Multilevel Monte Carlo method

Mike Giles

mike.giles@maths.ox.ac.uk

Oxford-Man Institute of Quantitative Finance

Mathematical Institute, University of Oxford

with Rob Scheichl, Aretha Teckentrup (Bath)

and Andrew Cliffe (Nottingham)

European Multigrid Conference 2010

Isola d'Ischia, September 19-23, 2010

Outline

- standard Monte Carlo simulation
- multilevel Monte Carlo simulation
- 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 driving Brownian motion in an SDE (stochastic differential equation)
  - 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 $= O(\varepsilon)$

In a d-dimensional SPDE application with grid spacing h, if the bias is $O(h^\alpha)$ then need $h = O(\varepsilon^{1/\alpha})$, and total cost is $O(\varepsilon^{-(2+d/\alpha)})$, assuming efficient multigrid solution

(very challenging because of very rough coefficients – Graham & Scheichl)

To get acceptable accuracy in 3D applications may need 10,000 simulations on a 128^3 grid $\Longrightarrow$ very expensive

Multilevel Monte Carlo

The multilevel objective is to greatly reduce this cost:

	$\alpha = 1$		$\alpha = 2$	
d	MC	MLMC	MC	MLMC
1	ε^{-3}	ε^{-2}	$\varepsilon^{-2.5}$	ε^{-2}
2	ε^{-4}	$\varepsilon^{-2}(\log \varepsilon)^2$	ε^{-3}	$\varepsilon^{-2}(\log \varepsilon)^2$
3	ε^{-5}	ε^{-3}	$\varepsilon^{-3.5}$	$\varepsilon^{-2.5}$

How does this compare to theoretical lower bound?

- ε^{-2} calculations on coarsest grid costs $O(\varepsilon^{-2})$
- 1 calculation on finest grid costs $O(\varepsilon^{-d/\alpha})$

so minimum cost is $O(\varepsilon^{-\max(2,d/\alpha)})$

Multilevel Monte Carlo

How can this be achieved?

Use multigrid philosophy:

- 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 multigrid.

Multilevel Monte Carlo

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

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

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

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 $\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$ for $l > 0$ is

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

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

Variance is $N_l^{-1} V_l$ where $V_l = \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}]$

Key point: V_l gets progressively smaller as l increases because $\widehat{P}_l, \widehat{P}_{l-1}$ 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 $\sum_{l=0}^{L} N_l^{-1} V_l$ and its computational cost is $\sum_{l=0}^{L} N_l \, C_l$ so the variance is minimised for fixed cost by choosing $N_l \propto \sqrt{V_l / C_l}$, and then the cost on level l is proportional to $N_l \, C_l \propto \sqrt{V_l \, C_l}$

To make RMS error ε

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

MLMC Theorem

If there exist independent estimators $\widehat{Y}_l$ 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) $\left|\mathbb{E}[\widehat{P}_l - P]\right| \leq c_1\, 2^{-\alpha\, l}$

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) $\mathbb{V}[\widehat{Y}_l] \leq c_2\, N_l^{-1} 2^{-\beta\, l}$

iv) $C_l \leq c_3\, 2^{\gamma\, l}$

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

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

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 better discretisation
- third paper with D. Higham & X. Mao (*Finance and Stochastics, 2009*) performed numerical analysis of discretisation in first paper
- new paper with K. Debrabant and A. Rößler analyses discretisation in second paper

Other work

- Yuan Xia, G – jump-diffusion models
- Sylvestre Burgos, G – Greeks (sensitivities)
- 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

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:

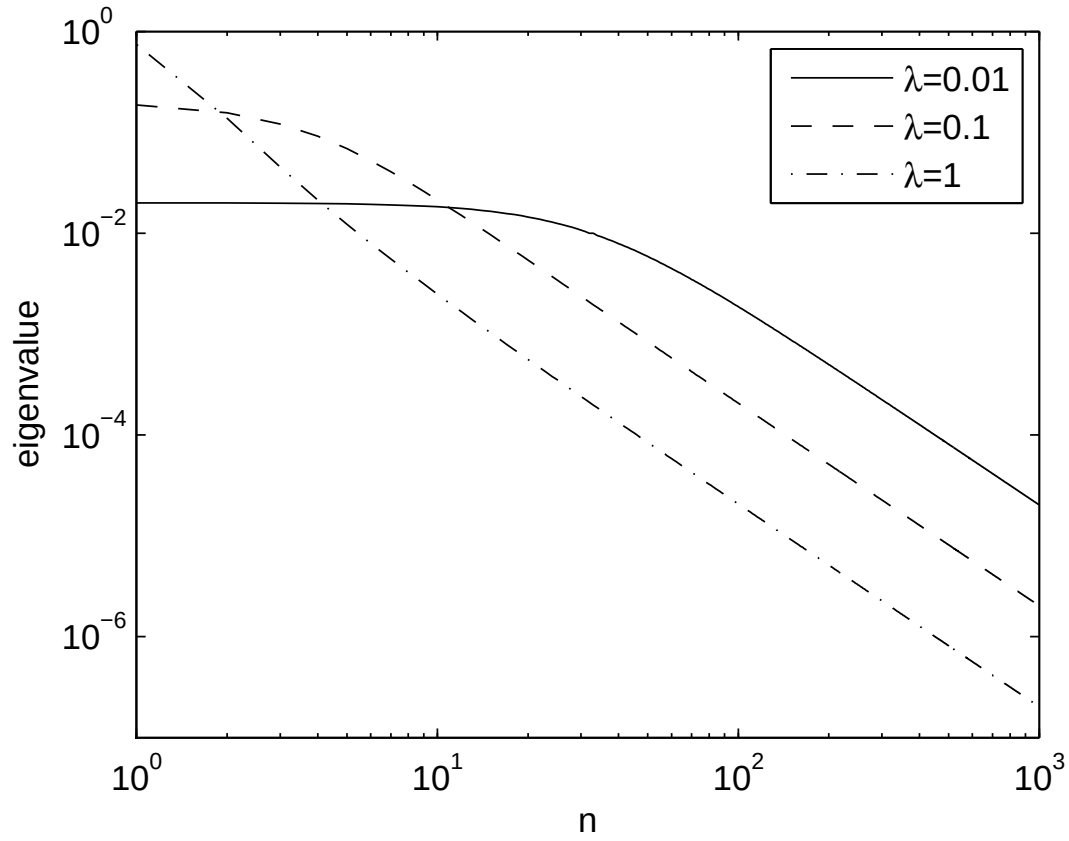
$$\int R(\mathbf{x}, \mathbf{y})\; f_n(\mathbf{y})\; \mathrm{d}\mathbf{y} = \theta_n\, f_n(\mathbf{x})$$

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: $-k\,\dfrac{\mathrm{d}p}{\mathrm{d}x}$

