

## Multilevel Estimation of Expected Exit Times and Other Functionals of Stopped Diffusions\*

Michael B. Giles<sup>†</sup> and Francisco Bernal<sup>‡</sup>

**Abstract.** This paper proposes and analyzes a new multilevel Monte Carlo method for the estimation of mean exit times for multidimensional Brownian diffusions and associated functionals which correspond to solutions to high-dimensional parabolic PDEs through the Feynman–Kac formula. In particular, it is proved that the complexity to achieve an  $\varepsilon$  root-mean-square error is  $O(\varepsilon^{-2} |\log \varepsilon|^3)$ .

**Key words.** multilevel, Monte Carlo, stochastic, numerical analysis, stopped diffusions

**AMS subject classifications.** 60H10, 60H35, 65C05, 65C20, 65C30

**DOI.** 10.1137/17M1116660

**1. Introduction.** In this paper we are concerned with solutions of a stochastic differential equation (SDE)

$$(1) \quad dX_t = a(X_t, t) dt + b(X_t, t) dW_t, \quad 0 < t \leq T,$$

with deterministic initial condition  $X_0 = x_0$ . We assume that  $X_t \in \mathbb{R}^d$  and  $W = \{W_t : t \geq 0\}$  is a standard  $d'$ -dimensional Brownian motion. Furthermore, we assume that  $a : \mathbb{R}^d \times [0, T] \rightarrow \mathbb{R}^d$  and  $b : \mathbb{R}^d \times [0, T] \rightarrow \mathbb{R}^d \times \mathbb{R}^{d'}$  are Lipschitz continuous with Lipschitz constants  $L_a, L_b$  such that

$$\begin{aligned} \|a(x, t) - a(y, s)\|_2 &\leq L_a (\|x - y\|_2 + |t - s|), \\ \|b(x, t) - b(y, s)\|_2 &\leq L_b (\|x - y\|_2 + |t - s|) \quad \forall (x, t), (y, s) \in \mathbb{R}^d \times [0, T], \end{aligned}$$

and also that  $bb^T$  is everywhere nonsingular.

Given a simply connected bounded open domain  $D \subset \mathbb{R}^d$ , and a fixed point  $x_0 \in D$ , we are interested in estimating the expected exit time  $\mathbb{E}^{x_0, 0}[\tau]$ , where the suffix on  $\mathbb{E}^{x, t}$  for time  $t : 0 \leq t \leq T$  indicates it is an expectation conditional on the path  $X_s, s > t$ , starting at  $X_t = x \in D$ , and the corresponding exit time  $\tau$  is defined as  $\tau = \min(T, \inf\{s > t : X_s \notin D\})$ .

In addition, we are interested in estimating  $u(x_0, 0)$  which is defined by

$$(2) \quad u(x, t) = \mathbb{E}^{x, t} \left[ \int_t^\tau E(t, s) f(X_s, s) ds + E(t, \tau) g(X_\tau, \tau) \right],$$

\*Received by the editors February 15, 2017; accepted for publication (in revised form) August 13, 2018; published electronically October 18, 2018.

<http://www.siam.org/journals/juq/6-4/M111666.html>

**Funding:** The work of the first author was partially supported by EPSRC grant EP/H05183X/1. The work of the second author was supported by Portuguese FCT funding under grant SFRH/BPD/79986/2011.

<sup>†</sup>Mathematical Institute, Oxford University, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK ([mike.giles@maths.ox.ac.uk](mailto:mike.giles@maths.ox.ac.uk)).

<sup>‡</sup>Ecole Polytechnique, CMAP, Route de Saclay, Palaiseau Cedex, 91128, France ([francisco.bernal@polytechnique.edu](mailto:francisco.bernal@polytechnique.edu)).

where

$$E(t_0, t_1) = \exp\left(-\int_{t_0}^{t_1} V(X_t, t) dt\right).$$

We assume that the functions  $f : \mathbb{R}^d \times [0, T] \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^d \times [0, T] \rightarrow \mathbb{R}$ ,  $V : \mathbb{R}^d \times [0, T] \rightarrow \mathbb{R}$  are all Lipschitz continuous with Lipschitz constants  $L_f, L_g, L_V$ , respectively, and also that  $g \in C^{2,1}(D \times [0, T])$ , with its spatial Hessian  $H_g$  and time derivative  $\dot{g} \equiv \partial g / \partial t$  both bounded.

The Feynman–Kac formula [15, 16] establishes that if the boundary  $\partial D$  is sufficiently smooth,  $u(x, t)$  satisfies the PDE

$$(3) \quad \frac{\partial u}{\partial t} + \sum_j a_j(x, t) \frac{\partial u}{\partial x_j} + \frac{1}{2} \sum_{j,k,l} b_{j,k}(x, t) b_{l,k}(x, t) \frac{\partial^2 u}{\partial x_j \partial x_l} - V(x, t) u(x, t) + f(x, t) = 0,$$

inside the domain  $D \times [0, T)$ , subject to Dirichlet boundary conditions  $u(x, t) = g(x, t)$  when either  $x \in D, t = T$  or  $x \in \partial D, 0 < t < T$ .

When using an Euler–Maruyama discretization for the SDE, with a uniform timestep of size  $h$ , standard analysis gives an  $O(h^{1/2})$  strong error with regards to the path approximation, and Gobet and Menozzi [11] have proved that the weak error for the expected stopping time is also  $O(h^{1/2})$ . To achieve a root-mean-square error of  $\varepsilon$  requires  $O(\varepsilon^{-2})$  path samples, with  $h = O(\varepsilon^2)$  so that the average cost of each path sample is  $O(\varepsilon^{-2})$ . Hence the total computational cost is  $O(\varepsilon^{-4})$ .

The  $O(h^{1/2})$  weak error is due to the  $O(h^{1/2})$  movement in the SDE solution within each timestep; this is not accounted for by the standard piecewise constant interpolation of the Euler–Maruyama discretization. One possible improvement would be the use of a brownian bridge interpolant [9] which would allow one to sample the minimum distance to smooth boundaries within each timestep, improving the weak order to  $O(h)$ . Another approach, developed by Gobet and Menozzi [11], based on an original idea by Broadie, Glasserman, and Kou [3], introduces a boundary correction, offsetting the boundary in the normal direction by a distance which is  $O(h^{1/2})$ . This too improves the weak error to  $O(h)$ , and so the overall computational cost is reduced to  $O(\varepsilon^{-3})$ .

Building on Giles’s multilevel Monte Carlo (MLMC) method for barrier options [6], Primozic [18] developed an MLMC algorithm for estimating the mean exit time for one-dimensional (1D) diffusions with planar boundaries. This used the Milstein discretization which gives an  $O(h)$  strong error, combined with a Brownian bridge estimate of the probability of crossing the boundary within each timestep. This gives an  $O(\varepsilon^{-2})$  computational complexity, but its generalization to multidimensional applications is limited to those cases in which the underlying SDE satisfies a commutativity condition, which means that the Lévy areas for each timestep are not required in the implementation of the Milstein discretization [9].

Higham et al. [13] instead proposed an MLMC method based on the Euler–Maruyama discretization and proved that its complexity is  $O(\varepsilon^{-3} |\log \varepsilon|^3)$ . If they had used either of the two methods mentioned previously to make the weak error first order, this would improve to  $O(\varepsilon^{-2.5} |\log \varepsilon|^3)$ .

The objective in this paper is to improve the numerical algorithm analyzed by Higham et al. to reduce the computational complexity to  $O(\varepsilon^{-2} |\log \varepsilon|^3)$ . This is achieved by addressing the problem identified in the analysis in [13], which is the poor decay in the variance of the

multilevel correction as  $h \rightarrow 0$ . This is approximately  $O(h^{1/2})$  because there is an  $O(h^{1/2})$  probability that a path which starts within  $O(h^{1/2})$  of the boundary will continue for a time which is  $O(1)$  before exiting the domain.

We fix this problem by splitting the fine or coarse path simulation into multiple independent copies once the other one has exited the domain. Averaging over these subpaths approximates the conditional expectation and reduces the multilevel variance to approximately  $O(h)$ . However, the expected value on each level is not changed, and therefore the telescoping summation which lies at the heart of MLMC remains valid.

The paper begins with a lemma bounding the variance of the Feynman–Kac functional, followed by a review of the Euler–Maruyama approximation of the SDE path and the path functional, the Gobet and Menozzi technique for improving the weak convergence, and the MLMC algorithm used by Higham et al.

The new algorithm is then presented and analyzed. The numerical analysis relies heavily on the theoretical results of Gobet and Menozzi [11] and is similar in structure to the analysis of Higham et al. [13]. The effectiveness of the new algorithm is demonstrated through numerical experiments, and directions for future research are discussed in the final conclusions.

**2. Exit times and Feynman–Kac functionals.** If we define  $u_{\text{exit}}(x, t)$  to correspond to the particular Feynman–Kac functional in which  $f_{\text{exit}}(x, t) \equiv 0$ ,  $g_{\text{exit}}(x, t) \equiv t$ ,  $V_{\text{exit}}(x, t) \equiv 0$ , then  $u_{\text{exit}}(x_0, 0)$  is equal to the expected exit time  $\mathbb{E}^{x_0, 0}[\tau]$  defined previously.

We now make a key assumption.

