SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 10, No. 1, pp. 200–226

Analysis of Nested Multilevel Monte Carlo Using Approximate Normal Random Variables*

Michael Giles† and Oliver Sheridan-Methven‡

Abstract. The multilevel Monte Carlo (MLMC) method has been used for a wide variety of stochastic applications. In this paper we consider its use in situations in which input random variables can be replaced by similar approximate random variables which can be computed much more cheaply. A nested MLMC approach is adopted in which a two-level treatment of the approximated random variables is embedded within a standard MLMC application. We analyze the resulting nested MLMC variance in the specific context of an SDE discretization in which normal random variables can be replaced by approximately normal random variables, and we provide numerical results to support the analysis.

Key words. multilevel Monte Carlo, random variables, approximations, SDEs, stochastic simulation, high performance computing

AMS subject classifications. 65C10, 65C05, 65Y20, 60H35, 65C30

DOI. 10.1137/21M1399385

1. Introduction.

Following the initial work of Heinrich [28] on parametric integration and Giles [15] on stochastic differential equations (SDEs), there has been huge development in the application of multilevel Monte Carlo (MLMC) methods to a wide variety of stochastic modeling contexts. This includes partial differential equations (PDEs) with stochastic coefficients [11, 4], stochastic PDEs [23], continuous-time Markov process models of biochemical reactions [1, 2], Markov chain Monte Carlo [30, 38], nested simulation [9, 18], particle filters [32], sequential Monte Carlo [6], ensemble Kalman filtering [31], probability density estimation [22, 7], and reliability estimation [41, 13]. A review of research on MLMC is provided by Giles [16].

In this paper we are concerned with the development and analysis of a new class of MLMC methods involving approximate probability distributions. Most numerical methods for simulating stochastic models start from random inputs from a variety of well-known distributions: normal, Poisson, binomial, noncentral χ^2, etc. Generating samples which have a distribution which matches the desired distribution to within the limits of finite precision arithmetic can be a significant part of the overall computational cost of the simulation. Here we consider what can be achieved if it is also possible to generate approximate random variables (random variables with a distribution which is only approximately correct) at a greatly reduced cost.

*Received by the editors February 17, 2021; accepted for publication (in revised form) September 27, 2021; published electronically February 8, 2022.
https://doi.org/10.1137/21M1399385
Funding: This research was supported by ICONIC EPSRC Programme grant EP/P020720/1 and the EPSRC Centre for Doctoral Training in Industrially Focused Mathematical Modelling grant EP/L015803/1 in collaboration with Arm Ltd.
†Mathematical Institute, Oxford University, Oxford, OX2 6GG, UK (mike.giles@maths.ox.ac.uk).
‡Mathematics, Oxford University, Oxford, Oxfordshire, OX2 6GG, UK (oliver.sheridan-methven@hotmail.co.uk).

We will show that a nested MLMC approach can be adopted in which a two-level treatment of the approximated random variables is embedded within a standard MLMC application. We then analyze the MLMC variance of the resulting treatment in the specific context of an SDE discretization in which normal random variables can be replaced by approximately normal random variables.

The most relevant prior research is the work of Giles, Hefter, Mayer, and Ritter [20, 19]. This research was in the context of *information based complexity*, working with a complexity model which counted the number of individual random bits, rather than viewing each standard uniformly distributed random variable as having a unit cost. Fundamental to the algorithms in these papers was the use of quantized normal random variables, which is the first of the three approximations to be discussed in the next section. Some elements of the analysis in this paper build on the ideas and analysis in those papers, but the context is quite different in aiming to minimize the real-world execution cost of MLMC algorithms on modern CPUs and graphics processing units (GPUs), and the specifics of the proposed method are quite different in using a nested MLMC approach.

Another relevant paper is by Müller, Scheichl, and shardlow [36]. In this work they use three- and four-point approximations to the normal distribution, equivalent to a piecewise constant approximation of $\Phi^{-1}(U)$ on 3 or 4 intervals of a nonuniform size, chosen so that the leading moments are the same as for the standard normal distribution. The way in which these are used within the MLMC construction violates, to a small extent, the usual telescoping summation which lies at the heart of the MLMC method, and therefore they have to be careful to bound the magnitude of this error. This new error is related to the fact that if Z_1 and Z_2 are unit normal random variables, then so too is their sum $(Z_1{+}Z_2)/\sqrt{2}$; in an SDE application, this is important in MLMC so that the sum of two Brownian increments from timesteps of size h corresponds to a Brownian increment from a timestep of size $2h$. However, this is no longer true when using approximate normal distributions; the sum $(\widetilde{Z}_1{+}\widetilde{Z}_2)/\sqrt{2}$ is still an approximation of a unit normal random number, but in general it does not come from exactly the same distribution as $\widetilde{Z}_1$ and $\widetilde{Z}_2$.

Similar ideas have also been investigated by Belomestny and Nagapetyan [5] who avoid errors in the telescoping summation by using different approximate distributions on each level of MLMC refinement. However, in the present paper we prefer to avoid these difficulties entirely by using the same approximate distribution throughout within a nested MLMC treatment; we think this will generalize better to different approximations well-suited to different computer hardware and to applications with more complex distributions. We also consider three different kinds of approximations, some of which will generalize better to distributions such as the noncentral χ^2-distribution in which there are additional parameters which may vary for each random number sample; in such cases a lookup table based on quantization could become unreasonably large. This ability to generalize our approximations to various distributions is a key factor specific to the inverse transform method, as opposed to other alternatives for sampling from the normal distribution.

2. Approximate normal distributions.

There are many ways in which approximate normal variables can be generated. In this paper we consider three methods, each of which can be viewed as an approximation of the generation of normal random variables through the inversion of the normal cumulative distribution function (CDF), $Z = \Phi^{-1}(U)$, with U being a uniform random variable on the unit interval $(0, 1)$. The corresponding approximations all have the form $\widetilde{Z} = \widetilde{Q}(U)$, so that it is possible to generate coupled pairs $(Z, \widetilde{Z})$ from the same random input U.

The three approximations are all motivated by the different hardware features of modern CPUs and GPUs (manycore graphics processing units). Their analysis, implementation details, and resulting execution performance are explained more fully in [24, 40], where [24] is a companion paper to this one.

Note that we are not concerned with the computational cost of generating the uniform random numbers U. This is because we can follow Giles et al. [19] in using a trick due to Bakhvalov [3] to very efficiently generate a set of uniformly distributed values $\{U_1, U_2, \ldots\}$ with pairwise independence (i.e., for any two indices $i \neq j$, U_i is independent of U_j) at a cost of less than 3 computer operations, on average. In essence, the procedure is very similar to the *digital shift* used to transform Sobol' points in a randomized quasi-Monte Carlo computation [34].

Before presenting the approximations, we should discuss the motivation surrounding the approximation of the inverse CDF, when other methods, such as the Box–Muller scheme [26, 39] or the Ziggurat method [35], are also popular. The first pitfall of the Ziggurat method is that it is a rejection based method, so it is ill suited to generating quasi-normal random variables, where samples can't be rejected so as to ensure the low discrepancy property is preserved. As for the Box–Muller scheme, this suffers from several of the same issues as the inverse transform method, namely, that it is evaluated through the evaluation of expensive intermediate functions that are ill suited to modern *single instruction multiple data* (SIMD) capable hardware. Furthermore, the Box–Muller method is specific to the normal distribution, whereas the inverse transform method immediately generalizes to numerous distributions. For the sake of analytic tractability, our analysis here is limited to the Euler–Maruyama scheme using the normal distribution, but in our companion piece [24] we construct and demonstrate the applicability of these same ideas to more complicated Milstein scheme [26, 33] and the more difficult Cox–Ingersoll–Ross process [12].

2.1. Quantized approximation. The first approximation is a simple piecewise constant approximation on $K = 2^q$ intervals of size 2^{-q}, with the value $\widetilde{Q}(U)$ on the kth interval $I_k = [(k-1)\, 2^{-q}, k\, 2^{-q}]$, given by either the average value of $\Phi^{-1}(U)$ on the interval or alternatively the value at the midpoint.

The index k corresponds to the first q bits of the binary expansion for U, and this quantized approximation is the one considered by Giles et al. [20, 19] since each random variable $\widetilde{Z}$ can be generated based directly on q random bits, each independently taking the value 0 or 1 with equal probability. Extending their analysis, it can be proved that

$$\mathbb{E}[\,|\widetilde{Z} - Z|^p] = o(2^{-q}).$$

Figure 2.1(a) illustrates the approximation for $K = 16$ and also has plots of $\mathbb{E}[(\widetilde{Z} - Z)^2]$ and $\mathbb{E}[(\widetilde{Z} - Z)^4]$ as a function of K, the number of intervals.

For $K = 1024$ the mean square error (MSE) is $\mathbb{E}[(\widetilde{Z} - Z)^2] \approx 1.5 \times 10^{-4}$. This size seems a good choice as the lookup table will fit inside the L1 cache of a current generation Intel CPU, leading to a very efficient scalar implementation.

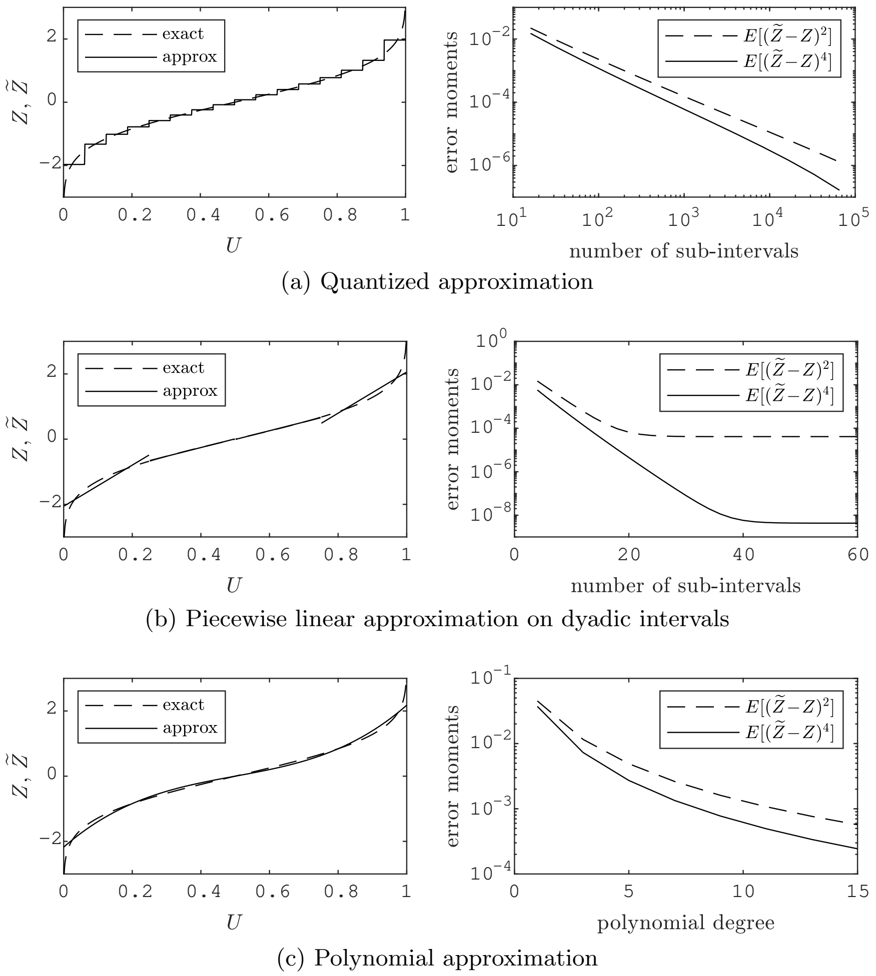


Figure 2.1. *Three approximations of the inverse normal CDF.*

2.2. Piecewise linear approximation on dyadic intervals. The second approximation uses a discontinuous piecewise linear approximation on a geometric sequence of subintervals. To be specific, for a given ratio $\frac{1}{2} \leq r < 1$, K subintervals on $[0, \frac{1}{2}]$ are defined by

$$I_k = [\tfrac{1}{2}r^k, \tfrac{1}{2}r^{k-1}], \quad k = 1, 2, \ldots K{-}1, \qquad I_K = [0, \tfrac{1}{2}r^{K-1}].$$

In the particular case $r = \frac{1}{2}$, given an input $0 < U < \frac{1}{2}$, the corresponding subinterval index k can be determined by computing the integer part of $\log_2 U$, which can be implemented very efficiently due to the floating point format of real numbers.

On each subinterval I_k, $\widetilde{Q}(U)$ is defined as the least-squares linear best fit approximation to $\Phi^{-1}(U)$, and the approximation on $[\frac{1}{2}, 1]$ is defined by $\widetilde{Q}(U) = -\widetilde{Q}(1{-}U)$. Standard analysis of the accuracy of piecewise linear interpolation leads to the result that

$$\mathbb{E}[\,|\widetilde{Z}{-}Z|^p] = o(r^K) + O((1{-}r)^{2p}),$$

where the first term comes from the two end intervals $[0, \frac{1}{2}r^{K-1}]$ and $[1-\frac{1}{2}r^{K-1}, 1]$, and the second term comes from the other intervals. Note that to achieve convergence to zero requires that both $r \to 1$ and $r^K \to 0$.

Figure 2.1(b) illustrates the approximation for $r = \frac{1}{2}$, $K = 2$ and also has plots of $\mathbb{E}[(\widetilde{Z}-Z)^2]$ and $\mathbb{E}[(\widetilde{Z}-Z)^4]$ as a function of $2K$, the number of intervals, with fixed $r = \frac{1}{2}$. Note that for both $\mathbb{E}[(\widetilde{Z}-Z)^2]$ and $\mathbb{E}[(\widetilde{Z}-Z)^4]$ the $(1-r)^{2p}$ error term eventually dominates once r^K is sufficiently small.

