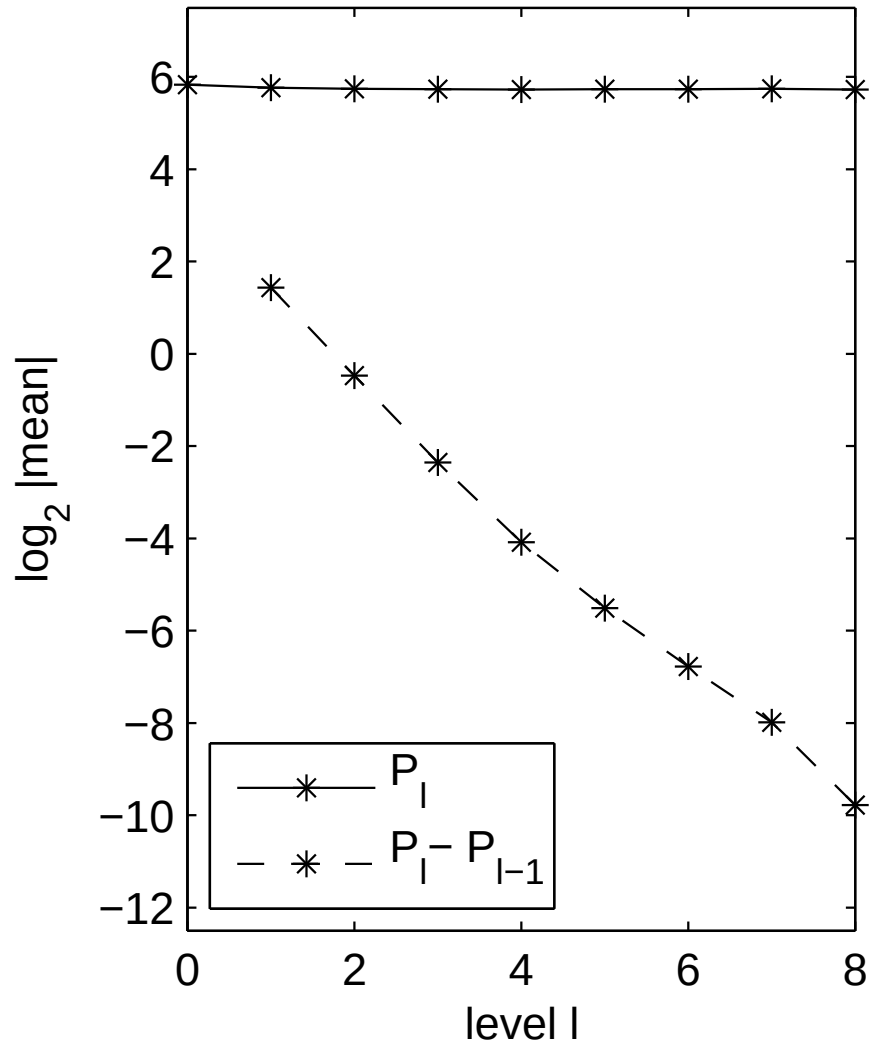
where $\Phi()$ is the cumulative Normal distribution function.

This leads to an $O(h_\ell^{3/2})$ variance.

MLMC Results

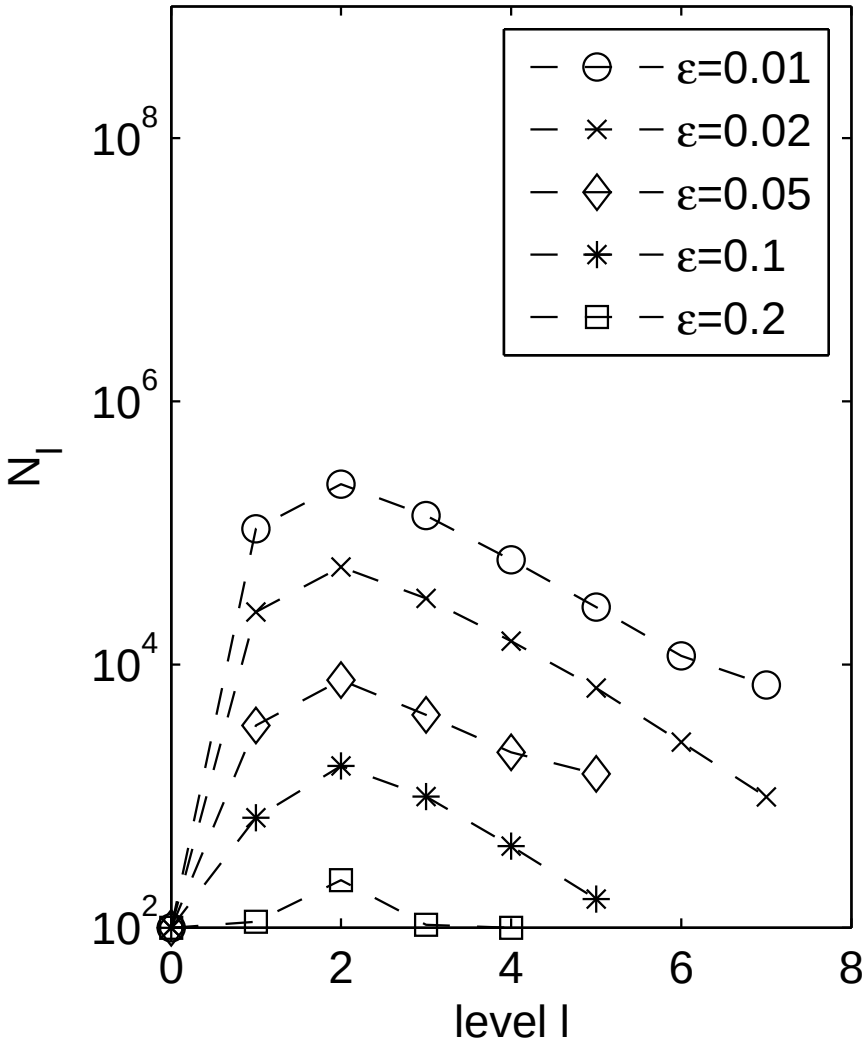
GBM: digital call $K\ \exp(-rT)\,\mathbf{1}\{S(T) > K\}$

