ICM 2018
RIO DE JANEIRO

Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 4 (3589–3608)

AN INTRODUCTION TO MULTILEVEL MONTE CARLO METHODS

Michael B. Giles

Abstract

In recent years there has been very substantial growth in stochastic modelling in many application areas, and this has led to much greater use of Monte Carlo methods to estimate expected values of output quantities from stochastic simulation. However, such calculations can be expensive when the cost of individual stochastic simulations is very high. Multilevel Monte Carlo greatly reduces the computational cost by performing most simulations with low accuracy at a correspondingly low cost, with relatively few being performed at high accuracy and a high cost.

This article reviews the key ideas behind the multilevel Monte Carlo method. Some applications are discussed to illustrate the flexibility and generality of the approach, and the challenges in its numerical analysis.

1 Introduction

Stochastic modelling and simulation is an important and growing area in applied mathematics and scientific computing. One large application area is in computational finance, in quantitative risk management and the pricing of financial derivatives. Another is Uncertainty Quantification in engineering and science, which has led to new journals and annual conferences.

When the dimensionality of the uncertainty (i.e. the number of uncertain input variables) is low, it can be appropriate to model the uncertainty using the Fokker-Planck PDE and use stochastic Galerkin, stochastic collocation or polynomial chaos methods Xiu and Karniadakis [2002], Babuška, Tempone, and Zouraris [2004], Babuška, Nobile, and Tempone [2010], and Gunzburger, Webster, and Zhang [2014]. When the level of uncertainty is low, and its effect is largely linear, then moment methods can be an efficient and accurate way in which to quantify the effects on uncertainty Putko, Taylor, Newman, and

The author acknowledges the financial support of the U.K. Engineering and Physical Sciences Research Council.

MSC2010: primary 65C05; secondary 65C30, 65C50, 60H35, 60H10, 60H15.

Green [2002]. However, when the uncertainty is high-dimensional and strongly nonlinear, Monte Carlo simulation often remains the preferred approach.

At its simplest, Monte Carlo simulation is extremely simple. To estimate $\mathbb{E}[P]$, the expected value of a scalar output quantity of interest, a simple Monte Carlo estimate is just an equally-weighted average of the values $P(\omega)$ for N independent samples ω coming from the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$,

$$N^{-1} \sum_{n=1}^{N} P(\omega^{(n)}).$$

The variance of this estimate is $N^{-1}\,\mathbb{V}[P]$, so the RMS (root-mean-square) error is $O(N^{-1/2})$ and an accuracy of ε requires $N = O(\varepsilon^{-2})$ samples. This is the weakness of Monte Carlo simulation; its computational cost can be very high, particularly when each sample $P(\omega)$ might require the approximate solution of a PDE, or a computation with many timesteps.

One approach to addressing this high cost is the use of Quasi-Monte Carlo (QMC) methods, in which the samples are not chosen randomly and independently, but are instead selected very carefully to reduce the error. In the best cases, the error may be $O(N^{-1})$, up to logarithmic terms, giving a very substantial reduction in the number of samples required for a given accuracy Dick, Kuo, and Sloan [2013].

In this article, we cover a different approach to improving the computational efficiency, the multilevel Monte Carlo (MLMC) method. This is important when the cost of computing the individual samples is very high, but it is possible to compute approximate values at a much lower cost. We also briefly discuss the combination of MLMC with QMC.

This article provides only a short introduction to the subject and some of the corresponding literature. For a more comprehensive overview of multilevel Monte Carlo methods, the author has recently written a 70-page review with a much more extensive list of references Giles [2015]. There is also a webpage web-page with a list of active research groups and their publications.

2 Multilevel Monte Carlo

2.1 MLMC with exact simulation. The key idea in Multilevel Monte Carlo is also very simple. Suppose we are interested in estimating $\mathbb{E}[P_L(\omega)]$, and it is possible to exactly simulate $P_L(\omega)$ but it is very costly. Suppose also that there is a sequence $P_0(\omega), \ldots, P_{L-1}(\omega)$ which approximates $P_L(\omega)$ with increasing accuracy, but also increasing cost. In this case, instead of directly estimating $\mathbb{E}[P_L]$ we can use the trivial identity

$$\mathbb{E}[P_L] = \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell - P_{\ell-1}],$$

to construct the following unbiased estimator for $\mathbb{E}[P_L]$,

$$N_0^{-1}\sum_{n=1}^{N_0} P_0^{(0,n)} + \sum_{\ell=1}^{L}\left\{N_\ell^{-1}\sum_{n=1}^{N_\ell}\left(P_\ell^{(\ell,n)} - P_{\ell-1}^{(\ell,n)}\right)\right\}$$

where $P_\ell^{(\ell,n)}$ is shorthand for $P_\ell(\omega^{(\ell,n)})$, with the inclusion of the level ℓ in the superscript (ℓ,n) indicating that independent samples are used at each level of correction. The important point is that by using the same $\omega^{(\ell,n)}$ we aim to ensure that $P_\ell^{(\ell,n)} - P_{\ell-1}^{(\ell,n)}$ is small for larger values of ℓ, so that relatively few samples are needed on finer levels to estimate $\mathbb{E}[P_\ell - P_{\ell-1}]$.

If we define C_0, V_0 to be the cost and variance of one sample of P_0, and C_ℓ, V_ℓ to be the cost and variance of one sample of $P_\ell - P_{\ell-1}$, then the overall cost and variance of the multilevel estimator is $\sum_{\ell=0}^{L} N_\ell\, C_\ell$ and $\sum_{\ell=0}^{L} N_\ell^{-1}\, V_\ell$, respectively. Ignoring the fact that the N_ℓ are integers, for a fixed cost the variance is minimised by choosing N_ℓ to minimise

$$\sum_{\ell=0}^{L}\left(N_\ell^{-1}\, V_\ell + \mu^2 N_\ell\, C_\ell\right)$$

for some value of the Lagrange multiplier μ^2, which gives

$$N_\ell = \mu\sqrt{V_\ell / C_\ell}. \tag{1}$$

To achieve an overall variance of ε^2 then requires that $\mu = \varepsilon^{-2}\ \sum_{\ell=0}^{L}\sqrt{V_\ell\, C_\ell}$.

Rounding up (1) to the nearest integer improves the overall variance and increases the cost by at most $\sum_{\ell=0}^{L} C_\ell$, so that the variance of ε^2 can be achieved at a total cost which is bounded by

$$C = \varepsilon^{-2}\left(\sum_{\ell=0}^{L}\sqrt{V_\ell\, C_\ell}\right)^2 + \sum_{\ell=0}^{L} C_\ell. \tag{2}$$

It is important to note whether the product $V_\ell\, C_\ell$ increases or decreases with ℓ. If it increases, then the dominant contribution to the cost comes from $V_L\, C_L$ and we have $C \approx \varepsilon^{-2} V_L\, C_L$, whereas if it decreases then the dominant contribution comes from $V_0\, C_0$ and $C \approx \varepsilon^{-2} V_0\, C_0$. This contrasts with the standard Monte Carlo cost of approximately $\varepsilon^{-2} V_0\, C_L$, assuming that the cost of computing P_L is similar to the cost of computing $P_L - P_{L-1}$ and $\mathbb{V}[P_L] \approx \mathbb{V}[P_0]$.