For $K = 16$ the MSE is $\mathbb{E}[(\widetilde{Z}-Z)^2] \approx 4 \times 10^{-5}$. This size is significant because in single precision it is possible to perform the table lookup within a single 512-bit AVX vector register on current generation Intel Xeon CPUs, giving a very efficient vector implementation.

2.3. Polynomial approximation. The final method is a simple polynomial approximation,

$$\widetilde{Q}(U) \equiv \sum_{k=1}^{K} a_k \, (U - \tfrac{1}{2})^{2k-1}$$

with the coefficients a_k determined by a least-squares best fit to $\Phi^{-1}(U)$. This method is particularly efficient on GPUs as it avoids the need for a table lookup; however, it is also the least accurate of the three approximations for realistic polynomial sizes.

Figure 2.1(c) illustrates the approximation for $K = 2$ and also has plots of $\mathbb{E}[(\widetilde{Z}-Z)^2]$ and $\mathbb{E}[(\widetilde{Z}-Z)^4]$ as a function of $2K-1$, the degree of the polynomial.

For $K = 4$ the MSE is $\mathbb{E}[(\widetilde{Z}-Z)^2] \approx 2.6 \times 10^{-3}$; this seems a good balance between computational cost and accuracy and will be used in the numerical experiments later.

2.4. Computational efficiency. The companion paper [24] details the high performance implementation of these approximations on the latest SIMD capable hardware, and so we will only restate some of the key findings of [24, Table 2] here. These are that for the exact inverse transform method, on modern Intel hardware, sampling takes on average 3.4 clock cycles, and the quantized approximation and the piecewise linear approximation on dyadic intervals both provide speed ups by a factor approximately 6 and 7, respectively, compared to the highly optimized Intel Math Kernel Library.

Furthermore, as the inverse transform method generalizes to other distributions, for the more complicated noncentral inverse χ^2 distribution arising in the Cox–Ingersoll–Ross process, identical approximations yield [24, Table 5] speed ups by factors of 300 up to 6000!

These results demonstrate how much more efficiently the normal distribution can be sampled from when using our approximation schemes and how well these generalize to other numerical schemes and more difficult stochastic processes (where a reference solution is not known a priori).

3. MLMC algorithms. In this section we begin with a quick recap of the MLMC method and then discuss how a nested version of MLMC can be used with approximate distributions. The third part then applies the ideas to the simulation of SDE solutions, using approximate normal random variables.

3.1. Standard MLMC. If P is a random variable which is a function of a set of random inputs ω, then the Monte Carlo estimate for the expected value $\mathbb{E}[P]$ is the simple average

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

where the $\omega^{(n)}, n=1,2,3,\ldots,N$, are independent and identically distributed (i.i.d.) samples of ω. To achieve a root-mean-square (RMS) error of ε requires $N\approx\varepsilon^{-2}V$ samples, where $V=\mathbb{V}[P]$ is the variance. If each sample costs C, then the total cost is approximately $\varepsilon^{-2}V\,C$.

Suppose now that $\widetilde{P}\approx P$; then since $\mathbb{E}[P]=\mathbb{E}[\widetilde{P}]+\mathbb{E}[P-\widetilde{P}]$ we can instead use the estimator

$$N_0^{-1}\sum_{n=1}^{N_0}\widetilde{P}(\omega^{(0,n)}) + N_1^{-1}\sum_{n=1}^{N_1}(P(\omega^{(1,n)})-\widetilde{P}(\omega^{(1,n)})).$$

The cost of this estimator is $N_0C_0+N_1C_1$, and the variance is $V_0/N_0+V_1/N_1$, where $V_0\equiv\mathbb{V}[\widetilde{P}]$, $V_1\equiv\mathbb{V}[P-\widetilde{P}]$ if all of the $\omega^{(\ell,n)}$ are independent. Minimizing the cost subject to the same accuracy requirement gives the total cost $\varepsilon^{-2}(\sqrt{V_0C_0}+\sqrt{V_1C_1})^2$ which is significantly less than $\varepsilon^{-2}V\,C$ if $C_0\ll C$ and $V_1\ll V$.

To give a quantitative example, suppose that $C_0=10^{-1}C$, $C_1=C$, so the cost of evaluating $\widetilde{P}$ is 10 times less than evaluating P, and $V_0=V$, $V_1=10^{-3}V$. In that case, the total cost is approximately $0.12\,\varepsilon^{-2}V\,C$, a factor of 8 savings compared to the original Monte Carlo computation.

This two-level calculation is easily generalized to a multilevel treatment. Given a sequence of increasingly accurate (and costly) approximations $\widehat{P}_0,\widehat{P}_1,\widehat{P}_2,\ldots\longrightarrow P$, for example, from the approximation of an SDE using 2^ℓ timesteps on level ℓ, then

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

and so the MLMC estimate for $\mathbb{E}[\widehat{P}_L]$ is

$$\widehat{Y}\equiv N_0^{-1}\sum_{n=1}^{N_0}\widehat{P}_0^{(n)}+\sum_{\ell=1}^{L}N_\ell^{-1}\sum_{n=1}^{N_\ell}(\widehat{P}_\ell^{(\ell,n)}-\widehat{P}_{\ell-1}^{(\ell,n)}).$$

The expected value of the estimator is $\mathbb{E}[\widehat{P}_L]$, and the MSE can be decomposed into the sum of the variance and the square of the bias to give

$$\mathrm{MSE}\;=\;\mathbb{V}[\widehat{Y}]+\Big(\mathbb{E}[\widehat{P}_L-P]\Big)^2=\sum_{\ell=0}^{L}N_\ell^{-1}V_\ell+\Big(\mathbb{E}[\widehat{P}_L-P]\Big)^2,$$

where $V_0\equiv\mathbb{V}[\widehat{P}_0]$ and, for $\ell\geq 1$, $V_\ell\equiv\mathbb{V}[\widehat{P}_\ell-\widehat{P}_{\ell-1}]$.

If level L is chosen so that $|\mathbb{E}[\widehat{P}_L-P]|<\varepsilon/\sqrt{2}$, then an overall RMS error of ε can be achieved at a total cost of approximately

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

where C_ℓ is the cost of a single sample of $\widehat{P}_\ell - \widehat{P}_{\ell-1}$. If the product $V_\ell C_\ell$ decreases exponentially with level, then the overall cost is $O(\varepsilon^{-2})$, corresponding to an $O(1)$ cost per sample on average, much less than for the standard Monte Carlo method. For further details see [15, 16].

3.2. Nested MLMC for approximate distributions. When using random variables from approximate distributions, the first possibility is to use the two-level treatment discussed before, with

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

In this case, each sample $P - \widetilde{P}$ would use the same underlying stochastic sample for both P and $\widetilde{P}$, for example, using the same input uniform random variable U and then applying the inverse of either the true CDF or an approximate CDF to produce the random variables required to compute P and $\widetilde{P}$, respectively. As stated before, this can give considerable savings if the cost of computing $\widetilde{P}$ is much less than the cost of computing P, and $\mathbb{V}[P - \widetilde{P}] \ll \mathbb{V}[P]$.

However, what can we do if our starting point is an MLMC expansion in some other quantity, such as the timestep? In that case we can use nested MLMC, an idea discussed by Giles, Kuo, and Sloan [21] which is a generalization of multi-index Monte Carlo (MIMC) by Haji-Ali, Nobile, and Tempone [27]. We start from the usual MLMC expansion and then split each of the required expectations into two pieces, one using the approximate random variables and the other computing the required correction,

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

The pair $(\widetilde{P}_\ell, \widetilde{P}_{\ell-1})$ are generated in the same way as $(P_\ell, P_{\ell-1})$, based on the same underlying uniform random variables U, but converting them differently into the random variables required for the calculation of $\widehat{P}$ and $\widetilde{P}$.

As explained in the previous section, the standard MLMC cost to achieve an RMS accuracy of ε is approximately

$$C_{\mathrm{MLMC}} = 2\,\varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{C_\ell V_\ell} \right)^2 ,$$

where C_ℓ is the cost of a single sample of $\widehat{P}_\ell - \widehat{P}_{\ell-1}$ on level ℓ, and V_ℓ is its variance. The complexity analysis extends naturally to the nested *approximate* MLMC algorithm described above, giving a cost of approximately

$$C_{\mathrm{AMLMC}} = 2\,\varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{\widetilde{C}_\ell V_\ell} + \sqrt{(C_\ell + \widetilde{C}_\ell)\widetilde{V}_\ell} \right)^2 ,$$

where $\widetilde{C}_\ell$ is the cost of one sample of $\widetilde{P}_\ell - \widetilde{P}_{\ell-1}$, which has a variance approximately equal to V_ℓ, and $C_\ell + \widetilde{C}_\ell$ is the cost of one sample of $(\widehat{P}_\ell - \widehat{P}_{\ell-1}) - (\widetilde{P}_\ell - \widetilde{P}_{\ell-1})$, and $\widetilde{V}_\ell$ is its variance.

Note that

$$C_{\text{AMLMC}} = 2\,\varepsilon^{-2}\left(\sum_{\ell=0}^{L}\sqrt{C_\ell V_\ell}\sqrt{\frac{\widetilde{C}_\ell}{C_\ell}}\left(1+\sqrt{\left(\frac{C_\ell}{\widetilde{C}_\ell}{+}1\right)\frac{\widetilde{V}_\ell}{V_\ell}}\right)\right)^2$$
$$\le C_{\text{MLMC}}\max_{0\le\ell\le L}\frac{\widetilde{C}_\ell}{C_\ell}\left(1+\sqrt{\left(\frac{C_\ell}{\widetilde{C}_\ell}{+}1\right)\frac{\widetilde{V}_\ell}{V_\ell}}\right)^2$$

so that if $\widetilde{V}_\ell/V_\ell \ll \widetilde{C}_\ell/C_\ell \ll 1$, then the cost is reduced by a factor of approximately $\max_\ell \widetilde{C}_\ell/C_\ell$.

3.3. Application to SDE simulation. To make the ideas in the preceding section more concrete, we consider an application involving the solution of a scalar autonomous SDE,

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + b(X_t)\,\mathrm{d}W_t,$$

on the time interval $[0,T]$ subject to fixed initial data X_0. Furthermore, we suppose that we are interested in the expected value of a function of the final path value $\mathbb{E}[f(X_T)]$.

If $\widehat{X}_n$ is an approximation to X_{nh} using a uniform timestep of size h, then the simplest numerical approximation is the Euler–Maruyama discretization,

$$\widehat{X}_{n+1} = \widehat{X}_n + a(\widehat{X}_n)\,h + b(\widehat{X}_n)\,\Delta W_n,$$

in which the Brownian increment ΔW_n is a normal random variable with mean 0 and variance h, so it can be simulated as $\Delta W_n \equiv \sqrt{h}\,Z_n$, where Z_n is a unit normal random variable. In turn Z_n can be generated from a uniform $(0,1)$ random variable U_n through $Z_n = \Phi^{-1}(U_n)$. Using this discretization the output quantity of interest would be $\widehat{P} \equiv f(\widehat{X}_N)$, where $N = T/h$ is assumed to be an integer.

The simplest way in which we could use approximate normal random variables would be to keep to a fixed number of timesteps and generate an approximate output quantity $\widetilde{P}$ by replacing the normals Z_n by approximate normals $\widetilde{Z}_n \equiv \widetilde{Q}(U_n)$ generated using the same U_n.

However, we would like to combine the benefits of MLMC for the time discretization with the reduced execution cost of approximate normals and so can instead use the nested MLMC approach. To do this, the key question is, how do we compute each sample of $(\widehat{P}_\ell{-}\widehat{P}_{\ell-1}) - (\widetilde{P}_\ell{-}\widetilde{P}_{\ell-1})$?

Let $h \equiv h_\ell$ be the timestep for a fine path on level ℓ, with $\widehat{X}^{\mathrm{f}}_n$ representing the fine path approximation to the SDE solution $X(nh)$ which is computed using the discrete equations

$$\widehat{X}^{\mathrm{f}}_{n+1} = \widehat{X}^{\mathrm{f}}_n + a(\widehat{X}^{\mathrm{f}}_n)\,h + b(\widehat{X}^{\mathrm{f}}_n)\,\sqrt{h}\,Z_n, \tag{3.1}$$

based on the true normals, Z_n. The corresponding coarse path approximation $\widehat{X}^{\mathrm{c}}_n$ using timesteps of $2h$ and combined Brownian increments $\Delta W_n{+}\Delta W_{n+1}$ is given by

$$\widehat{X}^{\mathrm{c}}_{n+2} = \widehat{X}^{\mathrm{c}}_n + 2\,a(\widehat{X}^{\mathrm{c}}_n)\,h + b(\widehat{X}^{\mathrm{c}}_n)\,(\Delta W_n + \Delta W_{n+1})$$
$$= \widehat{X}^{\mathrm{c}}_n + 2\,a(\widehat{X}^{\mathrm{c}}_n)\,h + b(\widehat{X}^{\mathrm{c}}_n)\,(\sqrt{h}\,Z_n + \sqrt{h}\,Z_{n+1})$$

for even integers n. Alternatively, it is more convenient to write it equivalently as

$$\widehat{X}^{\mathrm{c}}_{n+1} = \widehat{X}^{\mathrm{c}}_n + a(\widehat{X}^{\mathrm{c}}_{\underline{n}})\, h + b(\widehat{X}^{\mathrm{c}}_{\underline{n}})\, \sqrt{h}\, Z_n, \tag{3.2}$$

where $\underline{n} \equiv 2\lfloor n/2 \rfloor$ is n rounded down to the nearest even number. This gives the same values for $\widehat{X}^{\mathrm{c}}_n$ at the even timesteps.

