Monte Carlo Methods for Uncertainty Quantification

Mike Giles

Mathematical Institute, University of Oxford

ERCOFTAC course on Mathematical Methods and Tools in Uncertainty Management and Quantification

October 25, 2013

Lecture outline

Lecture 3:

- weak and strong convergence
- mean square error decomposition
- a PDE example
- multilevel Monte Carlo
- an SPDE example

Weak convergence

So far, we have assumed that given a random input X we are able to exactly compute the random output $f(X)$.

However, we often have to introduce an approximation $\widehat{f}(X)$, with $\widehat{f}(X) \to f(X)$ as $h \to 0$, where h represents some numerical discretisation parameter, such as the timestep or grid spacing.

We then have two notions of *c*onvergence.

For most applications, what matters is the **weak** order of convergence, defined by the error in the expected value. The weak order is m if

$$\mathbb{E}[f] - \mathbb{E}[\widehat{f}] = O(h^m)$$

Strong convergence

In other applications, what matters is the **strong** order of convergence, defined by the average difference in approximating each sample.

The strong error is defined in terms of an intermediate quantity S (usually the solution of an SDE or SPDE) and its numerical approximation $\widehat{S}$.

The strong order is m if

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

or

$$\mathbb{E}\left[|S-\widehat{S}|\right] = O(h^m)$$

If $f(S)$ is Lipschitz, the weak order of convergence can be higher than the strong order, because positive and negative errors $S-\widehat{S}$ cancel out.

Mean Square Error

We have two kinds of errors: Monte Carlo sampling error and approximation errors.

This then leads to a question: is it better to increase the number of samples, or to improve the accuracy of the samples

If the true value is $V = \mathbb{E}[f]$

and the discrete approximation is $\widehat{V} = \mathbb{E}[\widehat{f}]$

and the Monte Carlo estimate is $\widehat{Y} = \frac{1}{N} \sum_{n=1}^{N} \widehat{f}^{(n)}$

then . . .

Mean Square Error

. . . the Mean Square Error is

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

- first term is due to the variance of estimator
- second term is square of bias due to weak error

Mean Square Error

Suppose the weak error is proportional to h^α, and the computational cost is proportional to $C = N\,h^{-\gamma}$. Then the MSE is approximately

$$a\,N^{-1} + b\,h^{2\alpha} = a\,C^{-1}h^{-\gamma} + b\,h^{2\alpha}$$

For a fixed computational cost, this is a minimum when

$$\gamma\,a\,C^{-1}h^{-\gamma} = 2\,\alpha\,b\,h^{2\alpha}$$

and the two error terms have a similar magnitude.

To achieve a RMS error of ε requires $h = O(\varepsilon^{1/\alpha})$ and $N = O(\varepsilon^{-2})$, so the total cost is $O(\varepsilon^{-2-\gamma/\alpha})$.

PDE example

Consider Application 2 from Lecture 1, in which the trampoline displacement is defined by

$$T\left(\frac{\partial^2 Z}{\partial x^2}+\frac{\partial^2 Z}{\partial y^2}\right)=L(x,y), \qquad 0<x<1,\ 0<y<1$$

with uncertainty in the boundary conditions.

If we use a second order finite element approximation with grid spacing h, then $Z-\widehat{Z}=O(h^2)$, and the cost is $O(h^{-2})$, assuming a perfect multigrid solver.

If the output $f(Z)$ is Lipschitz, then the weak error will also be $O(h^2)$, so we have $\alpha=2$ and $\gamma=2$, and hence the total cost to achieve a RMS accuracy of ε is $O(\varepsilon^{-3})$.

Multilevel Monte Carlo

Monte Carlo was considered impractical for PDE and SPDE applications — a single 3D calculation is expensive, so 10^5 calculations is too costly.

This perspective has changed completely due to the introduction of Multilevel Monte Carlo by Stefan Heinrich and myself.

Multilevel Monte Carlo uses a similar idea to multigrid, combining simulations with different levels of resolution — same accuracy as finest calculations, but at a much lower computational cost.

Can also be viewed as a recursive control variate strategy.