**Assumption 1.** *There are Lipschitz constants  $L_u$ ,  $L_{\text{exit}}$  such that*

$$\begin{aligned} |u(x, t) - u(y, s)| &\leq L_u (\|x - y\|_2 + |t - s|), \\ |u_{\text{exit}}(x, t) - u_{\text{exit}}(y, s)| &\leq L_{\text{exit}} (\|x - y\|_2 + |t - s|) \quad \forall (x, t), (y, s) \in D \times [0, T]. \end{aligned}$$

**Comment.** If the boundary  $\partial D$  is sufficiently smooth, the Lipschitz conditions for  $u(x, t)$  and  $u_{\text{exit}}(x, t)$  in Assumption 1 follow from the assumed Lipschitz properties for  $a, b, f, g, V$ . However, this assumption might not be satisfied if there is a reentrant corner within the domain.

The following lemma bounds the variance of the SDE functional:

$$P_{t_0} = \int_{t_0}^{\tau} E(t_0, s) f(X_s, s) \, ds + E(t_0, \tau) g(X_\tau, \tau).$$

The equivalent theorem for the Euler–Maruyama discretization will play a critical role in the later numerical analysis.

**Lemma 1.** *There exists a constant  $c$  such that for any  $x_0 \in D, 0 \leq t_0 < T$*

$$\mathbb{V}^{x_0, t_0}[P_{t_0}] \leq c \mathbb{E}^{x_0, t_0}[\tau - t_0].$$

Note that  $\mathbb{V}^{x_0, t_0}$  represents the variance conditional on  $X_{t_0} = x_0$ .

**Proof.** By the Itô calculus,

$$d(E(t_0, s) g(X_s, s)) = E(t_0, s) \left( (-V g + \dot{g} + (\nabla g)^T a + \frac{1}{2} \text{trace}(b^T H_g b)) \, ds + (\nabla g)^T b \, dW_s \right)$$

with  $a, b, g, \dot{g} \equiv \partial g / \partial t, \nabla g$ , the spatial gradient of  $g$ , and  $H_g$ , the Hessian of  $g$ , all evaluated at  $(X_s, s)$ . Hence,

$$P_{t_0} - g(x_0, t_0) = p^{(1)} + p^{(2)},$$

where

$$p^{(1)} = \int_{t_0}^{\tau} E(t_0, s) (f - Vg + \dot{g} + (\nabla g)^T a + \frac{1}{2} \text{trace}(b^T H_g b)) ds,$$

$$p^{(2)} = \int_{t_0}^{\tau} E(t_0, s) (\nabla g)^T b dW_s.$$

We then have

$$\mathbb{V}^{x_0, t_0}[P_{t_0}] \leq \mathbb{E}^{x_0, t_0}[(P_{t_0} - g(x_0, t_0))^2] \leq 2\mathbb{E}^{x_0, t_0}[(p^{(1)})^2 + (p^{(2)})^2].$$

Since  $E(t_0, s) \leq \exp(T\|V\|_\infty)$ , and every other term in the integrand for  $p^{(1)}$  is similarly bounded, there exists a constant  $c$ , independent of  $W, x_0, t_0$ , such that  $|p^{(1)}| \leq c(\tau - t_0) \leq cT$  and hence

$$\mathbb{E}^{x_0, t_0}[(p^{(1)})^2] \leq c^2 T \mathbb{E}^{x_0, t_0}[\tau - t_0].$$

In addition,

$$\begin{aligned} \mathbb{E}^{x_0, t_0}[(p^{(2)})^2] &= \mathbb{E}^{x_0, t_0} \left[ \int_{t_0}^{\tau} (E(t_0, s))^2 \|(\nabla g)^T b\|_2^2 ds \right] \\ &\leq \exp(2T\|V\|_\infty) \|\nabla g\|_{2, \infty}^2 \|b\|_{2, \infty}^2 \mathbb{E}^{x_0, t_0}[\tau - t_0], \end{aligned}$$

where  $\|b\|_{2, \infty}, \|\nabla g\|_{2, \infty}$  are the maximum values of  $\|b\|_2, \|\nabla g\|_2$  over  $D \times [0, T]$ .

This completes the proof. ■

**3. Euler–Maruyama approximation.** Using a uniform timestep of size  $h$ , the standard Euler–Maruyama approximation of the SDE is

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

where  $\Delta W_n \equiv W_{t_{n+1}} - W_{t_n}$  and each component of  $\Delta W_n$  is an  $N(0, h)$  independent and identically distributed (i.i.d.) random variable. Let  $\widehat{X}_t$  be the piecewise-constant interpolation in which  $\widehat{X}_t$  is taken to be constant on the open-ended time interval  $[t_n, t_{n+1})$ . Correspondingly, for any time interval  $[t_0, t]$  with  $t_0 = nh$  for some integer  $n$ , we can define

$$\widehat{E}(t_0, t) = \exp\left(-\int_{t_0}^t V(\widehat{X}_s, s) ds\right)$$

and let  $\widehat{u}(x_0, t_0) = \mathbb{E}^{x_0, t_0}[\widehat{P}_{t_0}]$ , where now the suffix on  $\mathbb{E}^{x_0, t_0}$  indicates that the expectation is conditional on  $\widehat{X}_{t_0} = x_0$ , and  $\widehat{P}_{t_0}$  is defined by

$$\widehat{P}_{t_0} = \int_{t_0}^{\widehat{\tau}} \widehat{E}(t_0, s) f(\widehat{X}_s, s) ds + \widehat{E}(t_0, \widehat{\tau}) g(\widehat{X}_{\widehat{\tau}}, \widehat{\tau})$$

with  $\widehat{\tau} = \min(T, \min\{t_n : \widehat{X}_{t_n} \notin D\})$  being the exit time of the numerical approximation.

$\widehat{u}_{\text{exit}}(x_0, t_0)$  can be defined analogously as the expected exit time for the numerical path approximation starting from  $\widehat{X}_{t_0} = x_0$ .

This Euler–Maruyama discretization is the basis for the MLMC method proposed in this paper. In discussing the computational complexity of the algorithm, we make the following assumption.

**Assumption 2.** *There is a unit computational cost in performing one timestep of the Euler–Maruyama discretization and in determining whether  $\widehat{X}_{t_{n+1}} \in D$ .*

**Comment.** The main point in this assumption is that the cost does not depend on the timestep  $h$  or the overall accuracy  $\varepsilon$  to be achieved. Regarding the dependence of the cost on the dimension  $d$ , the actual computing time required to generate the Brownian increments  $\Delta W_n$  for one timestep will be proportional to  $d$ , while the computation of  $b(\widehat{X}_{t_n}, t_n)$  and the product  $b(\widehat{X}_{t_n}, t_n) \Delta W_n$  will have a cost proportional to  $d d'$  if  $b(x, t)$  is dense.

Determining whether  $\widehat{X}_{t_{n+1}} \in D$  may not be easy in practice. It obviously depends on the way in which  $D$ , or its boundary  $\partial D$ , is specified. If  $\partial D$  is specified as a collection of patches, then a generalized octree data structure may be required to minimize the searching required to determine which boundary patches a point is near to and whether the move from  $\widehat{X}_{t_n}$  to  $\widehat{X}_{t_{n+1}}$  has crossed any patches. This cost will actually reduce slightly as  $h \rightarrow 0$  since the jumps  $\widehat{X}_{t_{n+1}} - \widehat{X}_{t_n}$  will become smaller and so less searching is likely to be required.

Standard theory on the numerical analysis of the Euler–Maruyama discretization [17] gives the following strong convergence result.

**Lemma 2.** *For  $q \geq 1$  we have*

$$\mathbb{E}^{x_0, 0} \left[ \sup_{[0, T]} \|X_t - \widehat{X}_t\|^q \right]^{1/q} = O(h^{1/2} |\log h|^{1/2}).$$

The following corollary bounds the step changes in the Euler–Maruyama solution.

**Corollary 3.** *For  $q \geq 1$  we have*

$$\mathbb{E}^{x_0, 0} \left[ \sup_{[0, T]} \|\widehat{X}_t - \lim_{s \rightarrow t^-} \widehat{X}_s\|^q \right]^{1/q} = O(h^{1/2} |\log h|^{1/2}).$$

**Proof.**  $\|\widehat{X}_t - \widehat{X}_s\| \leq \|X_t - \widehat{X}_t\| + \|X_s - \widehat{X}_s\| + \|X_t - X_s\|$ , and then the result follows by taking the limit  $s \rightarrow t^-$  and using the result from the previous lemma. ■

The final assumption concerns the weak error due to the Euler–Maruyama approximation.

**Assumption 3.** *There exist constants  $c_u$  and  $c_{\text{exit}}$  such that for all  $x \in D, 0 < t < T$*

$$\begin{aligned} |u(x, t) - \widehat{u}(x, t)| &\leq c_u h^{1/2}, \\ |u_{\text{exit}}(x, t) - \widehat{u}_{\text{exit}}(x, t)| &\leq c_{\text{exit}} h^{1/2}. \end{aligned}$$

**Comment.** Gobet and Menozzi [11] have proved this is true given Assumption 1 and conditions on the smoothness of the boundary  $\partial D$ . See also the related matched asymptotic

analysis by Howison and Steinberg [14] and a new paper by Bouchard, Geiss, and Gobet [2], which analyzes the expected  $L_1$  error in the exit time, under certain conditions.

Under these conditions, based on an original idea due to Broadie, Glasserman, and Kou [3], Gobet and Menozzi go on to develop a numerical discretization with an improved first order weak convergence, by defining the path to have exited the domain at time  $t_n$  if either  $\widehat{X}_{t_n} \notin D$  or  $\widehat{X}_{t_n} \in D$  and

$$\|\widehat{X}_{t_n} - \pi(\widehat{X}_{t_n})\|_2 \leq c_0 \|n^T(\pi(\widehat{X}_{t_n})) b(\widehat{X}_{t_n})\|_2 h^{1/2},$$

where  $\pi(x) \equiv \arg \min_{y \in \partial D} \|x - y\|_2$  is the projection of  $x$  onto the boundary  $\partial D$ ,  $n(x)$  is the unit normal on the boundary, and  $c_0 = -\zeta(1/2)/\sqrt{2\pi} \approx 0.5826$  (see (2.1) in [11]).

