Optimal selection of levels for multilevel Monte Carlo

Mike Giles

University of Oxford
Mathematical Institute

Monte Carlo and Quasi-Monte Carlo Methods

April 8, 2014

Motivation

Previous MLMC work has generally used a geometric sequence of levels:

- with SDEs, increasing the number of timesteps on each level by factor 2 or 4
- with SPDEs, doubling the grid resolution in each direction

But is this optimal?

Seemed natural originally because of multigrid inspiration.

Also, for original application (SDE with Euler-Maruyama discretisation and European payoff) it could be argued that it was best for each level to be the geometric mean of the two on either side.

Motivation

However, in some cases it is not obvious how to define the levels.

The work in this talk arose because of a discussion with Ngoc Cuong Nguyen at MIT on MLMC for reduced order modelling.

Interested to compute

$$\mathbb{E}[\, g(\, U(\omega)\,)\,]$$

where $U(\omega)$ is the solution of

$$(A_0 + \omega\, A_1)\, U = f$$

Using a reduced order model

$$U = C\, V$$

where V has much lower dimension leads to the reduced order model:

$$\left(C^T A_0\, C + \omega\, C^T A_1\, C\right) V = C^T f$$

Motivation

Problem: solution cost is $O(\dim(V)^3)$, so high accuracy requires a high cost per Monte Carlo sample

Solution: use MLMC, with level ℓ using first k_ℓ reduced order modes (i.e. first k_ℓ elements of V and first k_ℓ columns of C).

New problem: how many levels should be used, and how should k_ℓ be chosen?

Idea: start with too many levels, and then get rid of some.

MLMC

The standard MLMC estimator (using $\widehat{P}_{-1} \equiv 0$) is

$$\widehat{Y} = \sum_{\ell} \widehat{Y}_\ell, \qquad \widehat{Y}_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} (\widehat{P}_\ell^{(n)} - \widehat{P}_{\ell-1}^{(n)}).$$

The total variance is

$$\sum_{\ell} N_\ell^{-1} V_\ell, \qquad V_\ell = \mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}],$$

and the total cost is

$$\sum_{\ell} N_\ell C_\ell, \qquad C_\ell = \text{cost of one sample} .$$

Hence, a constrained optimisation gives $N_\ell \propto \sqrt{V_\ell / C_\ell}$

MLMC

Keeping level ℓ, its contribution to the overall estimator is

$$N_\ell^{-1} \sum_{n=1}^{N_\ell} (\widehat{P}_\ell^{(n)} - \widehat{P}_{\ell-1}^{(n)}) \;+\; N_{\ell+1}^{-1} \sum_{n=1}^{N_{\ell+1}} (\widehat{P}_{\ell+1}^{(n)} - \widehat{P}_\ell^{(n)})$$

The combined variance from the two terms is

$$N_\ell^{-1} V_\ell + N_{\ell+1}^{-1} V_{\ell+1} = N_{\ell+1}^{-1} V_{\ell+1} \left(1 + \sqrt{\frac{V_\ell \; C_\ell}{V_{\ell+1} \; C_{\ell+1}}} \right),$$

and the combined cost is

$$N_\ell \; C_\ell + N_{\ell+1} \; C_{\ell+1} = N_{\ell+1} \; C_{\ell+1} \left(1 + \sqrt{\frac{V_\ell \; C_\ell}{V_{\ell+1} \; C_{\ell+1}}} \right),$$

and so the product of the combined variance and combined cost is

$$V_{\ell+1} \; C_{\ell+1} \left(1 + \sqrt{\frac{V_\ell \; C_\ell}{V_{\ell+1} \; C_{\ell+1}}} \right)^2.$$

MLMC

Dropping level ℓ, the replacement estimator is

$$\widetilde{N}_{\ell+1}^{-1} \sum_{n=1}^{\widetilde{N}_\ell} (\widehat{P}_{\ell+1}^{(n)} - \widehat{P}_{\ell-1}^{(n)}).$$

Its variance is $\widetilde{N}_{\ell+1}^{-1} \widetilde{V}_{\ell+1}$, where $\widetilde{V}_{\ell+1} \equiv \mathbb{V}[\widehat{P}_{\ell+1} - \widehat{P}_{\ell-1}]$,

Its cost is $\widetilde{N}_{\ell+1} \widetilde{C}_{\ell+1}$, where $\widetilde{C}_{\ell+1}$ is the new cost of a single sample.

Hence, the product of the variance and the cost is $\widetilde{V}_{\ell+1}\ \widetilde{C}_{\ell+1}$,
so level ℓ should be dropped if

$$\widetilde{V}_{\ell+1}\ \widetilde{C}_{\ell+1} \ < \ V_{\ell+1}\, C_{\ell+1} \left(1 + \sqrt{\frac{V_\ell\ C_\ell}{V_{\ell+1}\, C_{\ell+1}}} \right)^2 .$$

MLMC

This analysis handles the general problem of the reduced order model.