Multilevel MC Approach

Consider multiple sets of simulations with different timesteps / grid spacing $h_\ell = 2^{-\ell}$, $\ell = 0, 1, \ldots, L$, and output $\widehat{f}_\ell$

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

Expected value is same – aim is to reduce variance of estimator for a fixed computational cost.

Key point: approximate $\mathbb{E}[\widehat{f}_\ell - \widehat{f}_{\ell-1}]$ using N_ℓ simulations with $\widehat{f}_\ell$ and $\widehat{f}_{\ell-1}$ obtained using same random inputs (e.g. same boundary conditions in Application 2).

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

Multilevel MC Approach

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

$$\mathbb{V}\left[\sum_{\ell=0}^{L} \widehat{Y}_\ell\right] = \sum_{\ell=0}^{L} N_\ell^{-1} V_\ell, \qquad V_\ell \equiv \mathbb{V}[\widehat{f}_\ell - \widehat{f}_{\ell-1}],$$

and the computational cost is proportional to $\displaystyle\sum_{\ell=0}^{L} N_\ell \, h_\ell^{-\gamma}$.

Hence, the variance is minimised for a fixed computational cost by choosing N_ℓ to be proportional to $V_\ell^{1/2} \, h_\ell^{\gamma/2}$.

The constant of proportionality can be chosen so that the combined variance is $O(\varepsilon^2)$.

General MLMC Theorem

Theorem: *Let f be an output functional of the solution of an SDE or SPDE, and $\widehat{f}_\ell$ the discrete approximation using $h_\ell = M^{-\ell}$.*

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

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

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

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

General 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 $\mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[f]\right)^2\right] < \varepsilon^2$

with a total computational cost C *with bound*

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

Discussion

If $\beta > \gamma$, the cost is $O(\varepsilon^{-2})$:

- Monte Carlo requires $O(\varepsilon^{-2})$ samples to get a RMS error of ε
- average cost per sample is $O(1)$
- can't do any better without going to Quasi-Monte Carlo

When $\beta < \gamma$, the cost is $O(\varepsilon^{-2-(\gamma-\beta)/\alpha})$:

- cost of one calculation on the finest level is $O(\varepsilon^{-\gamma/\alpha})$
- MLMC is equivalent to $O(\varepsilon^{-2+\beta/\alpha})$ calcs on finest level
- if $\beta = 2\,\alpha$, that's $O(1)$ calcs on finest level — can't do any better

Discussion

In the case of Application 2, we have

$$Z - \widehat{Z} = O(h^2) \quad \Longrightarrow \quad \widehat{Z}_\ell - \widehat{Z}_{\ell-1} = O(h_\ell^2)$$

so

$$\mathbb{V}[\widehat{f}_\ell - \widehat{f}_{\ell-1}] \;\le\; \mathbb{E}[(\widehat{f}_\ell - \widehat{f}_{\ell-1})^2] \;=\; O(h_\ell^4)$$

Hence we have $\alpha = 2$, $\beta = 4$.

A perfect multigrid solver gives $\gamma = 2$, so overall cost for RMS accuracy ε is $O(\varepsilon^{-2})$, while single sample deterministic cost is $O(\varepsilon^{-1})$.

A less perfect direct solver gives $\gamma = 4$, so cost is $O(\varepsilon^{-2}(\log \varepsilon)^2)$ compared to $O(\varepsilon^{-2})$ for single sample.

This is why MLMC has really made the Monte Carlo approach practical for uncertainty quantification with PDEs.

Multilevel Algorithm

Asymptotically,

$$\mathbb{E}[\widehat{f}_L - \widehat{f}_{L-1}] \approx (M-1)\,\mathbb{E}[f - \widehat{f}_L]$$

so this can be used to decide when the bias error is sufficiently small. In case the correction changes sign at some level, it is safer to use

$$\max\left\{M^{-1}\left|\widehat{Y}_{L-1}\right|, \left|\widehat{Y}_L\right|\right\} < (M-1)\,\frac{\varepsilon}{\sqrt{2}}.$$

1. start with $L=0$
2. estimate V_L using some initial samples
3. define optimal N_ℓ, $\ell = 0, \ldots, L$
4. evaluate extra samples as needed for new N_ℓ
5. if $L \geq 2$, test for convergence
6. if $L<2$ or not converged, set $L := L+1$ and go to 2.

