Multi-level Monte Carlo Methods for Option Pricing

Giorgio Milani
giorgio.milani@epfl.ch

Philipp Weder
philipp.weder@epfl.ch

Margot Coste-Sarguet
margot.coste-sarguet@epfl.ch

***Abstract*—This study investigates multi-level Monte Carlo (MLMC) techniques for pricing Asian and barrier options, comparing their effectiveness with traditional Monte Carlo (MC) methods. MLMC demonstrates significant variance reduction and enhanced computational efficiency for Asian options, achieving desired accuracy at lower costs. However, its advantages are less pronounced for barrier options due to their discontinuous payoff structure. Additionally, we examine importance sampling as a variance-reduction method for high strike prices, which corresponds to rare-event simulation. Our numerical results confirm that MLMC offers substantial benefits in computational efficiency, particularly for Lipschitz continuous payoff functions.**

I. Introduction

In this project, we examine the problem of option pricing, which entails determining the fair value of derivative financial instruments contingent on the behavior of their underlying assets. Options are financial contracts that confer upon the holder the right, but not the obligation, to buy or sell an asset at a pre-determined price, referred to as the strike price, on or before a specified expiration date. The valuation of an option is inherently driven by the stochastic behavior of the underlying asset, such as a stock, commodity, or index, as fluctuations in its price directly determine the option's payoff. For instance, the payoff of a call option increases with the extent to which the asset's price exceeds the strike price at maturity, while for a put option, it is contingent upon how far the asset's price falls below the strike price. These payoffs are further influenced by factors such as the volatility of the asset, prevailing interest rates, and the time to expiration, all of which underscore the complexity of the pricing process. As a result, option pricing requires estimating the expected value of the payoff at maturity, which represents the rational price a buyer would be willing to pay under current market conditions. This necessitates accurately modeling the underlying asset's price dynamics, often through stochastic differential equations, to ensure reliable and theoretically sound valuations.

Let S be the value of an asset. Then, the time evolution of S can be modeled using stochastic differential equations (SDEs). In this project, we consider the geometric Brownian motion model,

$$\begin{cases} dS(t) &= rS(t)dt + \sigma S(t)dW_t, t \in [0,1], \\ S(0) &= 1, \end{cases} \tag{1}$$

where dW_t is a standard Brownian motion. The parameters $r > 0$ and $\sigma > 0$ denote the *annualized risk free interest rate* and the *volatility of the asset* S, respectively. Note here that we assume the time of maturity to be 1 for simplicity.

We consider two different kinds of options: the payoff of the *Asian* option is given by:

$$Y^{(1)} = \exp(-r)\max(0, \bar{S} - K), \qquad \bar{S} \coloneqq \int_0^1 S(t)dt,$$

where $K > 0$ is referred to as the *strike price*. Therefore, the payoff of an Asian option is positive, whenever the average value of the asset between the initial time and maturity exceeds the strike price, and zero otherwise. The payoff is discounted by the risk-free interest rate to take into account the possibility of not buying the option. The *barrier call option* is defined by:

$$Y^{(2)} = \exp(-r)\max(0, S(1) - K)\mathbf{1}_{\{\max_{t\in[0,1]} S(t) < S_{\max}\}},$$

where K is again a pre-defined strike price and $S_{\max}$ is referred to as the barrier value. Hence, the payoff is positive, whenever the asset value at maturity exceeds the strike price contingent on the asset price never reaching the barrier value. The barrier option's payoff structure is thus discontinuous and more path-dependent than the Asian option's payoff. Consequently, we expect more difficulty in estimating the average payoff of the barrier option.

Although the geometric Brownian motion in eq. (1) can be solved explicitly, analytical solutions to SDEs are unavailable in more realistic cases. Therefore, we approximate its solution with the Euler-Maruyama (EM) scheme as follows. Let $M \in \mathbb{N}, h = M^{-1}$ and consider the discrete time points $t_m = mh, m = 1, \dots, M$. Then, we approximate $S(t_m)$ by S^m computed from

$$S^{m+1} = (1 + rh + \sigma\Delta W_m)S^m, \qquad m = 1, \dots, M, \tag{2}$$

where $\Delta W_m \overset{iid}{\sim} N(0,h), m = 1, \dots, M$. The numerical scheme allows for the stochastic sampling of the payoffs $Y^{(i)}, i = 1, 2$, which will depend on the time step size h. We denote the payoffs computed from a discrete approximation of S with time step size h by $Y_h^{(i)}, i = 1, 2$:

$$\begin{aligned} Y_h^{(1)} &= \exp(-r)\max(0, \bar{S}_h - K), \quad \bar{S}_h \coloneqq h\sum_{m=1}^{M}(S^m + S^{m-1})/2, \\ Y_h^{(2)} &= \exp(-r)\max(0, S^M - K)\mathbf{1}_{\{\max_{m\in[0,M]} S^m < S_{\max}\}}. \end{aligned} \tag{3}$$

For this simple form of the SDE and a Lipschitz continuous payoff function, one can show that [4]

$$\left|\mathbb{E}\left[Y_h^{(i)} - Y^{(i)}\right]\right| = \mathcal{O}(h).$$

Therefore, the numerical discretization of eq. (1) will introduce a bias term in the estimation of the average payoff that needs to be controlled via the time step size.

II. Standard Monte Carlo

To fix ideas, we first study the standard Monte Carlo estimator. For ease of exposition, we omit the superscripts $i = 1, 2$ in this section, i.e., we write Y for either $Y^{(1)}$ or $Y^{(2)}$ etc. To estimate the price of an option, we estimate the expected payoff

$$\mu = \mathbb{E}\left[Y\right].$$

The classical (crude) Monte Carlo estimator is given by

$$\hat{\mu}_h^{\mathrm{MC}} \coloneqq \frac{1}{N}\sum_{n=1}^{N} Y_h^{(n)},$$

where $Y_h^{(n)}$ are i.i.d. samples of the payoff obtained from the EM scheme with time step size h.

The classical Monte Carlo estimator is an unbiased estimator of

$$\mu_h \coloneqq \mathbb{E}\left[Y_h\right],$$

which is different from μ in general due to the discretization error in the approximation of the SDE in eq. (1). Therefore, this estimator is biased with respect to μ. Indeed,

$$\begin{aligned} b_h^{\mathrm{MC}} &\coloneqq \left|\mathbb{E}\left[\hat{\mu}_h^{\mathrm{MC}}\right] - \mu\right| = \left|\mu_h - \mu\right| \\ &= \left|\mathbb{E}\left[Y_h - Y\right]\right| = \mathcal{O}(h). \end{aligned}$$

The variance is given by

$$\mathbb{V}[\hat{\mu}_h^{\mathrm{MC}}] = \frac{\mathbb{V}[Y_h]}{N},$$

since $\hat{\mu}_h^{\mathrm{MC}}$ is a classical Monte Carlo estimator. The variance can be estimated by

$$\mathbb{V}[\hat{\mu}_h^{\mathrm{MC}}] \approx \frac{\hat{\sigma}_h^2}{N}, \qquad \hat{\sigma}_h^2 \coloneqq \frac{1}{N-1}\sum_{n=1}^{N}\left[Y_h^{(n)} - \hat{\mu}_h^{\mathrm{MC}}\right].$$

Whenever no confusion may arise, we write $\mathbb{V}[\hat{\mu}_h^{\mathrm{MC}}]$ instead of $\hat{\sigma}_h^2/N$. In particular, we have the bias-variance decomposition of the mean squared error (MSE):

$$\mathrm{MSE} = \mathbb{V}[\hat{\mu}_h^{\mathrm{MC}}] + (b_h^{\mathrm{MC}})^2. \tag{4}$$

III. Multi-level Monte Carlo

Let us introduce the multi-level Monte Carlo (MLMC) method. We again omit the superscripts $i = 1, 2$ for ease of notation. The idea of the multi-level Monte Carlo method is to use approximations Y_h of the target random variable Y at different discretization levels h to obtain an estimate of the mean μ with a lower variance at the same computational cost.