This shows that in the first case the MLMC cost is reduced by factor V_L / V_0, corresponding to the ratio of the variances $\mathbb{V}[P_L - P_{L-1}]$ and $\mathbb{V}[P_L]$, whereas in the second case it

is reduced by factor C_0/C_L, the ratio of the costs of computing P_0 and $P_L - P_{L-1}$. If the product $V_\ell\, C_\ell$ does not vary with level, then the total cost is approximately $\varepsilon^{-2} L^2\, V_0\, C_0 = \varepsilon^{-2} L^2\, V_L\, C_L$.

2.2 MLMC with inexact simulation. In almost all MLMC applications, it is not possible to exactly simulate the quantity of interest $P(\omega)$, often because the calculation of $P(\omega)$ requires the approximate solution of a PDE or SDE. Instead, what we have is an infinite sequence of approximations $P_\ell, \ell = 0, 1, \ldots$ which approximate P with increasing accuracy and cost. If Y is an approximation to $\mathbb{E}[P]$, then a standard piece of theory gives the MSE (mean square error) as

$$\text{MSE} \;\equiv\; \mathbb{E}[\,(Y - \mathbb{E}[P])^2] \;=\; \mathbb{V}[Y] + (\,\mathbb{E}[Y] - \mathbb{E}[P]\,)^2. \tag{3}$$

If Y is now the multilevel estimator

$$Y = \sum_{\ell=0}^{L} Y_\ell, \quad Y_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} (P_\ell^{(\ell,n)} - P_{\ell-1}^{(\ell,n)}), \tag{4}$$

with $P_{-1} \equiv 0$, then

$$\mathbb{E}[Y] = \mathbb{E}[P_L], \quad \mathbb{V}[Y] = \sum_{\ell=0}^{L} N_\ell^{-1} V_\ell, \quad V_\ell \equiv \mathbb{V}[P_\ell - P_{\ell-1}]. \tag{5}$$

To ensure that the MSE is less than ε^2, it is sufficient to ensure that $\mathbb{V}[Y]$ and $(\mathbb{E}[P_L - P])^2$ are both less than $\frac{1}{2}\varepsilon^2$. Combining this idea with a geometric sequence of levels in which the cost increases exponentially with level, while both the weak error $\mathbb{E}[P_L - P]$ and the multilevel correction variance V_ℓ decrease exponentially, leads to the following theorem:

Theorem 1. *Let P denote a random variable, and let P_ℓ denote the corresponding level ℓ numerical approximation. If there exist independent estimators Y_ℓ based on N_ℓ Monte Carlo samples, each with expected cost C_ℓ and variance V_ℓ, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \ge \frac{1}{2}\min(\beta, \gamma)$ and*

i) $\left|\mathbb{E}[P_\ell - P]\right| \;\le\; c_1\, 2^{-\alpha \ell}$

ii) $\mathbb{E}[Y_\ell] \;= \begin{cases} \mathbb{E}[P_0], & \ell = 0 \\ \mathbb{E}[P_\ell - P_{\ell-1}], & \ell > 0 \end{cases}$

iii) $V_\ell \;\le\; c_2\, 2^{-\beta \ell}$

iv) $C_\ell \;\le\; c_3\, 2^{\gamma \ell}$,

then there exists a positive constant c_4 *such that for any* $\varepsilon < e^{-1}$ *there are values* L *and* N_ℓ *for which the multilevel estimator*

$$Y = \sum_{\ell=0}^{L} Y_\ell,$$

has a mean-square-error with bound

$$MSE \equiv \mathbb{E}\left[\left(Y - \mathbb{E}[P]\right)^2\right] < \varepsilon^2$$

with an expected computational complexity C *with bound*

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

The statement of the theorem is a slight generalisation of the original theorem in Giles [2008b]. It corresponds to the theorem and proof in Cliffe, Giles, Scheichl, and Teckentrup [2011], except for the minor change to expected costs to allow for applications in which the simulation cost of individual samples is itself random. Note that if condition iii) is tightened slightly to be a bound on $\mathbb{E}[(P_\ell - P_{\ell-1})^2]$, which is usually the quantity which is bounded in numerical analysis, then it would follow immediately that $\alpha \ge \frac{1}{2}\beta$.

The essence of the proof is very straightforward. If we have $V_\ell = O(2^{-\beta\ell})$ and $C_\ell = O(2^{\gamma\ell})$, then the analysis in Section 2.1 shows that the optimal number of samples N_ℓ on level ℓ is proportional to $2^{-(\beta+\gamma)\ell/2}$, and therefore the total cost on level ℓ is proportional to $2^{(\gamma-\beta)\ell/2}$. The result then follows from the requirement that L is chosen so that $(\,\mathbb{E}[Y] - \mathbb{E}[P]\,)^2 < \frac{1}{2}\varepsilon^2$, and the constant of proportionality for N_ℓ is chosen so that $\mathbb{V}[Y] < \frac{1}{2}\varepsilon^2$.

The result of the theorem merits some discussion. In the case $\beta > \gamma$, the dominant computational cost is on the coarsest levels where $C_\ell = O(1)$ and $O(\varepsilon^{-2})$ samples are required to achieve the desired accuracy. This is the standard result for a Monte Carlo approach using i.i.d. samples; to do better would require an alternative approach such as the use of Latin hypercube sampling or quasi-Monte Carlo methods.

In the case $\beta < \gamma$, the dominant computational cost is on the finest levels. Because of condition i), we have $2^{-\alpha L} = O(\varepsilon)$, and hence $C_L = O(\varepsilon^{-\gamma/\alpha})$. If $\beta = 2\alpha$, which is usually the best that can be achieved since typically $\mathbb{V}[P_\ell - P_{\ell-1}]$ is similar in magnitude to $\mathbb{E}[(P_\ell - P_{\ell-1})^2]$ which is greater than $(\mathbb{E}[P_\ell - P_{\ell-1}])^2$, then the total cost is $O(C_L)$, corresponding to $O(1)$ samples on the finest level, which is the best that can be achieved.

The dividing case $\beta = \gamma$ is the one for which both the computational effort, and the contributions to the overall variance, are spread approximately evenly across all of the levels; the $|\log\varepsilon|^2$ term corresponds to the L^2 factor in the corresponding discussion at the end of Section 2.1.

One comment on the Theorem is that it assumes lots of properties, and then from these determines relatively easily some conclusions for the efficiency of the MLMC approach. In real applications, the tough challenge is in proving that the assumptions are valid, and in particular determining the values of the parameters α, β, γ. Furthermore, the Theorem assumes knowledge of the constants c_1, c_2, c_3. In practice, c_1 and c_2 are almost never known, and instead have to be estimated based on empirical estimates of the weak error and the multilevel correction variance.

Equation (4) gives the natural choice for the multilevel correction estimator Y_ℓ. However, the multilevel theorem allows for the use of other estimators, provided they satisfy the restriction of condition ii) which ensures that $\mathbb{E}[Y] = \mathbb{E}[P_L]$. Examples of this will be given later in this article. In each case, the objective in constructing a more complex estimator is to achieve a greatly reduced variance $\mathbb{V}[Y_\ell]$ so that fewer samples are required.