Elliptic SPDE

I'm working with Rob Scheichl (Bath) and Andrew Cliffe (Nottingham) on multilevel Monte Carlo for the modelling of oil reservoirs and groundwater contamination in nuclear waste repositories.

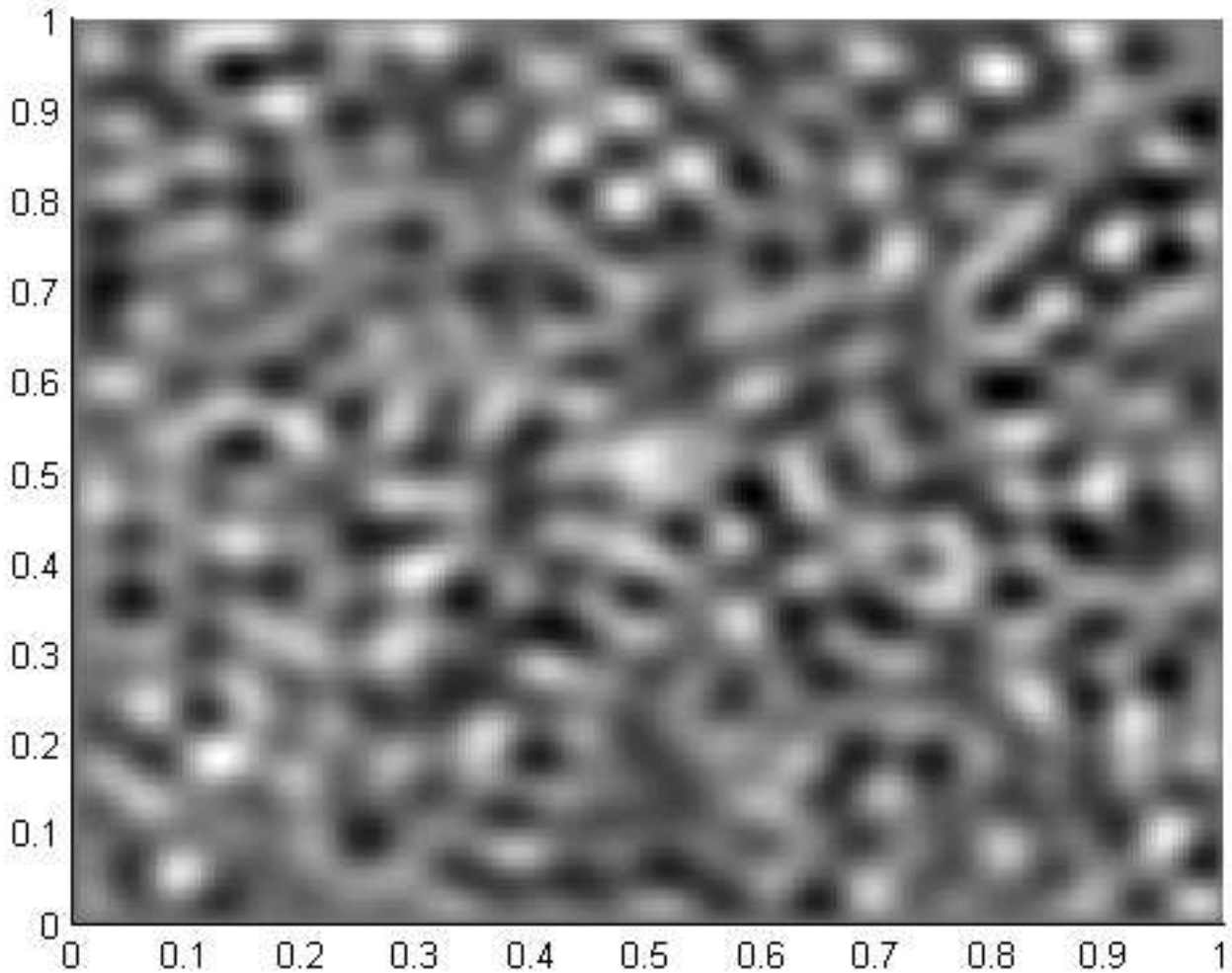
Here we have an elliptic SPDE coming from Darcy's law:

$$\nabla \cdot \Big(\kappa(x) \nabla p \Big) = 0$$

where the permeability $\kappa(x)$ is uncertain, and $\log \kappa(x)$ is often modelled as being Normally distributed with a spatial covariance such as

$$\mathrm{cov}(\log \kappa(x_1), \log \kappa(x_2)) = \sigma^2 \exp(-\|x_1 - x_2\|/\lambda)$$

Elliptic SPDE



A typical realisation of κ for $\lambda = 0.001,\ \sigma = 1$.

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:

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

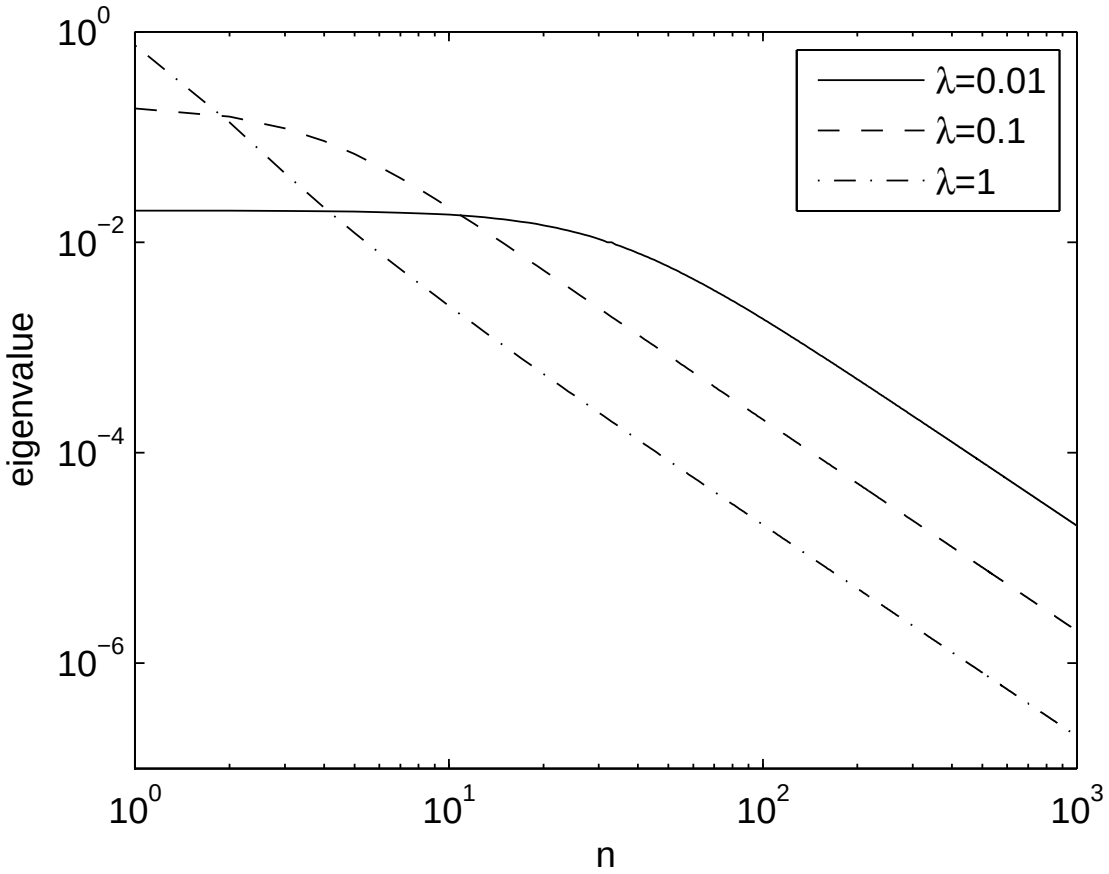
and $\xi_n(\omega)$ are standard Normal random variables.

Numerical experiments truncate the expansion.