Denote by $h > 0$ an initial, typically large, time step size. Then, we write $h_\ell = 2^{-\ell}h$ for $\ell \in \mathbb{N}_0$. Hereafter, ℓ is referred to as the level of the approximation Y_{h_ℓ} of Y. Given a maximal level $L \in \mathbb{N}$, we can rewrite μ_{h_L} as

$$\mu_{h_L} = \mathbb{E}\left[Y_{h_L}\right] = \sum_{\ell=0}^{L}\mathbb{E}\left[Y_\ell - Y_{\ell-1}\right],$$

with the convention $Y_{h_{-1}} = 0$. Motivated by the above expression, the multi-level Monte Carlo estimator is defined as

$$\hat{\mu}_h^{(L)} \coloneqq \sum_{\ell=0}^{L}\hat{\mu}_\ell, \qquad \hat{\mu}_\ell \coloneqq \frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(Y_{h_\ell}^{(n,\ell)} - Y_{h_{\ell-1}}^{(n,\ell)}\right),$$

where $N_\ell \in \mathbb{N}$ denotes the number of samples at each level $\ell = 0, \dots, L$. We stress the dependence of the samples on the level ℓ. Indeed, the samples across the level estimators $\hat{\mu}_\ell, \ell = 0, \dots, L$ must be independent, while the payoffs $Y_{h_\ell}^{(n,\ell)}$ and $Y_{h_{\ell-1}}^{(n,\ell)}$ for a given level ℓ must be computed with the same underlying noise.

Before we discuss how the multi-level idea can help with reducing the variance, let us note the following properties of the multi-level estimator. The expected value of the multi-level estimator satisfies

$$\mathbb{E}\left[\hat{\mu}_h^{(L)}\right] = \sum_{\ell=0}^{L}\mathbb{E}\left[\hat{\mu}_\ell\right] = \sum_{\ell=0}^{L}\mathbb{E}\left[Y_{h_\ell} - Y_{h_\ell - 1}\right] \tag{5}$$

$$= \mathbb{E}\left[\sum_{\ell=0}^{L} Y_{h_\ell} - Y_{h_\ell - 1}\right] = \mathbb{E}\left[Y_{h_L}\right] = \mathbb{E}\left[\hat{\mu}_{h_L}^{\mathrm{MC}}\right], \tag{6}$$

using again the convention $Y_{h_{-1}} = 0$. Furthermore, we have the usual bias-variance decomposition:

$$\mathrm{MSE} = \mathbb{E}\left[(\hat{\mu}_h^{(L)} - \mu)^2\right] = \mathbb{V}[\hat{\mu}_h^{(L)}] + (b_h^{(L)})^2, \tag{7}$$

where $b_h^{(L)}$ denotes the bias of the multi-level estimator with initial time step size h and L levels,

$$b_h^{(L)} \coloneqq \left|\mathbb{E}\left[\hat{\mu}_h^{(L)}\right] - \mu\right| = \left|\mathbb{E}\left[\hat{\mu}_{h_L}^{\mathrm{MC}}\right] - \mu\right| = b_{h_L}^{\mathrm{MC}},$$

due to eq. (5). Moreover, we define $V_\ell \coloneqq \mathbb{V}[Y_{h_\ell} - Y_{h_{\ell-1}}]$ for $\ell \in \mathbb{N}_0$. Since each level is estimated independently from the others, we thus have

$$\mathbb{V}[\hat{\mu}_h^{(L)}] = \sum_{\ell=0}^{L}\frac{V_\ell}{N_\ell}. \tag{8}$$

Therefore, the bias-variance decomposition can be expressed as

$$\mathrm{MSE} = \sum_{\ell=0}^{L}\frac{V_\ell}{N_\ell} + (b_{h_L}^{\mathrm{MC}})^2. \tag{9}$$

Finally, let C_ℓ denote the computational cost of computing a single sample Y_{h_ℓ}. The total cost of the MLMC estimator is then given by:

$$C_{\mathrm{MLMC}}^{(L)} = \sum_{\ell=0}^{L} C_\ell N_\ell. \tag{10}$$

It is clear that the computational cost will increase with the level of refinement ℓ. In the case of option pricing with the EM scheme, we approximately have $C_\ell = 2^\ell C_0$. The intuition is now that if the bias $b_{h_L}^{\mathrm{MC}}$ and the level variances $V_\ell, \ell = 0, \dots, L$ decrease rapidly enough with respect to the growth rate of the computational cost, then an estimate with the same MSE as a crude Monte Carlo estimate can be reached at a lower computational cost. This observation is formalized in the following theorem due to Giles [3] and generalized in [1]. We rephrase it in our notation for option pricing.

Theorem 1. *Let Y be a random variable representing the payoff of an option with numerical approximation Y_{h_ℓ} with a time step $h_\ell = 2^\ell h, \ell = 0,\dots,L$. Suppose there exist independent estimators $\hat{\mu}_\ell$ based on N_ℓ Monte Carlo samples, each with expected cost C_ℓ and variance V_ℓ, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$, such that $\alpha \geq \frac{1}{2}\min(\beta,\gamma)$, and*

(i) $b^{\mathrm{MC}}_{h_\ell} \leq c_1 2^{-\alpha\ell}$,

(ii) $\mathbb{E}[\hat{\mu}_\ell] = \begin{cases} \mathbb{E}[Y_h], & \ell = 0, \\ \mathbb{E}[Y_{h_\ell} - Y_{h_{\ell-1}}], & \ell > 0, \end{cases}$

(ii) $V_\ell \leq c_2 2^{-\beta\ell}$,

(iv) $C_\ell \leq c_3 2^{\gamma\ell}$.

Then, there exists a constant $c_4 > 0$ such that for any $\varepsilon < e^{-1}$ there are integers L and N_ℓ, for which the MSE of the MLMC estimator is bounded by ε^2, while the expected computational cost is bounded by

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

In the following subsections, we will investigate the properties of the two-level estimator in more detail.

A. Variance minimization of the two-level estimator

In this section, we sketch a proof for the optimality of allocating the number of samples N_0 and N_1 according to

$$\frac{N_1}{N_0} = \frac{\sqrt{V_1/C_1}}{\sqrt{V_0/C_0}}, \tag{11}$$

to minimize the variance eq. (8) at a fixed cost $C^{(2)}_{\mathrm{MLMC}}$.