2.3 Randomised MLMC for unbiased estimation. A very interesting extension was introduced by Rhee & Glynn in Rhee and Glynn [2015]. Rather than choosing the finest level of simulation L based on the desired accuracy, and then using the optimal number of samples on each level based on an estimate of the variance, the "single term" estimator in Rhee and Glynn [ibid.] instead uses N samples in total, and for each sample the level on which the simulation is performed is selected randomly, with level ℓ being chosen with probability p_ℓ.

The estimator is

$$Y = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{p_{\ell^{(n)}}} (P^{(n)}_{\ell^{(n)}} - P^{(n)}_{\ell^{(n)}-1})$$

with the level $\ell^{(n)}$ for each sample being selected randomly with the relevant probability. Alternatively, their estimator can be expressed as

$$Y = \sum_{\ell=0}^{\infty} \left(\frac{1}{p_\ell\, N} \sum_{n=1}^{N_\ell} (P^{(n)}_\ell - P^{(n)}_{\ell-1}) \right).$$

where N_ℓ, the number of samples from level ℓ, is a random variable with

$$\sum_{\ell=0}^{\infty} N_\ell = N, \quad \mathbb{E}[N_\ell] = p_\ell N.$$

Note that in this form it is very similar in appearance to the standard MLMC estimator.

The beauty of their estimator is that it is naturally unbiased, since

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{1}{p_{\ell'}}(P_{\ell'} - P_{\ell'-1})\right]$$
$$= \sum_{\ell=0}^{\infty} p_\ell \, \mathbb{E}\left[\frac{1}{p_{\ell'}}(P_{\ell'} - P_{\ell'-1}) \mid \ell' = \ell\right] = \sum_{\ell=0}^{\infty} \mathbb{E}\left[P_\ell - P_{\ell-1}\right] = \mathbb{E}[P].$$

The choice of probabilities p_ℓ is crucial. For both the variance and the expected cost to be finite, it is necessary that

$$\sum_{\ell=0}^{\infty} \frac{1}{p_\ell} V_\ell < \infty, \quad \sum_{\ell=0}^{\infty} p_\ell \, C_\ell < \infty.$$

Under the conditions of the MLMC Theorem, this is possible when $\beta > \gamma$ by choosing $p_\ell \propto 2^{-(\gamma+\beta)\ell/2}$, so that

$$\frac{1}{p_\ell} V_\ell \propto 2^{-(\beta-\gamma)\ell/2}, \quad p_\ell \, C_\ell \propto 2^{-(\beta-\gamma)\ell/2}.$$

It is not possible when $\beta \leq \gamma$, and for these cases the estimators constructed in Rhee and Glynn [ibid.] have infinite expected cost.

2.4 Multilevel Richardson-Romberg extrapolation. Richardson extrapolation is a very old technique in numerical analysis. Given a numerical approximation P_h based on a discretisation parameter h which leads to an error

$$P_h - P = a\, h^\alpha + O(h^{2\alpha}),$$

it follows that $P_{2h} - P = a\,(2h)^\alpha + O(h^{2\alpha})$, and hence the extrapolated value

$$\widetilde{P} = \frac{2^\alpha}{2^\alpha - 1} P_h - \frac{1}{2^\alpha - 1} P_{2h}$$

satisfies $\widetilde{P} - P = O(h^{2\alpha})$. Lemaire & Pagès take this approach much further Lemaire and Pagès [2017]. Assuming that the weak error has a regular expansion

$$\mathbb{E}[P_\ell] - \mathbb{E}[P] = \sum_{n=1}^{L} a_n 2^{-n\alpha\ell} + o(2^{-\alpha\ell L}),$$

they first determine the unique set of weights w_ℓ, $\ell = 0, 1, \ldots, L$ such that

$$\sum_{\ell=0}^{L} w_\ell = 1, \quad \sum_{\ell=0}^{L} w_\ell \, 2^{-n\alpha\ell} = 0, \quad n = 1, \ldots, L,$$

so that

$$\left(\sum_{\ell=0}^{L} w_\ell\, \mathbb{E}[P_\ell]\right) - \mathbb{E}[P] \;\equiv\; \sum_{\ell=0}^{L} w_\ell\, (\mathbb{E}[P_\ell] - \mathbb{E}[P]) \;=\; o(2^{-\alpha L^2}).$$

Next, they re-arrange terms to give

$$\sum_{\ell=0}^{L} w_\ell\, \mathbb{E}[P_\ell] = \sum_{\ell=0}^{L} v_\ell\, \mathbb{E}[P_\ell - P_{\ell-1}]$$

where as usual $P_{-1} \equiv 0$, and the coefficients v_ℓ are defined by $w_\ell = v_\ell - v_{\ell+1}$, with $v_{L+1} \equiv 0$, and hence

$$v_\ell = \sum_{\ell'=\ell}^{L} w_{\ell'}.$$

This leads to their Multilevel Richardson-Romberg extrapolation estimator,

$$Y = \sum_{\ell=0}^{L} Y_\ell, \quad Y_\ell = v_\ell N_\ell^{-1} \sum_n (P_\ell^{(\ell,n)} - P_{\ell-1}^{(\ell,n)}).$$

Because the remaining error is $o(2^{-\alpha L^2})$, rather than the usual $O(2^{-\alpha L})$, it is possible to obtain the usual $O(\varepsilon)$ weak error with a value of L which is approximately the square root of the usual value. Hence, in the case $\beta = \gamma$ they prove that the overall cost is reduced to $O(\varepsilon^{-2} |\log \varepsilon|)$, while for $\beta < \gamma$ the cost is reduced much more to $O(\varepsilon^{-2} 2^{(\gamma-\beta)\sqrt{|\log_2 \varepsilon|/\alpha}})$. This analysis is supported by numerical results which demonstrate considerable savings Lemaire and Pagès [2017], and therefore this is a very useful extension to the standard MLMC approach when $\beta \le \gamma$.

2.5 Multi-Index Monte Carlo. In standard MLMC, there is a one-dimensional set of levels, with a scalar level index ℓ, although in some applications changing ℓ can change more than one aspect of the computation (such as both timestep and spatial discretisation in a parabolic SPDE application, or timestep and number of sub-samples in a nested simulation). Multi-Index Monte Carlo developed by Haji-Ali, Nobile, and Tempone [2016] generalises this to "levels" being defined in multiple directions, so that the level "index" $\boldsymbol{\ell}$ is now a vector of integer indices. This is illustrated in Figure 1 for a 2D MIMC application.