The Lipschitz assumption for  $u_{\text{exit}}$ , together with the bound in Assumption 3, leads immediately to the following bound on the expected exit time of numerical path simulations, based on the initial distance from the boundary.

**Lemma 4.** *Under Assumptions 1 and 3, for any  $x_0 \in D$ ,  $t_0 = nh < T$  for some integer  $n$ , and  $y \in \partial D$ ,*

$$\mathbb{E}^{x_0, t_0} [\widehat{\tau} - t_0] \leq L_{\text{exit}} (\|y - x_0\|_2 + |s - t_0|) + c_{\text{exit}} h^{1/2}.$$

This lemma is important because of the following theorem.

**Theorem 5.** *There exists a constant  $c$  such that for any  $x_0 \in D$  with  $t_0 = nh < T$  for some integer  $n$ ,*

$$\mathbb{V}^{x_0, t_0} [\widehat{P}_{t_0}] \leq c \mathbb{E}^{x_0, t_0} [\widehat{\tau} - t_0].$$

*Proof.* The proof is similar to that for Lemma 1. If we define  $\mathbf{1}_n$  to be the indicator function for  $t_n < \widehat{\tau}$ , then by considering  $d(\widehat{E}(t_0, s) g(\widehat{X}_s, s))$  we obtain

$$\begin{aligned} \widehat{E}(t_0, \widehat{\tau}) g(\widehat{X}_{\widehat{\tau}}, \widehat{\tau}) - g(x_0, t_0) &= \int_{t_0}^{\widehat{\tau}} \widehat{E}(t_0, s) \left( -V(\widehat{X}_s, s) g(\widehat{X}_s, s) + \dot{g}(\widehat{X}_s, s) \right) ds \\ &\quad + \sum_{n \geq 0} \mathbf{1}_n \widehat{E}(t_0, t_{n+1}) \left( g(\widehat{X}_{t_{n+1}}, t_{n+1}) - g(\widehat{X}_{t_n}, t_{n+1}) \right). \end{aligned}$$

If the function  $F : \mathbb{R} \rightarrow \mathbb{R}$  is twice continuously differentiable, then there exists some  $0 < \xi < 1$  such that  $F(1) - F(0) = F'(0) + \frac{1}{2} F''(\xi)$ . Applying this to  $F(s) \equiv g(\widehat{X}_{t_n} + s \Delta \widehat{X}, t_{n+1})$  with  $\Delta \widehat{X} \equiv \widehat{X}_{t_{n+1}} - \widehat{X}_{t_n} = a(\widehat{X}_{t_n}, t_n) h + b(\widehat{X}_{t_n}, t_n) \Delta W_n$  gives

$$\begin{aligned} g(\widehat{X}_{t_{n+1}}, t_{n+1}) - g(\widehat{X}_{t_n}, t_{n+1}) &= (\nabla g)_n^T (a_n h + b_n \Delta W_n) \\ &\quad + \frac{1}{2} (a_n h + b_n \Delta W_n)^T H_{g,n} (a_n h + b_n \Delta W_n), \end{aligned}$$

where  $a_n \equiv a(\widehat{X}_{t_n}, t_n)$ ,  $b_n \equiv b(\widehat{X}_{t_n}, t_n)$ ,  $(\nabla g)_n \equiv \nabla g(\widehat{X}_{t_n}, t_{n+1})$ , and  $H_{g,n}$ , the Hessian of  $g$ , is evaluated at  $(\widehat{X}_{t_n} + \xi \Delta \widehat{X}, t_{n+1})$ .

Hence,

$$\widehat{P}_{t_0} - g(x_0, t_0) = p^{(1)} + p^{(2)} + p^{(3)},$$

where

$$\begin{aligned} p^{(1)} &= \int_{t_0}^{\hat{\tau}} \hat{E}(t_0, s) \left( f(\hat{X}_s, s) - V(\hat{X}_s, s) g(\hat{X}_s, s) + \dot{g}(\hat{X}_s, s) \right) ds, \\ p^{(2)} &= \sum_{n \geq 0} \mathbf{1}_n \hat{E}(t_0, t_{n+1}) (\nabla g)_n^T b_n \Delta W_n, \\ p^{(3)} &= \sum_{n \geq 0} \mathbf{1}_n \hat{E}(t_0, t_{n+1}) \left( (\nabla g)_n^T a_n h + \frac{1}{2} (a_n h + b_n \Delta W_n)^T H_{g,n} (a_n h + b_n \Delta W_n) \right). \end{aligned}$$

The variance therefore has the bound

$$\begin{aligned} \mathbb{V}^{x_0, t_0} [\hat{P}_{t_0}] &\leq \mathbb{E}^{x_0, t_0} \left[ (\hat{P}_{t_0} - g(x_0, t_0))^2 \right] \\ &\leq 3 \mathbb{E}^{x_0, t_0} [(p^{(1)})^2 + (p^{(2)})^2 + (p^{(3)})^2]. \end{aligned}$$

Since the integrand for  $p^{(1)}$  is bounded, as before, there exists a constant  $c$ , independent of  $W, x_0, t_0$ , and  $h$ , such that  $|p^{(1)}| \leq c(\hat{\tau} - t_0) \leq cT$  and hence

$$\mathbb{E}^{x_0, t_0} [(p^{(1)})^2] \leq c^2 T \mathbb{E}^{x_0, t_0} [\hat{\tau} - t_0].$$

In addition,

$$\begin{aligned} \mathbb{E}^{x_0, t_0} [(p^{(2)})^2] &= \mathbb{E}^{x_0, t_0} \left[ \sum_{n \geq 0} \mathbf{1}_n (\hat{E}(t_0, t_{n+1}))^2 ((\nabla g)_n^T b_n \Delta W_n)^2 \right] \\ &= \mathbb{E}^{x_0, t_0} \left[ \sum_{n \geq 0} \mathbf{1}_n (\hat{E}(t_0, t_{n+1}))^2 \|(\nabla g)_n^T b_n\|_2^2 h \right] \\ &\leq \exp(2T\|V\|_\infty) \|\nabla g\|_{2, \infty}^2 \|b\|_{2, \infty}^2 \mathbb{E}^{x_0, t_0} [\hat{\tau} - t_0]. \end{aligned}$$

Finally, we need to bound  $\mathbb{E}^{x_0, t_0} [(p^{(3)})^2]$ . There are at most  $T/h$  timesteps before the computed path leaves the domain, and hence

$$\mathbb{E}^{x_0, t_0} [(p^{(3)})^2] \leq \mathbb{E}^{x_0, t_0} \left[ \sum_{n \geq 0} \mathbf{1}_n h S_n \right],$$

where

$$S_n = Th^{-2} (\hat{E}(t_0, t_{n+1}))^2 \left( (\nabla g)_n^T a_n h + \frac{1}{2} (a_n h + b_n \Delta W_n)^T H_{g,n} (a_n h + b_n \Delta W_n) \right)^2.$$

Repeatedly using the inequality  $(u+v)^2 \leq 2(u^2+v^2)$ , and bounding the various terms, gives  $\mathbf{1}_n S_n \leq \mathbf{1}_n S'_n$ , where

$$\begin{aligned} S'_n &= 2T \exp(2T\|V\|_\infty) \left( \|\nabla g\|_{2, \infty}^2 \|a\|_{2, \infty}^2 \right. \\ &\quad \left. + \|H_g\|_{2, \infty}^2 (T^2 \|a\|_{2, \infty}^4 + h^{-2} \|b\|_{2, \infty}^4 \|\Delta W_n\|_2^4) \right). \end{aligned}$$

The  $S'_n$  are i.i.d. and have a finite expected value which is independent of  $h$ . In addition, for each  $n$ ,  $S'_n$  is independent of  $\mathbf{1}_n$ . Hence

$$\mathbb{E}^{x_0, t_0} [(p^{(3)})^2] \leq \mathbb{E}^{x_0, t_0} \left[ \sum_{n \geq 0} \mathbf{1}_n h S'_n \right] = \mathbb{E}[S'_0] \mathbb{E}^{x_0, t_0} [\widehat{\tau} - t_0].$$

This completes the proof. ■

**4. MLMC algorithm.** The MLMC method is well documented [7, 8]. The following is the version of the MLMC theorem given in [8].

**Theorem 6.** *Let  $P$  denote a random variable, and let  $\widehat{P}_\ell$  denote the corresponding level  $\ell$  numerical approximation.*

*If there exist independent estimators  $Y_\ell$  each of which is an average of  $N_\ell$  i.i.d. Monte Carlo samples, each with expected cost  $C_\ell$  and variance  $V_\ell$ , and positive constants  $\alpha, \beta, \gamma, c_1, c_2, c_3$  such that  $\alpha \geq \frac{1}{2} \min(\beta, \gamma)$  and*

- (i)  $|\mathbb{E}[\widehat{P}_\ell - P]| \leq c_1 2^{-\alpha \ell},$
- (ii)  $\mathbb{E}[Y_\ell] = \begin{cases} \mathbb{E}[\widehat{P}_0], & \ell = 0, \\ \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}], & \ell > 0, \end{cases}$
- (iii)  $V_\ell \leq c_2 2^{-\beta \ell}$
- (iv)  $C_\ell \leq c_3 2^{\gamma \ell},$

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

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

*has a mean-square-error with bound*

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

*with an expected computational complexity  $C$  with bound*

$$\mathbb{E}[C] \leq \begin{cases} c_4 \varepsilon^{-2}, & \beta > \gamma, \\ c_4 \varepsilon^{-2} (\log \varepsilon)^2, & \beta = \gamma, \\ c_4 \varepsilon^{-2 - (\gamma - \beta)/\alpha}, & \beta < \gamma. \end{cases}$$

In the setting of this paper,

$$(4) \quad P = \int_0^\tau E(0, s) f(X_s, s) ds + E(0, \tau) g(X_\tau, \tau),$$

with each path starting from  $X_0 = x_0$ , and  $\widehat{P}_\ell$  represents the corresponding numerical approximation on level  $\ell$  using a uniform timestep of size  $h_\ell$  defined by  $h_\ell = K^{-\ell} h_0$  for some fixed  $h_0 > 0$  and integer  $K > 1$ . Based on the observations in [7], we will use  $K = 4$  in the algorithm description and in the numerical experiments presented later, but the numerical analysis will allow for the possibility of other values.