(Latest 2D/3D work uses an efficient FFT construction based on a circulant embedding.)

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 3D will use an efficient AMG multigrid solver with a cost roughly proportional to the total number of grid points

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: $\lambda = 0.2$

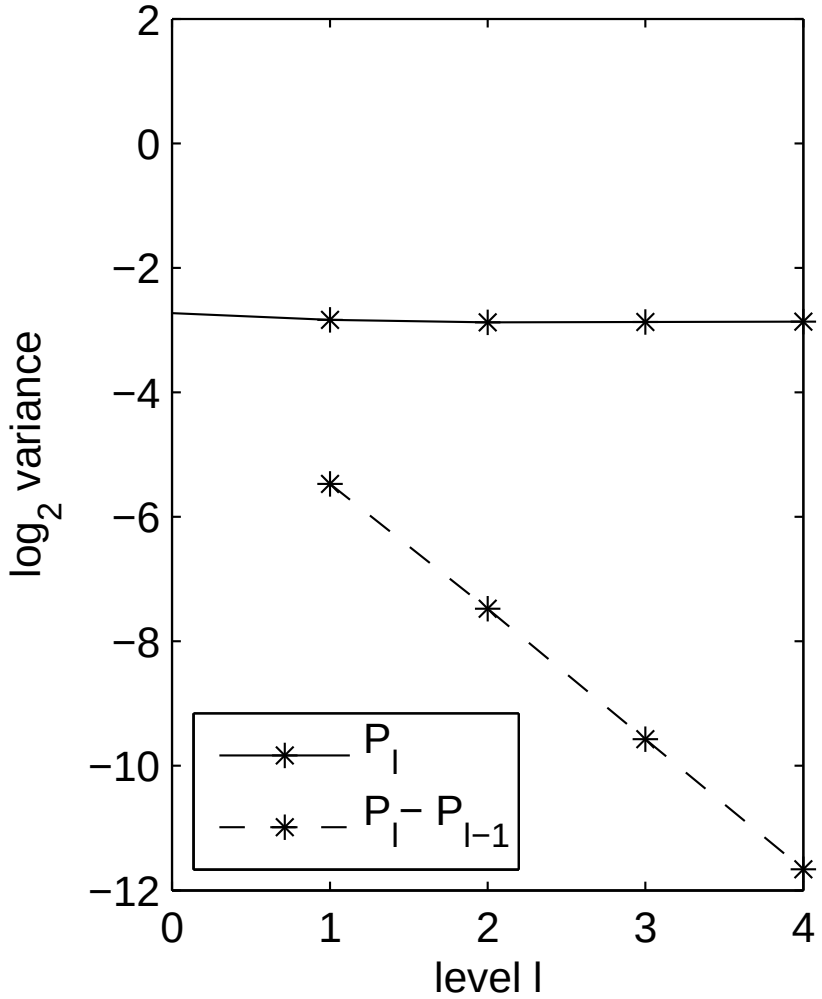
Coarsest grid: $h = 1/8$ (comparable to λ)