In MLMC, if we define the backward difference $\Delta P_\ell \equiv P_\ell - P_{\ell-1}$ with $P_{-1} \equiv 0$, as usual, then the telescoping sum which lies at the heart of MLMC is

$$\mathbb{E}[P] = \sum_{\ell \ge 0} \mathbb{E}[\Delta P_\ell].$$

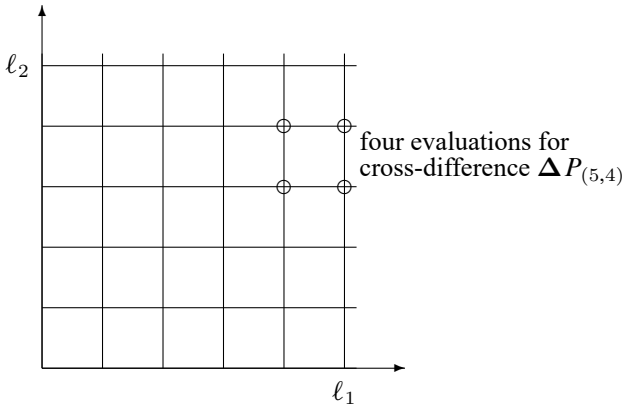

Figure 1: "Levels" in 2D multi-index Monte Carlo application

Generalising this to D dimensions, we can first define a backward difference operator in one particular dimension, $\Delta_d P_{\boldsymbol{\ell}} \equiv P_{\boldsymbol{\ell}} - P_{\boldsymbol{\ell}-\boldsymbol{e}_d}$ where $\boldsymbol{e}_d$ is the unit vector in direction d. Then defining the cross-difference

$$\boldsymbol{\Delta} P_{\boldsymbol{\ell}} \equiv \left(\prod_{d=1}^{D} \Delta_d \right) P_{\boldsymbol{\ell}}$$

the telescoping sum becomes

$$\mathbb{E}[P] = \sum_{\boldsymbol{\ell} \geq 0} \mathbb{E}[\boldsymbol{\Delta} P_{\boldsymbol{\ell}}].$$

As an example, Figure 1 marks the four locations at which $P_{\boldsymbol{\ell}}$ must be computed to determine the value of $\boldsymbol{\Delta} P_{(5,4)}$ in the 2D application.

Instead of summing $\mathbb{E}[\boldsymbol{\Delta} P_{\boldsymbol{\ell}}]$ over the full domain $\boldsymbol{\ell} \geq 0$, the sum is instead truncated to a summation region $\mathfrak{L}$. It might seem natural that this should be rectangular, as illustrated on the left in Figure 2, so that

$$\sum_{\boldsymbol{\ell} \in \mathfrak{L}} \boldsymbol{\Delta} P_{\boldsymbol{\ell}} = P_{\boldsymbol{L}}$$

where $\boldsymbol{L}$ is the outermost point on the rectangle. However, Haji-Ali, Nobile, and Tempone [ibid.] proves that it is often better to use a region $\mathfrak{L}$ of the form $\boldsymbol{\ell} \cdot \mathbf{n} \leq L$ for a particular choice of direction vector $\mathbf{n}$ with strictly positive components. In 2D, this corresponds to a triangular region, as illustrated on the right in Figure 2. This is very similar to the use of the sparse grid combination technique in high-dimensional PDE approximations Bungartz

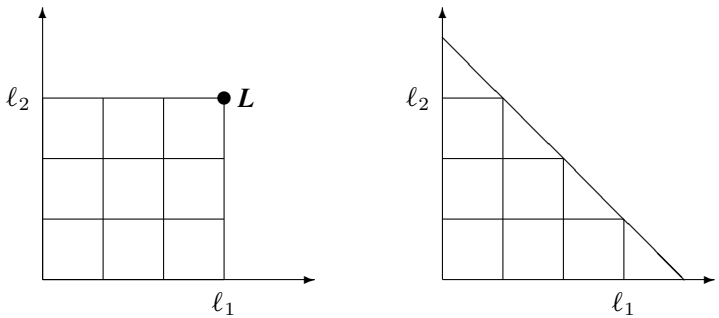

Figure 2: Two choices of 2D MIMC summation region $\mathfrak{L}$.

and Griebel [2004], and indeed MIMC can be viewed as a combination of this approach with Monte Carlo sampling.

The benefits of MIMC over the standard MLMC can be very substantial. They are perhaps best illustrated by an elliptic PDE or SPDE example, in which D corresponds to the number of spatial dimensions. Using the standard MLMC approach, β, the rate of convergence of the multilevel variance, will usually be independent of D, but γ, the rate of increase in the computational cost, will increase at least linearly with D. Therefore, in a high enough dimension we will have $\beta \le \gamma$ and therefore the overall computational complexity will be less (often much less) than the optimal $O(\varepsilon^{-2})$. However, using MIMC it is possible to achieve the optimal complexity independent of the value of D. Hence, in the same way that sparse grids offer the possibility of dimension-independent complexity for deterministic PDE applications, MIMC offers the possibility of dimension-independent complexity for SPDEs and other high-dimensional stochastic applications.

MIMC is the first multi-dimensional generalisation of MLMC, but it is not the only one. Other possibilities include using sparse grid samples within a standard MLMC formulation, and nested MLMC in which there is an outer MLMC telescoping sum in one "direction", and then each of its expectations is expressed as an MLMC telescoping sum in a second "direction". These ideas, and the inclusion of quasi-Monte Carlo sampling, are discussed in Giles, Kuo, and Sloan [2018].

2.6 MLQMC. The final part of this theory section concerns the use of quasi-Monte Carlo (QMC) sampling in place of standard Monte Carlo. The key change in MLQMC is that N_ℓ is now the size of a set of QMC points used on level ℓ. This set of points is not constructed randomly and independently, but is instead constructed very carefully to provide a relatively uniform coverage of a unit hypercube integration region which is then mapped into the required domain, for example by mapping unit interval uniformly

distributed random variables to standard Normal random variables Dick, Kuo, and Sloan [2013]. In the best cases, this results in the numerical integration error being approximately $O(N_\ell^{-1})$ rather than the usual $O(N_\ell^{-1/2})$ error which comes from Monte Carlo sampling.

Using just one set of N_ℓ points gives good accuracy, but no confidence interval. To regain a confidence interval one uses randomised QMC in which the set of points is collectively randomised in a way which ensures that the averages obtained from each set of points are independent. Using 32 randomisations, for example, yields 32 set averages for the quantity of interest, Y_ℓ, and from these the variance of their average, V_ℓ, can be estimated in the usual way.

Since V_ℓ is now defined to be the variance of the average of the set averages, the aim now is to choose the N_ℓ in a way which ensures that

$$\sum_{\ell=0}^{L} V_\ell \le \tfrac{1}{2}\varepsilon^2. \tag{6}$$

We can not use the same Lagrange multiplier approach as before to determine the optimal N_ℓ. Instead, we note that many QMC methods work naturally with N_ℓ as a power of 2. Doubling N_ℓ will usually eliminate a large fraction of the variance, so the greatest reduction in total variance relative to the additional computational effort is achieved by doubling N_ℓ on the level ℓ^* given by

$$\ell^* = \arg\max_\ell \frac{V_\ell}{N_\ell\, C_\ell}. \tag{7}$$

This approach was first developed in Giles and Waterhouse [2009], with application to stochastic differential equations (SDEs) using QMC samples based on extensible rank-1 lattices Dick, Pillichshammer, and Waterhouse [2008]. QMC is known to be most effective for low-dimensional applications, and the numerical results were very encouraging for SDE applications in which the dominant computational cost was on the coarsest levels of resolution. However, there was no supporting theory for this research. More recently, there has been considerable research on applications and the underlying theoretical foundations for MLQMC methods applied to PDEs with stochastic coefficients Niu, Hickernell, Müller-Gronbach, and Ritter [2011], Kuo, Schwab, and Sloan [2015], and Dick, Kuo, and Sloan [2013]. These theoretical developments are very encouraging, showing that under certain conditions they lead to multilevel methods with a complexity which is $O(\varepsilon^{-p})$ with $p < 2$.