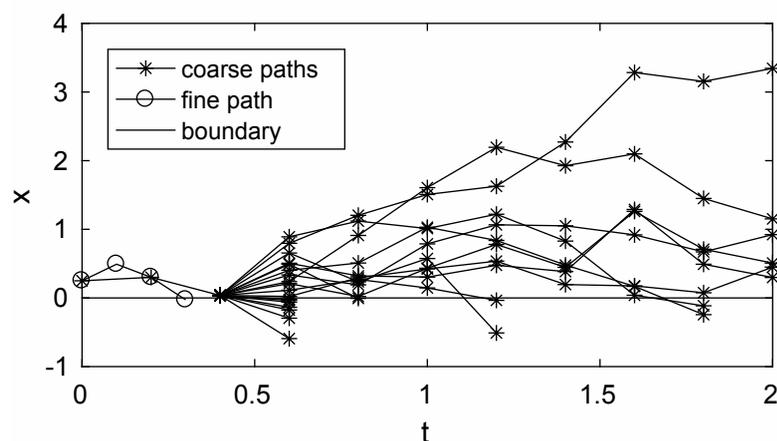
The question now is how to define  $Y_\ell$ ? Higham et al. [13] define it as the average of the differences in numerical approximations on levels  $\ell$  and  $\ell - 1$  from  $N_\ell$  independent path simulations,

$$Y_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} \left( \widehat{P}_\ell(W^{(n)}) - \widehat{P}_{\ell-1}(W^{(n)}) \right).$$

Here the notation  $\widehat{P}_\ell(W^{(n)}) - \widehat{P}_{\ell-1}(W^{(n)})$  indicates the difference of two numerical approximations with different timesteps but the same driving Brownian motion. The use of the same Brownian path is key to the multilevel treatment; it is this which ensures that the variance  $V_\ell$  decays to zero as  $\ell \rightarrow \infty$ .

However, the variance does not decay to zero as rapidly as one would hope. The problem is illustrated in Figure 1. The line with circles represents a fine Brownian path which after three timesteps crosses the boundary. The other line shows that after two coarse timesteps (each corresponding to two fine timesteps) the coarse path has not crossed the boundary, but it is extremely close to the boundary. Continuing the simulation from that point, the multiple lines correspond to multiple independent simulations of the extension of the Brownian path. Many of these quickly cross the boundary, but some fail to cross until much later. In fact, speaking rather loosely, for any fixed time  $T$  there is approximately an  $O(h^{1/2})$  probability that the coarse path continues for more than time  $T$  before crossing the boundary. Higham et al. [13] prove that as a consequence,  $V_\ell = O(h_\ell^{1/2} |\log h_\ell|^{1/2})$ , and this leads to an MLMC computational complexity bound which is  $O(\varepsilon^{-3} |\log \varepsilon|^3)$ .

Higham et al. did not use a boundary correction, and therefore their approximation had only  $O(h^{1/2})$  weak convergence. In the case of smooth boundaries  $\partial D$ , modifying their MLMC treatment to use Gobet and Menozzi's boundary shift [11] would improve the weak convergence to first order and reduce the overall complexity to  $O(\varepsilon^{-2.5} |\log \varepsilon|^3)$  by approximately halving the number of levels required to achieve the desired accuracy.



**Figure 1.** Illustration of fine and coarse 1D Brownian paths with termination when they cross a boundary at  $x=0$ . Note that the actual algorithm and analysis are based on a piecewise constant approximation, but a continuous piecewise linear interpolation is used here for visual clarity.

The new MLMC algorithm has the same complexity regardless of whether it uses the standard boundary treatment or the Gobet and Menozzi treatment. The numerical analysis will be given in detail for the former but has a natural extension to the latter. The numerical results for both methods indicate that the latter is 4 – 10× more efficient in practice.

**5. New MLMC algorithm.** Figure 1 illustrated not only the problem with the existing MLMC algorithm but also the solution. The key idea in the new algorithm is that when simulating a pair of coarse and fine paths, driven by the same Brownian motion, once one path exits the domain the other path is split (or duplicated) into multiple copies, each driven from that point onward by different independent Brownian path realizations. Averaging the output from these different copies gives an approximation to the conditional expectation for that path, and the variance for the multilevel estimator based on the conditional expectation is much lower. This is similar to the use of conditional expectations for digital options and barrier options in [6, 8].

By introducing the splitting, it is clear that the computational cost increases. This is important in deciding the number of split paths. Ignoring factors proportional to  $\log h_\ell$ , it will be shown that the split paths start at a distance which is  $O(h_\ell^{1/2})$  from the boundary, and therefore the expected exit time for each of the split paths is  $O(h_\ell^{1/2})$ , due to Assumptions 1 and 3. By making the number of split paths  $O(h_\ell^{-1/2})$ , the cost of the split paths is  $O(h_\ell^{-1})$ , which is the same as the order of cost of the coarse and fine paths before the split. Thus, the overall cost has been increased by some factor, but its order has not changed. It will be shown later, after the proof of Theorem 10, that this number of split paths is the minimum required to reduce the multilevel variance to the same order as using the true conditional expectation, which is equivalent to using an infinite number of split paths.

The algorithm to compute  $Y_\ell$  is given in Algorithm 5.1. The timestep on level  $\ell$  is taken to be  $h_\ell = 4^{-\ell} h_0$ . For each of the  $N_\ell$  samples, the fine and coarse paths are computed using the same driving Brownian path  $W^{(n)}$  until the end of the first coarse timestep for which at least one of the two paths has either left the domain  $D$  or reached the terminal time  $T$ . If the other path has not yet exited  $D$  or reached time  $T$ , it is split into  $M_\ell$  copies, and each is continued using independent Brownian paths represented by  $Z^{(n,m)}$  until it finally exits the domain or reaches time  $T$ . The final value for  $Y_\ell$  is given by

$$Y_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} \left( \bar{P}_\ell(W^{(n)}) - \bar{P}_{\ell-1}(W^{(n)}) \right),$$

where  $\bar{P}_\ell(W^{(n)}) = M_\ell^{-1} \sum_{m=1}^{M_\ell} \hat{P}_\ell(W^{(n)}, Z^{(n,m)})$ , and  $\bar{P}_{\ell-1}(W^{(n)})$  is defined similarly. With a slight abuse of notation, each of the quantities  $\hat{P}_\ell(W^{(n)}, Z^{(n,m)})$  is the appropriate approximation of the Feynman–Kac functional (4) as defined in section 4, and the averaging over  $m$  is trivial for the path which did not split and therefore has no dependence on  $Z^{(n,m)}$ .

## 6. Numerical analysis.

**6.1. Variance analysis.** For completeness, the numerical analysis begins with this general result for estimators with splitting. This use of splitting is related to the conditional Monte

---

**Algorithm 1.** Algorithm to compute  $Y_\ell$  using  $N_\ell$  pairs of coarse/fine paths, with a split into  $M_\ell$  subpaths after one exits.

---

**for**  $n = 1, \dots, N_\ell$  **do**  
  for given  $W^{(n)}$ , compute  $\widehat{X}_\ell^{(n)}$  and  $\widehat{X}_{\ell-1}^{(n)}$  using Euler–Maruyama discretization with Brownian motion  $W^{(n)}$  and timesteps  $h_\ell, h_{\ell-1}$ , respectively, until the end of the first coarse timestep for which at least one of the two has exited domain  $D$  or reached the terminal time  $T$   
  **if**  $\widehat{X}_\ell^{(n)}$  has exited  $D$  or the terminal time  $T$  has been reached **then**  
     $\overline{P}_\ell^{(n)} := \widehat{P}_\ell(W^{(n)})$  based on path already computed  
  **else**  
     $\overline{P}_\ell^{(n)} := M_\ell^{-1} \sum_{m=1}^{M_\ell} \widehat{P}_\ell(W^{(n)}, Z^{(n,m)})$  based on  $M_\ell$  independent path continuations using Brownian paths represented by  $Z^{(n,m)}$   
  **end if**  
  **if**  $\widehat{X}_{\ell-1}^{(n)}$  has exited  $D$  or the terminal time  $T$  has been reached **then**  
     $\overline{P}_{\ell-1}^{(n)} := \widehat{P}_{\ell-1}(W^{(n)})$  based on path already computed  
  **else**  
     $\overline{P}_{\ell-1}^{(n)} := M_\ell^{-1} \sum_{m=1}^{M_\ell} \widehat{P}_{\ell-1}(W^{(n)}, Z^{(n,m)})$  based on  $M_\ell$  independent path continuations using Brownian paths represented by  $Z^{(n,m)}$   
  **end if**  
**end for**  
 $Y_\ell := N_\ell^{-1} \sum_{n=1}^{N_\ell} (\overline{P}_\ell^{(n)} - \overline{P}_{\ell-1}^{(n)})$

---

Carlo method (see pp. 145–148 in [1]) but quite different from the splitting technique in rare event simulation.

