Two new developments in Multilevel Monte Carlo methods

Mike Giles

Abdul-Lateef Haji-Ali
Oliver Sheridan-Methven

Mathematical Institute, University of Oxford

Rob Scheichl's Farewell Conference

University of Bath, November 13th, 2018

MLMC – the early days

- Jan 2006 – the idea for MLMC for SDEs came while on holiday
- paper written by July 2006 – MCQMC conference presentation in Sept
- summer 2006
  - already thinking about PDEs, and had preliminary discussions with Tom Hou (Caltech) but that didn't lead to anything
  - also talked to Ian Sloan (UNSW) – that led to a visit and a collaboration on MCQMC
- by 2007 very keen to pursue MLMC for PDEs so approached Rob at ICIAM 2007 in Zurich
- a year later we started working together, obtained EPSRC funding with Andrew Cliffe, then Aretha joined in 2009 and the rest is history

My current MLMC research

- Matteo Croci, Patrick Farrell – coupled generation of stochastic fields on non-nested grids by solving elliptic PDE driven by spatial white noise

  $$(-\nabla^2 + k^2)^q \, u = W$$

  This is an extension of nested grid research by Rob based on prior research which Chris Farmer told us about – working well for MLMC, currently being extended to MLQMC

- Abdul-Lateef Haji-Ali – nested expectation

- Oliver Sheridan-Methven – approximate random variables and reduced precision computing

Nested expectations

The concern here is estimation of $\mathbb{E}\left[f\left(\mathbb{E}[X|Y]\right)\right]$ where

- Y is multi-dimensional
- X can be either scalar or multi-dimensional

We will assume that $X|Y$ can be simulated exactly at average $O(1)$ cost – in complex cases this requires use of Rhee & Glynn's randomised MLMC for unbiased estimation.

Applications differ in the smoothness of f:

- C^2 – the nice case, e.g. x^2
- continuous, piecewise C^2 – the more typical case. e.g. $\max(x, 0)$
- discontinuous – the nasty case, e.g. Heaviside $H(x)$

Nested simulation

MLMC for nice case is quite natural:

- on level ℓ use 2^ℓ samples to compute $\overline{X}^{(f)} \approx \mathbb{E}[X|Y]$
- on coupled level $\ell-1$, split into two subsets of $2^{\ell-1}$ and average each to get $\overline{X}^{(c1)}, \overline{X}^{(c2)}$
- MLMC "antithetic" correction estimator is then

$$Z = f(\overline{X}^{(f)}) - \tfrac{1}{2}\left(f(\overline{X}^{(c1)}) + f(\overline{X}^{(c2)}) \right)$$

- if $f(x)$ is linear then $Z=0$; with a bounded second derivative then

$$Z = O\left((\overline{X}^{(c1)} - \overline{X}^{(c2)})^2 \right) = O(2^{-\ell})$$

  so the variance is $O(2^{-2\ell})$ while the cost is $O(2^\ell)$
- hence $\beta=2, \gamma=1$ in MLMC theorem $\Longrightarrow O(\varepsilon^{-2})$ complexity

Nested simulation

What if $f(x) = \max(x, 0)$?

- $Z = 0$ if $\overline{X}^{(c1)}, \overline{X}^{(c2)}$ have same sign
- if $\mathbb{E}[X|\,Y]$ has bounded density near 0, then $O(2^{-\ell/2})$ probability of $\overline{X}^{(c1)}, \overline{X}^{(c2)}$ having different signs, and in that case $Z = O(2^{-\ell/2})$
- MLMC correction variance is $O(2^{-3\ell/2})$ while the cost is $O(2^{\ell})$
- hence $\beta = 3/2, \gamma = 1$ in MLMC theorem $\Longrightarrow O(\varepsilon^{-2})$ complexity

Nested simulation

Three major applications:

- Financial credit modelling – Bujok, Hambly, Reisinger (2013)

  continous, piecewise linear $f(x)$ – first proof of $O(2^{-3\ell/2})$ variance

- Expected Value of Partial Perfect Information – G, Goda (2018)

$$\mathbb{E}\left[\max_d \mathbb{E}[f_d(X)|Y]\right] - \max_d \mathbb{E}[f_d(X)]$$

- Conditional value-at-risk – G, Haji-Ali (2018)

$$\min_x \left\{ x + \frac{1}{\eta} \max\left(\mathbb{E}[X-x], 0\right) \right\}$$

Nested simulation

Finally, the nasty case $f(x) = H(x)$

- $Z = 0$ if $\overline{X}^{(c1)}, \overline{X}^{(c2)}$ have same sign
- if $\mathbb{E}[X|Y]$ has bounded density near 0, then $O(2^{-\ell/2})$ probability of $\overline{X}^{(c1)}, \overline{X}^{(c2)}$ having different signs, and in that case $Z = O(1)$
- MLMC correction variance is $O(2^{-\ell/2})$ while the cost is still $O(2^\ell)$
- hence $\beta = 1/2, \gamma = 1$ in MLMC theorem – additional analysis gives $\alpha = 1$ so $O(\varepsilon^{-5/2})$ complexity
- By adaptively changing the number of inner samples (with minimum 2^ℓ and maximum $2^{2\ell}$ on level ℓ, can improve to get $\beta = 1, \gamma = 1$ and $O(\varepsilon^{-2}|\log \varepsilon|^2)$ complexity
- Analysis was challenging but satisfying – G, Haji-Ali (2018)

Approximate random variables

In some applications, generating the random numbers can be a significant cost, especially with QMC when inverting the CDF.

e.g. Poisson distribution, increments of a Lévy process, non-central chi-squared distribution (CIR model)

Even with Normal random variables, cost of conversion from uniform r.v. to Normal is non-trivial for vector implementation.

This has led to previous research by Hefter, Mayer, Ritter at TU Kaiserslautern.

Approximate random variables

Simplest example: Euler-Maruyama approximation of a scalar SDE:

$$\widehat{X}_{t_{m+1}} = \widehat{X}_{t_m} + a(\widehat{X}_{t_m})\,h + b(\widehat{X}_{t_m})\sqrt{h}\,Z_m$$

where Z_m is a unit Normal r.v. generated as

$$Z_m = \Phi^{-1}(U_m)$$

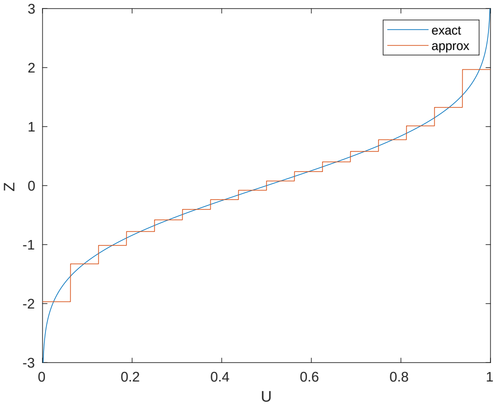
where U_m is a uniform $(0,1)$ r.v., and this leads to a Monte Carlo approximation of $\mathbb{E}[f(X_T)]$.

Suppose instead we use approximate Normals $\widetilde{Z}_m$ generated by

$$\widetilde{Z}_m = \widetilde{Q}(U_m)$$

where $\widetilde{Q}$ is a piecewise constant approximation to Φ^{-1} on 2^q uniformly-sized sub-intervals.

Approximate random variables



Note: only need first q bits of U_m and a lookup table to calculate $\widetilde{Z}_m$, and analysis proves that $\mathbb{E}[\,(Z-\widetilde{Z})^2] = O(q^{-1}2^{-q})$

(Alternatively, could use a polynomial approximation to Φ^{-1})

Approximate random variables

For an SDE approximation with a specified number of timesteps, a two-level MLMC approach gives

$$\mathbb{E}[\widehat{P}] \;=\; \mathbb{E}[\widetilde{P}] \;+\; \mathbb{E}[\widehat{P}-\widetilde{P}]$$

Further analysis proves that for a Lipschitz function $P \equiv f(X_T)$

$$\mathbb{E}[\,(\widehat{P}-\widetilde{P})^2] = O\left(\mathbb{E}[\,(Z-\widetilde{Z})^2]\right)$$

so this can lead to big savings if $\mathbb{E}[\,(Z-\widetilde{Z})^2] \ll 1$ and $\widetilde{Q}(U_m)$ is much cheaper to evaluate than $\Phi^{-1}(U_m)$.