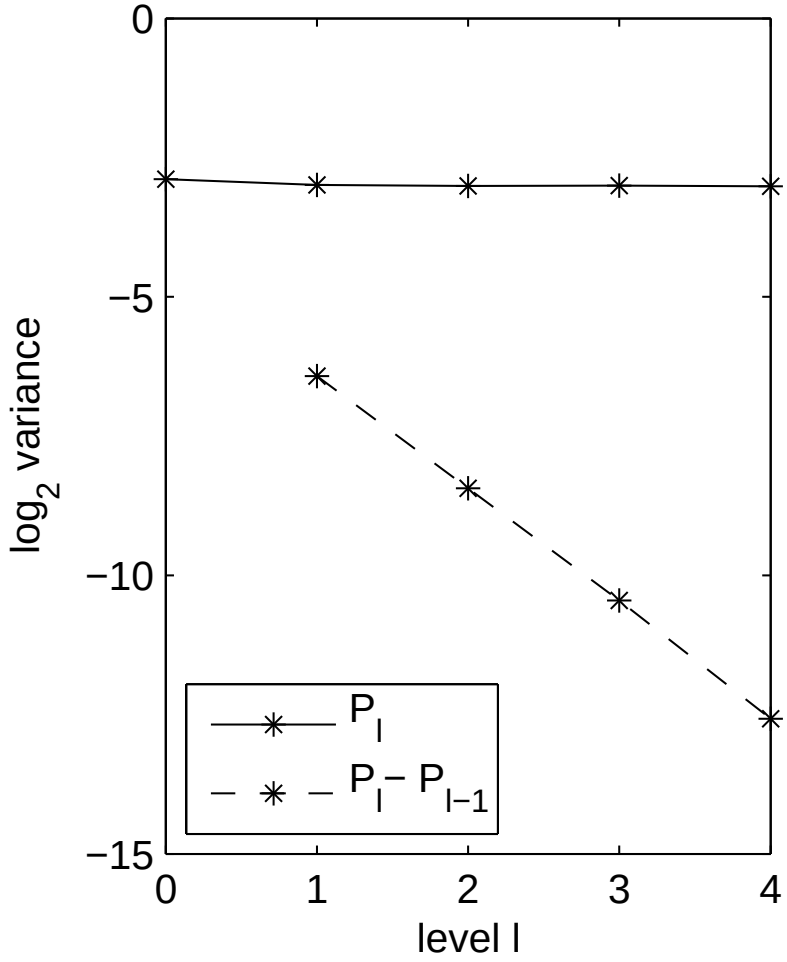
Correlation length: $\lambda = 0.1$

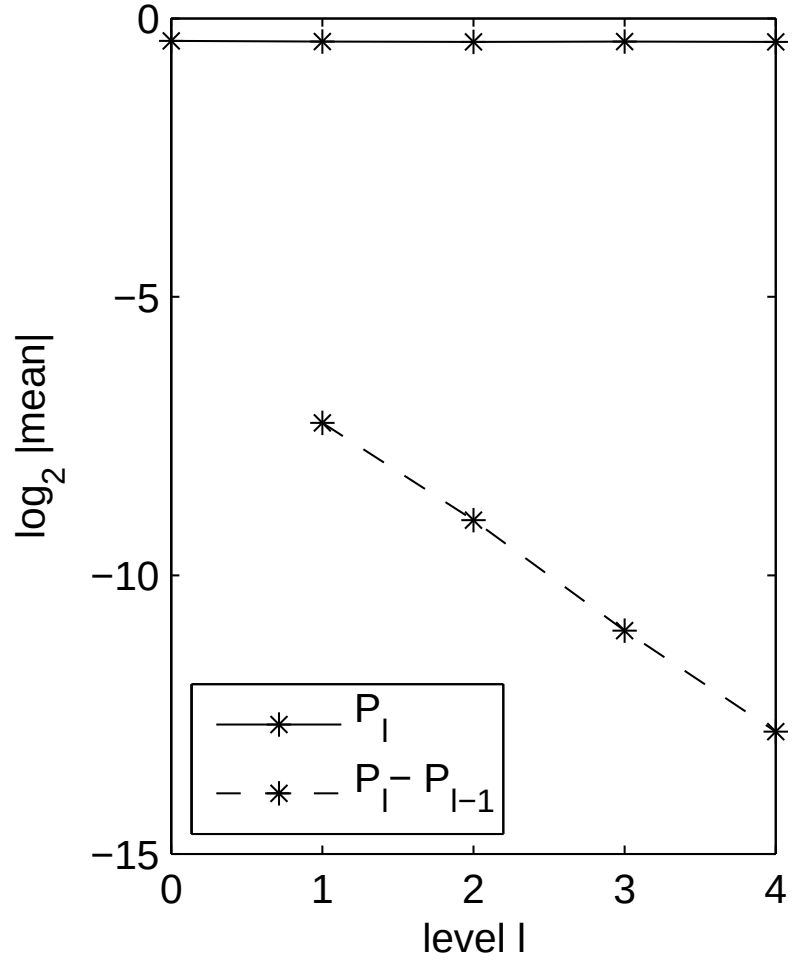
Coarsest grid: $h = 1/16$ (comparable to λ)
Finest grid: $h = 1/256$