**Lemma 7.** *If  $W$  and  $Z^{(1)}, \dots, Z^{(M)}$  are independent random variables such that  $Z^{(1)}, \dots, Z^{(M)}$  are identically distributed to a random variable  $Z$ , then*

$$\bar{f} = M^{-1} \sum_{m=1}^M f(W, Z^{(m)})$$

*is an unbiased estimator for  $\mathbb{E}[f(W, Z)]$  and its variance is*

$$\mathbb{V}[\bar{f}] = \mathbb{V}[\mathbb{E}[f(W, Z) | W]] + M^{-1} \mathbb{E}[\mathbb{V}[f(W, Z) | W]].$$

*Proof.* Conditional on a value for  $W$ , the expected value of  $\bar{f}$  is  $\mathbb{E}[f(W, Z) | W]$  and its variance is  $M^{-1} \mathbb{V}[f(W, Z) | W]$ , and therefore

$$\mathbb{E} [\bar{f}^2 | W] = \left( \mathbb{E}[f(W, Z) | W] \right)^2 + M^{-1} \mathbb{V}[f(W, Z) | W].$$

Taking the expectation over the distribution for  $W$  then gives

$$\begin{aligned} \mathbb{E}[\bar{f}] &= \mathbb{E} \left[ \mathbb{E}[f(W, Z) | W] \right] = \mathbb{E}[f(W, Z)], \\ \mathbb{E} [\bar{f}^2] &= \mathbb{E} \left[ \left( \mathbb{E}[f(W, Z) | W] \right)^2 \right] + M^{-1} \mathbb{E} \left[ \mathbb{V}[f(W, Z) | W] \right], \end{aligned}$$

from which it follows that

$$\mathbb{V}[\bar{f}] = \mathbb{E} [\bar{f}^2] - \left( \mathbb{E}[\bar{f}] \right)^2 = \mathbb{V} \left[ \mathbb{E}[f(W, Z) | W] \right] + M^{-1} \mathbb{E} \left[ \mathbb{V}[f(W, Z) | W] \right]. \quad \blacksquare$$

Applying this result to the multilevel estimator in section 5 using  $M_\ell$  split paths on level  $\ell$  gives us

$$(5) \quad V_\ell = \mathbb{V} [\bar{P}_\ell - \bar{P}_{\ell-1}] = \mathbb{V} \left[ \mathbb{E}[\hat{P}_\ell - \hat{P}_{\ell-1} | W] \right] + M_\ell^{-1} \mathbb{E} \left[ \mathbb{V}[\hat{P}_\ell - \hat{P}_{\ell-1} | W] \right].$$

The main theorem will bound  $\mathbb{V}[\mathbb{E}[\hat{P}_\ell - \hat{P}_{\ell-1} | W]]$  and  $\mathbb{E}[\mathbb{V}[\hat{P}_\ell - \hat{P}_{\ell-1} | W]]$ , and we will then bound  $C_\ell$ , the expected cost of each sample when using the splitting. To prepare for it, we begin by defining  $\underline{\tau}$  to be the exit time of the first of a pair of coarse/fine paths and  $\bar{\tau}$  to be  $\underline{\tau}$  rounded up to the end of a coarse timestep (i.e., a multiple of  $h_{\ell-1}$ ), which is the time at which the splitting occurs if one of the two paths remains within the domain. Note that  $\bar{\tau} \neq \underline{\tau}$  only when the fine path exits before the coarse path, partway through a coarse timestep.

The following lemma bounds the difference between the coarse and fine paths up to time  $\underline{\tau}$  and also the difference between the fine path at time  $\underline{\tau}$  and the coarse path at the possibly different time  $\bar{\tau}$ .

**Lemma 8.** *Given the definitions above, for  $q \geq 1$  we have*

$$\mathbb{E} \left[ \sup_{[0, \underline{\tau}]} \|\hat{X}_{\ell, t} - \hat{X}_{\ell-1, t}\|^q \right]^{1/q} = O(h_{\ell-1}^{1/2} |\log h_{\ell-1}|^{1/2})$$

and

$$\mathbb{E} \left[ \|\hat{X}_{\ell, \underline{\tau}} - \hat{X}_{\ell-1, \bar{\tau}}\|^q \right]^{1/q} = O(h_{\ell-1}^{1/2} |\log h_{\ell-1}|^{1/2}).$$

*Proof.* These follow directly from Lemma 2 and Corollary 3. \blacksquare

For a particular pair  $(W, Z)$ , we define  $\hat{\tau}_\ell$  and  $\hat{\tau}_{\ell-1}$  to be the exit times of the fine and coarse paths. The next lemma bounds the expected absolute difference between these two exit times. This is unaffected by the path splitting, so it is sufficient to consider what happens when there is no splitting.

**Lemma 9.** *Given the definitions above and Assumptions 1 and 3,*

$$\mathbb{E} [ |\hat{\tau}_\ell - \hat{\tau}_{\ell-1}| ] = O \left( h_{\ell-1}^{1/2} |\log h_{\ell-1}|^{1/2} \right).$$

*Proof.* Each  $W$  leads to one of three possibilities:

- $\widehat{X}_{\ell, \tau} \notin D$  and  $\widehat{X}_{\ell-1, \bar{\tau}} \notin D$ ,
- $\widehat{X}_{\ell, \tau} \notin D$  and  $\widehat{X}_{\ell-1, \bar{\tau}} \in D$ ,
- $\widehat{X}_{\ell-1, \tau} \notin D$  and  $\widehat{X}_{\ell, \tau} \in D$ .

In the first case,  $|\widehat{\tau}_\ell - \widehat{\tau}_{\ell-1}| = \bar{\tau} - \tau < h_{\ell-1}$ .

In the second case, since  $\widehat{X}_{\ell-1, \bar{\tau}} \in D$  but  $\widehat{X}_{\ell, \tau} \notin D$ , there is a point  $y \in \partial D$  on the line between the two. Hence, noting that  $u_{\text{exit}}(y, \bar{\tau}) = \bar{\tau}$ , we have

$$\begin{aligned} \mathbb{E}[\widehat{\tau}_{\ell-1} - \bar{\tau} | W] &= \left( \widehat{u}_{\ell-1, \text{exit}}(\widehat{X}_{\ell-1, \bar{\tau}}, \bar{\tau}) - u_{\text{exit}}(\widehat{X}_{\ell-1, \bar{\tau}}, \bar{\tau}) \right) \\ &\quad + \left( u_{\text{exit}}(\widehat{X}_{\ell-1, \bar{\tau}}, \bar{\tau}) - u_{\text{exit}}(y, \bar{\tau}) \right). \end{aligned}$$

The first term is bounded by Assumption 3 and the second term by the Lipschitz property in Assumption 1, giving

$$\mathbb{E}[\widehat{\tau}_{\ell-1} - \bar{\tau} | W] \leq c_{\text{exit}} h_{\ell-1}^{1/2} + L_{\text{exit}} \left( \|\widehat{X}_{\ell, \tau} - \widehat{X}_{\ell-1, \bar{\tau}}\|_2 + h_{\ell-1} \right).$$

Adding in the fact that  $\bar{\tau} - \widehat{\tau}_\ell \leq h_{\ell-1}$ , this can be simplified to there being a constant  $c$  for which

$$\mathbb{E}[\widehat{\tau}_{\ell-1} - \widehat{\tau}_\ell | W] \leq c \left( h_{\ell-1}^{1/2} + \|\widehat{X}_{\ell, \tau} - \widehat{X}_{\ell-1, \bar{\tau}}\|_2 \right).$$

A similar result holds in the third case, and hence there is a constant  $c$  for which

$$\mathbb{E}[\widehat{\tau}_\ell - \widehat{\tau}_{\ell-1} | W] \leq c \left( h_{\ell-1}^{1/2} + \|\widehat{X}_{\ell, \tau} - \widehat{X}_{\ell-1, \bar{\tau}}\|_2 \right).$$

Combining these two results, taking an expectation over  $W$ , and using the result in Lemma 8 gives the desired result that

$$\mathbb{E}[|\widehat{\tau}_\ell - \widehat{\tau}_{\ell-1}|] = O\left(h_{\ell-1}^{1/2} |\log h_{\ell-1}|^{1/2}\right). \quad \blacksquare$$

A new paper by Bouchard, Geiss, and Gobet [2] analyzes the expected  $L_1$  error in the exit time and proves that under certain conditions  $\mathbb{E}[|\widehat{\tau} - \tau|] = O(h^{1/2})$ . It follows from this that we have the slightly stronger result  $\mathbb{E}[|\widehat{\tau}_\ell - \widehat{\tau}_{\ell-1}|] = O(h_{\ell-1}^{1/2})$ .

We are now ready to prove the main theorem of this section.