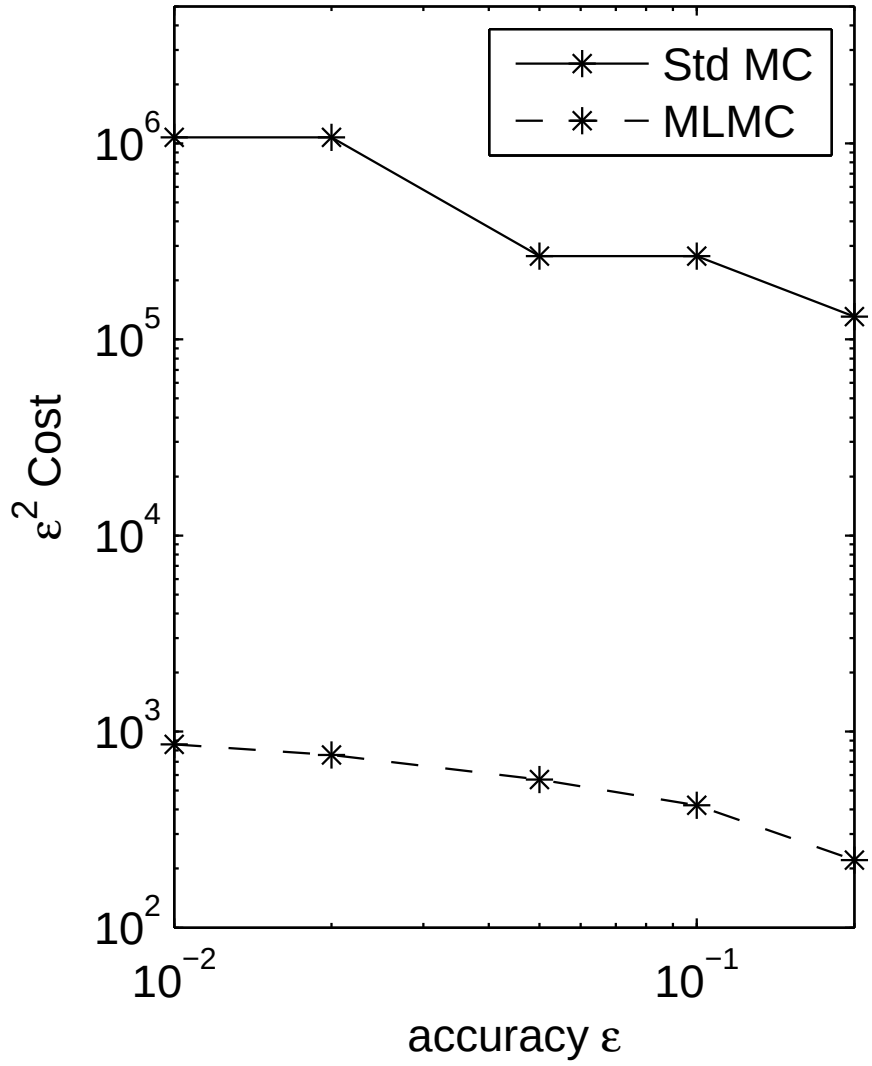




MLMC Results

GBM: digital call $K\ \exp(-rT)\,\mathbf{1}\{S(T) > K\}$





Basket Option

The techniques extend naturally to multivariate cases.

For example, the analytic conditional expectation can be used for a basket option in which the payoff is based on a weighted average of several stocks.