If we treat N_0 and N_1 as real numbers, we can formulate the problem as a continuous minimization problem under constraints. Write $\boldsymbol{N} \coloneqq [N_0, N_1]^\top \in \mathbb{R}^2_{\geq 0}$ and let $C > 0$ be the fixed computational cost. Then, define the variance function $V : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{\geq 0}$ by

$$V(\boldsymbol{N}) \coloneqq \frac{V_0}{N_0} + \frac{V_1}{N_1}.$$

The constraint is given by

$$g(\boldsymbol{N}) \coloneqq N_0 C_0 + N_1 C_1 = C.$$

The minimization problem reads:

$$\begin{cases} \text{Find } \boldsymbol{N}^\star = \arg\min_{\boldsymbol{N}\in\mathbb{R}^2_{\geq 0}} V(\boldsymbol{N}), \text{ subject to} \\ g(\boldsymbol{N}^\star) = 0. \end{cases}$$

The corresponding Lagrangian function $L : \mathbb{R}^2_{\geq 0} \times \mathbb{R} \to \mathbb{R}$ is defined by

$$L(\boldsymbol{N}, \lambda) \coloneqq V(\boldsymbol{N}) + \lambda g(\boldsymbol{N}).$$

Omitting the constraint $\boldsymbol{N} \in \mathbb{R}^2_{\geq 0}$, the optimality conditions for an optimal point $(\boldsymbol{N}^\star, \lambda^\star)$ read

$$\frac{\partial L}{N_0}(\boldsymbol{N}^\star, \lambda^\star) = -\frac{V_0}{(N_0^\star)^2} + \lambda^\star C_0 = 0\,, \tag{12}$$

$$\frac{\partial L}{N_1}(\boldsymbol{N}^\star, \lambda^\star) = -\frac{V_1}{(N_1^\star)^2} + \lambda^\star C_1 = 0\,, \tag{13}$$

$$\frac{\partial L}{\partial\lambda}(\boldsymbol{N}^\star, \lambda^\star) = N_0 C_0 + N_1 C_1 - C = 0. \tag{14}$$

From eq. (12) and eq. (13), we find that

$$\lambda^\star = \frac{V_0}{N_0^2 C_0} = \frac{V_1}{N_1^2 C_1}. \tag{15}$$

Rearranging the latter equation then yields eq. (11). A rigorous proof for continuous N_ℓ would require writing the complete Karush-Kuhn-Tucker (KKT) conditions and verifying that the constraints are qualified.

B. Cost minimization of the multi-level estimator

In this section, we sketch the proof for the quasi-optimality of allocating the samples per level $N_0, N_1, \dots, N_L$ according to

$$N_\ell = \varepsilon^{-2}\left(\sum_{\ell'=0}^{L}\sqrt{V'_\ell C'_\ell}\right)\sqrt{\frac{V_\ell}{C_\ell}}, \tag{16}$$

with respect to the computational cost for a fixed variance ε^2. We treat the N_ℓ again as natural numbers for simplicity. The cost function $C(\boldsymbol{N}) : \mathbb{R}^{L+1}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is given by

$$C(\boldsymbol{N}) \coloneqq \sum_{\ell=0}^{L} N_\ell C_\ell.$$

The constraint is given by

$$h(\boldsymbol{N}) \coloneqq V(\boldsymbol{N}) - \varepsilon^2 = \sum_{\ell=0}^{L}\frac{V_\ell}{N_\ell} - \varepsilon^2 = 0.$$

The minimization problem reads:

$$\begin{cases} \text{Find } \boldsymbol{N}^\star = \arg\min_{\boldsymbol{N}\in\mathbb{R}^{L+1}_{\geq 0}} C(\boldsymbol{N}), \text{ subject to} \\ h(\boldsymbol{N}^\star) = 0. \end{cases}$$

The corresponding Lagrangian function $L : \mathbb{R}^{L+1}_{\geq 0} \times \mathbb{R} \to \mathbb{R}$ is defined by

$$L(\boldsymbol{N}, \lambda) \coloneqq C(\boldsymbol{N}) + \lambda h(\boldsymbol{N}).$$

In this case, the simplified optimality conditions are given by

$$\frac{\partial L}{\partial N_\ell}(\boldsymbol{N}^\star, \lambda^\star) = C_\ell - \lambda^\star\frac{V_\ell}{N_\ell^2} = 0, \ell = 0, \dots, L, \tag{17}$$

$$\frac{\partial L}{\partial\lambda}(\boldsymbol{N}^\star, \lambda^\star) = \sum_{\ell=0}^{L}\frac{V_\ell}{N_\ell} - \varepsilon^2 = 0. \tag{18}$$

From eq. (17) , we deduce that

$$N_\ell^\star = \sqrt{-\lambda^\star}\sqrt{\frac{V_\ell}{C_\ell}}. \tag{19}$$

Inserting this expression for $N_\ell^\star$ into the constrain, we find

$$\varepsilon^2 = \sum_{\ell=0}^{L} V_\ell (N_\ell^\star)^{-1} = \sqrt{-\lambda^\star} \sum_{\ell=0}^{L} \sqrt{C_\ell V_\ell}.$$

This implies that

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

which justifies eq. (19), and in particular eq. (16). For a rigorous proof for continuous N_ℓ, one would again have to verify the full set of KKT conditions.

IV. Asian option

We begin by investigating the Asian option. First, we examine the convergence behavior of the EM discretization and an adaptive strategy for the standard estimator. Next, we analyze the variance reduction achieved by the two-level estimator. Finally, we apply the multi-level estimator and compare it to the standard one.

A. The standard estimator

We start our experiments by verifying the convergence rates predicted by the theory.

1) Convergence behavior: We first vary the number of samples N for a fixed time step size h. We expect the convergence rates

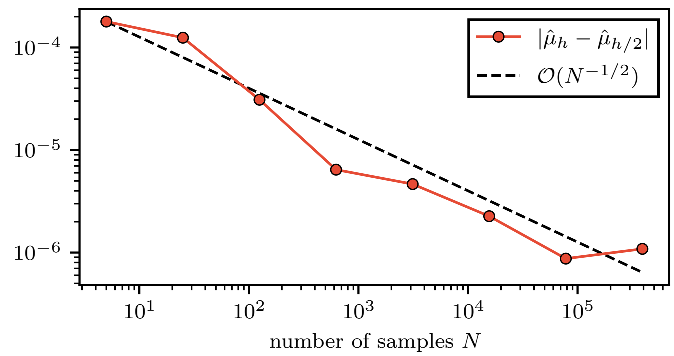
$$|\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}}| = \mathcal{O}(N^{-1/2}),$$
$$\mathbb{V}[\hat{\mu}_h^{\mathrm{MC}}] = \mathcal{O}(N^{-1}),$$
$$\mathbb{V}[\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}}] = \mathcal{O}(N^{-1}).$$

The convergence rate for the difference is due to the fact that $\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}}$ is an unbiased estimator of $\mathbb{E}\left[Y_h^{(1)} - Y_{h/2}^{(1)}\right]$. The convergence rate for the variances is due to the fact that $\hat{\mu}_h^{\mathrm{MC}}$ and $\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}}$ are standard Monte Carlo estimator, respectively. Figure 1 shows the results for the experiments with varying numbers of samples N and a fixed time step size of $h = 10^{-3}$. In fig. 1a, we observe the convergence rate $\mathcal{O}(N^{-1/2})$ for the bias as expected. Figure 1b shows the expected convergence rates for the variances of $\hat{\mu}_h^{\mathrm{MC}}$ and $\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}}$, respectively. In particular, we note that the variance of the latter is orders of magnitude smaller than the one of the former. This indicates that the multi-level strategy should work well for this example.