In practice, a few preliminary samples would be used on the different levels to decide which levels to keep.

Then the main calculation would be carried out with the final set of levels.

Remainder of the talk makes some additional assumptions to address the optimality of geometric sequences for SDE/SPDE applications.

MLMC

Will assume $C_{\ell+1} \approx \widetilde{C}_{\ell+1}$ since the main cost of both is computing $\widehat{P}_{\ell+1}$.

Also, for a geometric sequence of levels, the ratios $C_\ell / C_{\ell+1}$ and $V_\ell / V_{\ell+1}$ are generally known, at least asymptotically as $\ell \to \infty$.

The key question is how $\widetilde{V}_{\ell+1}$ compares to $V_{\ell+1}$?

$$\widetilde{V}_{\ell+1} \;=\; \mathbb{V}[\widehat{P}_{\ell+1} - \widehat{P}_{\ell-1}] \;=\; V_{\ell+1} + 2\rho\sqrt{V_\ell \; V_{\ell+1}} + V_\ell,$$

where ρ is the correlation between $\widehat{P}_{\ell+1} - \widehat{P}_\ell$ and $\widehat{P}_\ell - \widehat{P}_{\ell-1}$.

We now consider two extremes, $\rho = 1$ and $\rho = 0$.

MLMC

If $\rho=1$, then

$$\begin{aligned}\widetilde{V}_{\ell+1} &= V_{\ell+1}\left(1+\sqrt{\frac{V_\ell}{V_{\ell+1}}}\right)^2 \\ &> V_{\ell+1}\left(1+\sqrt{\frac{V_\ell\ C_\ell}{V_{\ell+1}\ C_{\ell+1}}}\right)^2\end{aligned}$$

so the decision is not to remove the level.

Near-perfect correlation is likely to arise in SPDE applications with finite dimensional uncertainty in the initial data, boundary data, or PDE coefficients.

In this case, the increments $\widehat{P}_\ell - \widehat{P}_{\ell-1}$ obtained in going from one level to the next come from the progressive elimination of truncation error due to the discretisation of the PDE.

MLMC

If $\rho=0$, then $\widetilde{V}_{\ell+1} = V_{\ell+1} + V_\ell$ and so we drop level ℓ if

$$1 + \frac{V_\ell}{V_{\ell+1}} \;<\; \left(1 + \sqrt{\frac{V_\ell\; C_\ell}{V_{\ell+1}\; C_{\ell+1}}}\;\right)^2 .$$

For the case of an SDE with an Euler discretisation with refinement factor K, and a Lipschitz European payoff, we have

$$\frac{C_\ell}{C_{\ell+1}} \approx K^{-1}, \qquad \frac{V_\ell}{V_{\ell+1}} \approx K$$

so the condition for dropping level ℓ becomes $K < 3$.

This is consistent with the analysis in my original Operations Research article which concluded that it was better to use $K=4$ rather than $K=2$.

MLMC

On the other hand, switching to the Milstein discretisation gives

$$\frac{C_\ell}{C_{\ell+1}} \approx K^{-1}, \qquad \frac{V_\ell}{V_{\ell+1}} \approx K^2$$

so the condition becomes

$$1 + K^2 < (1 + \sqrt{K}\,)^2 \quad \Longrightarrow \quad K < 2.3 \;\; \text{(approximately)}$$

This suggest the optimal refinement factor is less than for the Euler discretisation, which matches what is found in practice.

MLMC

In either case, provided the correlation ρ does not vary with level ℓ, then the test indicates refinement if $K < K_0$ for some K_0.

This leads to a sequence of levels with a refinement factor $K > K_0$.

i.e. even if the initial sequence is not geometric, it ends up being approximately geometric.

MLMC

This conclusion may seem in conflict with recent analysis by Haji-Ali, Nobile, von Schwerin, Tempone on "Optimization of mesh hierarchies in multilevel Monte Carlo samplers" (`http://arxiv.org/abs/1403.2480`)

However, it is not:

- when $C_\ell V_\ell \approx$ const, the computational effort is spread equally across all levels, and we agree that a geometric sequence is optimal
- if $C_\ell V_\ell \to 0$ as $\ell \to \infty$, the dominant effort is on the coarsest levels, so all that matters is that they have the near-optimal ratio K
- if $C_\ell V_\ell \to \infty$ as $\ell \to \infty$, the dominant effort is on the finest levels, so all that matters is that they have the near-optimal ratio K

Conclusions

- have presented a simple analysis showing how to decide whether it is beneficial to eliminate levels in a MLMC computation
- particularly useful in applications where the choice of levels is not obvious, such as in reduced order modelling
- also suggests that a geometric sequence of levels is optimal for SDE / SPDE applications, with the refinement ratio dependent on the correlation between corrections at successive levels