**Theorem 10.** *Under Assumptions 1 and 3,*

$$\begin{aligned} \mathbb{V} \left[ \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1} | W] \right] &= O(h_{\ell-1} |\log h_{\ell-1}|), \\ \mathbb{E} \left[ \mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1} | W] \right] &= O(h_{\ell-1}^{1/2} |\log h_{\ell-1}|^{1/2}). \end{aligned}$$

*Proof.* For the fine path,

$$\begin{aligned} \widehat{P}_\ell &= \int_0^{\widehat{\tau}_\ell} \widehat{E}_\ell(0, s) f(\widehat{X}_{\ell, s}, s) ds + \widehat{E}_\ell(0, \widehat{\tau}_\ell) g(\widehat{X}_{\ell, \widehat{\tau}_\ell}, \widehat{\tau}_\ell) \\ &= p_\ell^{(1)}(W) + p_\ell^{(2)}(W) p_\ell^{(3)}(W, Z), \end{aligned}$$

where

$$\begin{aligned}
 p_\ell^{(1)}(W) &= \int_0^\tau \widehat{E}_\ell(0, s) f(\widehat{X}_{\ell, s}, s) \, ds, \\
 p_\ell^{(2)}(W) &= \widehat{E}_\ell(0, \tau), \\
 p_\ell^{(3)}(W, Z) &= \int_\tau^{\widehat{\tau}_\ell} \widehat{E}_\ell(\tau, s) f(\widehat{X}_{\ell, s}, s) \, ds + \widehat{E}_\ell(\tau, \widehat{\tau}_\ell) g(\widehat{X}_{\ell, \widehat{\tau}_\ell}, \widehat{\tau}_\ell),
 \end{aligned}$$

and similarly for the coarse path,

$$\begin{aligned}
 \widehat{P}_{\ell-1} &= \int_0^{\widehat{\tau}_{\ell-1}} \widehat{E}_{\ell-1}(0, s) f(\widehat{X}_{\ell-1, s}, s) \, ds + \widehat{E}_{\ell-1}(0, \widehat{\tau}_{\ell-1}) g(\widehat{X}_{\ell-1, \widehat{\tau}_{\ell-1}}, \widehat{\tau}_{\ell-1}) \\
 &= p_{\ell-1}^{(1)}(W) + p_{\ell-1}^{(2)}(W) p_{\ell-1}^{(3)}(W, Z),
 \end{aligned}$$

where

$$\begin{aligned}
 p_{\ell-1}^{(1)}(W) &= \int_0^{\bar{\tau}} \widehat{E}_{\ell-1}(0, s) f(\widehat{X}_{\ell-1, s}, s) \, ds, \\
 p_{\ell-1}^{(2)}(W) &= \widehat{E}_{\ell-1}(0, \bar{\tau}), \\
 p_{\ell-1}^{(3)}(W, Z) &= \int_{\bar{\tau}}^{\widehat{\tau}_{\ell-1}} \widehat{E}_{\ell-1}(\bar{\tau}, s) f(\widehat{X}_{\ell-1, s}, s) \, ds + \widehat{E}_{\ell-1}(\bar{\tau}, \widehat{\tau}_{\ell-1}) g(\widehat{X}_{\ell-1, \widehat{\tau}_{\ell-1}}, \widehat{\tau}_{\ell-1}).
 \end{aligned}$$

Noting that  $\exp(v) - \exp(w) = (v - w) \exp(\xi)$ , for some  $\xi$  in the interval with endpoints  $v, w$ , and defining

$$\|\Delta \widehat{X}\|_\infty = \sup_{t \in [0, \tau]} \|\widehat{X}_{\ell, t} - \widehat{X}_{\ell-1, t}\|_2,$$

then for  $s \in [0, \tau]$  we have

$$\left| \widehat{E}_\ell(0, s) - \widehat{E}_{\ell-1}(0, s) \right| \leq \exp(T\|V\|_\infty) TL_V \|\Delta \widehat{X}\|_\infty.$$

We also have

$$\left| f(\widehat{X}_{\ell, s}, s) - f(\widehat{X}_{\ell-1, s}, s) \right| \leq L_f \|\Delta \widehat{X}\|_\infty,$$

and hence there exists a constant  $c^{(1)}$  (independent of  $W$ ) such that

$$(6) \quad \left| p_\ell^{(1)} - p_{\ell-1}^{(1)} \right| \leq c^{(1)} \|\Delta \widehat{X}\|_\infty + h_{\ell-1} \exp(T\|V\|_\infty) \|f\|_\infty$$

with the second term on the right being due to the possible difference in the upper limits of integration in the definitions of  $p_\ell^{(1)}, p_{\ell-1}^{(1)}$ .

Similarly, there exists a constant  $c^{(2)}$  (independent of  $W$ ) such that

$$(7) \quad \left| p_\ell^{(2)} - p_{\ell-1}^{(2)} \right| \leq c^{(2)} \|\Delta \widehat{X}\|_\infty + h_{\ell-1} \exp(T\|V\|_\infty) \|V\|_\infty.$$

Taking an expectation over  $Z$ , we have

$$\begin{aligned} \mathbb{E}[p_\ell^{(3)} | W] &= \hat{u}_\ell(\hat{X}_{\ell, \tau}, \tau), \\ \mathbb{E}[p_{\ell-1}^{(3)} | W] &= \hat{u}_{\ell-1}(\hat{X}_{\ell-1, \bar{\tau}}, \bar{\tau}), \end{aligned}$$

and therefore, using Assumption 3 together with Assumption 1’s Lipschitz property for  $u(x, t)$ , we obtain

$$(8) \quad \left| \mathbb{E}[p_\ell^{(3)} - p_{\ell-1}^{(3)} | W] \right| \leq c_u(h_\ell^{1/2} + h_{\ell-1}^{1/2}) + L_u \left( \|\hat{X}_{\ell, \tau} - \hat{X}_{\ell-1, \tau}\|_2 + h_{\ell-1} \right).$$

Combining the results in (6)–(8), there exist constants  $c_1, c_2, c_3$ , independent of  $W$ , such that

$$\left| \mathbb{E}[\hat{P}_\ell - \hat{P}_{\ell-1} | W] \right| \leq c_1 \|\Delta \hat{X}\|_\infty + c_2 \|\hat{X}_{\ell, \tau} - \hat{X}_{\ell-1, \tau}\|_2 + c_3 h_{\ell-1}^{1/2}.$$

Hence, using the results from Lemma 8, there exists a constant  $c$  such that

$$\mathbb{V} \left[ \mathbb{E}[\hat{P}_\ell - \hat{P}_{\ell-1} | W] \right] \leq \mathbb{E} \left[ \left( \mathbb{E}[\hat{P}_\ell - \hat{P}_{\ell-1} | W] \right)^2 \right] \leq c h_{\ell-1} |\log h_{\ell-1}|$$

for sufficiently small  $h_{\ell-1}$ .

This proves the first bound in the statement of the theorem. For the second bound, we note that

$$\begin{aligned} \mathbb{V} \left[ \hat{P}_\ell - \hat{P}_{\ell-1} | W \right] &= \mathbb{V} \left[ p_\ell^{(2)}(W) p_\ell^{(3)}(W, Z) - p_{\ell-1}^{(2)}(W) p_{\ell-1}^{(3)}(W, Z) | W \right] \\ &= \left( p_\ell^{(2)}(W) \right)^2 \mathbb{V} \left[ p_\ell^{(3)}(W, Z) | W \right] + \left( p_{\ell-1}^{(2)}(W) \right)^2 \mathbb{V} \left[ p_{\ell-1}^{(3)}(W, Z) | W \right]. \end{aligned}$$

The first line in the above is because  $p_\ell^{(1)}(W)$  and  $p_{\ell-1}^{(1)}(W)$  do not vary with  $Z$  and so do not contribute to the variance. The second line comes from the fact that at most only one of the coarse and fine paths is split, and so at most only one of  $p_\ell^{(3)}(W, Z)$  and  $p_{\ell-1}^{(3)}(W, Z)$  has a nonzero variance.

The terms  $p_\ell^{(2)}(W)$  and  $p_{\ell-1}^{(2)}(W)$  are each bounded by  $\exp(T\|V\|_\infty)$ . Using Theorem 5 to bound  $\mathbb{V}[p_\ell^{(3)}(W, Z)]$  and  $\mathbb{V}[p_{\ell-1}^{(3)}(W, Z)]$  gives the result that there exists a constant  $c$  such that

$$\mathbb{V} \left[ \hat{P}_\ell - \hat{P}_{\ell-1} | W \right] \leq c \mathbb{E}[\max(\hat{\tau}_\ell, \hat{\tau}_{\ell-1}) - \bar{\tau} | W] \leq c \mathbb{E}[|\hat{\tau}_\ell - \hat{\tau}_{\ell-1}| | W].$$

The above inequalities hold for each value of the random variable  $W$ . Taking an expectation and using the result in Lemma 9 then give the final result that

$$\mathbb{E} \left[ \mathbb{V} \left[ \hat{P}_\ell - \hat{P}_{\ell-1} | W \right] \right] = O \left( h_{\ell-1}^{1/2} |\log h_{\ell-1}|^{1/2} \right). \quad \blacksquare$$

Using the bounds from this theorem, we note that the two terms in the bound for the variance  $V_\ell$  in (5) are balanced when  $M_\ell = O(h_{\ell-1}^{-1/2} |\log h_{\ell-1}|^{-1/2})$ .

**6.2. Expected complexity.** Finally, given the previous results we can obtain the following result on the expected complexity to achieve an  $\varepsilon$  root-mean-square error.

**Corollary 11.** *Under the given assumptions, a root-mean-square error of  $\varepsilon$  can be achieved with an  $O(\varepsilon^{-2}|\log \varepsilon|^3)$  expected computational cost.*

*Proof.* The proof is very similar to the standard multilevel complexity analysis in [7] but has to be modified slightly because the additional  $\log h_\ell$  factors in the variance mean it does not quite have the usual form expected in condition (iii) in Theorem 6.