3 SDEs

The original multilevel path simulation paper Giles [2008b] treated stochastic differential equations

$$\mathrm{d}S_t = a(S_t,t)\,\mathrm{d}t + b(S_t,t)\,\mathrm{d}W,$$

using the simple Euler-Maruyama discretisation with a uniform timestep h and Brownian increments ΔW_n,

$$\widehat{S}_{(n+1)h} = \widehat{S}_{nh} + a(\widehat{S}_{nh},nh)\,h + b(\widehat{S}_{nh},nh)\,\Delta W_n.$$

The multilevel Monte Carlo implementation is very simple. On level ℓ, the uniform timestep is taken to be $h_\ell = M^{-\ell}h_0$, for some integer M. The timestep h_0 on the coarsest level is often taken to be the interval length T, so that there is just one timestep for the entire interval, but this is not required, and in some applications using such a large timestep may lead to numerical results which are so inaccurate that they are not helpful in reducing the variance.

The multilevel coupling is achieved by using the same underlying driving Brownian path for the coarse and fine paths; this is accomplished by summing the Brownian increments for the fine path timesteps to obtain the Brownian increments for the coarse timesteps. The multilevel estimator is then the natural one defined in (4), with the specific payoff approximation P_ℓ depending on the particular application.

Provided the SDE satisfies the usual conditions (see Theorem 10.2.2 in Kloeden and Platen [1992]), the strong error for the Euler discretisation with timestep h is $O(h^{1/2})$, so that

$$\mathbb{E}\left[\sup_{[0,T]} \|S_t - \widehat{S}_t\|^2\right] = O(h),$$

where $\widehat{S}_t$ is a piecewise constant interpolation of the discrete values $\widehat{S}_{nh}$.

For financial options for which the payoff is a Lipschitz function of S_t, with constant K, we have

$$\mathbb{V}[P-P_\ell] \;\le\; \mathbb{E}[(P-P_\ell)^2] \;\le\; K^2\,\mathbb{E}\left[\sup_{[0,T]} \|S_t - \widehat{S}_t\|^2\right],$$

where K is the Lipschitz constant, and

$$V_\ell \;\equiv\; \mathbb{V}[P_\ell - P_{\ell-1}] \;\le\; 2\left(\mathbb{V}[P-P_\ell] + \mathbb{V}[P-P_{\ell-1}]\right),$$

and hence $V_\ell = O(h_\ell)$.

	Euler-Maruyama		Milstein	
option	numerics	analysis	numerics	analysis
Lipschitz	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
Asian	$O(h)$	$O(h)$	$O(h^2)$	$O(h^2)$
lookback	$O(h)$	$O(h)$	$O(h^2)$	$o(h^{2-\delta})$
barrier	$O(h^{1/2})$	$o(h^{1/2-\delta})$	$O(h^{3/2})$	$o(h^{3/2-\delta})$
digital	$O(h^{1/2})$	$O(h^{1/2}\lvert\log h\rvert)$	$O(h^{3/2})$	$o(h^{3/2-\delta})$

Table 1: Observed and theoretical convergence rates for the multilevel correction variance for scalar SDEs, using the Euler-Maruyama and Milstein discretisations. δ is any strictly positive constant.

If $h_\ell = 4^{-\ell} h_0$, as in the numerical examples in Giles [2008b], then this gives $\alpha = 2$, $\beta = 2$ and $\gamma = 2$. This is found to be better than using $h_\ell = 2^{-\ell} h_0$ with twice as many timesteps on each successive level, which gives $\alpha = 1$, $\beta = 1$ and $\gamma = 1$. In either case, Theorem 1 gives the complexity to achieve a RMS error of ε to be $O(\varepsilon^{-2} |\log \varepsilon|^2)$, which has been proved to be optimal for a class of Lipschitz path-dependent functions Creutzig, Dereich, Müller-Gronbach, and Ritter [2009].

The more accurate Milstein approximation achieves first order strong convergence, giving $V_\ell = O(h_\ell^2)$ for certain Lipschitz payoff functions Giles [2008a]. Further challenges are encountered with digital and barrier options for which the payoff functions are a discontinuous function of the path S_t. In such cases, a small difference between the coarse and fine path approximations can nevertheless produce a large value for $P_\ell - P_{\ell-1}$. Techniques have been developed to partially address this Giles [ibid.]. Table 1 summarises the observed variance convergence rate in numerical experiments for a number of different financial options; the Asian option is based on the average value of the underlying asset, the lookback is based on its maximum or minimum value, the barrier is a discontinuous function of the maximum or minimum, and the digital is a discontinuous function of the final value. The table also displays the theoretical numerical analysis results which have been obtained Avikainen [2009], Giles, Higham, and Mao [2009], and Giles, Debrabant, and Rößler [2013].

There is insufficient space in this article to discuss in detail the many other extensions and generalisations in applying MLMC to SDEs. We will simply mention a few, and further details and references can be obtained from Giles [2015].

There are difficulties in implementing the Milstein approximation for multi-dimensional SDEs when they require the simulation of Lévy areas. In this case, there is a special "antithetic" MLMC estimator which eliminates to leading order the error due to the omission of the Lévy areas.

The classic analysis of SDE approximations assumes that the drift function $a(S_t, t)$ and volatility $b(S_t, t)$ are both globally Lipschitz functions of S_t. Some important applications have drift functions, such as $-S_t - S_t^3$, which are only locally Lipschitz. These require adaptive timestepping, or some other technique, to maintain numerical stability, and this causes additional difficulties for MLMC.

There are also extensions to jump-diffusion SDEs, in which there is an additional Poisson jump process, and SDEs driven by increments of a more general Lévy process instead of a Brownian motion.

Finally, in some applications the output quantity of interest is not the expected value of a scalar quantity but a function such as the density of that output, or the cumulative distribution function, $\text{CDF}(x) = \mathbb{P}[X < x] = \mathbb{E}[H(x-X)]$ where $H(x)$ is the Heaviside function.

4 PDEs and SPDEs

Applying MLMC to stochastic PDEs and PDEs with random data or stochastic coefficients was a natural follow-on to the use for SDEs. Indeed, there was more scope for computational savings because the cost of a single sample increases more rapidly with grid resolution for SPDEs with higher space-time dimension. There has been a variety of papers on elliptic Barth, Schwab, and Zollinger [2011] and Cliffe, Giles, Scheichl, and Teckentrup [2011], parabolic Barth, Lang, and Schwab [2013] and Giles and Reisinger [2012] and hyperbolic Mishra, Schwab, and Šukys [2012] PDEs and SPDEs, as well as for mixed elliptic-hyperbolic systems Efendiev, Iliev, and Kronsbein [2013] and Müller, Jenny, and Meyer [2013].

In almost all of this work, the construction of the multilevel estimator is quite natural, using a geometric sequence of grids and the natural estimators for $P_\ell - P_{\ell-1}$. It is the numerical analysis of the variance of the multilevel estimator which is often very challenging, but in the simplest cases it can be straightforward.