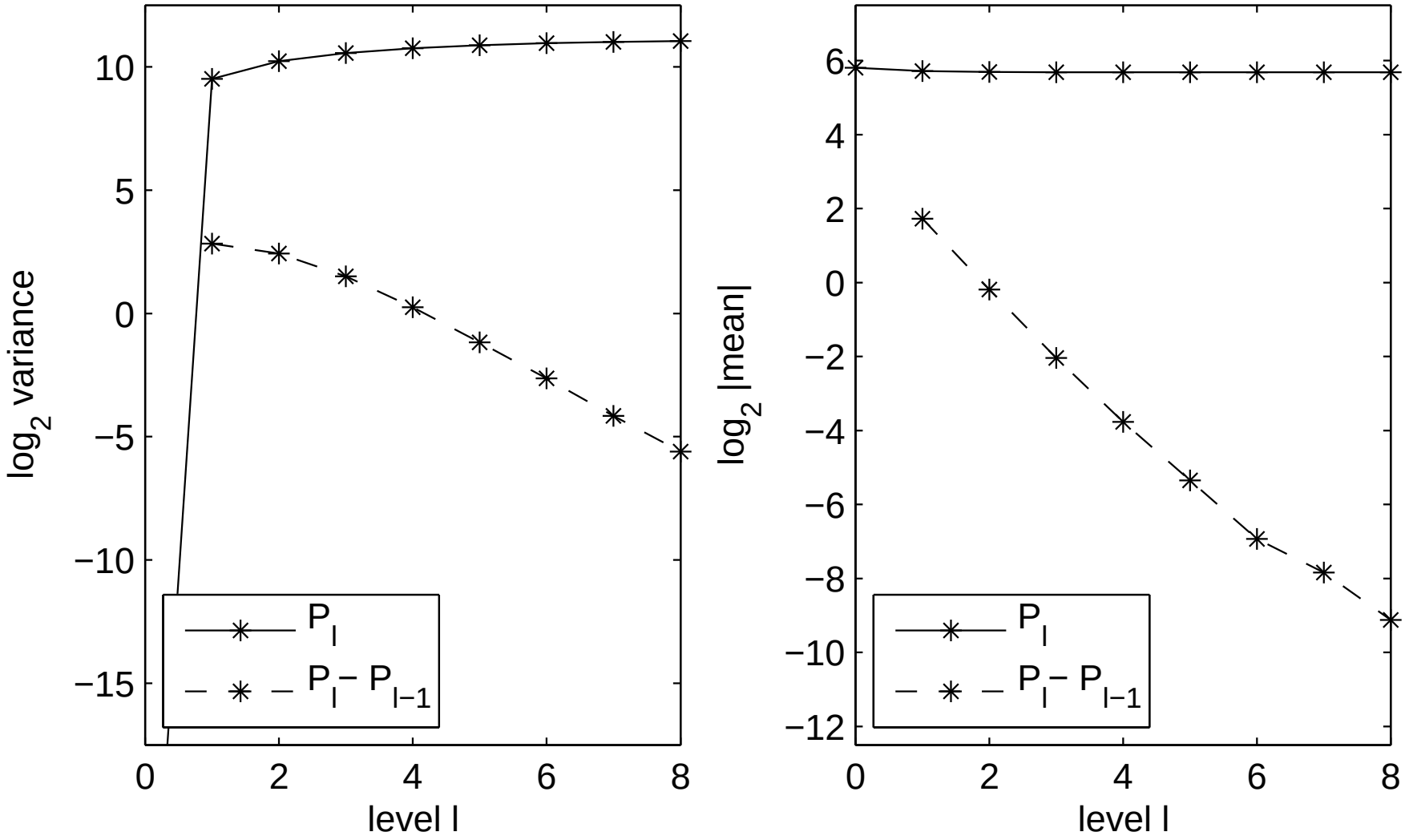
Basket of 5 underlying assets, each GBM with

$$r = 0.05, \quad T = 1, \quad S_i(0) = 100, \quad \sigma = (0.2,\, 0.25,\, 0.3,\, 0.35,\, 0.4),$$

and correlation $\rho = 0.25$ between each of the driving Brownian motions.