The corresponding discrete equations for the fine and coarse paths computed using the approximate normal random variables are

$$\widetilde{X}^{\mathrm{f}}_{n+1} = \widetilde{X}^{\mathrm{f}}_n + a(\widetilde{X}^{\mathrm{f}}_n)\, h + b(\widetilde{X}^{\mathrm{f}}_n)\, \sqrt{h}\, \widetilde{Z}_n \tag{3.3}$$

and

$$\widetilde{X}^{\mathrm{c}}_{n+1} = \widetilde{X}^{\mathrm{c}}_n + a(\widetilde{X}^{\mathrm{c}}_{\underline{n}})\, h + b(\widetilde{X}^{\mathrm{c}}_{\underline{n}})\, \sqrt{h}\, \widetilde{Z}_n, \tag{3.4}$$

and then finally we obtain

$$(\widehat{P}_\ell - \widehat{P}_{\ell-1}) - (\widetilde{P}_\ell - \widetilde{P}_{\ell-1}) \;=\; (f(\widehat{X}^{\mathrm{f}}_N) - f(\widehat{X}^{\mathrm{c}}_N)) - (f(\widetilde{X}^{\mathrm{f}}_N) - f(\widetilde{X}^{\mathrm{c}}_N)).$$

This description is for the case in which the timestep h_ℓ is halved on each successive level. There is a natural extension to other geometric sequences such as $h_\ell = 4^{-\ell} h_0$, with the values of the coarse path drift and volatility being updated at the end of each coarse timestep, while the fine path values are updated after each fine path timestep.

3.4. MLMC but not MIMC. To end this section, we must highlight that the MLMC setup using approximate random variables discussed thus far is not expressible as an MIMC [27]. A key aspect of the standard MLMC for SDEs is that the Brownian increment for a timestep of $2h$ is equal to the sum of Brownian increments for two timesteps of size h. Thus, if Z_1, Z_2 are standard normal random variables, then $\sqrt{h}\, Z_1 + \sqrt{h}\, Z_2 = \sqrt{2h}\, Z_3$, where Z_3 is also a standard normal random variable. However, if $\widetilde{Z}_1, \widetilde{Z}_2$ are approximate standard normal random variables and $\sqrt{h}\, \widetilde{Z}_1 + \sqrt{h}\, \widetilde{Z}_2 = \sqrt{2h}\, \widetilde{Z}_3$, then while $\widetilde{Z}_3$ is an approximate standard normal random variable, it is from a different distribution to $\widetilde{Z}_1$ and $\widetilde{Z}_2$. This is why we must use the nested MLMC approach, and what prohibits us from utilizing MIMC.

To emphasize this, consider the nested multilevel framework at the $\ell = 2$ discretization level. The $\ell = 2$ level will be the coarse level when coupled with the $\ell = 3$ level and conversely will be the fine level when coupled with the $\ell = 1$ level. This means that the nested MLMC framework will compute $\{\widetilde{P}^{\mathrm{f}}_2 - \widetilde{P}^{\mathrm{c}}_1, (\widehat{P}^{\mathrm{f}}_2 - \widehat{P}^{\mathrm{c}}_1) - (\widetilde{P}^{\mathrm{f}}_2 - \widetilde{P}^{\mathrm{c}}_1)\}$ and $\{\widetilde{P}^{\mathrm{f}}_3 - \widetilde{P}^{\mathrm{c}}_2, (\widehat{P}^{\mathrm{f}}_3 - \widehat{P}^{\mathrm{c}}_2) - (\widetilde{P}^{\mathrm{f}}_3 - \widetilde{P}^{\mathrm{c}}_2)\}$, where we have made explicitly clear which are the fine and coarse levels. MIMC would assume (and require) that for an underlying Brownian motion path sample ω the associated payoff $P|_\omega$ and its approximate Euler–Maruyama approximation $\widetilde{P}_2|_\omega$ on the $l = 2$ level must be equal when the $\ell = 2$ is either the fine or coarse path and that $\widetilde{P}^{\mathrm{f}}_2|_\omega = \widetilde{P}^{\mathrm{c}}_2|_\omega$. As outlined, this is not the case in general, and our nested MLMC framework does not require this.

4. Numerical analysis.

The numerical analysis in this section has been split into several subsections and is built up incrementally. The analysis culminates in the two main results of this paper, which are Lemmas 4.10 and 4.11, which provide bounds for the variance of the four way differences in the nested MLMC correction term for Lipschitz and differentiable, and Lipschitz but nondifferentiable, functions, respectively. To achieve this, we begin by outlining our assumptions on the SDE in section 4.1 and in section 4.2 present Lemma 4.3, which will be the key tool in proving all other subsequent results. The remaining sections are then split up into the analysis of the simpler two way difference terms in the nested MLMC framework in section 4.3 and the more involved four way differences in section 4.4. Lastly, the analysis of the four way difference is subdivided into the Lipschitz and differentiable case in section 4.4.1 and the Lipschitz but nondifferentiable case in section 4.4.2, where the former setting is particularly relevant to mathematical finance and the analysis and simulation of instruments such as call and put options.

4.1. Assumptions about the SDE.

We begin by making two sets of assumptions which will be assumed to hold throughout the analysis.

The first concerns the drift and volatility functions and assumes a greater degree of smoothness than the usual assumptions used to prove half-order strong convergence for the Euler–Maruyama discretization [33].

Assumption 4.1. *The drift function $a:\mathbb{R}\to\mathbb{R}$ and volatility function $b:\mathbb{R}\to\mathbb{R}$ are both $C^1(\mathbb{R})$, and both they and their derivatives are Lipschitz continuous so that there exist constants L_a, L_b, L'_a, L'_b such that*

$$\begin{aligned} |a(x)-a(y)| &\le L_a|x-y|, & |b(x)-b(y)| &\le L_b|x-y|,\\ |a'(x)-a'(y)| &\le L'_a|x-y|, & |b'(x)-b'(y)| &\le L'_b|x-y|. \end{aligned}$$

The second concerns approximate normal random variables $\widetilde{Z}$ and their relationship to corresponding exact normal random variables Z.

Assumption 4.2. *Random variable pairs $(Z,\widetilde{Z})$ can be generated such that $Z\sim N(0,1)$, $\mathbb{E}[\widetilde{Z}]=0$, and $\mathbb{E}[\,|\widetilde{Z}-Z|^p]\le \mathbb{E}[\,|Z|^p]$ for all $p\ge 2$.*

In most cases, the pairs will be generated as $(\Phi^{-1}(U),\widetilde{Q}(U))$ for a common uniform random variable U, but we leave open the possibility that they may be generated in an alternative way; this is needed later for Lemma 4.7. Note also that $|\widetilde{Z}|\le|Z|+|\widetilde{Z}-Z|$, so a consequence of this assumption is that $\mathbb{E}[\,|\widetilde{Z}|^p]\le 2^p\,\mathbb{E}[\,|Z|^p]$; this is also used later.

4.2. Bounding the differences arising from the Euler–Maruyama scheme.

By using the Euler–Maruyama scheme to approximate the SDE, in our nested MLMC setting we will frequently be required to bound the differences of numerous terms arising from the scheme. These bounds, and thus the proofs of the main results, are greatly simplified by first proving the following lemma using the discrete Burkholder–Davis–Gundy inequality [10].

Lemma 4.3. *Suppose that for a time interval $[0,T]$ with $T=Nh$ and $t_n=nh$, we have the discrete equations*

$$D_{n+1}=D_n+A_nD_nh+B_nD_nh^{1/2}\widetilde{Z}_n+\Theta_nh+\Psi_nh^{1/2}$$

subject to initial data $D_0 = 0$, where $(Z_n, \widetilde{Z}_n), n \geq 0$ are i.i.d. random variable pairs satisfying Assumption 4.2, and using the standard filtration $\mathcal{F}_{t_n}$ generated by $(Z_n, \widetilde{Z}_n)$; then A_n, B_n, Θ_n are $\mathcal{F}_{t_n}$-adapted, depending only on X_0 and $(Z_m, \widetilde{Z}_m), m<n$, whereas Ψ_n is $\mathcal{F}_{t_{n+1}}$-adapted with $\mathbb{E}[\Psi_n|\mathcal{F}_{t_n}] = 0$.

Furthermore, suppose that $|A_n| \leq L_a$, $|B_n| \leq L_b$, where L_a, L_b are as defined in Assumption 4.1, and for some $p \geq 2$ there are constants c_1, c_2, which do not depend on h, such that

$$\max_{0\leq n<N} \mathbb{E}[\,|\Psi_n|^p] \leq c_1, \qquad \max_{0\leq n<N} \mathbb{E}[\,|\Theta_n|^p] \leq c_2.$$

Then, there is a constant c_3 depending only on L_a, L_b, p, T such that

$$\mathbb{E}\left[\max_{0\leq n\leq N} |D_n|^p\right] \leq c_3(c_1 + c_2).$$

Note. The lemma includes as a special case $\widetilde{Z}_n = Z_n$, i.e., with normal random variables rather than approximate normal random variables. Crucially though, the approximate random variable $\widetilde{Z}$ comes from the random variable pair $(Z, \widetilde{Z})$ such that Assumption 4.2 is satisfied, which is necessary for our later results.

Proof. Summing over the first n timesteps gives

$$D_n = \sum_{m=0}^{n-1} \left\{ A_m D_m h + B_m D_m h^{1/2} \widetilde{Z}_m + \Theta_m h + \Psi_m h^{1/2} \right\},$$

and therefore if we define $E_n \equiv \mathbb{E}[\max_{0<n'\leq n} |D_{n'}|^p]$ we obtain, through Jensen's inequality,

$$\begin{aligned} E_n \leq\ & 4^{p-1}\,\mathbb{E}\left[\max_{n'\leq n}\left|\sum_{m=0}^{n'-1} A_m D_m h\right|^p\right] + 4^{p-1}\,\mathbb{E}\left[\max_{n'\leq n}\left|\sum_{m=0}^{n'-1} B_m D_m h^{1/2}\widetilde{Z}_m\right|^p\right] \\ & + 4^{p-1}\,\mathbb{E}\left[\max_{n'\leq n}\left|\sum_{m=0}^{n'-1} \Theta_m h\right|^p\right] + 4^{p-1}\,\mathbb{E}\left[\max_{n'\leq n}\left|\sum_{m=0}^{n'-1} \Psi_m h^{1/2}\right|^p\right]. \end{aligned}$$

For the first term, using Jensen's inequality again gives

$$\mathbb{E}\left[\max_{n'\leq n}\left|\sum_{m=0}^{n'-1} A_m D_m h\right|^p\right] \ \leq\ h^p n^{p-1} \sum_{m=0}^{n-1} \mathbb{E}\left[\,|A_m D_m|^p\right] \ \leq\ h\, T^{p-1} L_a^p \sum_{m=0}^{n-1} E_m.$$

For the second term the discrete time Burkholder–Davis–Gundy inequality [10], together with Jensen's inequality and Assumption 4.2, gives the bound

$$\begin{aligned} \mathbb{E}\left[\max_{n'\leq n}\left|\sum_{m=0}^{n'-1} B_m D_m h^{1/2}\widetilde{Z}_m\right|^p\right] &\leq C_p\ \mathbb{E}\left[\left|\sum_{m=0}^{n-1} B_m^2 D_m^2 h \widetilde{Z}_m^2\right|^{p/2}\right] \\ &\leq C_p\ h^{p/2} n^{p/2-1} \sum_{m=0}^{n-1} \mathbb{E}\left[\,|B_m D_m|^p |\widetilde{Z}_m|^p\right] \\ &\leq C_p\, h\, T^{p/2-1} L_b^p\, 2^p\, \mathbb{E}[\,|Z|^p] \sum_{m=0}^{n-1} E_m \end{aligned}$$

with the constant C_p depending only on p. Similarly, the third term has the bound

$$\mathbb{E}\left[\max_{n'\le n}\left|\sum_{m=0}^{n'-1}\Theta_m h\right|^p\right] \;\le\; T^p c_2,$$

and the fourth term has the bound

$$\mathbb{E}\left[\max_{n'\le n}\left|\sum_{m=0}^{n'-1}\Psi_m h^{1/2}\right|^p\right] \;\le\; C_p\, T^{p/2}\, c_1.$$

Combining these four bounds we obtain

$$E_n \le 4^{p-1}\left(T^{p-1}L_a^p + C_p T^{p/2-1} L_b^p\, 2^p\, \mathbb{E}[\,|Z|^p]\right)\, h \sum_{m=0}^{n-1} E_m + 4^{p-1}\left(T^p c_2 + C_p\, T^{p/2}\, c_1\right),$$

and therefore by Grönwall's inequality we obtain

$$E_n \le 4^{p-1}(T^p\, c_2 + C_p\, T^{p/2}\, c_1)\, \exp\!\left(4^{p-1}(T^p L_a^p + C_p\, 2^p\, \mathbb{E}[\,|Z|^p]\, T^{p/2} L_b^p)\right).$$

Setting $c_3 = 4^{p-1}\max(T^p, C_p\, T^{p/2})\;\exp(4^{p-1}(T^p L_a^p + C_p\, 2^p\, \mathbb{E}[\,|Z|^p]\, T^{p/2} L_b^p))$ completes the proof. ■

4.3. Analysis of the two way difference terms. In the nested MLMC setting, there is a two way and a four way difference, and we begin our analysis in this section by considering the simpler. The first result we require is a generalization of Lemma 2 in [20].

