# Ideas, tricks and techniques for reduced MLMC variance

Mike Giles

Mathematical Institute, University of Oxford

BIRS Workshop on

Approximation of High-Dimensional Numerical Problems: Algorithms, Analysis and Applications

Sept 28, 2015

# Outline

- multilevel Monte Carlo
- conditional expectation
- change-of-measure
- splitting
- zero-mean control variate
- antithetic samples
- other techniques

# Multilevel Monte Carlo

MLMC is based on the telescoping sum

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

where $\widehat{P}_\ell$ represents an approximation using on level $\ell$.

If the weak convergence is

$$\mathbb{E}[\widehat{P}_\ell - P] = O(2^{-\alpha\,\ell}),$$

and $\widehat{Y}_\ell$ is an unbiased estimator for $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$, based on $N_\ell$ samples, with variance

$$\mathbb{V}[\widehat{Y}_\ell] = O(N_\ell^{-1}\, 2^{-\beta\,\ell}),$$

and expected cost

$$\mathbb{E}[C_\ell] = O(N_\ell\; 2^{\gamma\,\ell}), \quad \ldots$$

# Multilevel Monte Carlo

... then the finest level $L$ and the number of samples $N_\ell$ on each level can be chosen to achieve an RMS error of $\varepsilon$ at an expected cost

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

This talk is all about ways of increasing $\beta$ without increasing $\gamma$, so that we can improve the expected cost.

# Multilevel Monte Carlo

The standard estimator for most applications is

$$\widehat{Y}_\ell = N_\ell^{-1} \sum_{n=0}^{N_\ell} \left( \widehat{P}_\ell(\omega^{(n)}) - \widehat{P}_{\ell-1}(\omega^{(n)}) \right)$$

where $\omega^{(n)}$ indicates that the same stochastic sample is used on both the fine and coarse levels.

Exactly what this means depends on the application:

- same $W_t$ in SDE simulations
- same random numbers in PDEs with random initial/boundary data

Sometimes, as with continuous-time Markov processes, it is not clear what it means – in this case, Anderson & Higham constructed a very good coupling, but there are others which could be viewed as just as "natural".

# Multilevel Monte Carlo

The behaviour of the multilevel variance $V_\ell \equiv \mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ also depends on the output functional $P$.

There are particular problems when small differences between the fine and coarse simulations can sometimes lead to a large difference $\widehat{P}_\ell - \widehat{P}_{\ell-1}$.

The classic example of this is a digital option in finance, in which $P$ is a discontinuous function of the SDE path solution.

As well as leading to a slower decay in $V_\ell$, it also leads to a high kurtosis, so estimation of $V_\ell$ is often inaccurate.

# Conditional expectation

A digital option payoff is 0 or 1, depending on whether $S_T$ is above or below the strike $K$.

Using the Milstein approximation the strong error is $O(h)$, so

$$\widehat{S}_\ell - \widehat{S}_{\ell-1} = O(h_\ell).$$

An $O(h)$ fraction of fine/coarse pairs straddles the strike

$\Longrightarrow \; V_\ell = O(h_\ell)$ and the kurtosis is $O(h_\ell^{-1})$.

Both aspects are poor.

# Conditional expectation

The "fix" is to use E-M approximation for final timestep, then take conditional expectation over final fine path Brownian increment $\Delta W_N$.

For fine path

$$\widehat{S}_T = \widehat{S}_{T-h} + a_{T-h}\, h_\ell + b_{T-h}\, \Delta W_N,$$

$$\Longrightarrow \quad \widehat{P}_\ell = \Phi\left(\frac{\widehat{S}_{T-h} + a_{T-h}\, h_\ell - K}{b_{T-h}\sqrt{h_\ell}}\right)$$

while for the coarse path,

$$\widehat{S}_T = \widehat{S}_{T-2h} + 2\, a_{T-2h}\, h_\ell + b_{T-2h}\, (\Delta W_{N-1} + \Delta W_N),$$

$$\Longrightarrow \quad \widehat{P}_{\ell-1} = \Phi\left(\frac{\widehat{S}_{T-2h} + 2\, a_{T-2h}\, h_\ell + b_{T-2h}\, \Delta W_{N-1} - K}{b_{T-2h}\sqrt{h_\ell}}\right)$$

Further analysis proves $V_\ell = O(h_\ell^{3/2})$ and the kurtosis is $O(h_\ell^{-1/2})$.