Approximate random variables

How does this work in combination with standard MLMC?

Answer: nested MLMC

$$
\begin{aligned}
\mathbb{E}[P_L] &= \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell - P_{\ell-1}] \\
&= \mathbb{E}[\widetilde{P}_0] + \mathbb{E}[P_0 - \widetilde{P}_0] \\
&+ \sum_{\ell=1}^{L} \left\{ \mathbb{E}[\widetilde{P}_\ell - \widetilde{P}_{\ell-1}] + \mathbb{E}\left[(P_\ell - P_{\ell-1}) - (\widetilde{P}_\ell - \widetilde{P}_{\ell-1}) \right] \right\}
\end{aligned}
$$

The pair $(\widetilde{P}_\ell, \widetilde{P}_{\ell-1})$ are generated in the same way as $(P_\ell, P_{\ell-1})$, just replacing exact Z_m by approximate $\widetilde{Z}_m$.

Approximate random variables

Numerical analysis – a work-in-progress

Path differences?

$$\begin{aligned}
\widehat{X}_\ell - \widehat{X}_{\ell-1} &\sim h^{1/2} \\
\widetilde{X}_\ell - \widetilde{X}_{\ell-1} &\sim h^{1/2} \\
\widehat{X}_\ell - \widetilde{X}_\ell &\sim \sqrt{\mathbb{E}[(Z-\widetilde{Z})^2]} \\
(\widehat{X}_\ell - \widehat{X}_{\ell-1}) - (\widetilde{X}_\ell - \widetilde{X}_{\ell-1}) &\sim h^{1/2}\sqrt{\mathbb{E}[(Z-\widetilde{Z})^2]}
\end{aligned}$$

Approximate random variables

MLMC variances?

For smooth payoff functions $f(x)$, looks possible to prove

$$\mathbb{V}[(\widehat{P}_\ell - \widehat{P}_{\ell-1}) - (\widetilde{P}_\ell - \widetilde{P}_{\ell-1})] \sim h \; \mathbb{E}[(Z - \widetilde{Z})^2]$$

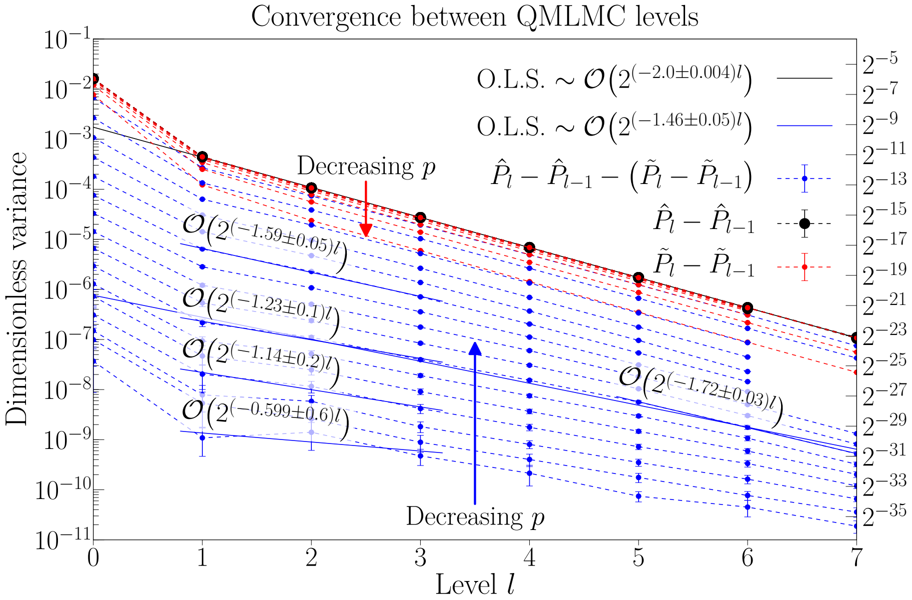
For continuous piecewise smooth functions (e.g. put/call functions) the analysis is much tougher. Numerical results suggest