Lemma 4.4. *For a fixed time interval $T = Nh$, for any $p \ge 2$ there exists a constant c which depends on X_0, a, b, T, p but not on h or $\mathbb{E}[\,|\widetilde{Z}-Z|^p]$ such that*

$$\mathbb{E}\left[\max_{0\le n\le N}|\widetilde{X}_n-\widehat{X}_n|^p\right] \le c\; \mathbb{E}[\,|\widetilde{Z}-Z|^p].$$

Proof. Taking the difference between

$$\widetilde{X}_{n+1} = \widetilde{X}_n + a(\widetilde{X}_n)\, h + b(\widetilde{X}_n)\, h^{1/2}\widetilde{Z}_n \tag{4.1}$$

and

$$\widehat{X}_{n+1} = \widehat{X}_n + a(\widehat{X}_n)\, h + b(\widehat{X}_n)\, h^{1/2} Z_n \tag{4.2}$$

and defining $D_n \equiv \widetilde{X}_n - \widehat{X}_n$, we obtain

$$D_{n+1} = D_n + a'(\xi_{1,n})\, D_n h + b'(\xi_{2,n}) D_n h^{1/2}\widetilde{Z}_n + b(\widehat{X}_n)\, h^{1/2}(\widetilde{Z}_n - Z_n) \tag{4.3}$$

for suitably defined $\xi_{1,n}$, $\xi_{2,n}$ arising from the mean value theorem, Lemma A.1.

This is in the correct form to apply Lemma 4.3 since $|a'(\xi_{1,n})| \leq L_a, \quad |b'(\xi_{2,n})| \leq L_b,$

$$\mathbb{E}\left[b(\widehat{X}_n)\,(\widetilde{Z}_n - Z_n) \mid \widehat{X}_n\right] \;=\; b(\widehat{X}_n)\;\mathbb{E}[\widetilde{Z}_n - Z_n] \;=\; 0,$$

and

$$\mathbb{E}\left[|b(\widehat{X}_n)\,(\widetilde{Z}_n - Z_n)|^p\right] \;=\; \mathbb{E}[\,|b(\widehat{X}_n)|^p]\;\mathbb{E}[\,|\widetilde{Z}_n - Z_n|^p].$$

The result then follows from Lemma 4.3 after noting that the standard analysis of the Euler–Maruyama method (e.g., see [33]) proves that $\mathbb{E}[\,|\widehat{X}_n|^p]$, $\mathbb{E}[\,|a(\widehat{X}_n)|^p]$, and $\mathbb{E}[\,|b(\widehat{X}_n)|^p]$ are all uniformly bounded on the time interval $[0,T]$. ∎

Corollary 4.5. *For a fixed time interval $T = Nh$, for any $p \geq 2$ there exists a constant c which depends on X_0, a, b, T, p but not on h such that for any $0 < n \leq N$,*

$$\mathbb{E}[\,|\widetilde{X}_n|^p] \leq c, \quad \mathbb{E}[\,|a(\widetilde{X}_n)|^p] \leq c, \quad \mathbb{E}[\,|b(\widetilde{X}_n)|^p] \leq c.$$

Proof. The standard analysis of the Euler–Maruyama method proves that $\mathbb{E}[\,|\widehat{X}_n|^p]$ is uniformly bounded on $[0,T]$, and so it follows from Lemma 4.4 that there exist constants c_1, c_2 such that

$$\mathbb{E}[\,|\widetilde{X}_n|^p] \;\leq\; c_1 + c_2\,\mathbb{E}[\,|\widetilde{Z} - Z|^p] \;\leq\; c_1 + c_2\,\mathbb{E}[\,|Z|^p]$$

due to Assumption 4.2. Since $|a(\widetilde{X}_n)| \leq |a(0)| + L_a|\widetilde{X}_n|$, it follows that

$$|a(\widetilde{X}_n)|^p \leq 2^{p-1}\left(|a(0)|^p + L_a^p\,|\widetilde{X}_n|^p\right),$$

and therefore $\mathbb{E}[\,|a(\widetilde{X}_n)|^p]$ can be uniformly bounded, and similarly $\mathbb{E}[\,|b(\widetilde{X}_n)|^p]$. ∎

Corollary 4.6. *For a fixed time interval $T = Nh$, for any $p \geq 2$ there exists a constant c which depends on X_0, a, b, T, p but not on h or $\mathbb{E}[\,|\widetilde{Z} - Z|^p]$ such that*

$$\max_{0 \leq n < N} \mathbb{E}[\,|\widetilde{X}_{n+1} - \widetilde{X}_n|^p] \leq c\,h^{p/2}, \qquad \max_{0 \leq n < N} \mathbb{E}[\,|\widehat{X}_{n+1} - \widehat{X}_n|^p] \leq c\,h^{p/2},$$

and

$$\max_{0 \leq n < N} \mathbb{E}\left[\,|(\widetilde{X}_{n+1} - \widetilde{X}_n) - (\widehat{X}_{n+1} - \widehat{X}_n)|^p\right] \leq c\,h^{p/2}\,\mathbb{E}[\,|\widetilde{Z} - Z|^p].$$

Proof. Since

$$\mathbb{E}[\,|\widetilde{X}_{n+1} - \widetilde{X}_n|^p] \;\leq\; 2^{p-1}\left(\mathbb{E}[\,|a(\widetilde{X}_n)|^p]\,h^p + \mathbb{E}[\,|b(\widetilde{X}_n)|^p]\,h^{p/2}\mathbb{E}[\,|\widetilde{Z}_n|^p]\right),$$

the first assertion follows from the uniform boundedness of $\mathbb{E}[\,|a(\widetilde{X}_n)|^p]$ and $\mathbb{E}[\,|b(\widetilde{X}_n)|^p]$, together with the trivial inequality $h^p \leq h^{p/2}T^{p/2}$ and the $\mathbb{E}[\,|\widetilde{Z}_n|^p]$ bound due to Assumption 4.2. The second assertion follows similarly.

Rearranging (4.3) gives

$$D_{n+1} - D_n = a'(\xi_{1,n})\,D_n h + b'(\xi_{2,n}) D_n h^{1/2}\,\widetilde{Z}_n + b(\widehat{X}_n)\,h^{1/2}(\widetilde{Z}_n - Z_n)$$

with $D_n \equiv \widetilde{X}_n - \widehat{X}_n$. Hence,

$$\begin{aligned}&\mathbb{E}[\,|(\widetilde{X}_{n+1} - \widetilde{X}_n) - (\widehat{X}_{n+1} - \widehat{X}_n)|^p]\\ &\leq 3^{p-1}\left(L_a^p\,\mathbb{E}[\,|D_n|^p]\,h^p + L_b^p\,\mathbb{E}[\,|D_n|^p]\,h^{p/2}\,\mathbb{E}[\,|\widetilde{Z}_n|^p] + \mathbb{E}[\,|b(\widehat{X}_n)|^p]\,h^{p/2}\,\mathbb{E}[\,|\widetilde{Z} - Z|^p]\right),\end{aligned}$$

and the third assertion follows from the bounds on $\mathbb{E}[\,|\widetilde{X}_n - \widehat{X}_n|^p]$ and $\mathbb{E}[\,|b(\widehat{X}_n)|^p]$. ∎

We now have the first MLMC results involving level ℓ fine paths and level $\ell-1$ coarse paths, as defined in (3.1)–(3.4).

Lemma 4.7. *For a fixed time interval $T=Nh$, for any $p\geq 2$ there exists a constant c which depends on X_0, a, b, T, p but not on h or $\mathbb{E}[\,|\widetilde{Z}-Z|^p]$ such that*

$$\mathbb{E}\left[\max_{0<n\leq N}|\widetilde{X}^{\rm f}_n-\widehat{X}^{\rm f}_n|^p\right]\leq c\,\mathbb{E}[\,|\widetilde{Z}-Z|^p], \quad \mathbb{E}\left[\max_{0<n\leq N}|\widetilde{X}^{\rm c}_n-\widehat{X}^{\rm c}_n|^p\right]\leq c\,\mathbb{E}[\,|\widetilde{Z}-Z|^p].$$

Proof. The first assertion comes immediately from Lemma 4.4, but the second assertion requires the observation that if we set

$$Z_3=(Z_1+Z_2)/\sqrt{2}, \qquad \widetilde{Z}_3=(\widetilde{Z}_1+\widetilde{Z}_2)/\sqrt{2},$$

where the independent pairs $(Z_1,\widetilde{Z}_1)$ and $(Z_2,\widetilde{Z}_2)$ satisfy Assumption 4.2, then $Z_3\sim N(0,1)$, $\mathbb{E}[\widetilde{Z}_3]=0$, and by Jensen's inequality

$$\mathbb{E}\left[\,|\widetilde{Z}_3-Z_3|^p\right]\ \leq\ 2^{p/2-1}\left(\mathbb{E}[\,|\widetilde{Z}_1-Z_1|^p]+\mathbb{E}[\,|\widetilde{Z}_2-Z_2|^p]\right)\ \leq\ 2^{p/2}\,\mathbb{E}[\,|Z|^p].$$

Therefore the pair $(Z_3,\widetilde{Z}_3)$ also satisfies Assumption 4.2 apart from an increased bound on $\mathbb{E}[\,|\widetilde{Z}_3-Z_3|^p]$. This requires minor changes to the constants in the subsequent lemmas, but in the end the desired result follows from Lemma 4.4. ∎

Lemma 4.8. *For a fixed time interval $T=Nh$, for any $p\geq 2$ there exists a constant c which depends on X_0, a, b, T, p but not on h or $\mathbb{E}[\,|\widetilde{Z}-Z|^p]$ such that*

$$\mathbb{E}\left[\max_{0<n\leq N}|\widetilde{X}^{\rm f}_n-\widetilde{X}^{\rm c}_n|^p\right]\leq c\,h^{p/2}, \quad \mathbb{E}\left[\max_{0<n\leq N}|\widehat{X}^{\rm f}_n-\widehat{X}^{\rm c}_n|^p\right]\leq c\,h^{p/2},$$

Proof. Defining $D_n\equiv\widetilde{X}^{\rm f}_n-\widetilde{X}^{\rm c}_n$, and taking the difference between (3.3) and (3.4) gives

$$\begin{aligned}
D_{n+1} &= D_n+(a(\widetilde{X}^{\rm f}_n)-a(\widetilde{X}^{\rm c}_n))\,h+(b(\widetilde{X}^{\rm f}_n)-b(\widetilde{X}^{\rm c}_n))\,h^{1/2}\widetilde{Z}_n\\
&\qquad +(a(\widetilde{X}^{\rm c}_n)-a(\widetilde{X}^{\rm c}_{\underline{n}}))\,h+(b(\widetilde{X}^{\rm c}_n)-b(\widetilde{X}^{\rm c}_{\underline{n}}))\,h^{1/2}\widetilde{Z}_n\\
&= D_n+a'(\xi_{1,n})\,D_n h+b'(\xi_{2,n})D_n h^{1/2}\widetilde{Z}_n\\
&\qquad +a'(\xi_{3,n})\,(\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}})\,h+b'(\xi_{4,n})\,(\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}})\,h^{1/2}\widetilde{Z}_n
\end{aligned}$$

for suitably defined $\xi_{1,n}$, $\xi_{2,n}, \xi_{3,n}$, $\xi_{4,n}$ arising from the mean value theorem. Noting that

$$\mathbb{E}[\,|a'(\xi_{3,n})\,(\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}})|^p]\leq L_a^p\,\mathbb{E}[\,|\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}|^p]$$

and

$$\mathbb{E}[\,|b'(\xi_{4,n})\,(\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}})\widetilde{Z}_n|^p]\ \leq\ L_b^p\,\mathbb{E}[\,|\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}|^p]\,\mathbb{E}[\,|\widetilde{Z}|^p]\ \leq\ L_b^p\,2^p\,\mathbb{E}[\,|Z|^p]\,\mathbb{E}[\,|\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}|^p],$$

the first assertion then follows again from Lemma 4.3 after using the bounds for $\mathbb{E}[\,|\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}|^p]$ which come from Corollary 4.6.

The second assertion follows similarly. ∎

4.4. Analysis of the four way difference correction terms. We now come to the analysis of the four way cross-difference terms from the nested MLMC formulation. For this analysis, we will make use of Lemma A.2, which is a four way variant of the mean value theorem (Lemma A.1). However, we will see for the analysis of Lipschitz but nondifferentiable functions, which are especially important in mathematical finance in the form of call and put options, we will require a second and more involved analysis, presented later in section 4.4.2.