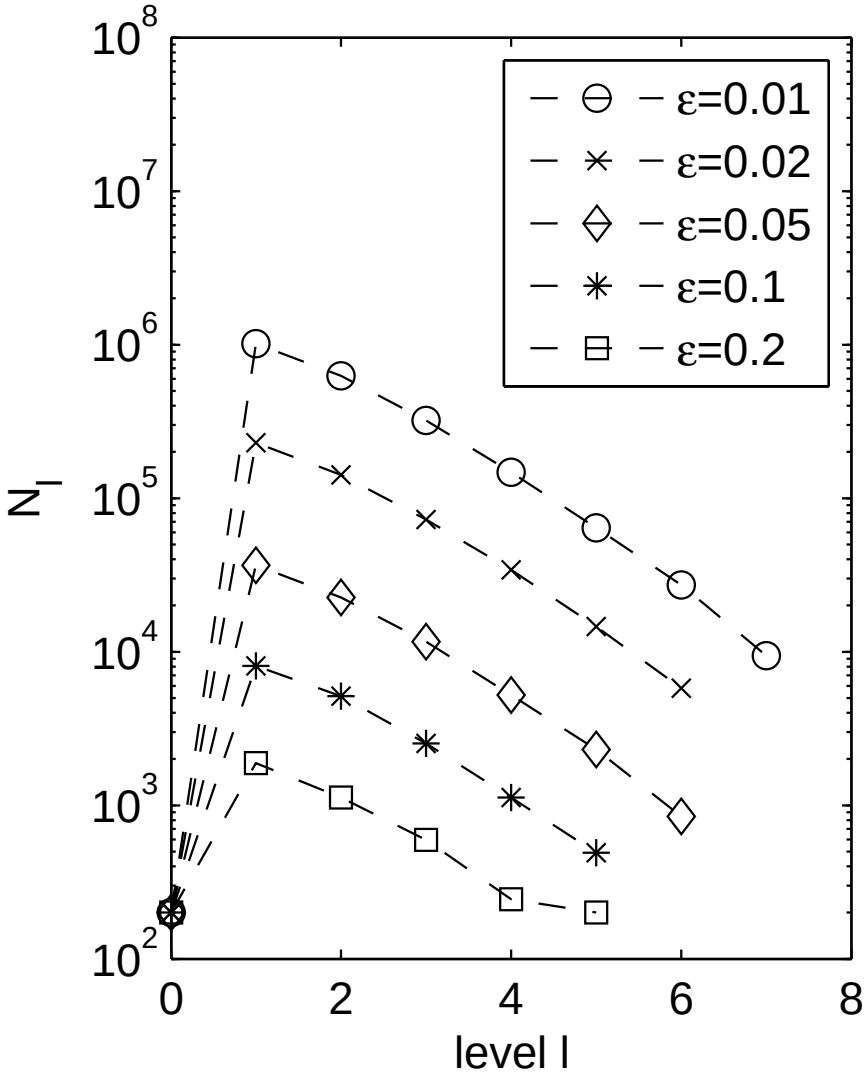
Basket Option

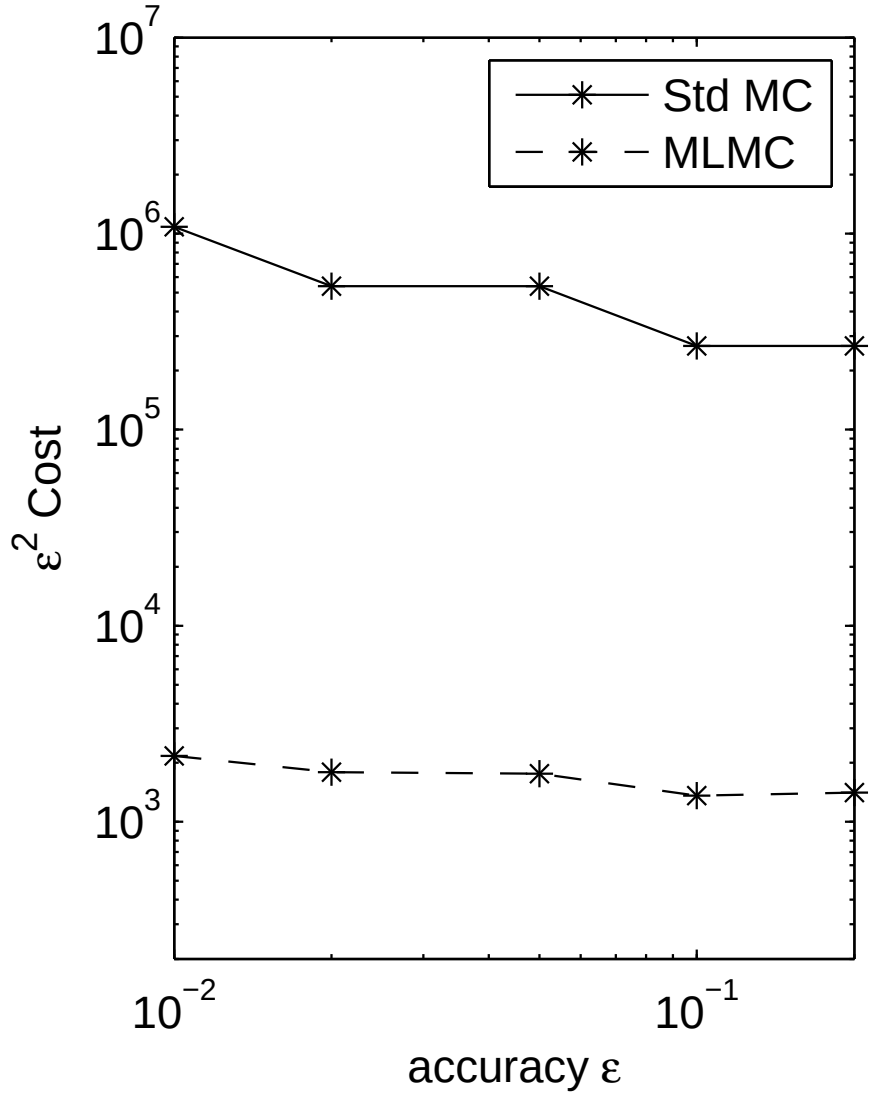
GBM: digital call on basket of 5 assets



Basket Option

GBM: digital call on basket of 5 assets





Elliptic SPDE

We consider the elliptic PDE

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

with **random coefficient** $k(\mathbf{x},\omega)$.

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

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

Numerical experiments use $\sigma = 1$ and

- in 1D, $\lambda = 0.1$ on unit interval $[0,1]$
- in 2D, $\lambda = 0.2$ on unit square $[0,1]^2$

Elliptic SPDE

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 θ_n, f_n are eigenvalues / eigenfunctions of the correlation function:

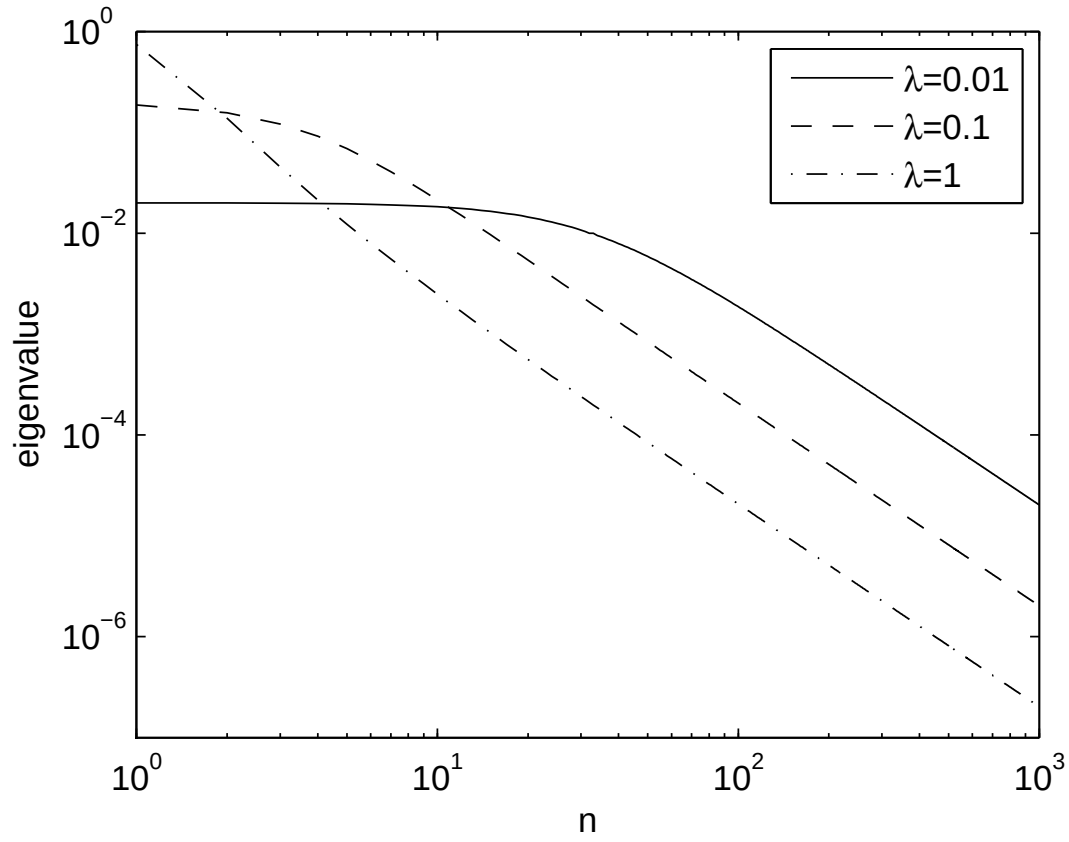
$$\int R(\mathbf{x}, \mathbf{y})\ f_n(\mathbf{y})\ \mathrm{d}\mathbf{y} = \theta_n\, f_n(\mathbf{x})$$