# Conditional expectation

Note that this conditional expectation is a standard Monte Carlo technique for smoothing the payoff to enable pathwise sensitivity calculations. (L'Ecuyer, Glasserman?)

Another example is a down-and-out barrier options, where the option is knocked out if the path drops below a certain value.

Payoff can be smoothed by computing probability of this happening, conditional on computed path at discrete timesteps.

Again, this works well for both pathwise sensitivity analysis and MLMC.

(G, 2008, G, Burgos, 2012)

# Change of measure

Let's re-consider digital option.

For both the fine and coarse paths, we have conditional Gaussian distributions for $\widehat{S}_T$, with different means and variances.

Can perform a change of measure to the same Gaussian distribution, and then pick the same sample for both paths.

$$\widehat{P}_\ell - \widehat{P}_{\ell-1} = \widehat{P}(\widehat{S}_T)\,(R_\ell - R_{\ell-1})$$

where $R_\ell, R_{\ell-1}$ are the respective Radon-Nikodym derivatives.

Advantage: works well in multiple dimensions where often cannot evaluate the analytic conditional expectation.

(Burgos, 2014)

# Change of measure

Another example is a Merton-style jump-diffusion SDE with path-dependent jump rate $\lambda(S,t)$.

Problem is that coarse and fine paths will jump at different times; one might jump just before $T$, the other just after $\implies$ large $\widehat{P}_\ell - \widehat{P}_{\ell-1}$

Solution: use Glasserman & Merener thinning technique, over-sampling possible jump times using rate $\lambda_{sup} > \lambda(S,t)$, and combine with change of measure for identical acceptance/rejection decision for fine/coarse paths.

Leads to an estimator which looks like

$$\widehat{P}_\ell \, R_\ell - \widehat{P}_{\ell-1} \, R_{\ell-1}$$

and gives $V_\ell = O(h_\ell^2)$ when combined with Milstein discretisation.

(G, Xia, 2012)

# Splitting

Back once more to the multi-dimensional digital option.

The conditional expectation can be estimated numerically by averaging over a number of independent samples for the final Brownian increment.

$O(h^{-1})$ samples can be used without increasing the path cost significantly.

This is sufficient to reduce $V_\ell$ to about the same level as using the analytic conditional expectation.

Bonus: can use more accurate Milstein method for final timestep.

G, Burgos (2012) have also used splitting for MLMC for pathwise sensitivity analysis for put/call options.

# Splitting

A new application of splitting is for Feynman-Kac functionals arising for stopped diffusions – SDE calculations which terminate when the path leaves the domain.

The issue here is that when a fine path exits, there is an $O(h^{1/2})$ probability that the corresponding coarse path does not leave until much later.

This is solved by estimating a conditional expectation by splitting the coarse path into $O(h^{-1/2})$ independent sub-simulations.

$V_\ell$ is improved from $O(h_\ell^{1/2})$ to approximately $O(h_\ell)$.

# Zero-mean control variate

Very simple idea: suppose numerical analysis reveals that

$$\widehat{P}_\ell(\omega) - \widehat{P}_{\ell-1}(\omega) \approx Z(\omega)$$

where $\mathbb{E}[Z] = 0$.

Then changing the MLMC estimator to use

$$\widehat{P}_\ell(\omega) - \widehat{P}_{\ell-1}(\omega) - Z(\omega)$$

will reduce the variance without changing the expected value.

# Zero-mean control variate

Applying MLMC with $2^\ell$ timesteps to multi-dimensional SDEs with Milstein discretisation without Lévy area term, numerical analysis shows

$$\widehat{P}_\ell - \widehat{P}_{\ell-1} \approx \sum_n \sum_{j,k} a_{n,jk}(W) \, \Delta W_{n,j} \; Z_{n,k}$$

where coarse path has Brownian increments $\Delta W_n$, and $Z_n$ is used in Brownian Bridge construction of fine path increments.

$a_{n,jk}$ doesn't depend on $Z$, so expected value of sum is zero.

I was about to implement the zero-control CV when I realised I could generate a second fine path with $-Z$, leading to ...