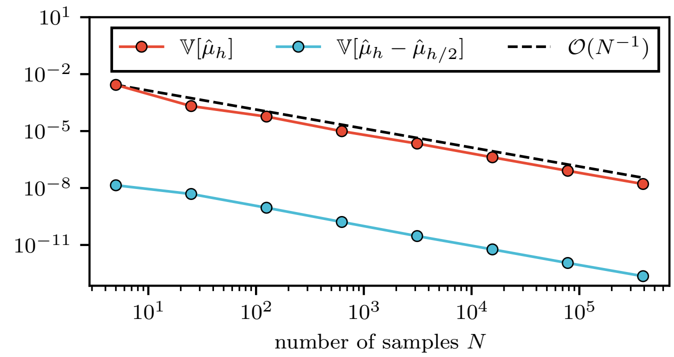
In a second experiment, we vary the time step size h. As mentioned in section I, we expect the convergence rate

$$|\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}}| = \mathcal{O}(h),$$

due to the strong convergence property of the Euler-Maruyama scheme. Figure 2 shows the results for the experiments with varying time step size h and constant number of samples $N = 10'000$. We observe the expected convergence rate for the difference between the discretization with h and $h/2$ in fig. 2a. Figure 2b shows, moreover, that we may assume the variance of Y_h to be approximately independent of the time step size.



(a) bias



(b) variance

Figure 1: Standard estimator: convergence rate of $\mathcal{O}(N^{-1/2})$ for the bias (a) and convergence rate of $\mathcal{O}(N^{-1})$ for the variance (b) for varying numbers of samples N.

2) An adaptive strategy: We now propose an adaptive algorithm to achieve a prescribed MSE based on the above findings. The idea is to use the bias-variance decomposition in eq. (4) and iteratively either increase the number of samples N or decrease the time step size h until both terms in the sum are bounded by $\varepsilon^2/2$. Thereby, the variance has to be estimated by $\hat{\sigma}_h^2/N$, while the bias is estimated by $b_h^{\mathrm{MC}} \approx \hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}}$. The adaptive strategy is summarized in algorithm 1.

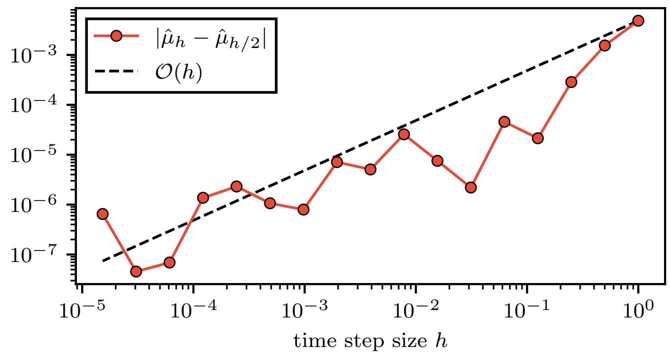
B. The two-level estimator

In this section, we investigate the variance reduction achieved by using just two levels, with a coarse time step $h_0 > 0$ and a fine one $h_1 = h_0/2$. In section III-A, we have established an optimal allocation formula in eq. (11) for the numbers of samples N_0 and N_1 to minimize the variance. Under the assumption that a sample on the fine level is twice as expensive than a sample on the coarse one, i.e. $C_1 = 2C_0$, the latter simplifies to

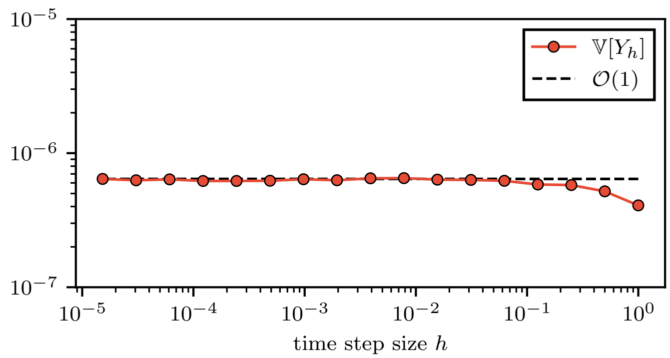
$$\frac{N_1}{N_0} = \sqrt{\frac{V_1}{2V_0}}.$$

The total cost in eq. (10) of a two-level estimate then simplifies to

$$C_{\mathrm{MLMC}}^{(2)} = C_0 N_0 \left(1 + 2\sqrt{\frac{V_1}{2V_0}}\right).$$



(a) bias



(b) variance

Figure 2: Standard estimator: Convergence rate of $\mathcal{O}(h)$ for the bias (a) and the asymptotically constant variance (b) for varying time step sizes h.

Algorithm 1 Adaptive Monte Carlo

Given: tolerance ε^2, number of initial samples $N_0 \in \mathbb{N}$, initial time step size $h_0 > 0$.
Initialize $N \leftarrow N_0$
Initialize $h \leftarrow h_0$
Generate N replicas $(Y_h^{(1)}, \dots, Y_h^{(N)})$ and compute $\hat{\mu}_h^{\mathrm{MC}}$ and $\hat{\sigma}_h^2$
Generate N replicas $(Y_{h/2}^{(1)}, \dots, Y_{h/2}^{(N)})$ and compute $\hat{\mu}_{h/2}^{\mathrm{MC}}$
Set MSE $\leftarrow \frac{\hat{\sigma}_h^2}{N} + (\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}})^2$
while MSE $> \varepsilon^2$ **do**
  if $\hat{\sigma}_h^2/N > \varepsilon^2/2$ **then**
    $N \leftarrow 2N$
  end if
  if $(\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}})^2 > \varepsilon^2/2$ **then**
    $h \leftarrow h/2$
  end if
  Generate N replicas $(Y_h^{(1)}, \dots, Y_h^{(N)})$ and compute $\hat{\mu}_h^{\mathrm{MC}}$ and $\hat{\sigma}_h^2$
  Generate N replicas $(Y_{h/2}^{(1)}, \dots, Y_{h/2}^{(N)})$ and compute $\hat{\mu}_{h/2}^{\mathrm{MC}}$
  Set MSE $\leftarrow \frac{\hat{\sigma}_h^2}{N} + (\hat{\mu}_h^{\mathrm{MC}} - \hat{\mu}_{h/2}^{\mathrm{MC}})^2$
end while
return $\hat{\mu}_h^{\mathrm{MC}}$

Consider now the crude Monte Carlo estimator with the fine time step h_1 and denote its number of samples and its cost by N_{MC} and C_{MC}, respectively. For C_{MC} to be comparable to $C_{\mathrm{MLMC}}^{(2)}$, we require

$$C_{\mathrm{MC}} = 2C_0 N_{\mathrm{MC}} \simeq C_0 N_0 \left(1 + 2\sqrt{\frac{V_1}{2V_0}}\right),$$

or equivalently

$$N_{\mathrm{MC}} \simeq N_0 \left(\frac{1}{2} + \sqrt{\frac{V_1}{2V_0}}\right).$$