and $\xi_n(\omega)$ are standard Normal random variables.

Numerical experiments truncate the expansion.

Elliptic SPDE

Decay of 1D eigenvalues



When $\lambda = 1$, can use a low-dimensional polynomial chaos approach, but it's impractical for smaller λ.

Elliptic SPDE

Discretisation:

- cell-centred finite volume discretisation on a uniform grid – for rough coefficients we need to make grid spacing very small on finest grid
- each level of refinement has twice as many grid points in each direction
- current numerical experiments use a direct solver for simplicity, but in the future will use an efficient multigrid solver and so “computational cost” is defined to be proportional to the total number of grid points

1D Results

Numerical results for unit interval $[0, 1]$.

Boundary conditions – fixed pressure: $p(0) = 1,\ p(1) = 0$

Output quantity – mass flux: $-k \dfrac{\mathrm{d}p}{\mathrm{d}x}$

Correlation length: $\lambda = 0.1$

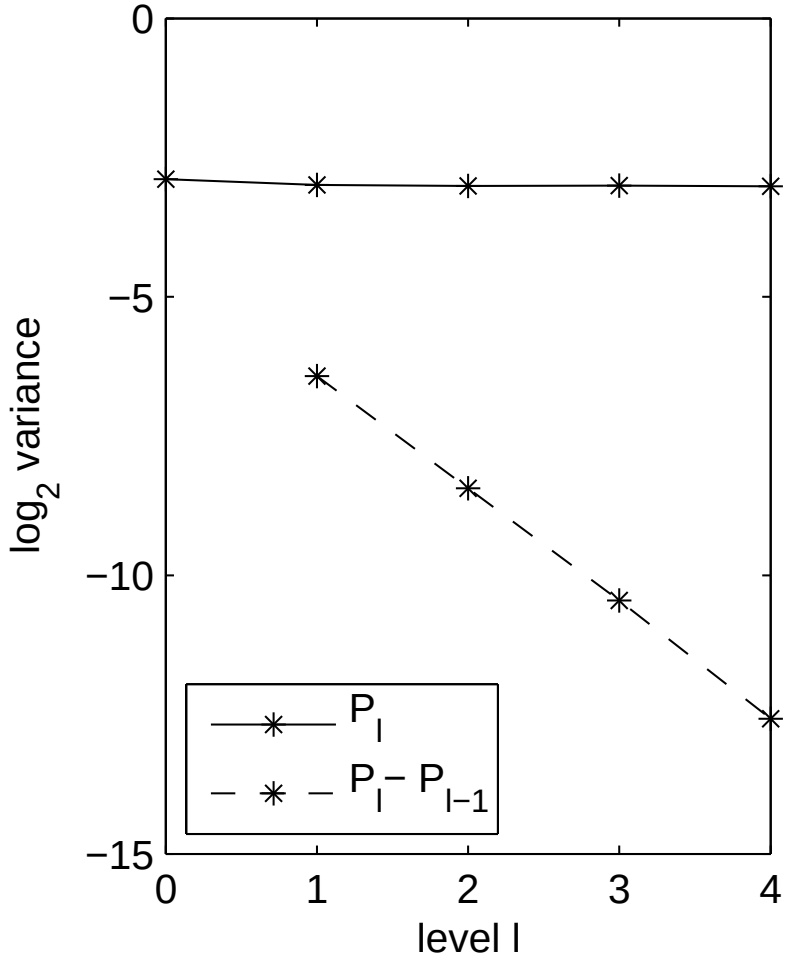
Coarsest grid: $h = 1/16$ (comparable to λ)