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

Cost taken to be proportional to number of nodes

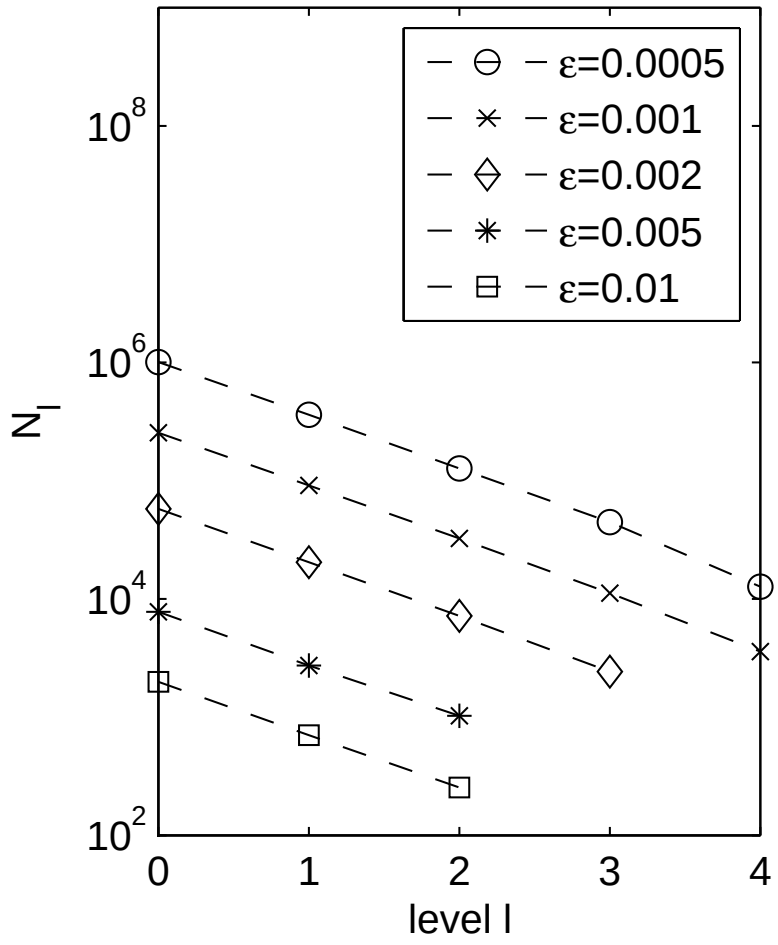
1D Results

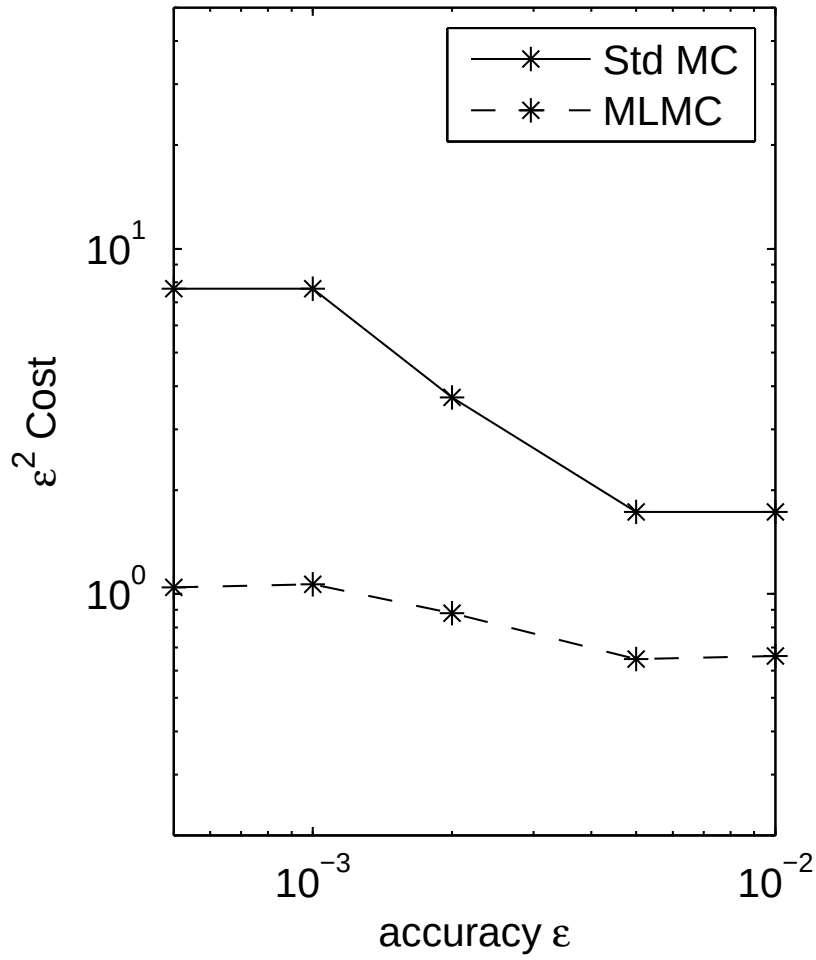




$$\mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] \sim h_l^2, \qquad \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}] \sim h_l^2$$