# Antithetic sampling

Generalising the previous analysis shows we can cancel out the leading order error by using an antithetic pair of fine paths, with estimator

$$\frac{1}{2}\left( f(\widehat{S}^{(f)}(W_t)) + f(\widehat{S}^{(f)}(W_t^{anti}))\right) - f(\widehat{S}^{(c)}(W_t))$$

where within each coarse timestep $[t_n, t_n{+}h]$ the antithetic Brownian path is given by

$$W_t^{anti} \;=\; W_{t_n} + W_{t_n+h} - W_{2t_n+h-t}$$

It's antithetic in the sense that

$$\widehat{S}^{(f)}(W_t^{anti}) - \widehat{S}^{(c)}(W_t) \;\approx\; -\left(\widehat{S}^{(f)}(W_t) - \widehat{S}^{(c)}(W_t)\right)$$

I demonstrated $O(h^2)$ variance in 2007, but took several years for the numerical analysis (G, Szpruch, 2014).

# Antithetic sampling

When estimating $\mathbb{E}[\, f(\mathbb{E}[g(X,Z)\,|\,Z])\,]$, it is natural to use $2^\ell$ samples on level $\ell$ to estimate $\mathbb{E}[g(X,Z)\,|\,Z]$.

An appropriate antithetic coupling is then

$$f\left(\sum_{n=1}^{2^\ell} g(X^{(m,n)}, Z^{(m)})\right) - \tfrac{1}{2} f\left(\sum_{n=1}^{2^{\ell-1}} g(X^{(m,n)}, Z^{(m)})\right) - \tfrac{1}{2} f\left(\sum_{n=2^{\ell-1}+1}^{2^\ell} g(X^{(m,n)}, Z^{(m)})\right)$$

If $f\in C^2$, this improves $V_\ell$ from $O(2^{-\ell})$ to $O(2^{-2\ell})$.

(Haji-Ali, 2012; Chen, 2012?; Bujok, Hambly, Reisinger, 2013)

# Antithetic sampling

In SPDEs, stochastic fields are given by the Karhunen-Loève expansion:

$$k(\mathbf{x},\omega) = \sum_{n=1}^{\infty} \sqrt{\theta_n}\, \xi_n(\omega)\ f_n(\mathbf{x}),$$

where $\theta_n$, $f_n$ are eigenvalues / eigenfunctions of the correlation function.

The summation is truncated at $K_\ell$ which can vary with level, leading to the possibility of an antithetic pair of fine samples using

$$k(\mathbf{x},\omega) \ = \ \sum_{n=1}^{K_{\ell-1}} \sqrt{\theta_n}\, \xi_n(\omega)\ f_n(\mathbf{x}) \ \pm \ \sum_{n=K_{\ell-1}+1}^{K_\ell} \sqrt{\theta_n}\, \xi_n(\omega)\ f_n(\mathbf{x})$$

However, in our numerical tests we found no major benefit in doing this (G, Scheichl, Teckentrup, unpublished, 2012).

# Other techniques

- explicit smoothing of indicator function for CDF approximation,
  – tradeoff between bias and variance (G, Nagapetyan, Ritter, 2015)

- Malliavin integration by parts for digital options
  (Altmayer & Neuenkirch, 2015)

- adaptive timestepping (G, Lester, Whittle, 2015)
  - long-chain molecules
  - continuous-time Markov processes
  - reflected diffusions
  - CIR / Heston SDEs

- importance sampling
  - deep out-of-the-money put option (Gasparotto, 2015)

- Richard-Romberg extrapolation (Lemaire, Pagès, 2014)

- MIMC (Haji-Ali, Nobile, Tempone, 2015)

- MLQMC (G, Waterhouse, 2009, and others)

# Conclusions

Most applications use "plain" MLMC with no need for any of these tricks.

If you already have $\beta > \gamma$, why worry about improving $\beta$?

For applications needing an improved $\beta$, there is a growing toolkit of techniques to consider, with the same technique being used in widely differing applications.

In some cases, ideas have been taken from pathwise sensitivity analysis which also needs to smooth payoffs.

Webpages:
http://people.maths.ox.ac.uk/gilesm/mlmc.html
http://people.maths.ox.ac.uk/gilesm/mlmc/
http://people.maths.ox.ac.uk/gilesm/acta/
http://people.maths.ox.ac.uk/gilesm/mlmc_community.html