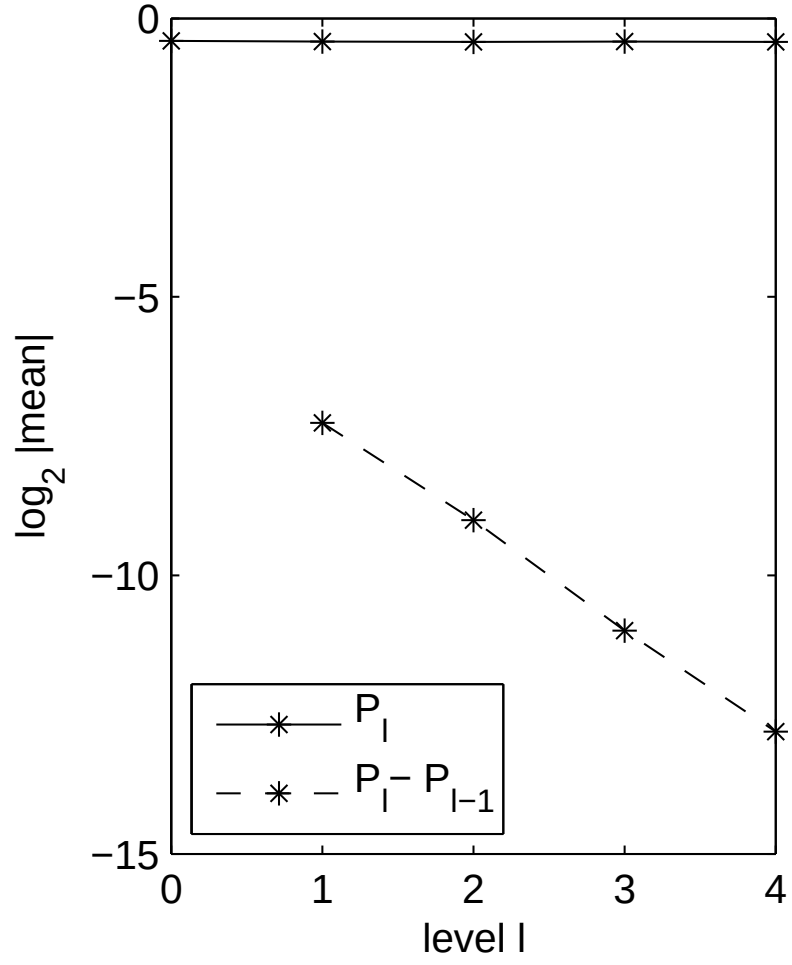
Finest grid: $h = 1/256$

Karhunen-Loève truncation: $m_{KL} = 800$

Cost taken to be proportional to number of nodes

1D Results

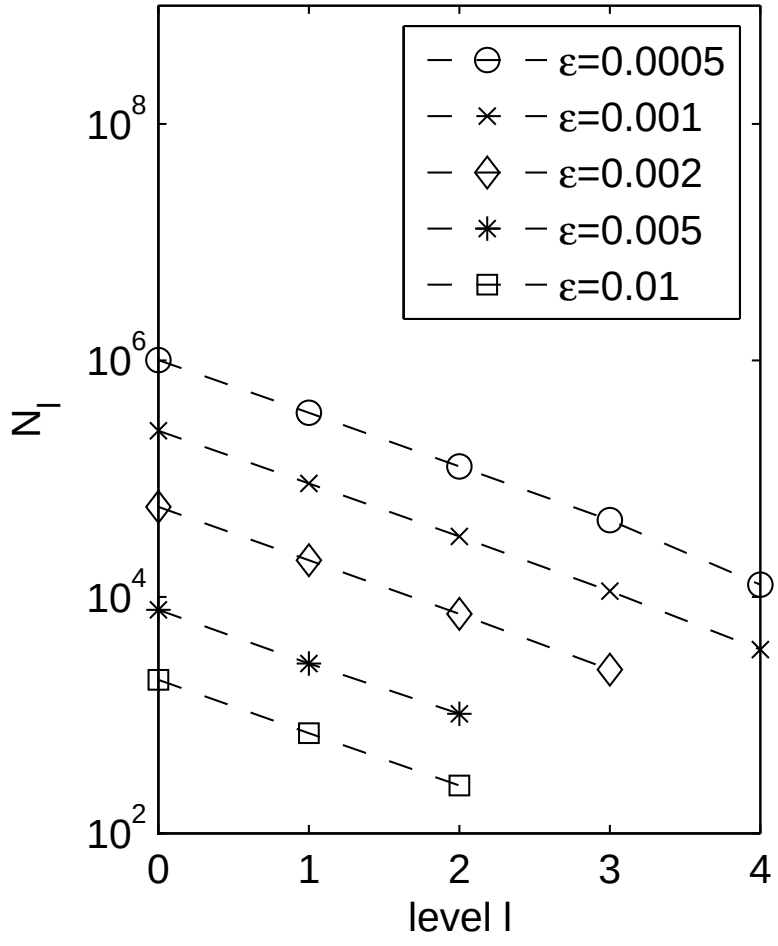


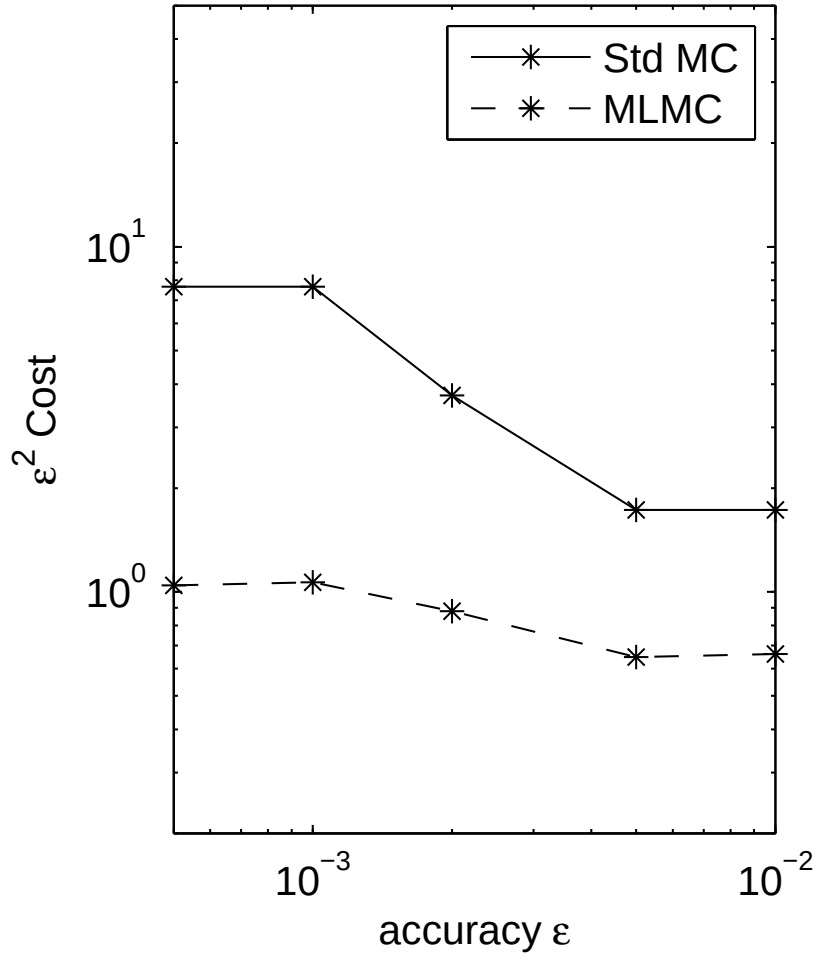


$$\mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] \sim h_\ell^2,$$

$$\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] \sim h_\ell^2$$

1D Results


ε=0.0005
ε=0.001
ε=0.002
ε=0.005
ε=0.01
10^8
10^6
10^4
10^2