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: $\lambda = 0.2$

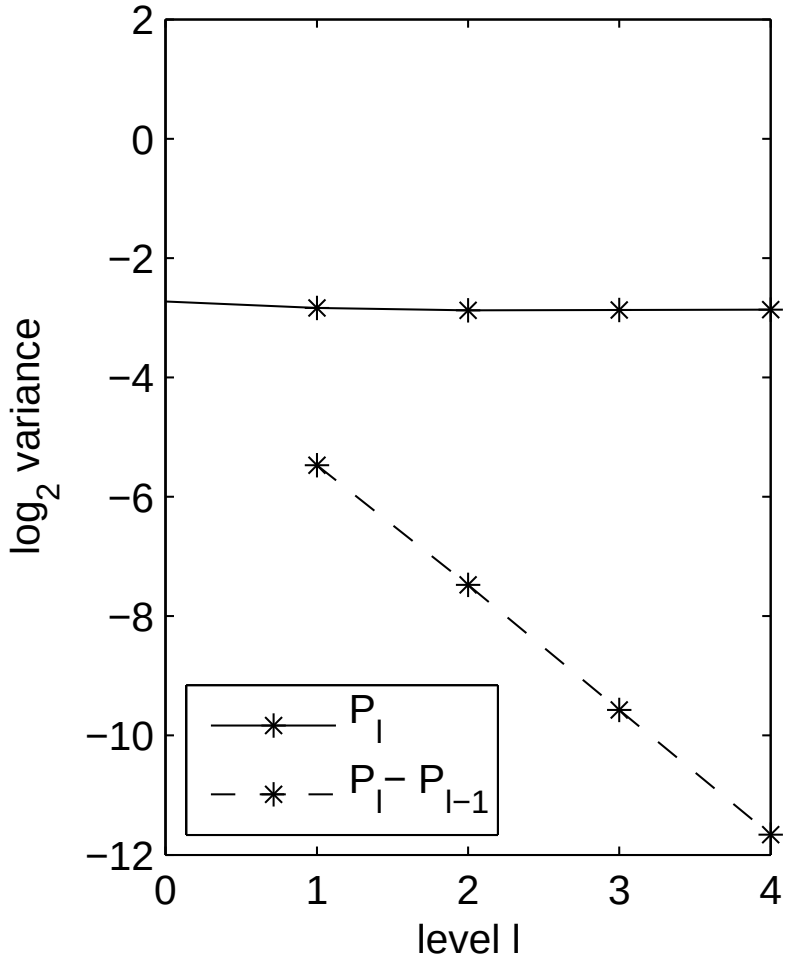
Coarsest grid: $h = 1/8$ (comparable to λ)