4.4.1. Lipschitz and differentiable functions.

Lemma 4.9. *For a fixed time interval $T = Nh$, for any p, q with $2 \le p < q$ there exists a constant c which depends on X_0, a, b, T but not on h or $\mathbb{E}[\,|\widetilde{Z}-Z|^q]$ such that*

$$\mathbb{E}\left[\max_{0\le n<N}\left|\widetilde{X}^{\rm f}_n-\widetilde{X}^{\rm c}_n-\widehat{X}^{\rm f}_n+\widehat{X}^{\rm c}_n\right|^p\right] \le c\, h^{p/2}\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{p/q}.$$

Proof. Define $D_n \equiv \widetilde{X}^{\rm f}_n-\widetilde{X}^{\rm c}_n-\widehat{X}^{\rm f}_n+\widehat{X}^{\rm c}_n$; then the difference of (3.1)–(3.4) together with Lemma A.2 gives

$$\begin{aligned}
&D_{n+1}\\
&= D_n + \left(a(\widetilde{X}^{\rm f}_n)-a(\widetilde{X}^{\rm c}_n)-a(\widehat{X}^{\rm f}_n)+a(\widehat{X}^{\rm c}_n)\right) h + \left(b(\widetilde{X}^{\rm f}_n)-b(\widetilde{X}^{\rm c}_n)-b(\widehat{X}^{\rm f}_n)+b(\widehat{X}^{\rm c}_n)\right) h^{1/2}\widetilde{Z}_n\\
&\qquad + \left(a(\widetilde{X}^{\rm c}_n)-a(\widetilde{X}^{\rm c}_{\underline{n}})-a(\widehat{X}^{\rm c}_n)+a(\widehat{X}^{\rm c}_{\underline{n}})\right) h + \left(b(\widetilde{X}^{\rm c}_n)-b(\widetilde{X}^{\rm c}_{\underline{n}})-b(\widehat{X}^{\rm c}_n)+b(\widehat{X}^{\rm c}_{\underline{n}})\right) h^{1/2}\widetilde{Z}_n\\
&\qquad\qquad + \left(b(\widehat{X}^{\rm f}_n)-b(\widehat{X}^{\rm c}_{\underline{n}})\right) h^{1/2}(\widetilde{Z}_n-Z_n)\\
&= D_n + a'(\xi_{1,n})\, D_n\, h + b'(\xi_{2,n})\, D_n\, h^{1/2}\widetilde{Z}_n\\
&\qquad + a'(\xi_{3,n})\left(\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}-\widehat{X}^{\rm c}_n+\widehat{X}^{\rm c}_{\underline{n}}\right) h + b'(\xi_{4,n})\left(\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}-\widehat{X}^{\rm c}_n+\widehat{X}^{\rm c}_{\underline{n}}\right) h^{1/2}\widetilde{Z}_n\\
&\qquad + (R_{1,n}+R_{3,n})\, h + (R_{2,n}+R_{4,n})\, h^{1/2}\widetilde{Z}_n + b'(\xi_{5,n})\left(\widehat{X}^{\rm f}_n-\widehat{X}^{\rm c}_{\underline{n}}\right) h^{1/2}(\widetilde{Z}_n-Z_n)
\end{aligned}$$

for suitably defined $\xi_{1,n}, \xi_{2,n}, \xi_{3,n}, \xi_{4,n}, \xi_{5,n}$ arising from Lemma A.2 and the mean value theorem, and with

$$\begin{aligned}
|R_{1,n}| &\le \tfrac{1}{2}L'_a\left(|\widetilde{X}^{\rm f}_n-\widetilde{X}^{\rm c}_n| + |\widehat{X}^{\rm f}_n-\widehat{X}^{\rm c}_n|\right)\left(|\widetilde{X}^{\rm f}_n-\widehat{X}^{\rm f}_n| + |\widetilde{X}^{\rm c}_n-\widehat{X}^{\rm c}_n|\right),\\
|R_{2,n}| &\le \tfrac{1}{2}L'_b\left(|\widetilde{X}^{\rm f}_n-\widetilde{X}^{\rm c}_n| + |\widehat{X}^{\rm f}_n-\widehat{X}^{\rm c}_n|\right)\left(|\widetilde{X}^{\rm f}_n-\widehat{X}^{\rm f}_n| + |\widetilde{X}^{\rm c}_n-\widehat{X}^{\rm c}_n|\right),\\
|R_{3,n}| &\le \tfrac{1}{2}L'_a\left(|\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}| + |\widehat{X}^{\rm c}_n-\widehat{X}^{\rm c}_{\underline{n}}|\right)\left(|\widetilde{X}^{\rm c}_n-\widehat{X}^{\rm c}_n| + |\widetilde{X}^{\rm c}_{\underline{n}}-\widehat{X}^{\rm c}_{\underline{n}}|\right),\\
|R_{4,n}| &\le \tfrac{1}{2}L'_b\left(|\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}| + |\widehat{X}^{\rm c}_n-\widehat{X}^{\rm c}_{\underline{n}}|\right)\left(|\widetilde{X}^{\rm c}_n-\widehat{X}^{\rm c}_n| + |\widetilde{X}^{\rm c}_{\underline{n}}-\widehat{X}^{\rm c}_{\underline{n}}|\right).
\end{aligned}$$

This equation is in the correct form for the application of Lemma 4.3 with

$$\begin{aligned}
\Theta_n &= a'(\xi_{3,n})\left(\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}-\widehat{X}^{\rm c}_n+\widehat{X}^{\rm c}_{\underline{n}}\right) + (R_{1,n}+R_{3,n}),\\
\Psi_n &= b'(\xi_{4,n})\left(\widetilde{X}^{\rm c}_n-\widetilde{X}^{\rm c}_{\underline{n}}-\widehat{X}^{\rm c}_n+\widehat{X}^{\rm c}_{\underline{n}}\right)\widetilde{Z}_n + (R_{2,n}+R_{4,n})\,\widetilde{Z}_n\\
&\qquad + b'(\xi_{5,n})\left(\widehat{X}^{\rm f}_n-\widehat{X}^{\rm c}_{\underline{n}}\right)(\widetilde{Z}_n-Z_n).
\end{aligned}$$

Corollary 4.6 and Lemma 4.4 together with the Hölder inequality imply that there exists a constant c such that

$$\begin{aligned}\mathbb{E}[\,|\widetilde{X}^{\mathrm{f}}_n-\widetilde{X}^{\mathrm{c}}_n|^p|\widetilde{X}^{\mathrm{f}}_n-\widehat{X}^{\mathrm{f}}_n|^p] &\le \left(\mathbb{E}[\,|\widetilde{X}^{\mathrm{f}}_n-\widetilde{X}^{\mathrm{c}}_n|^{p/(1-p/q)}]\right)^{1-p/q}\left(\mathbb{E}[\,|\widetilde{X}^{\mathrm{f}}_n-\widehat{X}^{\mathrm{f}}_n|^q]\right)^{p/q}\\ &\le c\,h^{p/2}\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{p/q}.\end{aligned}$$

Bounding the other terms similarly, there is a different constant c such that

$$\mathbb{E}[\,|R_{i,n}|^p]\le c\;h^{p/2}\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{p/q},\quad i=1,2,3,4.$$

Due to Corollary 4.6 we also have

$$\mathbb{E}[\,|\widetilde{X}^{\mathrm{c}}_n-\widetilde{X}^{\mathrm{c}}_{\underline{n}}-\widehat{X}^{\mathrm{c}}_n+\widehat{X}^{\mathrm{c}}_{\underline{n}}|^p]\le c\,h^{p/2}\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{p/q}$$

for some constant c, and finally, for another constant c,

$$\mathbb{E}[\,|(\widehat{X}^{\mathrm{f}}_n-\widehat{X}^{\mathrm{c}}_{\underline{n}})(\widetilde{Z}_n-Z_n)|^p]=\mathbb{E}[\,|\widehat{X}^{\mathrm{f}}_n-\widehat{X}^{\mathrm{c}}_{\underline{n}}|^p]\;\mathbb{E}[\,|\widetilde{Z}-Z|^p]\le c\,h^{p/2}\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{p/q}.$$

Hence, we end up concluding that there exists another constant c such that

$$\mathbb{E}[\,|\Psi_n|^p]\le c\,h^{p/2}\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{p/q},\quad \mathbb{E}[\,|\Theta_n|^p]\le c\,h^{p/2}\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{p/q},$$

and then Lemma 4.3 gives us the desired final result. ■

We now obtain a lemma for $C^1(\mathbb{R})$ output functions $f(x)$ which are locally Lipschitz with at worst a polynomial growth as $|x|\to\infty$.

Lemma 4.10. *For a fixed time interval $T=Nh$, if the function $f:\mathbb{R}\to\mathbb{R}$ is $C^1(\mathbb{R})$ and there is an exponent $r>0$ and Lipschitz constants L_f, L'_f such that*

$$|f(x)-f(y)|\le L_f\,(1+|x|^r+|y|^r)\;|x-y|,\qquad |f'(x)-f'(y)|\le L'_f\,(1+|x|^r+|y|^r)\;|x-y|,$$

then for any $q>2$ there exists a constant c which depends on X_0,a,b,f,T but not on h or $\mathbb{E}[\,|\widetilde{Z}-Z|^q]$ such that

$$\mathbb{V}\left[f(\widehat{X}^{\mathrm{f}}_N)-f(\widehat{X}^{\mathrm{c}}_N)-f(\widetilde{X}^{\mathrm{f}}_N)+f(\widetilde{X}^{\mathrm{c}}_N)\right]\le c\,h\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{2/q}.$$

Proof. Given the assumptions on f, we are able to follow the proof of Lemma A.2 to obtain

$$f(\widehat{X}^{\mathrm{f}}_N)-f(\widehat{X}^{\mathrm{c}}_N)-f(\widetilde{X}^{\mathrm{f}}_N)+f(\widetilde{X}^{\mathrm{c}}_N)=f'(\xi)\;(\widehat{X}^{\mathrm{f}}_N-\widehat{X}^{\mathrm{c}}_N-\widetilde{X}^{\mathrm{f}}_N+\widetilde{X}^{\mathrm{c}}_N)+R,$$

where

$$|f'(\xi)|\le L_f\left(1+|\widehat{X}^{\mathrm{f}}_N|^r+|\widehat{X}^{\mathrm{c}}_N|^r+|\widetilde{X}^{\mathrm{f}}_N|^r+|\widetilde{X}^{\mathrm{c}}_N|^r\right)$$

and

$$|R| \le \tfrac{1}{2}L'_f \left(1 + |\widehat{X}^{\rm f}_N|^r + |\widehat{X}^{\rm c}_N|^r + |\widetilde{X}^{\rm f}_N|^r + |\widetilde{X}^{\rm c}_N|^r\right)$$
$$\times \left(|\widehat{X}^{\rm f}_N - \widehat{X}^{\rm c}_N| + |\widetilde{X}^{\rm f}_N - \widetilde{X}^{\rm c}_N|\right) \left(|\widehat{X}^{\rm f}_N - \widetilde{X}^{\rm f}_N| + |\widehat{X}^{\rm c}_N - \widetilde{X}^{\rm c}_N|\right).$$

Hence,

$$\mathbb{V}\left[f(\widehat{X}^{\rm f}_N) - f(\widehat{X}^{\rm c}_N) - f(\widetilde{X}^{\rm f}_N) + f(\widetilde{X}^{\rm c}_N)\right] \le 2\,\mathbb{E}\left[(f'(\xi))^2 |\widehat{X}^{\rm f}_N - \widehat{X}^{\rm c}_N - \widetilde{X}^{\rm f}_N + \widetilde{X}^{\rm c}_N|^2\right] + 2\,\mathbb{E}[R^2].$$

Due to Hölder's inequality,

$$\mathbb{E}\left[|\widehat{X}^{\rm f}_N|^{2r}\,|\widehat{X}^{\rm f}_N - \widehat{X}^{\rm c}_N|^2\,|\widehat{X}^{\rm f}_N - \widetilde{X}^{\rm f}_N|^2\right]$$
$$\le \left(\mathbb{E}[\,|\widehat{X}^{\rm f}_N|^{2r/(1/2-1/q)}]\right)^{1/2-1/q} \left(\mathbb{E}[\,|\widehat{X}^{\rm f}_N - \widehat{X}^{\rm c}_N|^{2/(1/2-1/q)}]\right)^{1/2-1/q} \left(\mathbb{E}[\,|\widehat{X}^{\rm f}_N - \widetilde{X}^{\rm f}_N|^q]\right)^{2/q}.$$

Note that $\mathbb{E}[\,|\widehat{X}^{\rm f}_N|^{2r/(1/2-1/q)}]$ is finite and uniformly bounded due to Corollary 4.5. The other terms in $\mathbb{E}[R^2]$ can be bounded similarly, and therefore due to the bounds from Lemmas 4.4 and 4.8 there exists a constant c, not depending on h or $\mathbb{E}[\,|\widetilde{Z} - Z|^q]$, such that

$$\mathbb{E}[R^2] \le c\,h \left(\mathbb{E}[\,|\widetilde{Z} - Z|^q]\right)^{2/q}.$$

Similarly, choosing p such that $2 < p < q$, due to Hölder's inequality,

$$\mathbb{E}\left[(f'(\xi))^2 |\widehat{X}^{\rm f}_N - \widehat{X}^{\rm c}_N - \widetilde{X}^{\rm f}_N + \widetilde{X}^{\rm c}_N|^2\right]$$
$$\le \left(\mathbb{E}[\,|f'(\xi)|^{2/(1-2/p)}]\right)^{1-2/p} \left(\mathbb{E}\left[\,|\widehat{X}^{\rm f}_N - \widehat{X}^{\rm c}_N - \widetilde{X}^{\rm f}_N + \widetilde{X}^{\rm c}_N|^p\right]\right)^{2/p},$$

$\mathbb{E}[\,|f'(\xi)|^{2/(1-2/p)}]$ is finite and uniformly bounded due to Corollary 4.5, and therefore the bound for $\mathbb{E}[\,|\widehat{X}^{\rm f}_N - \widehat{X}^{\rm c}_N - \widetilde{X}^{\rm f}_N + \widetilde{X}^{\rm c}_N|^p]$ from Lemma 4.9 completes the proof. ■

4.4.2. Lipschitz and nondifferentiable functions. In finance applications, put and call options correspond to $f(x) \equiv \max(K - x, 0)$ and $\max(x - K, 0)$, respectively, with $K > 0$ being the "strike." More generally, we can consider functions f which are globally Lipschitz with a derivative which exists and is continuous everywhere except at a single point K.