Consider, for example, a D-dimensional elliptic PDE, $\nabla^2 u = f(x, \omega)$, where the r.h.s. forcing term is stochastic, depending on a number of random variables. If f is sufficiently smooth, then a standard piecewise linear finite element method might achieve second order accuracy for a large class of output functionals. Hence, if the computational grid has spacing proportional to $2^{-\ell}$ in each direction then the error would be $O(2^{-2\ell})$ and the MLMC variance V_ℓ would be $O(2^{-4\ell})$. Using an efficient multigrid solver, the computational cost would be approximately proportional to the total number of grid points, which is $O(2^{D\ell})$. Hence this MLMC application has $\alpha = 2$, $\beta = 4$, $\gamma = D$.

This means that a RMS accuracy of ε can be achieved at $O(\varepsilon^{-2})$ cost for $D < 4$, while the cost is $O(\varepsilon^{-2} |\log \varepsilon|^2)$ for $D = 4$, and $O(\varepsilon^{-D/2})$ for $D > 4$. By comparison, to achieve

an accuracy of ε for a single deterministic calculation requires $2^{-2\ell} \sim \varepsilon$, and hence the cost is $O(\varepsilon^{-D/2})$, which for $D > 4$ is the same order as the cost of estimating the expectation in the random setting.

The largest amount of research on multilevel for SPDEs has been for elliptic PDEs with random coefficients. The PDE typically has the form

$$-\nabla \cdot (\kappa(\mathbf{x}, \omega) \nabla p(\mathbf{x}, \omega)) = 0, \qquad \mathbf{x} \in D.$$

with Dirichlet or Neumann boundary conditions on the boundary ∂D. For sub-surface flow problems, such as the modelling of groundwater flow in nuclear waste repositories, the diffusivity (or permeability) κ is often modelled as a lognormal random field, i.e. $\log \kappa$ is a Gaussian field with a uniform mean and a covariance function $R(\mathbf{x}, \mathbf{y})$. Samples of $\log \kappa$ can be provided by a Karhunen-Loève expansion:

$$\log \kappa(\mathbf{x}, \omega) = \sum_{n=0}^{\infty} \sqrt{\theta_n}\, \xi_n(\omega)\, f_n(\mathbf{x}),$$

where θ_n and f_n are the eigenvalues and eigenfunctions defined by

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

and ξ_n are independent unit Normal random variables. However, it can be more efficient to generate them using a circulant embedding technique based on the use of FFTs.

There is no space to detail the huge range of other applications. The one other point to note here is that they are not all based on a geometric hierarchy of approximations. A non-geometric example is the use of a reduced basis approximation in which the approximate solution u at a set of discrete grid points for an arbitrary set of random inputs ω is written as

$$u(\omega) = \sum_{k=1}^{K} c_k(\omega)\, u_k$$

where the c_k are a set of scalar coefficients, and the u_k are a fixed set of vectors, for example corresponding to solutions of the PDE for particular values of the random inputs. In such a reduced basis approximation, the accuracy improves if one increases the number of basis functions, K, but so too does the cost. Neither behaves in a simple way such that there is an obvious way in which to prescribe K_ℓ as a function of level, and therefore numerical optimisation can be used instead Vidal-Codina, Nguyen, Giles, and Peraire [2015].

5 Continuous-time Markov chains

A very interesting and important application of MLMC has been to continuous-time Markov Chain simulation Anderson and Higham [2012]. Such models arise in the context of stochastic chemical reactions, when species concentrations are extremely low and so stochastic effects become significant. When there is just one chemical species which is being spontaneously created at a rate which depends on the current number of molecules x, the "tau-leaping" method (which is essentially the Euler-Maruyama method, approximating the reaction rate as being constant throughout the timestep) gives the discrete equation

$$x_{n+1} = x_n + P(h\,\lambda(x_n)),$$

where the state x_n is an integer, h is the timestep, $\lambda(x_n)$ is the reaction rate (or propensity function), and $P(t)$ represents a unit-rate Poisson random variable over time interval $[0,t]$. If this equation defines the fine path in the multilevel simulation, then the coarse path, with double the timestep, is given by

$$x^c_{n+2} = x^c_n + P(2h\,\lambda(x^c_n))$$

for even timesteps n.

The question then is how to couple the coarse and fine path simulations in a MLMC calculation. The key observation in Anderson and Higham [ibid.], is that for any $t_1, t_2 > 0$, the sum of two independent Poisson variates $P(t_1)$, $P(t_2)$ is equivalent in distribution to $P(t_1+t_2)$. Based on this, the first step is to express the coarse path Poisson variate as the sum of two independent Poisson variates, $P(h\,\lambda(\mathbf{x}^c_n))$ corresponding to the first and second fine path timesteps. For the first of the two fine timesteps, the coarse and fine path Poisson variates are coupled by defining two Poisson variates based on the minimum of the two reactions rates, and the absolute difference,

$$P_1 = P\left(h\min(\lambda(\mathbf{x}_n), \lambda(\mathbf{x}^c_n))\right), \quad P_2 = P\left(h\,|\lambda(\mathbf{x}_n) - \lambda(\mathbf{x}^c_n)|\right),$$

and then using P_1 as the Poisson variate for the path with the smaller rate, and P_1+P_2 for the path with the larger rate. This elegant approach naturally gives a small difference in the Poisson variates when the difference in rates is small, and leads to a very effective multilevel algorithm with a correction variance which is $O(h)$, leading to an $O(\varepsilon^{-2}|\log\varepsilon|^2)$ complexity.

In their paper Anderson and Higham [ibid.], Anderson & Higham treat more general systems with multiple species and multiple reactions. They also include an additional coupling at the finest level to the exact Stochastic Simulation Algorithm developed by Gillespie [1976] which updates the reaction rates after every single reaction. Hence, their

overall multilevel estimator is unbiased, unlike the estimators discussed earlier for SDEs, and the complexity is reduced to $O(\varepsilon^{-2})$ because the number of levels remains fixed as $\varepsilon \to 0$. They give a complete numerical analysis of the variance of their multilevel algorithm; this has been further sharpened in more recent work Anderson, Higham, and Sun [2014]. Because stochastic chemical simulations typically involve 1000's of reactions, the multilevel method is particularly effective in this context, providing computational savings in excess of a factor of 100 Anderson and Higham [2012].

They also give an interesting numerical example in which an approximate model with fewer reactions/reactants is used as a control variate for the full system. This kind of multilevel modelling is another possibility which could be considered in a wide variety of circumstances.

6 Nested simulation

In nested simulations we are interested in estimating quantities of the form

$$\mathbb{E}_Z\left[\, f\Big(\mathbb{E}_W[g(Z,W)]\Big)\,\right]$$

where $\mathbb{E}_Z$ represents an expectation with respect to Z, an outer random variable, and $\mathbb{E}_W[g(Z,W)]$ is a conditional expectation with respect to an independent inner random variable W. For example, in some financial applications, Z represents different risk *scenarios*, $\mathbb{E}_W[g(Z,W)]$ represents the conditional value of a portfolio, and f corresponds to the loss in excess of a certain level, so that $\mathbb{E}_Z\left[\, f\Big(\mathbb{E}_W[g(Z,W)]\Big)\,\right]$ is the expected shortfall.

This can be simulated using nested Monte Carlo simulation with N outer samples $Z^{(n)}$, M inner samples $W^{(m,n)}$ and a standard Monte Carlo estimator:

$$Y = N^{-1}\sum_{n=1}^{N} f\left(M^{-1}\sum_{m=1}^{M} g(Z^{(n)}, W^{(m,n)})\right)$$

Note that to improve the accuracy of the estimate we need to increase both M and N, and this will significantly increase the cost.