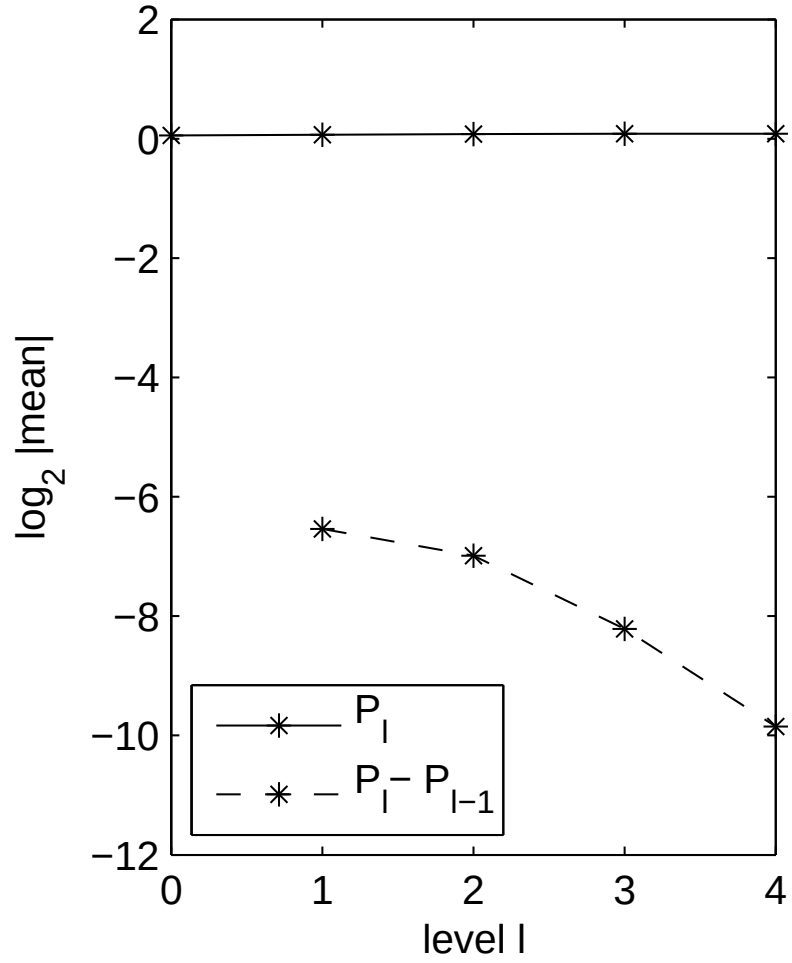
Finest grid: $h = 1/128$

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

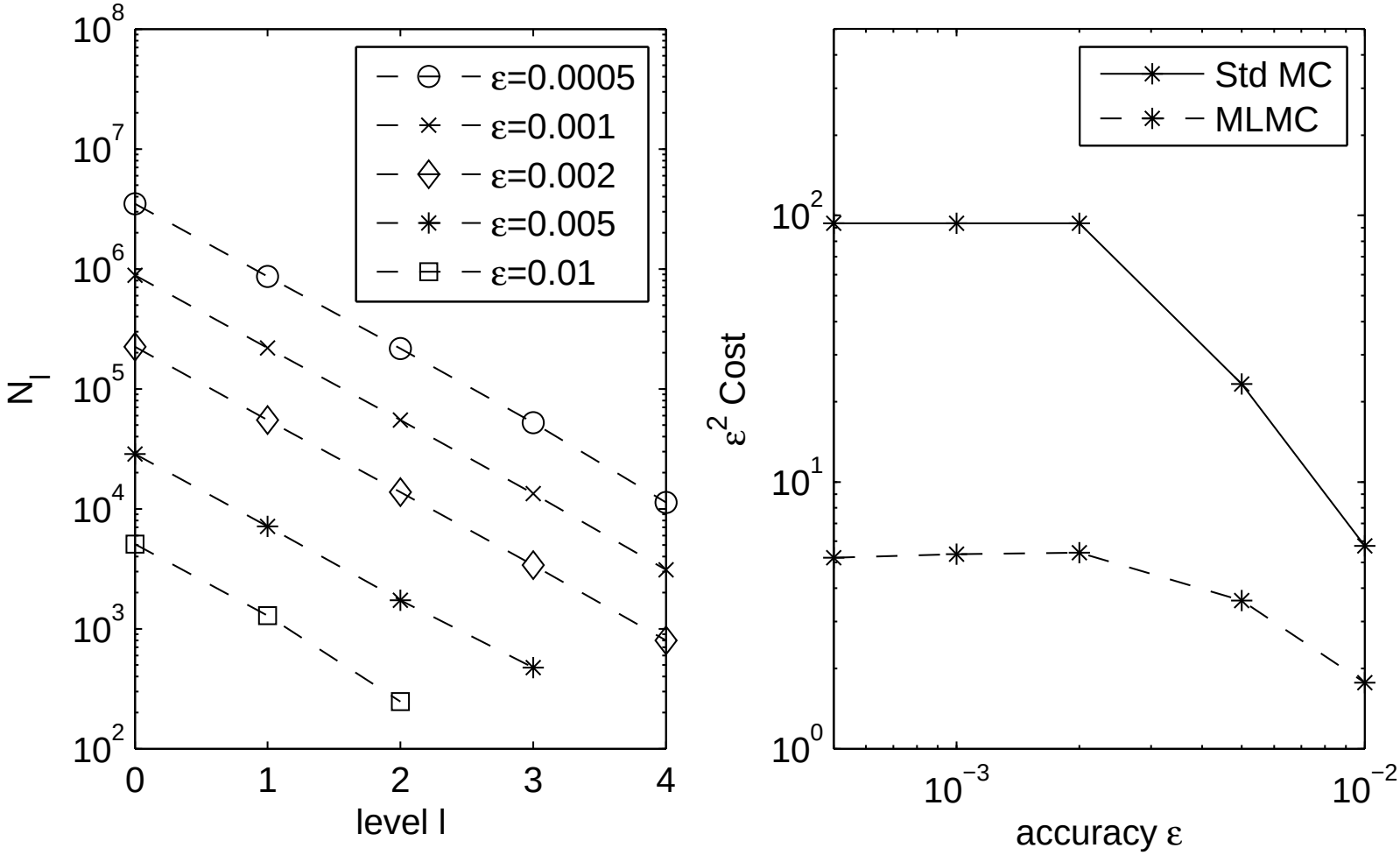
Cost still taken to be proportional to number of nodes

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

- standard Monte Carlo is prohibitively expensive for 2D and 3D elliptic SPDE applications
- multilevel Monte Carlo greatly reduces the cost, making this feasible for engineering applications
- should be superior to the polynomial chaos approach for applications with minimal spatial correlation
- numerical analysis is tough, but making some headway with finite element analysis to gain insight into its effectiveness