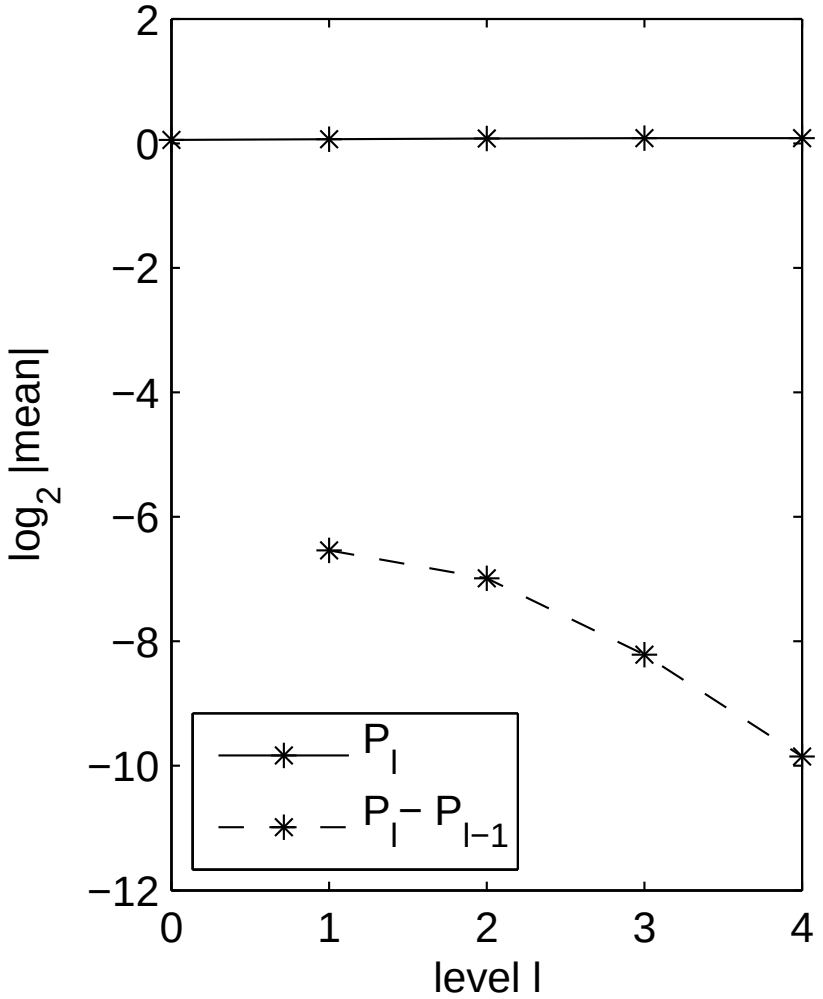
Finest grid: $h = 1/128$

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

Cost taken to be proportional to number of nodes

2D Results

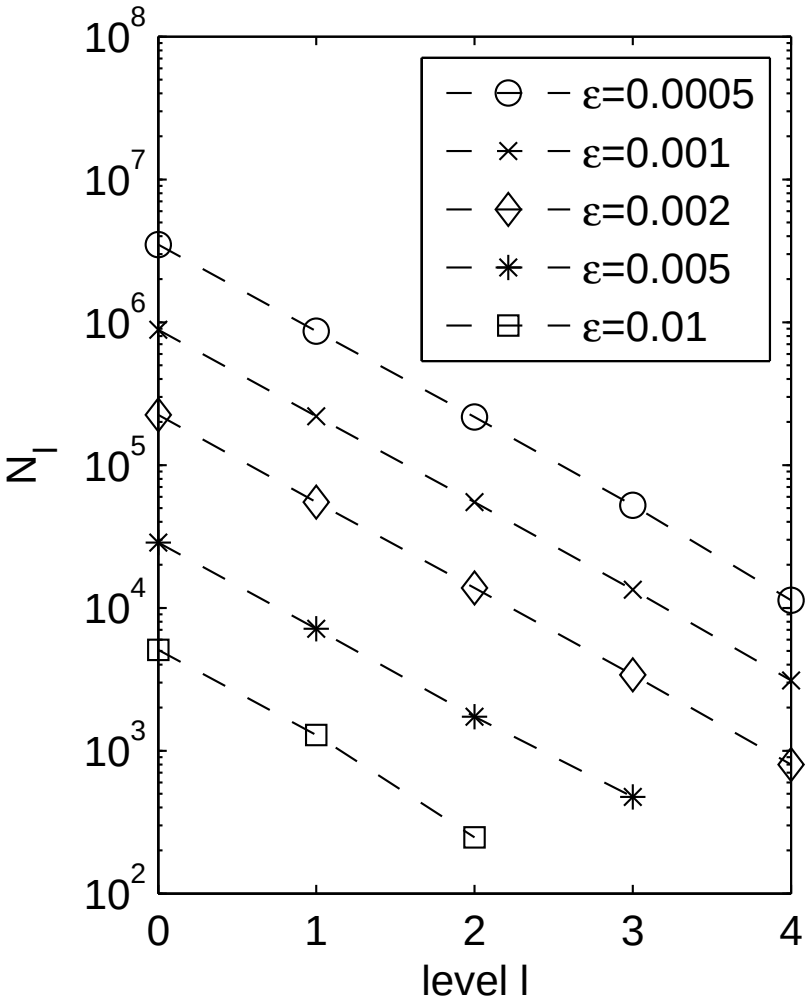


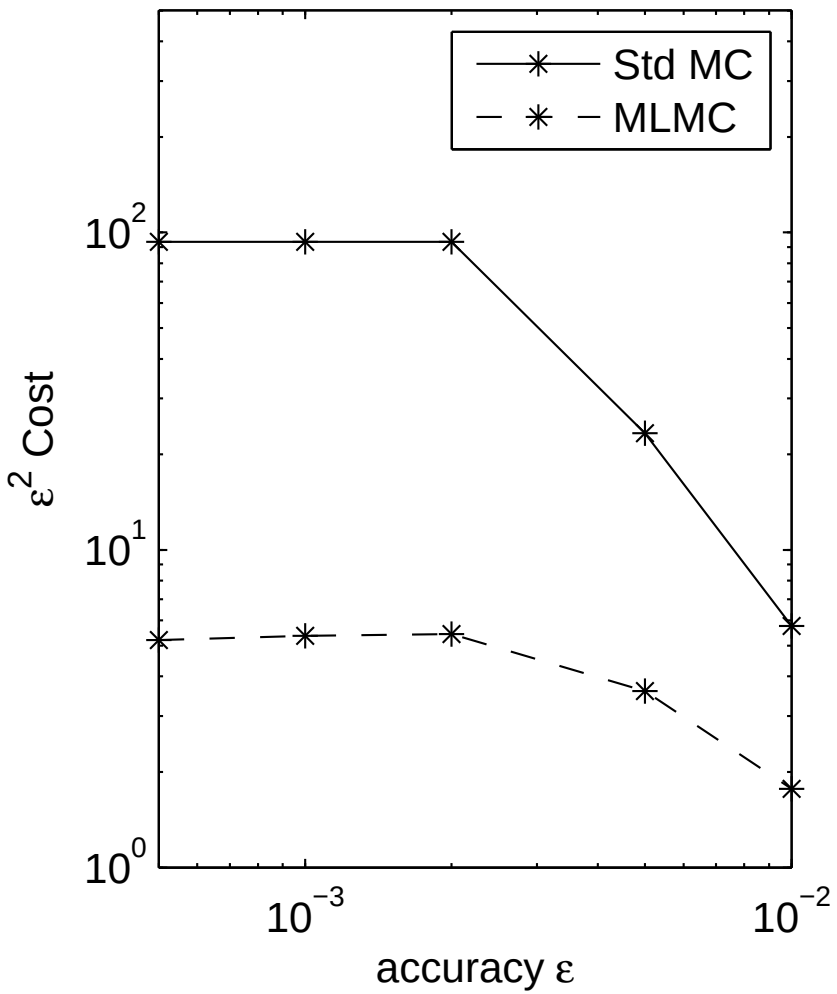


$$\mathbb{V}[\widehat{f}_\ell - \widehat{f}_{\ell-1}] \sim h_\ell^2$$

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

2D Results


ε=0.0005
ε=0.001
ε=0.002
ε=0.005
ε=0.01
N_l
level l


Std MC
MLMC
ε^2 Cost
accuracy ε

Complexity analysis

Relating things back to the MLMC theorem:

$$\begin{aligned} \mathbb{E}[\widehat{f}_\ell - f] &\sim 2^{-2\ell} \implies \alpha = 2 \\ V_\ell &\sim 2^{-2\ell} \implies \beta = 2 \\ C_\ell &\sim 2^{d\ell} \implies \gamma = d \quad \text{(dimension of PDE)} \end{aligned}$$

To achieve r.m.s. accuracy ε requires finest level grid spacing $h \sim \varepsilon^{1/2}$ and hence we get the following complexity:

dim	MC	MLMC
1	$\varepsilon^{-2.5}$	ε^{-2}
2	ε^{-3}	$\varepsilon^{-2}(\log \varepsilon)^2$
3	$\varepsilon^{-3.5}$	$\varepsilon^{-2.5}$

Other SPDE applications

For more on multilevel for SPDEs, see the work of Christoph Schwab and his group (ETH Zurich):

`http://www.math.ethz.ch/~schwab/`

- elliptic, parabolic and hyperbolic PDEs
- stochastic coefficients, initial data, boundary data

Schwab used to work on alternative techniques such as "polynomial chaos" but has now largely switched to MLMC because of its superior efficiency for many applications.