N_l
0
1
2
3
4
level l


Std MC
MLMC
10^1
10^0
ε^2 Cost
10^{-3}
10^{-2}
accuracy ε

2D Results

Boundary conditions for unit square $[0,1]^2$:
– fixed pressure: $p(0,x_2)=1,\ p(1,x_2)=0$
– Neumann b.c.: $\partial p/\partial x_2(x_1,0)=\partial p/\partial x_2(x_1,1)=0$

Output quantity – mass flux: $-\int k\,\dfrac{\partial p}{\partial x_1}\,\mathrm{d}x_2$

Correlation length: $\lambda = 0.2$

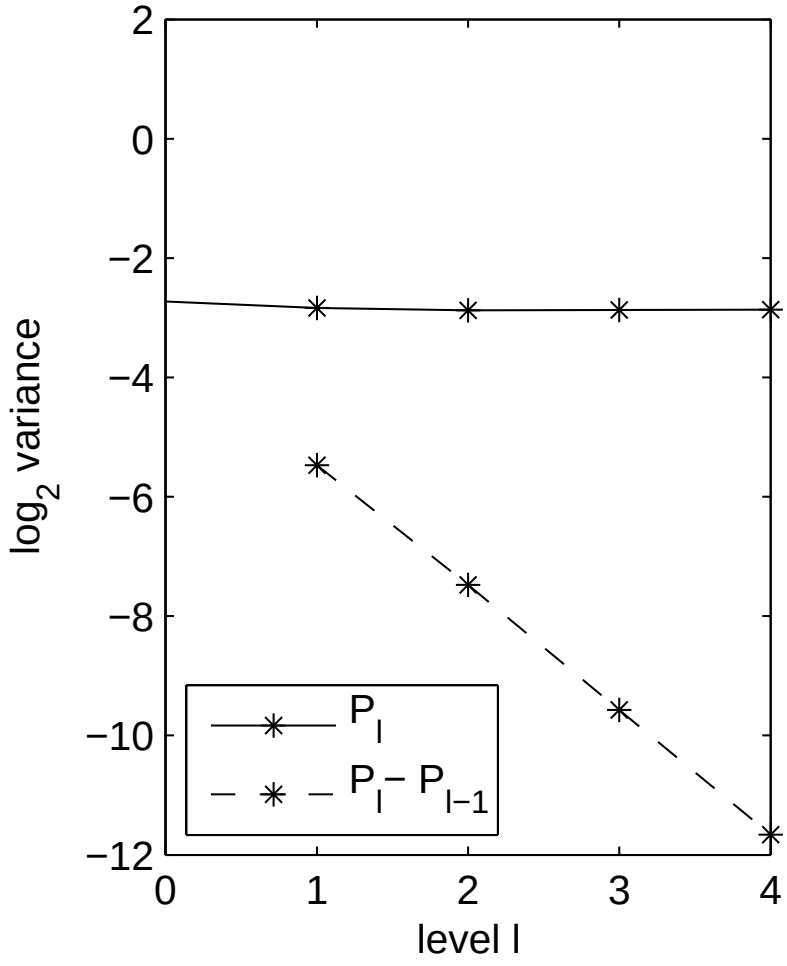
Coarsest grid: $h = 1/8$ (comparable to λ)