Heuristically, the four values $\widehat{X}^{\rm f}_N, \widehat{X}^{\rm c}_N, \widetilde{X}^{\rm f}_N, \widetilde{X}^{\rm c}_N$ do not differ from each other, or from X_T, by more than $O(\max\{h^{1/2}, (\mathbb{E}[\,|\widetilde{Z} - Z|^2])^{1/2}\})$. If X_T has a bounded probability density, then the probability that X_T is within this distance of K is $O(\max\{h^{1/2}, (\mathbb{E}[\,|\widetilde{Z} - Z|^2])^{1/2}\})$, and in this case the global Lipschitz property for f gives

$$f(\widehat{X}^{\rm f}_N) - f(\widehat{X}^{\rm c}_N) - f(\widetilde{X}^{\rm f}_N) + f(\widetilde{X}^{\rm c}_N) = O\left(\min\left\{h^{1/2}, (\mathbb{E}[\,|\widetilde{Z} - Z|^2])^{1/2}\right\}\right),$$

with the first term in the minimum coming from

$$f(\widehat{X}^{\rm f}_N) - f(\widehat{X}^{\rm c}_N) - f(\widetilde{X}^{\rm f}_N) + f(\widetilde{X}^{\rm c}_N) = \left(f(\widehat{X}^{\rm f}_N) - f(\widehat{X}^{\rm c}_N)\right) - \left(f(\widetilde{X}^{\rm f}_N) + f(\widetilde{X}^{\rm c}_N)\right)$$

together with Lemma 4.8, while the second term comes from

$$f(\widehat{X}^{\mathrm{f}}_N) - f(\widehat{X}^{\mathrm{c}}_N) - f(\widetilde{X}^{\mathrm{f}}_N) + f(\widetilde{X}^{\mathrm{c}}_N) = \left(f(\widehat{X}^{\mathrm{f}}_N) - f(\widetilde{X}^{\mathrm{f}}_N) \right) - \left(f(\widehat{X}^{\mathrm{c}}_N) + f(\widetilde{X}^{\mathrm{c}}_N) \right)$$

together with Lemma 4.4.

On the other hand, if X_T is more than this distance from K, then all four values will be on the same side of K, and then

$$f(\widehat{X}^{\mathrm{f}}_N) - f(\widehat{X}^{\mathrm{c}}_N) - f(\widetilde{X}^{\mathrm{f}}_N) + f(\widetilde{X}^{\mathrm{c}}_N) = O\left(h^{1/2}\, (\mathbb{E}[\,|\widetilde{Z} - Z|^2])^{1/2} \right).$$

Consequently,

$$\begin{aligned}
&\mathbb{V}[f(\widehat{X}^{\mathrm{f}}_N) - f(\widehat{X}^{\mathrm{c}}_N) - f(\widetilde{X}^{\mathrm{f}}_N) + f(\widetilde{X}^{\mathrm{c}}_N)] \\
&= O\left(\max\left\{ h^{1/2},\, (\mathbb{E}[\,|\widetilde{Z} - Z|^2])^{1/2} \right\} \right) \times O\left(\min\left\{ h,\, \mathbb{E}[\,|\widetilde{Z} - Z|^2] \right\} \right) + O\left(h\; \mathbb{E}[\,|\widetilde{Z} - Z|^2] \right) \\
&= O\left(\min\left\{ h\, (\mathbb{E}[\,|\widetilde{Z} - Z|^2])^{1/2},\, h^{1/2}\, \mathbb{E}[\,|\widetilde{Z} - Z|^2] \right\} \right).
\end{aligned}$$

The bound in the following lemma (for the case $q \approx 2$) is slightly weaker, but the proof follows along similar lines in establishing that the dominant contribution to the variance comes from samples with X_T near K.

In the proof, we will use the notation

$$g_1(h, \mathbb{E}[\,|\widetilde{Z} - Z|^q]) \prec g_2(h, \mathbb{E}[\,|\widetilde{Z} - Z|^q])$$

for any two strictly positive functions g_1, g_2 to mean that there exists a constant $c > 0$ which does not depend on h or $\mathbb{E}[\,|\widetilde{Z} - Z|^q]$ such that

$$g_1(h, \mathbb{E}[\,|\widetilde{Z} - Z|^q]) < c\; g_2(h, \mathbb{E}[\,|\widetilde{Z} - Z|^q]).$$

Note that if $0 < a < b$, then

$$h^b \le T^{b-a} h^a \quad \Longrightarrow \quad h^b \prec h^a,$$

and likewise, due to Assumption 4.2,

$$\left(\mathbb{E}[\,|\widetilde{Z} - Z|^q] \right)^b \le (\mathbb{E}[\,|Z|^q])^{b-a} \left(\mathbb{E}[\,|\widetilde{Z} - Z|^q] \right)^a \quad \Longrightarrow \quad \left(\mathbb{E}[\,|\widetilde{Z} - Z|^q] \right)^b \prec \left(\mathbb{E}[\,|\widetilde{Z} - Z|^q] \right)^a.$$

Lemma 4.11. *Suppose that the conditions of Lemma 4.10 are slightly modified so that* $f : \mathbb{R} \to \mathbb{R}$ *is* $C^1(\mathbb{R} \backslash K)$ *and there is an exponent* $r > 0$ *and constants* L_f, L'_f *such that*

$$\begin{aligned}
|f(x) - f(y)| &\le L_f (1 + |x|^r + |y|^r)\, |x - y| \quad \textit{for all } x, y, \\
|f'(x) - f'(y)| &\le L'_f (1 + |x|^r + |y|^r)\, |x - y| \quad \textit{if } \underline{\textit{either}}\ x > y > K\ \underline{\textit{or}}\ x < y < K,
\end{aligned}$$

and furthermore X_T *has a bounded probability density in the neighborhood of* K *and therefore there is a constant* $c_\rho > 0$ *such that for any* $D > 0$,

$$\mathbb{P}[\,|X_T - K| < D] \le c_\rho\, D.$$

Then for any $q>2$ and any $\delta>0$, there exists a constant c_δ which depends on X_0, a, b, f, T, q, and δ but not on h or $\mathbb{E}[\,|\widetilde{Z}-Z|^q]$ such that

$$\begin{aligned}&\mathbb{V}\left[f(\widehat{X}^{\mathrm{f}}_N)-f(\widehat{X}^{\mathrm{c}}_N)-f(\widetilde{X}^{\mathrm{f}}_N)+f(\widetilde{X}^{\mathrm{c}}_N)\right]\\ &\qquad\le c_\delta \min\left\{h\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{(1-\delta)/(q+1)},\ h^{(1-\delta)/2-1/q}\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{2/q}\right\}.\end{aligned}\tag{4.4}$$

A key point to take away from Lemma 4.11 is the transitional behavior from $O(h^{1/2})\to O(h)$ as h becomes sufficiently small, and conversely the diminished convergence rate of $O(h^{1/2})$ which arises at the coarser discretizations from the presence of points of nondifferentiability. (Note that the expected differences between the exact and approximate normal random variables are independent of h).

Proof. The proof follows an approach used previously in the analysis of MLMC variance for similar options in the context of multidimensional SDEs (Theorem 5.2 in [25]).

The proof is given for $0<\delta<1-1/q$. If the assertion is true for δ in this range, then it also holds for larger values.

If we define the events A and B as

$$A: |X_T-K|\le D,\qquad B:\max\left\{|\widehat{X}^{\mathrm{f}}_N-X_T|,|\widehat{X}^{\mathrm{c}}_N-X_T|,|\widetilde{X}^{\mathrm{f}}_N-\widehat{X}^{\mathrm{f}}_N|,|\widetilde{X}^{\mathrm{c}}_N-\widehat{X}^{\mathrm{c}}_N|,\right\}\ge D/2$$

for some choice of constant $D>0$, and define $\Delta f\equiv f(\widehat{X}^{\mathrm{f}}_N)-f(\widehat{X}^{\mathrm{c}}_N)-f(\widetilde{X}^{\mathrm{f}}_N)+f(\widetilde{X}^{\mathrm{c}}_N)$, then

$$\mathbb{V}[\Delta f]\le\mathbb{E}\left[(\Delta f)^2\mathbf{1}_{A\cup B}\right]+\mathbb{E}\left[(\Delta f)^2\mathbf{1}_{A^c\cap B^c}\right],$$

where A^c,B^c are the complements of A and B, and $\mathbf{1}_C$ is the indicator function which has value 1 if the random sample $\omega\in C$ and 0 otherwise. Note that if $\omega\in A^c\cap B^c$, then the four values $\widehat{X}^{\mathrm{f}}_N,\widehat{X}^{\mathrm{c}}_N,\widetilde{X}^{\mathrm{f}}_N,\widetilde{X}^{\mathrm{c}}_N$ are all on the same side of K, and therefore the proof in Lemma 4.10 means that

$$\mathbb{E}\left[(\Delta f)^2\mathbf{1}_{A^c\cap B^c}\right]\prec h\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{2/q}\prec h\left(\mathbb{E}[\,|\widetilde{Z}-Z|^q]\right)^{(1-\delta)/(q+1)},\tag{4.5}$$

so $\mathbb{E}\left[(\Delta f)^2\mathbf{1}_{A^c\cap B^c}\right]$ is not the dominant contributor to the bound in (4.4).

To address the other term, $\mathbb{E}\left[(\Delta f)^2\mathbf{1}_{A\cup B}\right]$, we begin by noting that the two terms in the bound on the right-hand side of (4.4) are equal when $h^{1/2}=(\mathbb{E}[\,|\widetilde{Z}-Z|^q])^{1/(q+1)}$.

Case A: $h^{1/2}\le(\mathbb{E}[\,|\widetilde{Z}-Z|^q])^{1/(q+1)}$.

In this case, we set $D=(\mathbb{E}[\,|\widetilde{Z}-Z|^q])^{(1-\delta/2)/(q+1)}$, and by Hölder's inequality we have

$$\mathbb{E}\left[(\Delta f)^2\mathbf{1}_{A\cup B}\right]\le\left(\mathbb{E}[|\Delta f|^{2/\delta'}]\right)^{\delta'}(\mathbb{E}[\mathbf{1}_{A\cup B}])^{1-\delta'}\le\left(\mathbb{E}[|\Delta f|^{2/\delta'}]\right)^{\delta'}(\mathbb{P}[A]+\mathbb{P}[B])^{1-\delta'},$$

where $\delta'=\delta/(2-\delta)$ so that $1-\delta'=(1-\delta)/(1-\delta/2)$.

Due to the assumed bounded density for X_T, we have $\mathbb{P}[A]\prec D$. Also,

$$\begin{aligned}\mathbb{P}[B]\le\ &\mathbb{P}[\,|\widehat{X}^{\mathrm{f}}_N-X_T|>D/2]+\mathbb{P}[\,|\widehat{X}^{\mathrm{c}}_N-X_T|>D/2]\\&+\mathbb{P}[\,|\widetilde{X}^{\mathrm{f}}_N-\widehat{X}^{\mathrm{f}}_N|>D/2]+\mathbb{P}[\,|\widetilde{X}^{\mathrm{c}}_N-\widehat{X}^{\mathrm{c}}_N|>D/2].\end{aligned}$$

By the Markov inequality, together with the standard strong convergence results,

$$\mathbb{P}[\,|\widehat{X}_N^{\mathrm{f}}-X_T|>D/2] \leq \frac{\mathbb{E}[\,|\widehat{X}_N^{\mathrm{f}}-X_T|^p]}{(D/2)^p} \prec \frac{h^{p/2}}{D^p} \prec (\mathbb{E}[\,|\widetilde{Z}-Z|^q])^{p\delta/(2q+2)} \prec D,$$

by choosing $p > 2/\delta - 1$. A similar bound follows for $\mathbb{P}[\,|\widehat{X}_N^{\mathrm{c}}-X_T|>D/2]$. In addition, the Markov inequality, together with Lemma 4.4, gives

$$\mathbb{P}[\,|\widetilde{X}_N^{\mathrm{f}}-\widehat{X}_N^{\mathrm{f}}|>D/2] \leq \frac{\mathbb{E}[\,|\widetilde{X}_N^{\mathrm{f}}-\widehat{X}_N^{\mathrm{f}}|^q]}{D^q} \prec \frac{\mathbb{E}[\,|\widetilde{Z}-Z|^q]}{D^q} \prec D,$$

and a similar bound holds for $\mathbb{P}[\,|\widetilde{X}_N^{\mathrm{c}}-\widehat{X}_N^{\mathrm{c}}|>D/2]$. The conclusion from this is that $\mathbb{P}[B] \prec D$.

Hence,

$$(\mathbb{P}[A]+\mathbb{P}[B])^{(1-\delta)/(1-\delta/2)} \prec D^{(1-\delta)/(1-\delta/2)} = (\mathbb{E}[\,|\widetilde{Z}-Z|^q])^{(1-\delta)/(q+1)}.$$

In addition, we have

$$|\Delta f|^{2/\delta'} \leq 2^{2/\delta'-1}\left(|f(\widehat{X}_N^{\mathrm{f}})-f(\widehat{X}_N^{\mathrm{c}})|^{2/\delta'} + |f(\widetilde{X}_N^{\mathrm{f}})-f(\widetilde{X}_N^{\mathrm{c}})|^{2/\delta'}\right),$$

and due to Hölder's inequality and the bounds in Corollary 4.5 and Lemma 4.8 we have

$$\begin{aligned}\mathbb{E}[\,|f(\widetilde{X}_N^{\mathrm{f}})-f(\widetilde{X}_N^{\mathrm{c}})|^{2/\delta'}] &\leq L_f^{2/\delta'}\left(\mathbb{E}[\,|1+c|\widetilde{X}_N^{\mathrm{f}}|^r + c|\widetilde{X}_N^{\mathrm{c}}|^r|^{4/\delta'}]\right)^{1/2}\left(\mathbb{E}[\,|\widetilde{X}_N^{\mathrm{f}}-\widetilde{X}_N^{\mathrm{c}}|^{4/\delta'}]\right)^{1/2}\\ &\prec h^{1/\delta'}.\end{aligned}$$

There is a similar bound for $\mathbb{E}[\,|f(\widehat{X}_N^{\mathrm{f}})-f(\widehat{X}_N^{\mathrm{c}})|^{2/\delta'}]$, and hence we have the result that

$$\mathbb{E}\left[(\Delta f)^2\mathbf{1}_{A\cup B}\right] \prec h\,(\mathbb{E}[\,|\widetilde{Z}-Z|^q])^{(1-\delta)/(q+1)}$$

when $h^{1/2} \leq (\mathbb{E}[\,|\widetilde{Z}-Z|^q])^{1/(q+1)}$.