$$\mathbb{V}[(\widehat{P}_\ell - \widehat{P}_{\ell-1}) - (\widetilde{P}_\ell - \widetilde{P}_{\ell-1})] \sim \min\left\{ h^{1/2}\, \mathbb{E}[(Z - \widetilde{Z})^2], h\, (\mathbb{E}[(Z - \widetilde{Z})^2])^{1/2} \right\}$$

but it looks tough to prove better than

$$\mathbb{V}[(\widehat{P}_\ell - \widehat{P}_{\ell-1}) - (\widetilde{P}_\ell - \widetilde{P}_{\ell-1})] \sim \min\left\{ \mathbb{E}[(Z - \widetilde{Z})^2], h\, (\mathbb{E}[(Z - \widetilde{Z})^2])^{1/3} \right\}.$$

Approximate random variables – numerical results



Reduced precision arithmetic

Further computational savings can be achieved by using reduced precision arithmetic.

Previous research at TU Kaiserslautern (Korn, Ritter, Wehn and others) and Imperial College (Luk and others) has used FPGAs with complete control over the precision, but I prefer GPUs.

In the latest NVIDIA GPUs, half-precision FP16 is twice as fast as single precision FP32 (which is 2-8 times faster than double precision FP64).

In most cases, single precision is perfectly sufficient for calculating $\widehat{P}$; half precision can be used for $\widetilde{P}$.

MC averaging should probably be done in double precision in both cases.

Reduced precision arithmetic

Very important: to ensure the telescoping sum is respected, must ensure that **exactly** the same computations are performed for $\widetilde{P}$ whether on its own or in calculating $\widehat{P} - \widetilde{P}$.

The effect of half-precision arithmetic can be modelled as

$$\widetilde{X}_{t_{m+1}} = \widetilde{X}_{t_m} + a(\widetilde{X}_{t_m})h + b(\widetilde{X}_{t_m})\sqrt{h}\,\widetilde{Z}_m + \delta\,\widetilde{X}_{t_m} V_m$$

where $\delta \approx 10^{-3}$ and the V_m are iid unit variance random variables.

Overall, this leads to an $O(\delta^2/h)$ increase in the variance; if this increases it too much, the reduced precision should not be used.

Conclusions

- I continue to look for new ways in which to use MLMC
- I'm very happy with new nested simulation MLMC – Abdul will continue this line of research at Heriot-Watt
- I think there's more to be done (probably by Abdul) with adaptive sampling within MLMC in other contexts
- I'm now excited about the new research using approximate distributions – I think this has interesting potential for the future

References

M. Croci, MBG, M.E. Rognes, P.E. Farrell. 'Efficient white noise sampling and coupling for multilevel Monte Carlo with non-nested meshes'. To appear in *SIAM/ASA Journal on Uncertainty Quantification*, 2018/9

MBG, A.-L. Haji-Ali. 'Multilevel nested simulation for efficient risk estimation'. arXiv pre-print 1802.05016.

MBG, T. Goda. 'Decision-making under uncertainty: using MLMC for efficient estimation of EVPPI'. To appear in *Statistics and Computing*, 2018/9.

MBG, M. Hefter, L. Mayer, K. Ritter. 'Random bit quadrature and approximation of distributions on Hilbert spaces'. *Foundations of Computational Mathematics*, online March 2018.

MBG, M. Hefter, L. Mayer, K. Ritter. 'Random bit multilevel algorithms for stochastic differential equations'. arXiv pre-print 1808.10623.