Under the above assumptions, we have

$$\mathbb{V}[\hat{\mu}_h^{(2)}] = \frac{V_0}{N_0} + \frac{V_1}{N_1} = \frac{1}{N_0}(V_0 + \sqrt{2V_0 V_1}). \tag{20}$$

We write $V_1^{\mathrm{MC}} = \mathbb{V}[Y_{h/2}^{(1)}]$ and we assume that $V_1^{\mathrm{MC}} \simeq V_0$, which is justified by the assumption that the variance of $Y_h^{(1)}$ is approximately independent of h; see fig. 2b. Furthermore, we have observed in section IV-A that $V_1 \ll V_0$. Hence, we find

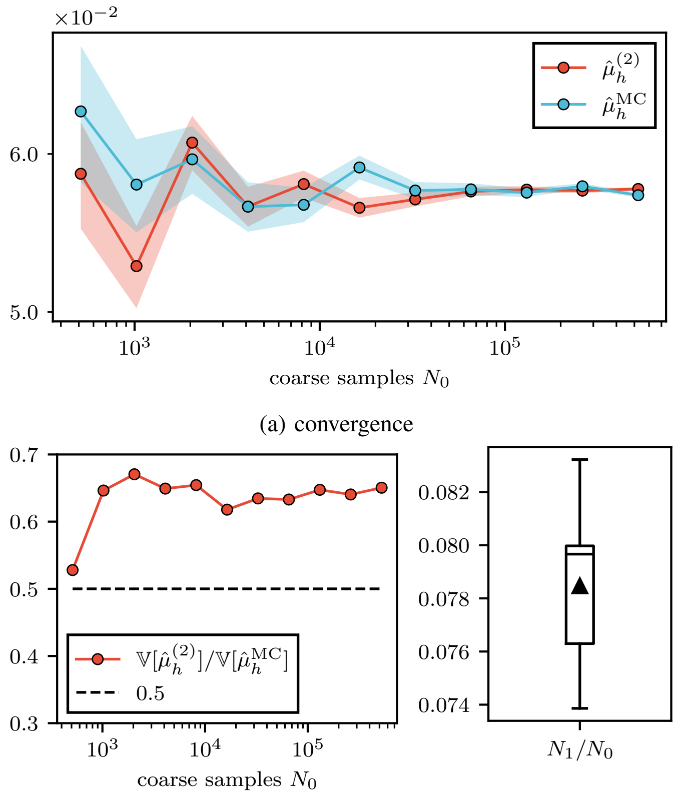
$$\frac{\mathbb{V}[\hat{\mu}_h^{\mathrm{MC}}]}{\mathbb{V}[\hat{\mu}_h^{(2)}]} \simeq \frac{1}{2} + \sqrt{\frac{2V_1}{V_0}\frac{V_1}{V_0}} \simeq \frac{1}{2}.$$

Therefore, we can hope for a variance reduction of approximately 50%.

In our numerical experiments, we now analyze the variance reduction achieved by the two-level estimator for varying numbers of coarse samples N_0. The optimal ratio N_1/N_0 is estimated in every run from a pilot run with 5'000 samples. Figure 3a shows that the standard estimator and the two-level estimator converge to the same mean value. The standard deviations are represented by the shaded areas. At least for the first values of N_0, one can see the effect of the variance reduction. For $N_0 = 524'288$ coarse samples, the two-level estimator yields a variance of $1.303 \cdot 10^{-8}$, while the crude estimator at the same cost yields a variance of $2.003 \cdot 10^{-8}$. In fig. 3b the variance ratios between the two-level and crude estimators are plotted for different values of N_0. We see that the variance reduction quickly reaches a plateau at around 0.63. Hence, the two-level estimator does not reach the lower bound derived in eq. (20). Figure 3c displays a boxplot for all the optimal ratios computed from the different pilot runs for each value of N_0. The mean value of the optimal ratio is 0.075 with a standard deviation of 0.0027. This showcases the effectiveness of the multi-level strategy, since only very few samples at the fine level are needed. Therefore, we may hope to achieve even better performance for a given mean squared error when using more than two levels.

C. The multi-level estimator

We continue by studying the computational efficiency of the MLMC algorithm compared to the classical Monte Carlo estimator. Based on theorem 1, whose assumptions we assume to hold with $\alpha = \beta = \gamma = 1$, we derive a strategy to optimally assign the number of levels L and the number of samples per



(a) convergence

(b) variance reduction

(c) optimal ratio

Figure 3: Two-level estimator: convergence of the standard and two-level estimators to the same mean value (a), the variance reduction (b) and the distribution of optimal ratios computed from a pilot run (c). The shaded areas in (a) show the standard deviations of the corresponding estimator.

level. Moreover, we estimate the number of samples needed for a crude Monte Carlo estimator of comparable accuracy and a way to compare the computational costs of both approaches.

1) Choice of L and N_l: Similar to the adaptive strategy in section IV-A2, we aim to bound each term in the bias-variance decomposition in eq. (9) by ε^2, to achieve $\mathrm{MSE}(\hat{\mu}_L^{(h)}) < 2\varepsilon^2$. Furthermore, we make the assumption that

$$E_\ell \simeq \tilde{E}_0 2^{-\ell}, \text{ and } V_\ell \simeq \tilde{V}_0 2^{-\ell},$$

where $E_\ell \coloneqq |\mathbb{E}\left[Y_{h_\ell} - Y_{h_{\ell-1}}\right]|$, for $\ell \in \mathbb{N}$, and the level variance V_ℓ is defined as in section III.

Note that the bias term in eq. (9) only depends on the highest level L. Hence, under the assumptions above, choosing L large enough is sufficient. Indeed, using eq. (5) and a telescopic sum, we obtain for $K \in \mathbb{N}$ that

$$\begin{aligned}
&\left|\mathbb{E}\left[\mu - \hat{\mu}_h^{(L)}\right]\right| = \left|\mathbb{E}\left[\mu - \hat{\mu}_{h_L}^{\mathrm{MC}}\right]\right| \\
&= \left|\sum_{k=0}^{K-1} \mathbb{E}\left[\hat{\mu}_{h_{L+k+1}}^{\mathrm{MC}} - \hat{\mu}_{h_{L+k}}^{\mathrm{MC}}\right] + \mathbb{E}\left[\mu - \hat{\mu}_{h_{L+K}}^{\mathrm{MC}}\right]\right| \\
&\le \sum_{k=1}^{K} E_{L+k} + \left|\mathbb{E}\left[\mu - \hat{\mu}_{h_{L+K}}^{\mathrm{MC}}\right]\right|.
\end{aligned}$$

Using the assumption $E_\ell \le \tilde{E}_0 2^{-\alpha\ell}$, with $\alpha = 1$, assumption (i) from theorem 1, and a geometric series in the first term, we obtain

$$\left|\mathbb{E}\left[\mu - \hat{\mu}_h^{(L)}\right]\right| \le \tilde{E}_0 \frac{2^{-\alpha L}}{1 - 2^{-\alpha}} + c_1 2^{-\alpha(L+K)} \to \tilde{E}_0 \frac{2^{-\alpha L}}{1 - 2^{-\alpha}},$$

for $K \to \infty$. Since $K \in \mathbb{N}$ was arbitrary, we can set

$$L = \left\lceil \frac{1}{\alpha} \log_2 \left(\frac{\tilde{E}_0}{\varepsilon(1 - 2^{-\alpha})} \right) \right\rceil, \tag{21}$$

to achieve $(b_h^{(L)})^2 < \varepsilon^2$.