Case B: $h^{1/2} \geq (\mathbb{E}[\,|\widetilde{Z}-Z|^q])^{1/(q+1)}$.

In this case we set $D = h^{(1-\delta)/2}$, and by Hölder's inequality we have

$$\begin{aligned}\mathbb{E}\left[(\Delta f)^2\mathbf{1}_{A\cup B}\right] &\leq \left(\mathbb{E}[|\Delta f|^{2/(2/q+\delta')}]\right)^{2/q+\delta'}(\mathbb{E}[\mathbf{1}_{A\cup B}])^{1-2/q-\delta'}\\ &\leq \left(\mathbb{E}[|\Delta f|^{2/(2/q+\delta')}]\right)^{2/q+\delta'}(\mathbb{P}[A]+\mathbb{P}[B])^{1-2/q-\delta'},\end{aligned}$$

where $\delta' = 2\delta/(q(1-\delta))$ so that $(1-2/q-\delta')(1-\delta)/2 = (1-\delta)/2 - 1/q$.

We again have $\mathbb{P}[A] \prec D$. By the Markov inequality, together with the standard strong convergence results,

$$\mathbb{P}[\,|\widehat{X}_N^{\mathrm{f}}-X_T|>D/2] \leq \frac{\mathbb{E}[\,|\widehat{X}_N^{\mathrm{f}}-X_T|^p]}{(D/2)^p} \prec \frac{h^{p/2}}{h^{p(1-\delta)/2}} = h^{p\delta/2} \prec D,$$

by choosing $p > 1/\delta$, and a similar bound follows for $\mathbb{P}[\,|\widehat{X}^{\rm c}_N - X_T| > D/2]$. In addition, the Markov inequality, together with Lemma 4.4, gives

$$\mathbb{P}[\,|\widetilde{X}^{\rm f}_N - \widehat{X}^{\rm f}_N| > D/2] \leq \frac{\mathbb{E}[\,|\widetilde{X}^{\rm f}_N - \widehat{X}^{\rm f}_N|^q]}{D^q} \prec \frac{\mathbb{E}[\,|\widetilde{Z} - Z|^q]}{D^q} \prec \frac{h^{(q+1)/2}}{h^{q(1-\delta)/2}} = h^{1/2+q\delta/2} \prec D,$$

and a similar bound holds for $\mathbb{P}[\,|\widetilde{X}^{\rm c}_N - \widehat{X}^{\rm c}_N| > D/2]$. The conclusion from this is that $\mathbb{P}[B] \prec D$, as before, and so

$$(\mathbb{P}[A] + \mathbb{P}[B])^{1-2/q-\delta'} \prec D^{1-2/q-\delta'} = h^{(1-2/q-\delta')(1-\delta)/2} = h^{(1-\delta)/2-1/q}.$$

In addition, defining $\delta'' = \delta'/(2/q+\delta')$ so that $2/(2/q+\delta') = q(1-\delta'')$, we have

$$|\Delta f|^{q(1-\delta'')} \leq 2^{q(1-\delta'')-1}\left(|f(\widetilde{X}^{\rm f}_N) - f(\widehat{X}^{\rm f}_N)|^{q(1-\delta'')} + |f(\widetilde{X}^{\rm c}_N) - f(\widehat{X}^{\rm c}_N)|^{q(1-\delta'')}\right),$$

and due to Hölder's inequality and the bounds in Corollary 4.5 and Lemma 4.8 we have

$$\begin{aligned}
&\mathbb{E}[\,|f(\widetilde{X}^{\rm f}_N) - f(\widehat{X}^{\rm f}_N)|^{q(1-\delta'')}] \\
&\qquad \leq L_f^{q(1-\delta'')}\left(\mathbb{E}[\,|1 + c|\widetilde{X}^{\rm f}_N|^r + c|\widehat{X}^{\rm f}_N|^r|^{q(1-\delta'')/\delta''}]\right)^{\delta''}\left(\mathbb{E}[\,|\widetilde{X}^{\rm f}_N - \widehat{X}^{\rm f}_N|^q]\right)^{1-\delta''} \\
&\qquad \prec (\mathbb{E}[\,|\widetilde{Z} - Z|^q])^{1-\delta''} \\
\Longrightarrow \quad &\left(\mathbb{E}[\,|f(\widetilde{X}^{\rm f}_N) - f(\widehat{X}^{\rm f}_N)|^{2/(2/q+\delta')}]\right)^{2/q+\delta'} \prec (\mathbb{E}[\,|\widetilde{Z} - Z|^q])^{2/q}.
\end{aligned}$$

There is a similar bound for $\mathbb{E}[\,|f(\widetilde{X}^{\rm c}_N) - f(\widehat{X}^{\rm c}_N)|^{q(1-\delta'')}]$, and hence we have the result that $\mathbb{E}\left[(\Delta f)^2 \mathbf{1}_{A\cup B}\right] \prec h^{(1-\delta)/2-1/q}(\mathbb{E}[\,|\widetilde{Z} - Z|^q])^{2/q}$ when $h^{1/2} \geq (\mathbb{E}[\,|\widetilde{Z} - Z|^q])^{1/(q+1)}$.

When we combine the bounds from cases A and B with (4.5), we obtain the desired final result. ∎

5. Numerical results.

Our numerical tests are for the simplest possible example of geometric Brownian motion,

$$\mathrm{d}X_t = \mu\, X_t\, \mathrm{d}t + \sigma\, X_t\, \mathrm{d}W_t.$$

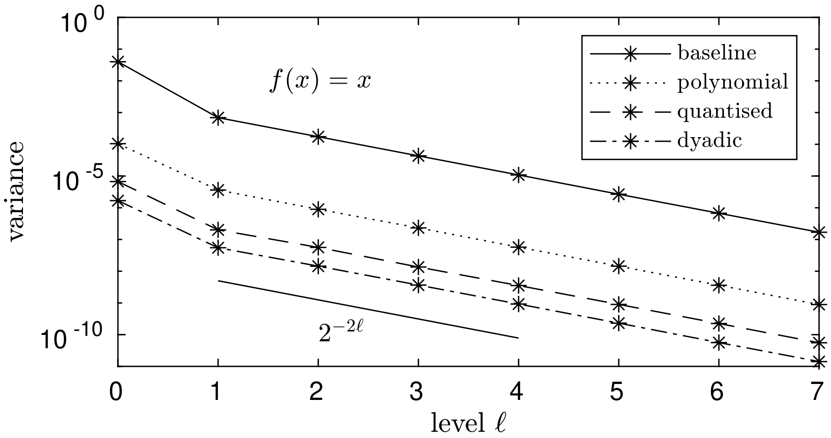
In our simulations we take $\mu = 0.05$, $\sigma = 0.2$, $T = 1$, and $X_0 = 1$. The coarsest level $\ell = 0$ uses a single timestep, and higher levels use 4^ℓ timesteps on level ℓ so that $h_\ell = 2^{-2l}$.

For the normal random variables we use the approximations discussed in section 2:

1. the quantized piecewise constant approximation using 1024 intervals;
2. the piecewise linear approximation on 16 dyadic intervals on $(0, 1/2)$;
3. a degree 7 polynomial approximation.

Note that the values of $\mathbb{E}[\,|\widetilde{Z} - Z|^2]$ for these are 1.5×10^{-4}, 4×10^{-5}, and 2.6×10^{-3}, respectively.

Figure 5.1 presents results for all three approximations for two different output functions, $f(x) \equiv x$ and $f(x) \equiv \max(x-1, 0)$. In all cases the variances for $\mathbb{V}[\widetilde{P}_\ell - \widetilde{P}_{\ell-1}]$ using the approximate normals are visually indistinguishable from $\mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ which is the line labeled as "baseline"; the other three lines are the variances $\mathbb{V}[(\widehat{P}_\ell - \widehat{P}_{\ell-1}) - (\widetilde{P}_\ell - \widetilde{P}_{\ell-1})]$ for the three approximations.



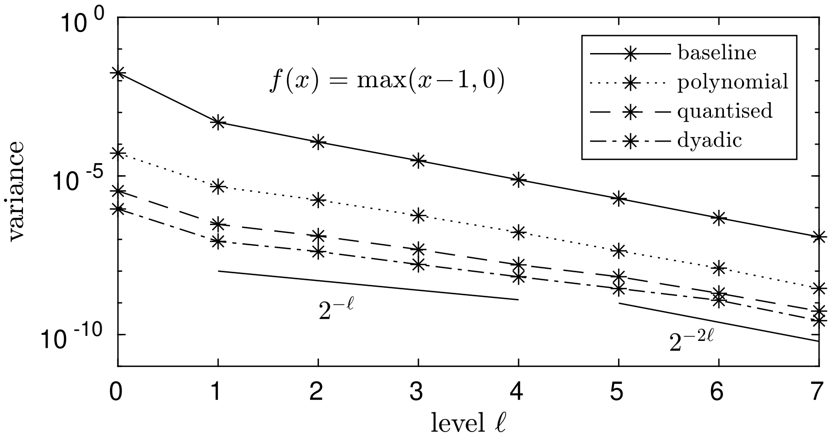


Figure 5.1. *MLMC variances for two different output functions, with reference lines proportional to* $2^{-2\ell}$ *and* $2^{-\ell}$.

For the first case, $f(x)\equiv x$, by choosing q close to 2, Lemma 4.10 gives

$$\widetilde{V}_\ell \equiv \mathbb{V}\left[(\widehat{P}_\ell - \widehat{P}_{\ell-1}) - (\widetilde{P}_\ell - \widetilde{P}_{\ell-1})\right] \approx O\left(2^{-2\ell}\, \mathbb{E}[\,|\widetilde{Z} - Z|^2]\right).$$

The numerical results appear to be consistent with this, with $\widetilde{V}_\ell$ decreasing with level approximately proportional to $2^{-2\ell}$, as indicated by the reference line which is proportional to $2^{-2\ell}$. For a fixed level ℓ, the variation in $\widetilde{V}_\ell$ between the three different approximations is roughly proportional to $\mathbb{E}[\,|\widetilde{Z} - Z|^2]$, with the piecewise linear approximation on dyadic intervals being the most accurate and hence giving the smallest values for $\widetilde{V}_\ell$, and the polynomial approximation being much less accurate leading to larger values for $\widetilde{V}_\ell$.

For the second case, $f(x)\equiv \max(x-1, 0)$, choosing δ close to zero, Lemma 4.11 gives

$$\widetilde{V}_\ell \approx O\left(\min\left\{2^{-2\ell}\, \mathbb{E}[\,|\widetilde{Z} - Z|^q]^{1/(q+1)},\, 2^{-(1-2/q)\ell}\, \mathbb{E}[\,|\widetilde{Z} - Z|^q]^{2/q}\right\}\right)$$

for any $q>2$, whereas the earlier heuristic analysis suggested

$$\widetilde{V}_\ell \approx O\left(\min\left\{2^{-2\ell}\,\mathbb{E}[\,|\widetilde{Z}-Z|^2]^{1/2}, 2^{-\ell}\,\mathbb{E}[\,|\widetilde{Z}-Z|^2]\right\}\right).$$

The numerical results are plotted with reference lines proportional to $2^{-\ell}$ and $2^{-2\ell}$. The results do show a slight change in the slope reflecting the switch from $O(2^{-\ell})$ to $O(2^{-2\ell})$ in the analysis.

Regarding the overall computational efficiency, as discussed in sections 3.2 and 2.4, the CPU implementations using the quantized and dyadic approximations are approximately 7 times more efficient, so $\widetilde{C}_\ell/C_\ell \approx 1/7$. The quantity $\sqrt{(C_\ell/\widetilde{C}_\ell+1)\,\widetilde{V}_\ell/V_\ell}$ is approximately 0.026 and 0.052 for the output function $f(x)=x$, using the dyadic and quantized approximations, respectively, and 0.14 and 0.19 for the output function $f(x)=\max(x-1,0)$, using the dyadic and quantized approximations. Therefore, in all four cases the total cost is reduced by a factor which is close to $\widetilde{C}_\ell/C_\ell$.

For more details regarding the computational aspects and implementation details we again refer the reader to the companion paper [24]. This also analyzes the convergence rates for the two piecewise approximations, although a suitable analytic bound for the single polynomial over the entire interval could not be found, and thus an optimized high performance implementation was not considered. Furthermore, it presents the exact speed ups that can be expected for a range of approximation fidelities (piecewise constant, linear, and cubic approximations), where there is a trade off between the speed of the approximation and the number of correction terms required, which is determined by the approximation fidelity. Table 3 in [24] enumerates these possibilities, finding the piecewise linear approximation using dyadic intervals to the optimal choice, achieving a computation speed up of 6.70 in a nested MLMC setting, and achieving a computational efficiency 95.7% of the speed improvements offered by the approximation.

5.1. Simulating the Cox–Ingersoll–Ross process. To briefly showcase the benefits from naively employing our approximate Gaussian random variables to more complicated SDEs (outside the strict remit of our analysis), we consider simulating from the Cox–Ingersoll–Ross (CIR) process [12] $\mathrm{d}X_t = \kappa(\theta - X_t)\mathrm{d}t + \sigma\sqrt{X_t}\mathrm{d}W_t$. This represents a significantly more difficult stochastic process than the geometric Brownian motion, and one which does not have a closed form solution in terms of the underlying Brownian motion (although the distribution is known to be a noncentral inverse χ^2). Choosing $\kappa = 0.5$ and $\theta = X_0 = \sigma = 1$ we can perform the same MLMC simulation using the Euler–Maruyama scheme. Noting that technically the Euler–Maruyama scheme is ill posed because of the $\sqrt{X_t}$ term, we consequently use the truncated Euler–Maruyama scheme from Higham, Mao, and Stuart [29], remarking that the chosen coefficients satisfy the Feller condition [14].