An MLMC implementation is straightforward; on level ℓ we can use $M_\ell = 2^\ell$ inner samples. To construct a low variance estimate for $\mathbb{E}[P_\ell - P_{\ell-1}]$ where

$$\mathbb{E}[P_\ell] \equiv \mathbb{E}_Z\left[\, f\left(M_\ell^{-1}\sum_m g(Z, W^{(m)})\right)\,\right],$$

we can use an *antithetic* approach and split the M_ℓ samples of W for the "fine" value into two subsets of size $M_{\ell-1}$ for the "coarse" value:

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

Note that this has the correct expectation, i.e. $\mathbb{E}[Y_\ell] = \mathbb{E}[P_\ell - P_{\ell-1}]$.

If we now define

$$M_{\ell-1}^{-1} \sum_{m=1}^{M_{\ell-1}} g(Z^{(n)}, W^{(m,n)}) = \mathbb{E}[g(Z^{(n)}, W)] + \Delta g_1^{(n)},$$

$$M_{\ell-1}^{-1} \sum_{m=M_{\ell-1}+1}^{M_\ell} g(Z^{(n)}, W^{(m,n)}) = \mathbb{E}[g(Z^{(n)}, W)] + \Delta g_2^{(n)},$$

then if f is twice differentiable a Taylor series expansion gives

$$Y_\ell \approx -\frac{1}{4\,N_\ell} \sum_{n=1}^{N_\ell} f''\left(\mathbb{E}[g(Z^{(n)}, W)]\right) \left(\Delta g_1^{(n)} - \Delta g_2^{(n)}\right)^2$$

By the Central Limit Theorem, $\Delta g_1^{(n)}, \Delta g_2^{(n)} = O(M_\ell^{-1/2})$ and therefore

$$f''\left(\mathbb{E}[g(Z^{(n)}, W)]\right) \left(\Delta g_1^{(n)} - \Delta g_2^{(n)}\right)^2 = O(M_\ell^{-1}).$$

It follows that $\mathbb{E}[Y_\ell] = O(M_\ell^{-1})$ and $V_\ell = O(M_\ell^{-2})$. For the MLMC theorem, this corresponds to $\alpha = 1$, $\beta = 2$, $\gamma = 1$, so the complexity is $O(\varepsilon^{-2})$.

This approach has been used for a financial credit derivative application Bujok, Hambly, and Reisinger [2015], but in that case the function f was piecewise linear, not twice differentiable, and so the rate of variance convergence was slightly lower, with $\beta = 1.5$. However, this is still sufficiently large to achieve an overall complexity which is $O(\varepsilon^{-2})$.

Current research in this area is addressing the challenges of functions f which are discontinuous, and Multi-Index Monte Carlo or nested MLMC for applications in which there are additional "dimensions" to the problem, such as the number of timesteps in an SDE simulation in the inner conditional expectation.

7 Variable precision arithmetic

This final category of applications is included to illustrate the flexibility and generality of the MLMC approach. In the latest Intel CPUs, each core has a vector unit which can perform 16 single precision or 8 double precision operations with one instruction. Hence, single precision computations can be twice as fast as double precision on CPUs. The latest GPUs (graphics processing units) take this idea even further, including a half-precision capability which is twice as fast as single-precision.

This leads naturally to the idea of a 2-level MLMC calculation on CPUs, or a 3-level calculation on GPUs, with the different levels corresponding to different levels of floating point precision. To ensure that the MLMC telescoping sum is correctly respected, all MLMC summations should be performed in double precision, which we can view as being "exact". It is also important that the random numbers are generated consistently, so that the distribution of half-precision random numbers used on level 0, is equivalent to the distribution of "coarse sample" half-precision random numbers obtained on level 1 by first generating single-precision random numbers are then truncating them down to half-precision.

This approach has been generalised in research which exploits FPGAs (field-programmable gate arrays) which can perform computations with a user-specified number of bits to represent floating-point or fixed-point numbers. Thus, it is possible to implement a multilevel treatment in which the number of bits used increases with level Brugger, de Schryver, Wehn, Omland, Hefter, Ritter, Kostiuk, and Korn [2014].

8 Conclusions

The last ten years has seen considerable progress in the theoretical development, application and analysis of multilevel Monte Carlo methods. On the theoretical side, the key extensions are to unbiased randomised estimators for applications with a rapid rate of variance convergence; Richardson-Romberg extrapolation for improved computational efficiency when the rate of variance convergence is low; and multi-index Monte Carlo (MIMC), generalising multilevel to multiple "directions" in which approximations can be refined. On the practical side, the range of applications is growing steadily, including the examples given in this article and others such as reliability and rare event simulation, and MCMC and Bayesian inverse methods. There has also been excellent progress on the numerical analysis of the MLMC variances for the full range of applications.

This review has attempted to emphasise the conceptual simplicity of the multilevel approach; in essence it is simply a recursive control variate strategy, using cheap inaccurate approximations to some random output quantity as a control variate for more accurate but more costly approximations. In practice, the challenge is first to develop a tight coupling

between successive approximation levels, to minimise the variance of the difference in the output obtained from each level, and then to develop a corresponding numerical analysis.

Acknowledgments. The author is very grateful to the many collaborators and students he has worked with on MLMC applications over the past 10 years, and in particular to Dr. Abdul-Lateef Haji-Ali for his comments on the paper.

References

David F. Anderson and Desmond J. Higham (2012). "Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics". *Multiscale Model. Simul.* 10.1, pp. 146–179. MR: 2902602 (cit. on pp. 3604, 3605).

David F. Anderson, Desmond J. Higham, and Yu Sun (2014). "Complexity of multilevel Monte Carlo tau-leaping". *SIAM J. Numer. Anal.* 52.6, pp. 3106–3127. MR: 3504598 (cit. on p. 3605).

Rainer Avikainen (2009). "On irregular functionals of SDEs and the Euler scheme". *Finance Stoch.* 13.3, pp. 381–401. MR: 2519837 (cit. on p. 3601).

Ivo Babuška, Fabio Nobile, and Raúl Tempone (2010). "A stochastic collocation method for elliptic partial differential equations with random input data". *SIAM Rev.* 52.2, pp. 317–355. MR: 2646806 (cit. on p. 3589).

Ivo Babuška, Raúl Tempone, and Georgios E. Zouraris (2004). "Galerkin finite element approximations of stochastic elliptic partial differential equations". *SIAM J. Numer. Anal.* 42.2, pp. 800–825. MR: 2084236 (cit. on p. 3589).

Andrea Barth, Annika Lang, and Christoph Schwab (2013). "Multilevel Monte Carlo method for parabolic stochastic partial differential equations". *BIT* 53.1, pp. 3–27. MR: 3029293 (cit. on p. 3602).

Andrea Barth, Christoph Schwab, and Nathaniel Zollinger (2011). "Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients". *Numer. Math.* 119.1, pp. 123–161. MR: 2824857 (cit. on p. 3602).