For the variance, we use eq. (16) which simplifies under the assumptions $C_\ell = C_0 2^\ell$ and $V_\ell = \tilde{V}_0 2^{-\ell}$ to

$$N_\ell = \left\lceil \varepsilon^{-2}(L+1)\tilde{V}_0 2^{-\ell} \right\rceil. \tag{22}$$

Indeed, this choice of N_ℓ guarantees that

$$\mathbb{V}[\hat{\mu}_h^{(L)}] = \sum_{\ell=0}^{L} \frac{V_\ell}{N_\ell} < \varepsilon^2,$$

again using a geometric series. In numerical computations, the constants $\tilde{E}_0$ and $\tilde{V}_0$ need to be estimated from a pilot run with few levels $L_p \in \mathbb{N}$.

To obtain a crude Monte Carlo estimate with an MSE smaller than ε^2, we can determine the step size h by using the fact that we have chosen L such that $\left|\mathbb{E}\left[\hat{\mu}_{h_L}^{\mathrm{MC}} - \mu\right]\right| < \varepsilon^2$. Hence, it is enough to bound again the variance term by ε^2 in eq. (4). Since can assume that $\mathbb{V}[Y_{h_\ell}]$ is approximately constant with respect to ℓ, we use the sample estimate of the variance $\hat{\sigma}_{h_{L_p}}^2$ of the finest level in the pilot run and set

$$N_{\mathrm{MC}} = \left\lceil \frac{\hat{\sigma}_{h_{L_p}}^2}{\varepsilon^2} \right\rceil.$$

Regarding the cost comparison, we proceed in the spirit of [3]: under the assumption that the cost scales like $C_\ell = C_0 2^\ell$, we can evaluate the *normalized costs* defined by

$$C_{\mathrm{MLMC}}^{(L)}/C_0 = N_0 + \sum_{\ell=1}^{L} N_\ell (2^\ell + 2^{\ell-1}),$$

and

$$C_{\mathrm{MC}}/C_0 = N_{\mathrm{MC}} 2^L,$$

for the MLMC and crude estimator, respectively. We clearly state that this measure of computational complexity does not take into account any overhead in the implementation of the multi-level method. Therefore, the raw CPU timings need to be analyzed too. While theorem 1 predicts a cost scaling for the MLMC estimator, we note that due to eq. (21), $2^L \simeq \varepsilon^{-1/\alpha}$, and therefore

$$C_M C/C_0 = \mathcal{O}(\varepsilon^{-(2+1/\alpha)}). \tag{23}$$

2) Results: The results of our experiments with the MLMC estimator for the Asian option are reported in fig. 4. Figure 4a shows our fit of the decay for E_ℓ and V_ℓ from a pilot run with 50'000 samples and $L_p = 6$ levels. Although we clearly observe an exponential decay for both quantities, we point out that the estimate for $\tilde{E}_0$ is subject to large uncertainty. In addition, both theoretical convergence rates α and β appear to be faster than the prescribed value 1. The estimates we obtain are $\tilde{E}_0 = 2.418 \cdot 10^{-4}$ and $\tilde{V}_0 = 7.292 \cdot 10^{-5}$, which are used in the following.

Figure 4b shows the distribution of the number of samples N_ℓ for different target accuracies ε. As expected, lower values of ε lead to more levels and more samples per level. However, the number of levels per sample decreases at the same rate across different target accuracies. Moreover, the bulk of the samples is assigned to the lowest levels.

The normalized costs for the MLMC and the crude estimators are plotted in fig. 4c. For the classical estimator, we retrieve a cost that scales as $\mathcal{O}(\varepsilon^{-3})$ as predicted by the theory, cf. eq. (23). The normalized cost of the multi-level estimator deviates from the predicted rate of $\mathcal{O}(\varepsilon^{-2}|\log(\varepsilon)|^2)$ at higher values of the target precision ε. This can be understood by considering our estimate for $\tilde{E}_0 \approx 2 \cdot 10^{-4}$. When ε approaches the same order of magnitude as $\tilde{E}_0$ the number of levels L tends to vanish, at which point the MLMC approach becomes ineffective. We should also note that we assumed the decay rates of E_ℓ and V_ℓ to be $\alpha = \beta = 1$. As previously discussed this introduces uncertainties in our estimates of $\tilde{E}_0$ and $\tilde{V}_0$, which results in a suboptimal assignment of the samples. Nevertheless, the normalized cost of the MLMC estimator is approximately two orders of magnitude lower than the one of the classical estimator.

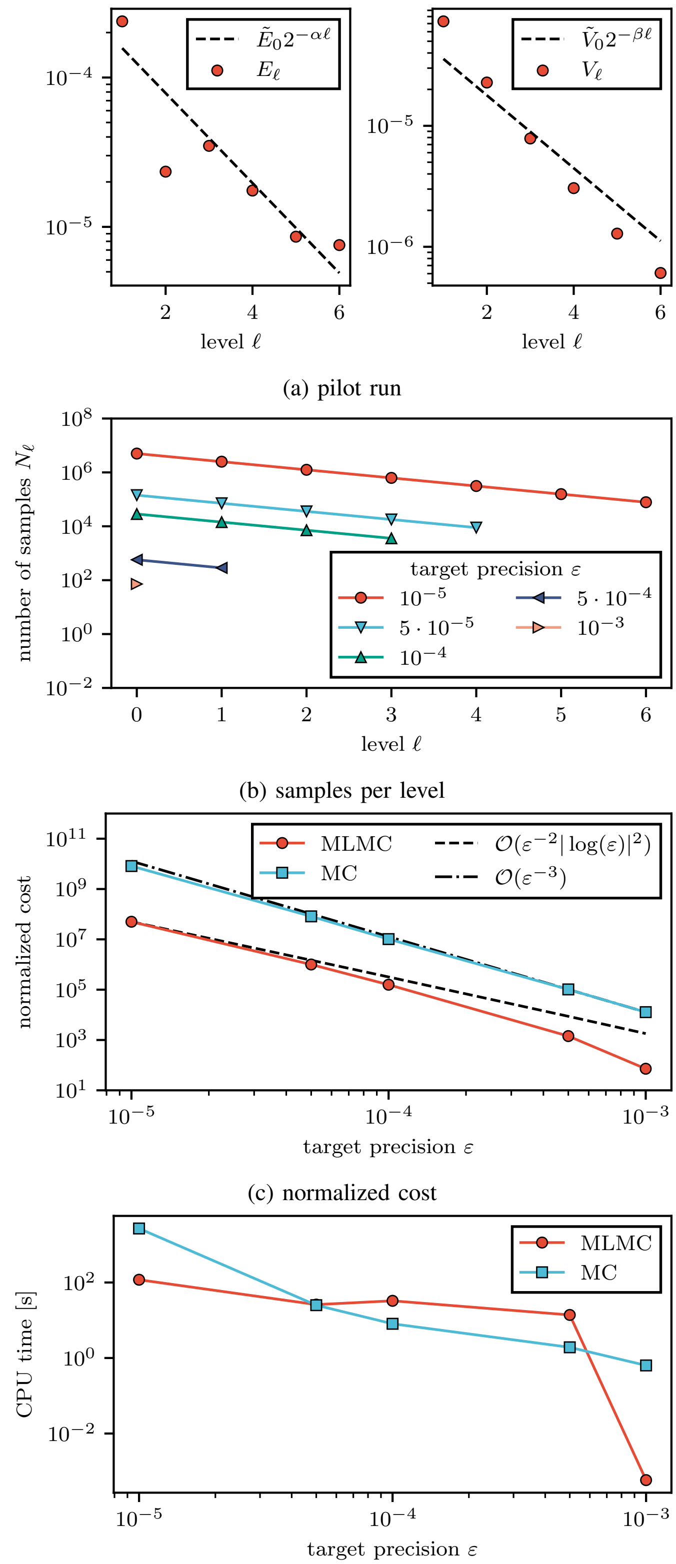
Figure 4d shows the CPU runtimes for the classical and the multi-level estimators for varying levels of accuracy. In contrast to the normalized costs, the classical estimator outperforms the MLMC estimator for the higher values of ε. This can be explained by the multi-level strategy introducing computational overhead in the implementation. For the highest value of ε, the MLMC cost suddenly drops. This is due to the approach nearing the end of its range of validity, with the number of levels approaching zero, as explained above.