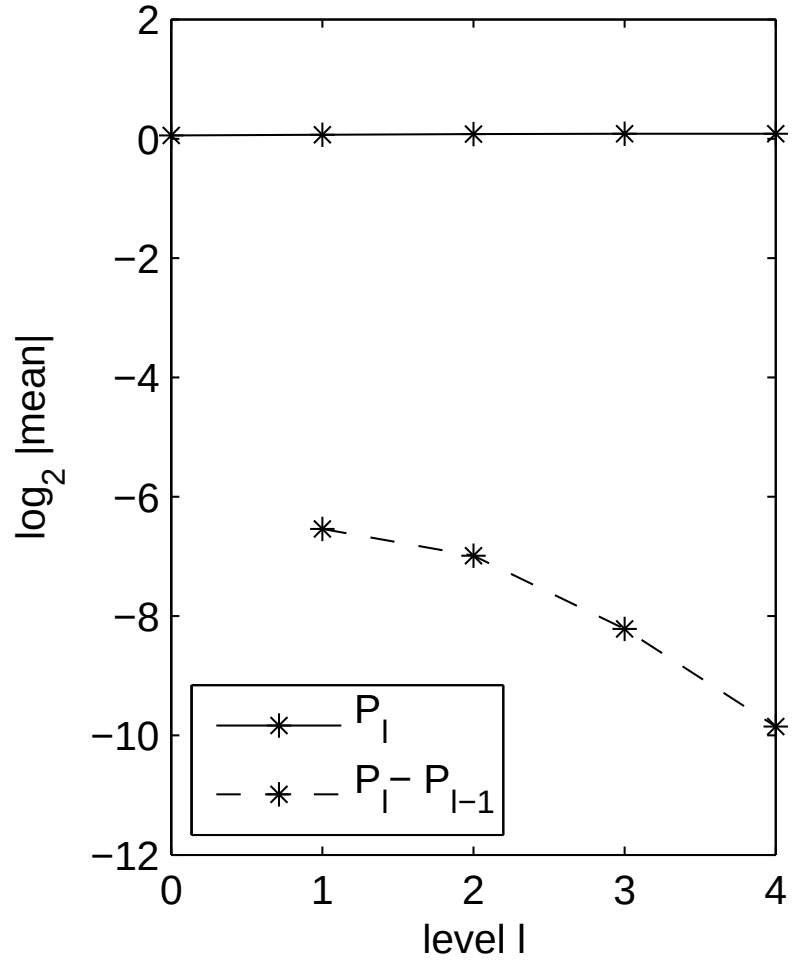
Finest grid: $h = 1/128$

Karhunen-Loève truncation: $m_{KL} = 4000$

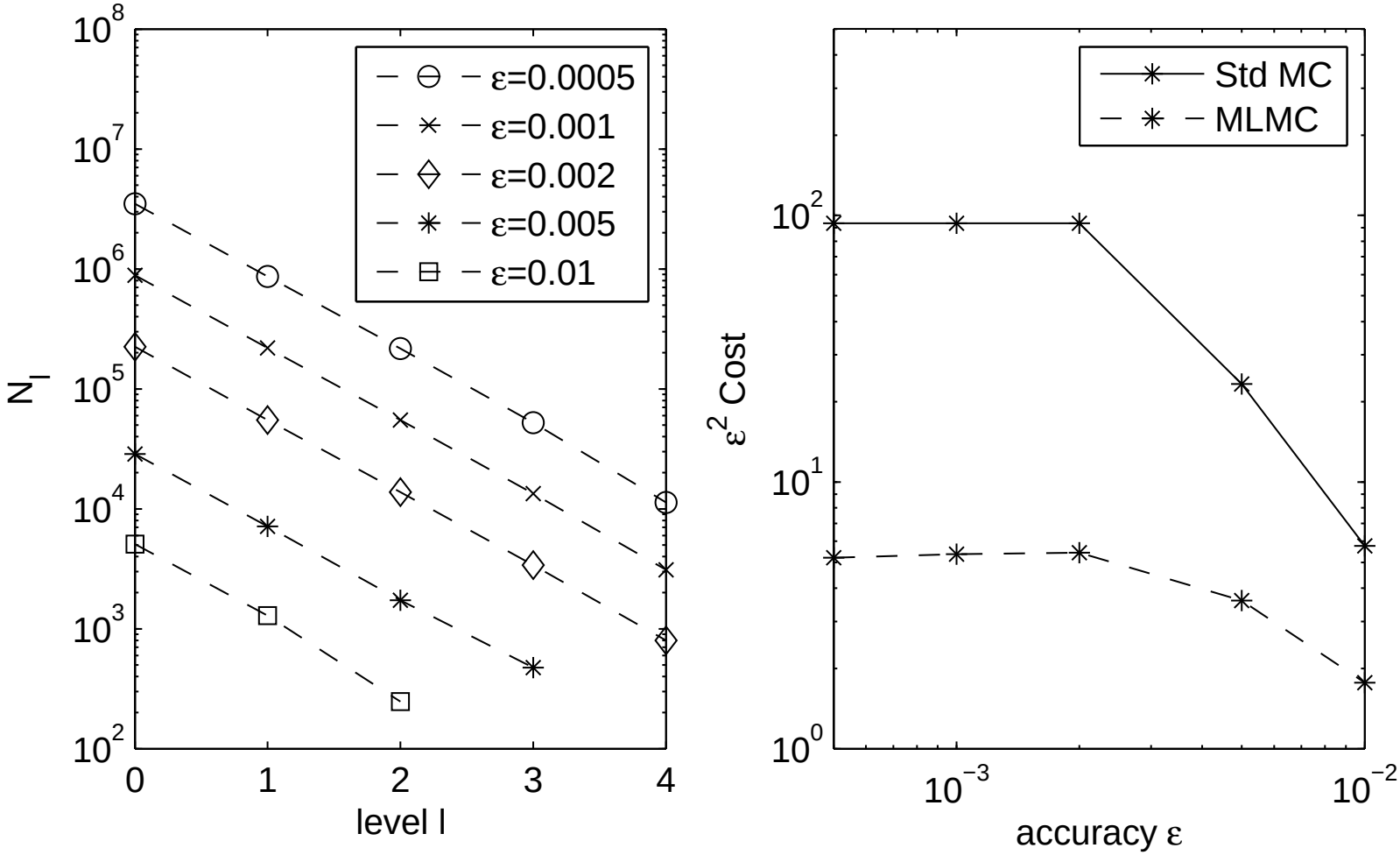
Cost still taken to be proportional to number of nodes

2D Results


2
0
−2
−4
−6
−8
−10
−12
0
1
2
3
4
log2 variance
level l
Pl
Pl − Pl−1


2
0
−2
−4
−6
−8
−10
−12
0
1
2
3
4
log2 |mean|
level l
Pl
Pl − Pl−1

2D Results



Greater savings because of greater cost on finer grids

Conclusions

- multilevel greatly reduces the cost of Monte Carlo simulation
- uses lots of cheap approximate simulations, and relatively few expensive accurate simulations
- makes Monte Carlo estimation a feasible technique for uncertainty quantification in engineering applications
- numerical analysis can be tough, but good progress is being made by various researchers