Christian Brugger, Christian de Schryver, Norbert Wehn, Steffen Omland, Mario Hefter, Klaus Ritter, Anton Kostiuk, and Ralf Korn (2014). "Mixed precision multilevel Monte Carlo on hybrid computing systems". In: *Computational Intelligence for Financial Engineering & Economics (CIFEr), 2104 IEEE Conference on*. IEEE, pp. 215–222 (cit. on p. 3607).

K. Bujok, B. M. Hambly, and C. Reisinger (2015). "Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives". *Methodol. Comput. Appl. Probab.* 17.3, pp. 579–604. MR: 3377850 (cit. on p. 3606).

Hans-Joachim Bungartz and Michael Griebel (2004). "Sparse grids". *Acta Numer.* 13, pp. 147–269. MR: 2249147 (cit. on p. 3597).

K. A. Cliffe, Michael B. Giles, R. Scheichl, and A. L. Teckentrup (2011). "Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients". *Comput. Vis. Sci.* 14.1, pp. 3–15. MR: 2835612 (cit. on pp. 3593, 3602).

Jakob Creutzig, Steffen Dereich, Thomas Müller-Gronbach, and Klaus Ritter (2009). "Infinite-dimensional quadrature and approximation of distributions". *Found. Comput. Math.* 9.4, pp. 391–429. MR: 2519865 (cit. on p. 3601).

Josef Dick, Frances Y. Kuo, and Ian H. Sloan (2013). "High-dimensional integration: the quasi-Monte Carlo way". *Acta Numer.* 22, pp. 133–288. MR: 3038697 (cit. on pp. 3590, 3599).

Josef Dick, Friedrich Pillichshammer, and Benjamin J. Waterhouse (2008). "The construction of good extensible rank-1 lattices". *Math. Comp.* 77.264, pp. 2345–2373. MR: 2429889 (cit. on p. 3599).

Y. Efendiev, O. Iliev, and C. Kronsbein (2013). "Multilevel Monte Carlo methods using ensemble level mixed MsFEM for two-phase flow and transport simulations". *Comput. Geosci.* 17.5, pp. 833–850. MR: 3104637 (cit. on p. 3602).

Michael B. Giles (2008a). "Improved multilevel Monte Carlo convergence using the Milstein scheme". In: *Monte Carlo and quasi-Monte Carlo methods 2006*. Springer, Berlin, pp. 343–358. MR: 2479233 (cit. on p. 3601).

– (2008b). "Multilevel Monte Carlo path simulation". *Oper. Res.* 56.3, pp. 607–617. MR: 2436856 (cit. on pp. 3593, 3600, 3601).

– (2015). "Multilevel Monte Carlo methods". *Acta Numer.* 24, pp. 259–328. MR: 3349310 (cit. on pp. 3590, 3601).

Michael B. Giles, Kristian Debrabant, and Andreas Rößler (Feb. 2013). "Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation". arXiv: 1302.4676 (cit. on p. 3601).

Michael B. Giles, Desmond J. Higham, and Xuerong Mao (2009). "Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff". *Finance Stoch.* 13.3, pp. 403–413. MR: 2519838 (cit. on p. 3601).

Michael B. Giles, Frances Y. Kuo, and I. H. Sloan (2018). "Combining sparse grids, multilevel MC and QMC for elliptic PDEs with random coefficients". In: *Monte Carlo and quasi-Monte Carlo methods 2016*. Ed. by P.W. Glynn and A. Owen. Springer Proc. Math. Stat. Springer, Heidelberg (cit. on p. 3598).

Michael B. Giles and Christoph Reisinger (2012). "Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance". *SIAM J. Financial Math.* 3.1, pp. 572–592. MR: 2968046 (cit. on p. 3602).

Michael B. Giles and Benjamin J. Waterhouse (2009). "Multilevel quasi-Monte Carlo path simulation". In: *Advanced financial modelling*. Vol. 8. Radon Ser. Comput. Appl. Math. Walter de Gruyter, Berlin, pp. 165–181. MR: 2648461 (cit. on p. 3599).

Daniel T. Gillespie (1976). "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". *J. Computational Phys.* 22.4, pp. 403–434. MR: 0503370 (cit. on p. 3604).

Max D. Gunzburger, Clayton G. Webster, and Guannan Zhang (2014). "Stochastic finite element methods for partial differential equations with random input data". *Acta Numer.* 23, pp. 521–650. MR: 3202242 (cit. on p. 3589).

Abdul-Lateef Haji-Ali, Fabio Nobile, and Raúl Tempone (2016). "Multi-index Monte Carlo: when sparsity meets sampling". *Numerische Mathematik* 132.4, pp. 767–806. MR: 3474489 (cit. on pp. 3596, 3597).

Peter E. Kloeden and Eckhard Platen (1992). *Numerical solution of stochastic differential equations*. Vol. 23. Applications of Mathematics (New York). Springer-Verlag, Berlin, pp. xxxvi+632. MR: 1214374 (cit. on p. 3600).

Frances Y. Kuo, Christoph Schwab, and Ian H. Sloan (2015). "Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients". *Found. Comput. Math.* 15.2, pp. 411–449. MR: 3320930 (cit. on p. 3599).

Vincent Lemaire and Gilles Pagès (2017). "Multilevel Richardson-Romberg extrapolation". *Bernoulli* 23.4A, pp. 2643–2692. MR: 3648041 (cit. on pp. 3595, 3596).

S. Mishra, Ch. Schwab, and J. Šukys (2012). "Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions". *J. Comput. Phys.* 231.8, pp. 3365–3388. MR: 2897628 (cit. on p. 3602).

Florian Müller, Patrick Jenny, and Daniel W. Meyer (2013). "Multilevel Monte Carlo for two phase flow and Buckley-Leverett transport in random heterogeneous porous media". *J. Comput. Phys.* 250, pp. 685–702. MR: 3079555 (cit. on p. 3602).

Ben Niu, Fred J. Hickernell, Thomas Müller-Gronbach, and Klaus Ritter (2011). "Deterministic multi-level algorithms for infinite-dimensional integration on $\mathbb{R}^{\mathbb{N}}$". *J. Complexity* 27.3-4, pp. 331–351. MR: 2793867 (cit. on p. 3599).

Michele M Putko, Arthur C Taylor, Perry A Newman, and Lawrence L Green (2002). "Approach for input uncertainty propagation and robust design in CFD using sensitivity derivatives". *Journal of Fluids Engineering* 124.1, pp. 60–69 (cit. on p. 3589).

Chang-Han Rhee and Peter W. Glynn (2015). "Unbiased estimation with square root convergence for SDE models". *Oper. Res.* 63.5, pp. 1026–1043. MR: 3422533 (cit. on pp. 3594, 3595).

F. Vidal-Codina, N. C. Nguyen, Michael B. Giles, and J. Peraire (2015). "A model and variance reduction method for computing statistical outputs of stochastic elliptic partial differential equations". *J. Comput. Phys.* 297, pp. 700–720. MR: 3361685 (cit. on p. 3603).

Dongbin Xiu and George Em Karniadakis (2002). "The Wiener-Askey polynomial chaos for stochastic differential equations". *SIAM J. Sci. Comput.* 24.2, pp. 619–644. MR: 1951058 (cit. on p. 3589).

Received 2017-11-30.
Michael B. Giles: Mathematical Institute, University of Oxford
Mike.Giles@maths.ox.ac.uk