On level  $\ell \geq 1$ , we choose to use  $h_\ell = 4^{-\ell}h_0$  with  $M_\ell = \lceil 2^\ell/\ell^{1/2} \rceil$  paths in the splitting estimator (with  $\lceil x \rceil$  denoting  $x$  rounded up to the nearest integer). Recalling from Assumption 2 that there is assumed to be a unit computational cost for each timestep, the computational cost of one sample is equal to the sum of the number of coarse and fine timesteps before the first of the two paths exits the domain, plus the combined sum of the timesteps for the subsequent  $M_\ell$  split coarse or fine paths.

Hence the expected cost of one sample with splitting on level  $\ell \geq 1$  is

$$C_\ell \leq (h_\ell^{-1} + h_{\ell-1}^{-1}) \mathbb{E}[\min(\widehat{\tau}_\ell, \widehat{\tau}_{\ell-1})] + M_\ell h_\ell^{-1} \mathbb{E}[|\widehat{\tau}_\ell - \widehat{\tau}_{\ell-1}|] = O(h_\ell^{-1}),$$

since  $M_\ell = O(h_\ell^{-1/2}|\log h_\ell|^{-1/2})$  and Lemma 9 gives  $\mathbb{E}[|\widehat{\tau}_\ell - \widehat{\tau}_{\ell-1}|] = O(h_\ell^{1/2}|\log h_\ell|^{1/2})$ .

Given  $O(h_\ell^{1/2})$  weak convergence, to reduce the bias to  $\varepsilon/\sqrt{2}$  requires  $h_L = O(\varepsilon^2)$  on the finest level, while for  $O(h_\ell)$  weak convergence using the Gobet and Menozzi correction we require  $h_L = O(\varepsilon)$ . In either case, we have  $L = O(|\log \varepsilon|)$ .

Combining (5) with the results from Theorem 10, the multilevel variance for levels  $\ell \geq 1$  is

$$V_\ell = O(h_{\ell-1}|\log h_{\ell-1}|) = O(h_{\ell-1} \ell).$$

Following the analysis in [8], choosing

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

to ensure that the overall variance is less than  $\frac{1}{2}\varepsilon^2$ , the expected total cost is bounded by

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

Since  $C_\ell V_\ell = O(\ell)$ , there exists a constant  $c$  such that

$$\sum_{\ell=0}^L \sqrt{C_\ell V_\ell} \leq c \sum_{\ell=0}^L \ell^{1/2} \leq c \int_0^{L+1} \ell^{1/2} d\ell = \frac{2c}{3} (L+1)^{3/2}.$$

Because  $\sum_{\ell=0}^L C_\ell = O(\varepsilon^{-2})$ , it follows that the total cost is  $O(\varepsilon^{-2}|\log \varepsilon|^3)$ . ■

*Comment.* The analysis above uses  $M_\ell = \lceil 2^\ell/\ell^{1/2} \rceil$ . If instead we use  $M_\ell = 2^\ell$ , then  $C_\ell = O(h_\ell^{-1}|\log h_\ell|^{1/2})$  and the overall complexity becomes  $O(\varepsilon^{-2}|\log \varepsilon|^{7/2})$ . This is poorer, but if the variance analysis is not sharp, and in fact  $\mathbb{V}[\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1} | W]] = O(h_\ell)$  and  $\mathbb{E}[\mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1} | W]] = O(h_\ell^{1/2})$ , then the choice  $M_\ell = 2^\ell$  is asymptotically optimal and would give a complexity of  $O(\varepsilon^{-2}|\log \varepsilon|^2)$ . Hence, the numerical experiments use  $M_\ell = 2^\ell$ .

**7. Numerical experiments.** The testcase is simple Brownian diffusion in a 3D cube,  $D = [-1, +1]^3$ , over a unit time interval,  $[0, 1]$ . The initial data is  $x_0 = \mathbf{0}$ , and the output quantity of interest is the expected exit time. For the PDE formulation this corresponds to the PDE

$$\frac{\partial u}{\partial t} + \frac{1}{2} \nabla^2 u + 1 = 0,$$

subject to homogeneous boundary data on  $\partial D$  and  $t = 1$ . Using a standard Fourier series expansion, adjusted to the homogeneous boundary conditions on  $\partial D$ , and noting the symmetry in each coordinate direction, the solution at earlier times  $0 \leq t < 1$  can be expressed as

$$u(x, t) = \sum_{\text{odd } i, j, k \geq 1} A_{i, j, k}(t) \cos(i\pi x_1/2) \cos(j\pi x_2/2) \cos(k\pi x_3/2),$$

where the amplitudes  $A_{i, j, k}(t)$  satisfy the ordinary differential equation

$$\frac{dA_{i, j, k}}{dt} - \frac{(i^2 + j^2 + k^2) \pi^2}{8} A_{i, j, k} + \frac{64 (-1)^{(i+j+k+1)/2}}{i j k \pi^3} = 0,$$

subject to terminal condition  $A_{i, j, k}(1) = 0$ . Therefore

$$A_{i, j, k}(t) = \frac{512 (-1)^{(i+j+k+1)/2}}{i j k (i^2 + j^2 + k^2) \pi^5} \left( 1 - \exp \left( -\frac{8 (1-t)}{(i^2 + j^2 + k^2) \pi^2} \right) \right),$$

which gives  $u(\mathbf{0}, 0) \approx 0.435930$ .

On the coarsest level the timestep is chosen to be  $h_0 = 0.1$ , so that the mean exit time corresponds to just four timesteps. On finer levels it is  $h_\ell = 4^{-\ell} h_0$ .  $M_\ell = 2^\ell$  subsamples are used when approximating the conditional expectations once either the coarse or the fine path has left the domain.

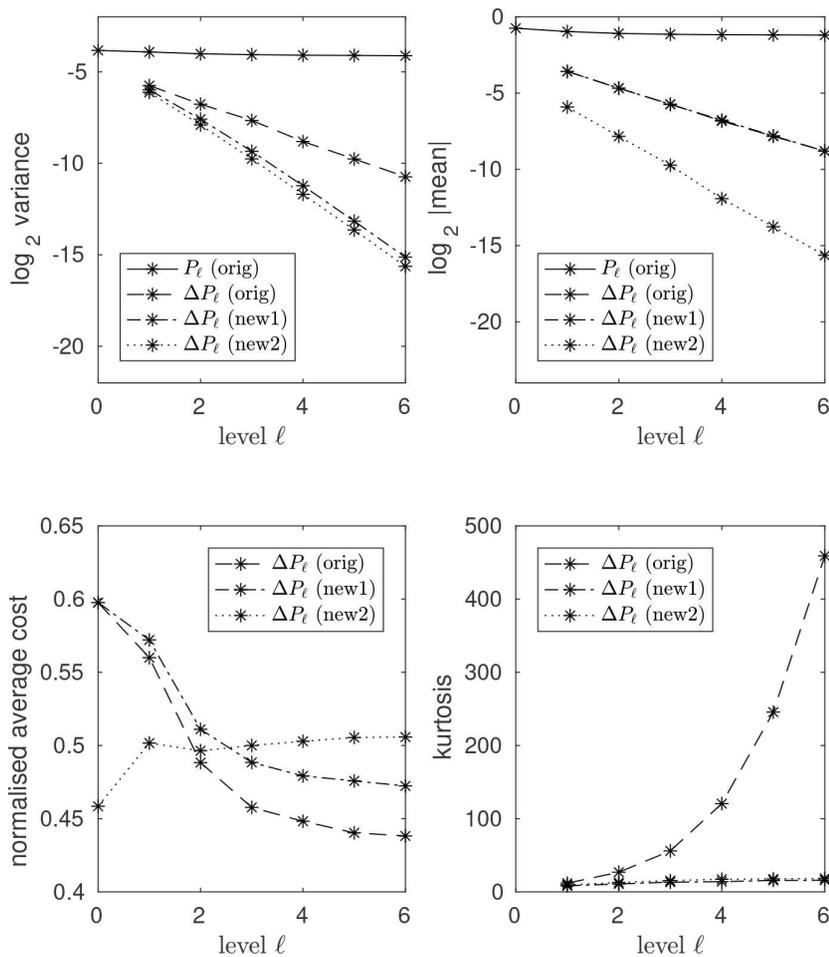
Numerical results are obtained with three schemes:

- (orig)—the original MLMC method proposed by Higham *et al.* [13];
- (new1)—the new method with subsampling to approximate the conditional expectations;
- (new2)—the new method which in addition uses the Gobet and Menozzi boundary shift, in which a path on level  $\ell$  is considered to have crossed the boundary if  $\|x\|_\infty > 1 - c_0 h_\ell^{1/2}$ , where  $c_0 = -\zeta(1/2)/\sqrt{2\pi} \approx 0.5826$ .

Figure 2 compares results from the three schemes and also a standard single-level Monte Carlo simulation without the Gobet and Menozzi boundary correction.

The top left plot shows the decay in the MLMC variance  $V_\ell$ . This improves from approximately  $O(h_\ell^{1/2})$  with the original method to approximately  $O(h_\ell)$  with subsampling. Adding the Gobet and Menozzi correction reduces the variance slightly. The top right plot shows the convergence of  $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ . As expected the introduction of subsampling does not affect this, to within Monte Carlo sampling error, but the Gobet and Menozzi correction improves the rate of convergence from  $O(h_\ell^{1/2})$  to  $O(h_\ell)$ .