The variance of the MLMC terms are shown in Figure 5.2, where similar to the geometric Brownian motion results, we can observe both a substantial reduction in the variance between the two and four way differences and a continued reduction in variance as the levels increase.

However, we do not discuss this example any further, as it is primarily illustrative of the applicability of these ideas to more complicated stochastic processes which currently fall considerably outside of the scope of our analysis. In our companion piece [24] we discuss the CIR

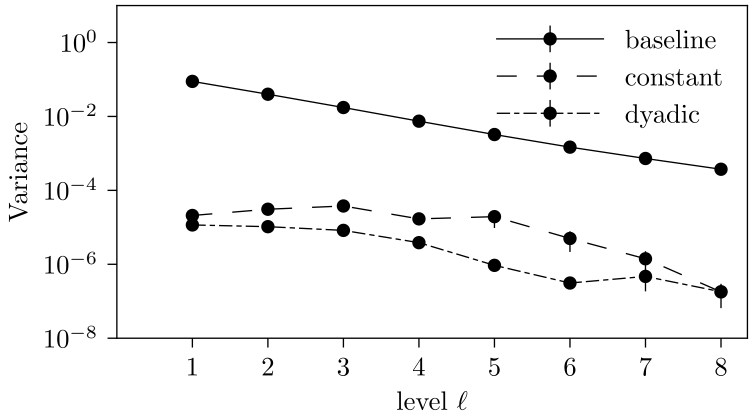


Figure 5.2. *The nested MLMC framework applied to the CIR process.*

process in much greater depth, showing how the approximation ideas developed here can be transferred to the underlying noncentral χ^2 distribution and that this gives superior convergence results in this setting, and we detail how speed ups by a factor of 300 can be achieved. Furthermore, we also explore utilizing the same techniques with the more complicated Milstein method.

6. Conclusions and future work.

In this paper we have presented a general nested MLMC framework which employs approximate random variables which can be sampled much more efficiently than the true distribution. As a specific example, we investigated the use of approximate normal random variables for an Euler–Maruyama discretization of a scalar SDE. A detailed error analysis bounds the variance of the differences in the SDE path approximations as a function of the error in the approximate inverse normal distribution. This analysis is supported by numerical results for the simplest possible case of geometric Brownian motion.

There are two directions in which we plan to extend this research. This first is to investigate approximations of other distributions. Two are of particular interest; one is the Poisson distribution, which is important for continuous-time Markov processes [1, 2] and is simulated using the inverse normal CDF [17], and the other is the noncentral χ^2-distribution which is important for simulating the CIR process which is used extensively in computational finance. In both cases, the computational savings may be greater, but it may prove to be very difficult to carry out a detailed numerical analysis of the resulting MLMC variances.

The second direction is to use reduced precision computer arithmetic in performing the calculations of $\widetilde{X}_n$, further reducing the cost of the approximate calculations. This builds on prior research by others, implementing MLMC methods on field-programmable gate arrays [8, 37]. The rounding error effect of finite precision arithmetic can be modeled as an additional random error at each timestep; it is expected that on the coarsest levels with few timesteps this additional error will be small, but on the finest levels it may become significant and so perhaps such levels should be computed using single precision.

Appendix A. Mean value theorem and a generalization.

Lemma A.1 (mean value theorem). *If $f : \mathbb{R} \to \mathbb{R}$ is $C^1(\mathbb{R})$, then there exists ξ which is a positively weighted average of x_1, x_2 (i.e., $\xi = s\, x_1 + (1-s)x_2$ for some $0<s<1$) such that*

$$f(x_1) - f(x_2) = (x_1 - x_2)\, f'(\xi).$$

Lemma A.2. *If $f : \mathbb{R} \to \mathbb{R}$ is $C^1(\mathbb{R})$ and f' is Lipschitz continuous with Lipschitz constant L'_f, then there exists ξ which is a positively weighted average of x_1, x_2, x_3, x_4 such that*

$$f(x_1) - f(x_2) - f(x_3) + f(x_4) = (x_1 - x_2 - x_3 + x_4) f'(\xi) + R,$$

where

$$|R| \le \tfrac{1}{2} L'_f (|x_1 - x_2| + |x_3 - x_4|)\; (|x_1 - x_3| + |x_2 - x_4|).$$

Proof. Without loss of generality, we can assume

$$|x_1 - x_2| + |x_3 - x_4| \;\le\; |x_1 - x_3| + |x_2 - x_4|, \tag{A.1}$$

since otherwise we can just swap x_2 and x_3.

Now, using Lemma A.1 we get

$$\begin{aligned} f(x_1) - f(x_2) &= (x_1 - x_2)\, f'(\xi_1), \\ f(x_3) - f(x_4) &= (x_3 - x_4)\, f'(\xi_2), \end{aligned}$$

where ξ_1 and ξ_2 are positively weighted averages of x_1, x_2 and x_3, x_4, respectively. Taking the difference gives

$$f(x_1) - f(x_2) - f(x_3) + f(x_4) = \tfrac{1}{2}(x_1 - x_2 - x_3 + x_4)(f'(\xi_1) + f'(\xi_2)) + R,$$

where

$$R = \tfrac{1}{2}(x_1 - x_2 + x_3 - x_4)(f'(\xi_1) - f'(\xi_2)).$$

Since f' is continuous, there exists an ξ which is a positively weighted of ξ_1 and ξ_2 and hence of x_1, x_2, x_3, x_4 such that

$$\tfrac{1}{2}(f'(\xi_1) + f'(\xi_2)) = f'(\xi).$$

Note that

$$\xi_1 - \xi_2 = (\xi_1 - \tfrac{1}{2}(x_1 + x_2)) + (\xi_2 - \tfrac{1}{2}(x_3 + x_4)) + (\tfrac{1}{2}(x_1 + x_2) - \tfrac{1}{2}(x_3 + x_4)),$$

and therefore, due to (A.1),

$$\begin{aligned} |\xi_1 - \xi_2| &\le \tfrac{1}{2}|x_1 - x_2| + \tfrac{1}{2}|x_3 - x_4| + \tfrac{1}{2}|x_1 - x_3| + \tfrac{1}{2}|x_2 - x_4| \\ &\le |x_1 - x_3| + |x_2 - x_4|. \end{aligned}$$

Hence, due to the Lipschitz property of f',

$$|R| \le \tfrac{1}{2} L'_f (|x_1 - x_2| + |x_3 - x_4|)\; (|x_1 - x_3| + |x_2 - x_4|). \qquad \blacksquare$$

REFERENCES

[1] D. Anderson and D. Higham, *Multi-level Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics*, SIAM Multiscale Model. Simul., 10 (2012), pp. 146–179.

[2] D. Anderson, D. Higham, and Y. Sun, *Complexity of multilevel Monte Carlo tau-leaping*, SIAM J. Numer. Anal., 52 (2014), pp. 3106–3127.

[3] N. Bakhvalov, *Optimal convergence bounds for quadrature processes and integration methods of Monte Carlo type for classes of functions*, Zh. Vychisl. Mat. Mat. Fiz., 4 (1964), pp. 5–63.

[4] A. Barth, C. Schwab, and N. Zollinger, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161.

[5] D. Belomestny and T. Nagapetyan, *Multilevel path simulation for weak approximation schemes with application to Lévy-driven SDEs*, Bernoulli, 23 (2017), pp. 927–950.

[6] A. Beskos, A. Jasra, K. Law, R. Tempone, and Y. Zhou, *Multilevel sequential Monte Carlo samplers*, Stochastic Process. Appl., 127 (2017), pp. 1417–1440.

[7] C. Bierig and A. Chernov, *Approximation of probability density functions by the multilevel Monte Carlo maximum entropy method*, J. Comput. Phys., 314 (2016), pp. 661–681.

[8] C. Brugger, C. de Schryver, N. Wehn, S. Omland, M. Hefter, K. Ritter, A. Kostiuk, and R. Korn, *Mixed precision multilevel Monte Carlo on hybrid computing systems*, in Proceedings of the Conference on Computational Intelligence for Financial Engineering and Economics, IEEE, 2014.

[9] K. Bujok, B. Hambly, and C. Reisinger, *Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives*, Methodol. Comput. Appl. Probab., 17 (2015), pp. 579–604.

[10] D. Burkholder, B. Davis, and R. Gundy, *Integral inequalities for convex functions of operators on martingales*, in Sixth Berkeley Symposium Math. Statist. Prob., Vol II., University of California Press, Berkeley, 1972, pp. 223–240.

[11] K. Cliffe, M. Giles, R. Scheichl, and A. Teckentrup, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15.

[12] J. Cox, J. Ingersoll, and S. Ross, *A theory of the term structure of interest rates*, Econometrica, 53 (1985), pp. 385–407.

[13] D. Elfverson, F. Hellman, and A. Målqvist, *A multilevel Monte Carlo method for computing failure probabilities*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 312–330.

[14] W. Feller, *Two singular diffusion problems*, Ann. of Math., (1951), pp. 173–182.

[15] M. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[16] M. Giles, *Multilevel Monte Carlo methods*, Acta Numer., 24 (2015), pp. 259–328.

[17] M. Giles, *Approximation of the inverse Poisson cumulative distribution function*, ACM Trans. Math. Software, 42 (2016), 7.

[18] M. Giles and T. Goda, *Decision-making under uncertainty: Using MLMC for efficient estimation of EVPPI*, Stat. Comput., 29 (2019), pp. 739–751.

[19] M. Giles, M. Hefter, L. Mayer, and K. Ritter, *Random bit multilevel algorithms for stochastic differential equations*, J. Complexity, 54 (2019), 101395.

[20] M. Giles, M. Hefter, L. Mayer, and K. Ritter, *Random bit quadrature and approximation of distributions on Hilbert spaces*, Found. Comput. Math., 19 (2019), pp. 205–238.

[21] M. Giles, F. Kuo, and I. Sloan, *Combining sparse grids, multilevel MC and QMC for elliptic PDEs with random coefficients*, in Monte Carlo and Quasi-Monte Carlo Methods 2016, P. Glynn and A. Owen, eds., Springer, Cham, 2018.

[22] M. Giles, T. Nagapetyan, and K. Ritter, *Multilevel Monte Carlo approximation of distribution functions and densities*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 267–295.

[23] M. Giles and C. Reisinger, *Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance*, SIAM J. Financial Math., 3 (2012), pp. 572–592.

[24] M. Giles and O. Sheridan-Methven, *Approximating Inverse Cumulative Distribution Functions to Produce Approximate Random Variables*, preprint, arXiv:2012.09715 [math.NA], 2020, https://arxiv.org/abs/2012.09715.

[25] M. Giles and L. Szpruch, *Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation*, Ann. Appl. Probab., 24 (2014), pp. 1585–1620.

[26] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004.
[27] A.-L. Haji-Ali, F. Nobile, and R. Tempone, *Multi index Monte Carlo: When sparsity meets sampling*, Numer. Math., 132 (2016), pp. 767–806.
[28] S. Heinrich, *Monte Carlo complexity of global solution of integral equations*, J. Complexity, 14 (1998), pp. 151–175.
[29] D. J. Higham, X. Mao, and A. M. Stuart, *Strong convergence of Euler-type methods for nonlinear stochastic differential equations*, SIAM J. Numer. Anal., 40 (2002), pp. 1041–1063.
[30] V. Hoang, C. Schwab, and A. Stuart, *Complexity analysis of accelerated MCMC methods for Bayesian inversion*, Inverse Problems, 29 (2013), 085010.
[31] H. Hoel, K. J. Law, and R. Tempone, *Multilevel ensemble Kalman filtering*, SIAM J. Numer. Anal., 54 (2016), pp. 1813–1839.
[32] A. Jasra, K. Kamatani, K. J. Law, and Y. Zhou, *Multilevel particle filters*, SIAM J. Numer. Anal., 55 (2017), pp. 3068–3096.
[33] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.
[34] P. l'Ecuyer, *Randomized quasi-Monte Carlo: An introduction for practitioners*, in International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Springer, Cham, 2016, pp. 29–52.
[35] G. Marsaglia and W. W. Tsang, *The Ziggurat method for generating random variables*, J. Statist. Softw., 5 (2000), pp. 1–7.
[36] E. Müller, R. Scheichl, and T. Shardlow, *Improving multilevel Monte Carlo for stochastic differential equations with application to the Langevin equation*, Proc. A, 471 (2015), 20140679.
[37] S. Omland, M. Hefter, K. Ritter, C. Brugger, C. de Schryver, N. Wehn, and A. Kostiuk, *Exploiting mixed-precision arithmetic in a multilevel Monte Carlo approach on FPGAs*, in FPGA Based Accelerators for Financial Applications, Springer, Cham, 2015.
[38] R. Scheichl, A. Stuart, and A. Teckentrup, *Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 466–492.
[39] R. Seydel, *Tools for Computational Finance*, 5th ed., Springer, Cham, 2012.
[40] O. Sheridan-Methven, *Nested Multilevel Monte Carlo Methods and a Modified Euler-Maruyama Scheme Utilising Approximate Gaussian Random Variables Suitable for Vectorised Hardware and Low-Precisions*, PhD thesis, University of Oxford, 2020.
[41] E. Ullmann and I. Papaioannou, *Multilevel estimation of rare events*, SIAM J. Uncertain. Quantif., 3 (2015), pp. 922–953.