For other papers, see my MLMC community homepage:

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

Extension to multiple outputs

Suppose we have M outputs and want

$$\sum_{\ell=0}^{L} N_\ell^{-1} V_{\ell,m} \leq \tfrac{1}{2} \varepsilon_m^2, \qquad m = 1, \ldots, M,$$

where $V_{\ell,m}$ is the variance of the multilevel correction for output m, and ε_m is the desired RMS accuracy for that output.

As usual the computational cost is $\sum_{\ell=0}^{L} N_\ell C_\ell$, and we can then do a constrained optimisation using M Lagrange multipliers.

However, this is a bit nasty – it's not clear how many of the Lagrange multipliers will be "active"

Extension to multiple outputs

A simpler approach is to define

$$V_\ell = \max_m \frac{V_{\ell,m}}{\varepsilon_m^2}$$

and make the variance constraint

$$\sum_{\ell=0}^{L} N_\ell^{-1} V_\ell \ \leq \ \tfrac{1}{2}.$$

This brings it back to a problem with a single Lagrange multiplier with the same optimal solution as before.

Klaus Ritter and Tigran Nagapetyan (Kaiserslautern) are using this to estimate the CDF (cumulative distribution function) of an exit time τ. We estimate the CDF at a set of exit times τ_k, and then use a cubic spline to approximate the full CDF.

Non-geometric multilevel

Almost all applications of multilevel in the literature so far use a geometric sequence of levels, refining the timestep (or the spatial discretisation for PDEs) by a constant factor when going from level ℓ to level $\ell + 1$.

Coming from a multigrid background, this is very natural, but it is **NOT** a requirement of the multilevel Monte Carlo approach.

All MLMC needs is a sequence of levels with

- increasing accuracy
- increasing cost
- increasingly small difference between outputs on successive levels

Non-geometric multilevel

At MIT, Nyugen is starting to look at an application based on reduced order modelling in which a solution can be expressed as a sum over a number of modes:

$$u = \sum_{m=1}^{M} a_m \; u_m$$

Here the u_m are fixed, and the amplitudes a_m are calculated by solving a reduced order problem which depends on some stochastic inputs.

Increasing M increases the accuracy, but also increases the cost.

MLMC may be very effective, but it's not at all clear how to choose the levels. Geometric might be OK:

$$M_\ell = \{1,\ 2,\ 4,\ 8,\ 16\}$$

but perhaps linear would be better?

$$M_\ell = \{2,\ 4,\ 6,\ 8,\ 10,\ 12\}$$

Final words

- multilevel Monte Carlo has made the Monte Carlo approach viable for engineering applications which require the solution of PDEs
- the implementation is often very straightforward, using the same stochastic inputs for simulations on two levels of refinement
- it is particularly effective for applications with strong nonlinearity and high stochastic dimensionality, where the alternative methods are not appropriate
- MATLAB code and test examples are available from `http://people.maths.ox.ac.uk/gilesm/mc2013/`
  – see the last part of lecture 4 for useful information on the code

References

M.B. Giles, "Multilevel Monte Carlo path simulation", *Operations Research*, 56(3):607-617, 2008.

K.A. Cliffe, M.B. Giles, R. Scheichl, A.L. Teckentrup, "Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients", *Computing and Visualization in Science*, 14(1):3-15, 2011

M.B. Giles and B.J. Waterhouse. "Multilevel quasi-Monte Carlo path simulation". pp.165-181 in *Advanced Financial Modelling*, in Radon Series on Computational and Applied Mathematics, de Gruyter, 2009.

M.B. Giles. 'Multilevel Monte Carlo methods', pp.79-98 in *Monte Carlo and Quasi-Monte Carlo Methods 2012*, Springer, 2013.

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

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