V. Barrier Call Option

Figure 5 displays the results of the MLMC estimator applied to the barrier option. In the first part, table I presents the outcomes from a pilot run involving 50,000 samples with eight levels ($L = 8$). Notably, the estimate for the decay of biases appears more stable than the results for the Asian option.

Figure 5b provides further analysis, showing the distribution of samples across different levels for various target accuracies ε. Similar to the Asian option, we observe consistent patterns in this distribution. However, we point out that in this case, since $\beta < \gamma = 1$, we allocate the N_ℓ according to

$$N_\ell = \left\lceil \varepsilon^{-2} \tilde{V}_0 2^{(\gamma-\beta)L/2} \left(1 - 2^{-(\gamma-\beta)/2}\right)^{-1} 2^{-(\beta+\gamma)\ell/2} \right\rceil,$$



(a) pilot run

(b) samples per level

(c) normalized cost

(d) CPU runtime

Figure 4: Multi-level (MLMC) vs. crude (MC) estimator for the Asian option: fitting of the biases and variances (a) to estimate $\tilde{E}_0$ and $\tilde{V}_0$, samples per level (b), normalized computational costs (c), and CPU runtimes (d) for different target accuracies ε, and time step $h = 0.01$.

Table I: MLMC pilot run for the barrier option: estimated parameter values.

parameter	value	std. deviation
$\tilde{E}_0$	3.71×10^{-3}	2.21×10^{-4}
α	4.29×10^{-1}	3.28×10^{-2}
$\tilde{V}_0$	1.48×10^{-3}	1.18×10^{-4}
β	3.86×10^{-1}	4.11×10^{-2}

as suggested by the proof of the complexity theorem in [1][Appendix A].

Figure 5c presents the normalized computational costs for both the MLMC and crude estimators at different accuracy levels. The normalized MC cost scales as $\mathcal{O}(\varepsilon^{-(2+1/\alpha)})$. The MLMC trend observed aligns with the predictions made in theorem 1. It is important to note that since $\beta < \gamma = 1$, we expect less reduction in computational cost, which our numerical results reflect. Furthermore, the normalized costs of both methods are growing faster for decreasing values of ε in comparison to the Asian option, due to the lower weak convergence rate α.

Finally, fig. 5d reports the CPU timings for the MLMC and crude Monte Carlo estimators. As with the findings for the Asian option, we again notice the overhead associated with the multi-level approach.

VI. Higher strike price and a variance reduction technique

High strike prices pose a challenge in option pricing because, when simulating the price paths of the underlying asset, only a limited number of them are likely to reach the strike price. In other words, high strike prices are associated with rare events, resulting in high variance. To address this issue, importance sampling can be used.

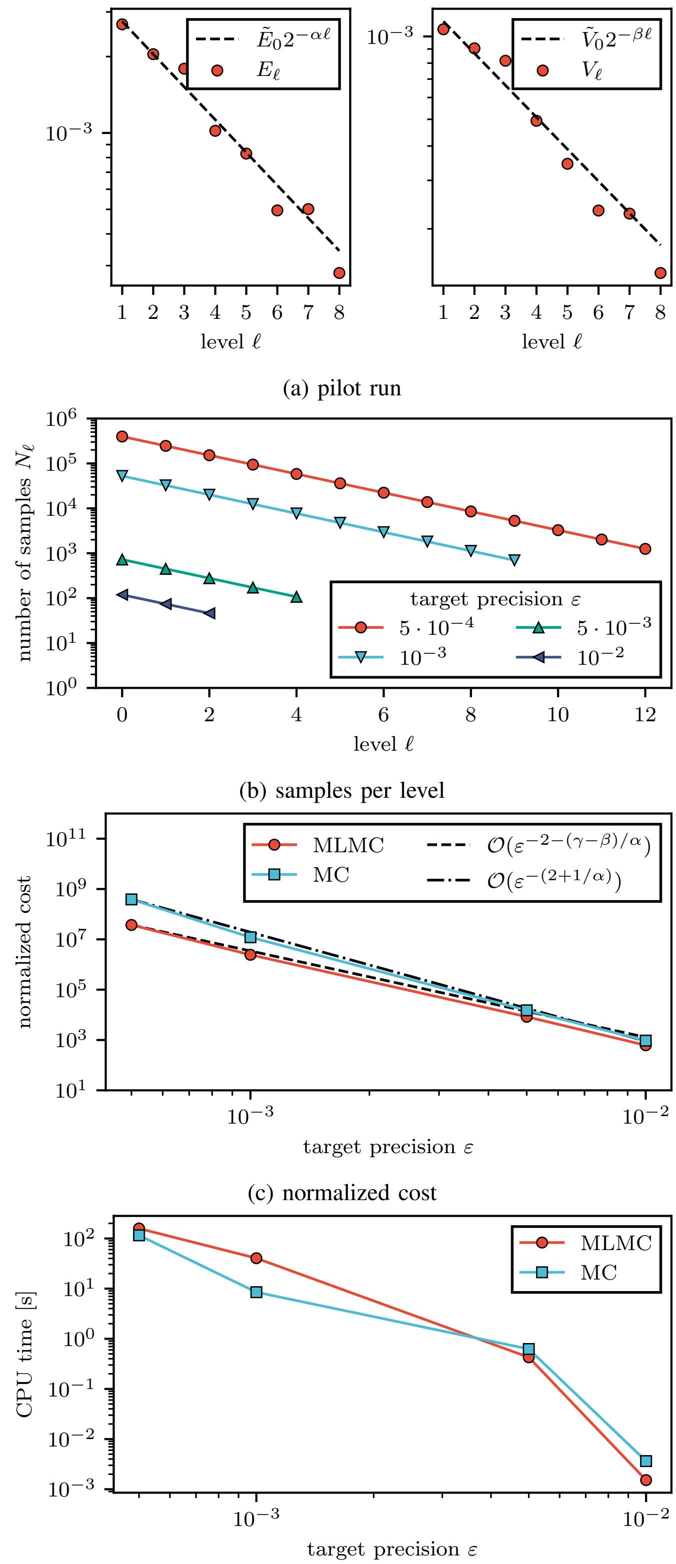
1) Importance sampling (IS) for option pricing: Importance sampling requires a dominant distribution. When dealing with discretized stochastic differential equations, a dominant distribution can be constructed by shifting the underlying Gaussian increments ΔW_m of the process (see in [5]). Indeed replacing ΔW_m by $\Delta \tilde{W}_m \overset{iid}{\sim} \mathcal{N}(\sigma^{-1}(\tilde{r}-r)h, h)$ in eq. (2) corresponds to modifying the risk-free interest rate r to $\tilde{r}$ in the SDE. We chose this method because the entire path is required to compute the payoffs, not just the value at maturity. We set $\tilde{r} = 10 \cdot r$ to increase the probability of paths hitting the strike price. The likelihood ratio w for each path $\Delta \tilde{W}^{(i)}_{0:M}$ can then be expressed as:

$$w(\Delta \tilde{W}^{(i)}_{0:M}) = \exp\left(\frac{1}{2} M h \phi^2 - \phi \sum_{m=1}^{M} \Delta \tilde{W}^{(i)}_m\right),$$

with $\phi = \sigma^{-1}(\tilde{r} - r)$. The estimator eventually reads:

$$\hat{\mu}_{\mathrm{IS}} \coloneqq \frac{1}{N} \sum_{i=1}^{N} \psi(\Delta \tilde{W}^{(i)}_{0:M}) w(\Delta \tilde{W}^{(i)}_{0:M}),$$

where $\psi(\Delta \tilde{W}^{(i)}_{0:M})$ is the approximated payoff (for Asian or barrier options) in terms of the discretized path, as given in eq. (3).



(a) pilot run

(b) samples per level

(c) normalized cost

(d) CPU runtime

Figure 5: Multi-level (MLMC) vs. crude (MC) estimator for the barrier option: fitting the biases and variances (a) to estimate $\tilde{E}_0$ and α, and $\tilde{V}_0$ and β, respectively; samples per level (b); normalized computational costs (c); and CPU runtimes (d) for different target accuracies ε.

Table II: Payoff estimators with strike price $K = 2$ and $S_{max} = 2.5$. The crude Monte Carlo (MC) is compared to importance sampling (IS), for $N = 10^5$ samples and time step $h = 0.001$ for the Asian (a) and barrier (b) call options .

(a) Asian option

	MC	IS
$\hat{\mu}$	0	$5.153 \cdot 10^{-10}$
$\mathbb{V}[\hat{\mu}]$	0	$1.746 \cdot 10^{-20}$

(b) barrier call option

	MC	IS
$\hat{\mu}$	$4.832 \cdot 10^{-5}$	$4.362 \cdot 10^{-5}$
$\mathbb{V}[\hat{\mu}]$	$7.137 \cdot 10^{-11}$	$1.540 \cdot 10^{-13}$

2) Results: Table II shows the payoff estimators for the Asian and barrier call options, comparing the expected payoffs and variances between the crude Monte Carlo (MC) approach and the importance sampling strategy. As expected, the variance is significantly reduced for the latter, by two orders of magnitude compared to MC for the barrier call option. This shows that even without optimizing the importance sampling distribution, our choice of $\tilde{r}$ was effective enough to drastically reduce the variance. In the Asian option case, no path had a mean value higher than the strike price, even with a substantial number of samples. As a result, the crude Monte Carlo estimator yields a zero mean and variance as shown in table IIa. Using importance sampling meaningful results are obtained, further highlighting the strength of this technique for simulating rare events.

3) IS in MLMC algorithms: Implementing importance sampling in an MLMC algorithm should be feasible. It could be implemented at each level by shifting the underlying Brownian motion as described above while being careful to use the same noise again for both the fine and coarse estimators. Another simple option would be to apply importance sampling to the first level only since its variance is orders of magnitude higher than the one of the consecutive levels. Finding the optimal modified interest rate $\tilde{r}$ remains an open question.

VII. Conclusions

This study illustrates the efficacy of multi-level Monte Carlo techniques in the pricing of options, particularly Asian and barrier options. MLMC demonstrates considerable computing efficiency and variance reduction relative to the conventional Monte Carlo approach. In the case of the Asian option, MLMC demonstrates exponential decay rates for bias and variance, resulting in a reduction of computational expenses by roughly two orders of magnitude compared to MC. The efficacy of MLMC for barrier options is diminished due to the discontinuous and path-dependent characteristics of the reward, exhibiting slower bias and variance decay rates, hence reducing computational advantages.

Importance sampling was employed as a variance reduction technique for the standard MC estimator in the context of high strike prices. It has proven effective in reducing the variance, even with an ad hoc choice of the dominant distribution. This simplified version, without optimization, could be applied at each level of a multi-level scheme with minimal changes to the implementation. This is particularly interesting for handling rare event scenarios.

Estimating the optimal number of levels and samples per level proved challenging in the MLMC framework, as uncertainties in the decay rates of bias and variance impacted sample allocation and overall efficiency. Additionally, the computational overhead introduced by the multi-level strategy diminished MLMC's efficiency.

Artificial Intelligence Disclosure

GitHub Copilot was used during the implementation process. Grammarly was used for spell-checking the report.

References

[1] K. A. Cliffe et al. "Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients". In: *Computing and Visualization in Science* 14.1 (Jan. 2011), pp. 3–15. ISSN: 1432-9360, 1433-0369. DOI: 10.1007/s00791-011-0160-x. URL: http://link.springer.com/10.1007/s00791-011-0160-x (visited on 12/12/2024).

[2] Michael B. Giles. "Multilevel Monte Carlo methods". In: *Acta Numerica* 24 (May 1, 2015), pp. 259–328. ISSN: 0962-4929, 1474-0508. DOI: 10.1017/S096249291500001X. URL: https://www.cambridge.org/core/product/identifier/S096249291500001X/type/journal_article (visited on 12/12/2024).

[3] Michael B. Giles. "Multilevel Monte Carlo Path Simulation". In: *Operations Research* 56.3 (June 2008), pp. 607–617. ISSN: 0030-364X, 1526-5463. DOI: 10.1287/opre.1070.0496. URL: https://pubsonline.informs.org/doi/10.1287/opre.1070.0496 (visited on 12/12/2024).

[4] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992. ISBN: 978-3-642-08107-1 978-3-662-12616-5. DOI: 10.1007/978-3-662-12616-5. URL: http://link.springer.com/10.1007/978-3-662-12616-5 (visited on 12/23/2024).

[5] Fabio Nobile. *Stochastic Simulation, Chap. 6*. Lecture notes for MATH-414, 2024–2025.