The bottom left plot shows the average cost per path at each level, defined as the number of generated Normal random numbers expressed as a fraction of the number  $3/h_\ell$  which would be required for a single fine path to reach time  $t = 1$ . As expected, the cost for the

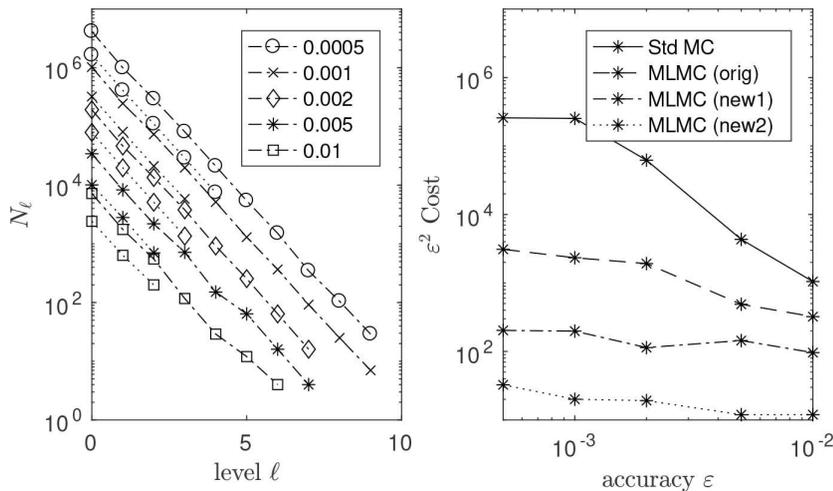


**Figure 2.** Testcase: MLMC variance  $V_\ell$ , mean correction  $\mathbb{E}[\bar{P}_\ell - \bar{P}_{\ell-1}]$ , normalized cost per path sample, and kurtosis for three MLMC methods, and standard single level Monte Carlo.

original method is approximately equal to the expected exit time. With subsampling, the cost is slightly higher, up to 15% when using the GM correction, but the additional cost is not significant and in particular does not increase with level  $\ell$ .

The bottom right plot shows the computed kurtosis of  $\bar{P}_\ell - \bar{P}_{\ell-1}$ . The increasing value for the original method is a clear symptom of the fact that for a few samples the exit times of the coarse and fine paths are very different. With subsampling, there is no asymptotic increase in the kurtosis.

Figure 3 presents MLMC results for a number of different user-specified accuracies  $\varepsilon$ . The MLMC methodology in selecting the near-optimal number of samples on each level, and finest level  $L$ , follows the description given in [8]. The plot on the left shows the number of path samples used on level  $\ell$  by the best two methods, (new1: dash-dot lines) and (new2: dotted lines). As is standard in MLMC, for larger values of  $\varepsilon$  it is not necessary to use such a large value for  $L$  to ensure that the weak error (or bias) contribution to the desired



**Figure 3.** Testcase: number of samples on different levels for different accuracies when using two new MLMC methods (dash-dot lines: new1; dotted lines: new2), and cost versus accuracy comparison of three MLMC methods and the standard single level Monte Carlo.

mean square error is sufficiently small. It is particularly noticeable that the superior weak convergence of the Gobet and Menozzi correction in (new2) greatly reduces the number of levels needed compared to (new1).

The plot on the right shows the total cost (defined as the number of Normal random numbers used) multiplied by  $\varepsilon^2$ . With the new subsampling, this scaled cost is almost independent of the desired accuracy, which is consistent with the theory which predicts the cost should be  $O(\varepsilon^{-2}L^3)$ . Because of the improved weak convergence, (new2) uses only half as many levels as (new1) and consequently the cost is reduced by a factor of 6 to 8.

**8. Conclusions and extensions.** In this paper we have developed an efficient multilevel Monte Carlo algorithm for the estimation of mean exit times and other associated functionals for stopped diffusions within a bounded domain. A root-mean-square accuracy of  $\varepsilon$  is achieved with a computational complexity which is  $O(\varepsilon^{-2}|\log \varepsilon|^3)$ . The key feature of the algorithm is the use of an approximated conditional expectation once either the coarse or the fine path has exited the domain. This greatly reduces the multilevel correction variance without significantly increasing the computational cost per path.

There are several directions in which this work can be extended. The first is to linear functionals of the PDE solution, such as

$$P = \int_0^T \int_D \rho(x, t) u(x, t) \, dx \, dt.$$

If we take  $\rho(x, t)$  to be normalized so that  $\int_0^T \int_D \rho(x, t) \, dx \, dt = 1$ , then this is very simply treated by randomizing the starting point for the SDE paths, taking the starting point  $(x_0, t_0)$  from the probability distribution with density  $\rho(x, t)$ . There are similar extensions for linear functionals of the solution on  $t = 0$ .

The next possible extension is to nonlinear functionals of the form

$$P = \int_0^T \int_D \rho(x, t) F(u(x, t)) \, dx \, dt$$

for some nonlinear function  $F(u)$ . This case can be tackled by representing it as a nested expectation

$$P = \mathbb{E}[F(u(x_0, t_0))]$$

with the outer expectation over the starting point  $(x_0, t_0)$  and  $u(x_0, t_0)$  given by the usual inner expectation from (2). Multilevel techniques for treating nested simulations are discussed in [8] based on prior research by Bujok, Hambly, and Reisinger [4]. In particular, the multi-index Monte Carlo method [12] may be very helpful in this context.

Another direction for extension is to Neumann boundary conditions. If an application has only homogeneous Neumann boundary conditions, so the Feynman–Kac functional depends only on the interior path, then the extension should be straightforward as it is easy to simulate the reflected diffusion. The difficulty will arise if there are mixed Neumann/Dirichlet boundary conditions. In this case, the MLMC variance could be larger because of the problem of one of the fine/coarse path pairs hitting the Dirichlet boundary while the other hits the Neumann boundary. In addition, the PDE solution may lack the necessary regularity assumed in this paper.

Finally, we could consider the solution of elliptic PDEs. This corresponds to parabolic solutions in the limit in which the time duration is extended to infinity. Multilevel techniques for handling this limit are presented and analyzed in [5], as well as being discussed in section 10.1 in [8], based on prior research by Glynn and Rhee on limiting distributions from Markov chains [10]. However, there is a difficulty in the numerical analysis, and possibly in practice, which is that the standard strong error analysis for a finite time interval  $[0, T]$  includes a factor which grows exponentially with respect to  $T$ . Only under certain conditions (see [5]), similar to the contracting Markov chains considered by Glynn and Rhee, does this not occur, and this may restrict the use of MLMC for this application.

*Note.* In compliance with EPSRC’s open access initiative, the data in this paper, and the codes which generated it, are available from [doi.org/10.5287/bodleian: XmZwK1zzv](https://doi.org/10.5287/bodleian: XmZwK1zzv).

**Acknowledgments.** The authors are very grateful to Abdul-Lateef Haji-Ali, Wei Fang, and the anonymous referees for their feedback, which led to significant improvements in the paper.

## REFERENCES

- [1] S. ASMUSSEN AND P. W. GLYNN, *Stochastic Simulation*, Springer, New York, 2007.
- [2] B. BOUCHARD, S. GEISS, AND E. GOBET, *First time to exit of a continuous Itô process: General moment estimates and  $L_1$ -convergence rate for discrete time approximations*, *Bernoulli*, 23 (2017), pp. 1631–1662.
- [3] M. BROADIE, P. GLASSERMAN, AND S. KOU, *A continuity correction for discrete barrier options*, *Math. Finance*, 7 (1997), pp. 325–348.

- [4] 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.
- [5] W. FANG AND M. B. GILES, *Adaptive Euler-Maruyama Method for SDEs with Non-globally Lipschitz drift: Part II, Infinite Time Interval*, preprint, [arXiv:1703.06743](https://arxiv.org/abs/1703.06743), 2017.
- [6] M. B. GILES, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, in Monte Carlo and Quasi-Monte Carlo Methods 2006, A. Keller, S. Heinrich, and H. Niederreiter, eds., Springer, 2008, pp. 343–358.
- [7] M. B. GILES, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.
- [8] M. B. GILES, *Multilevel Monte Carlo methods*, Acta Numer., 24 (2015), pp. 259–328.
- [9] P. GLASSERMAN, *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004.
- [10] P. W. GLYNN AND C.-H. RHEE, *Exact estimation for Markov chain equilibrium expectations*, J. Appl. Probab., 51 (2014), pp. 377–389.
- [11] E. GOBET AND S. MENOZZI, *Stopped diffusion processes: overshoots and boundary correction*, Stochastic Process. Appl., 120 (2010), pp. 130–162.
- [12] A.-L. HAJI-ALI, F. NOBILE, AND R. TEMPONE, *Multi Index Monte Carlo: when sparsity meets sampling*, Numer. Math., 132 (2016), pp. 767–806.
- [13] D. J. HIGHAM, X. MAO, M. ROJ, Q. SONG, AND G. YIN, *Mean exit times and the multi-level Monte Carlo method*, SIAM/ASA J. Uncertain. Quantif., 1 (2013), pp. 2–18.
- [14] S. HOWISON AND M. STEINBERG, *A matched asymptotic expansions approach to continuity corrections for discretely sampled options. Part 1: Barrier options*, Appl. Math. Finance, 14 (2007), pp. 63–89.
- [15] M. KAC, *On distributions of certain Wiener functionals*, Trans. Amer. Math. Soc., 65 (1949), pp. 1–13.
- [16] I. KARATZAS AND S. E. SHREVE, *Brownian Motion and Stochastic Calculus*, Grad. Texts in Math. 113. Springer, New York, 1991.
- [17] T. MÜLLER-GRONBACH, *Strong Approximation of Systems of Stochastic Differential Equations*, Habilitation thesis, TU Darmstadt, 2002.
- [18] T. PRIMOZIC, *Estimating Expected First Passage Times Using Multilevel Monte Carlo Algorithm*, M.Sc. thesis, University of Oxford, 2011.