École polytechnique fédérale de Lausanne

Semester project

Multi-level Monte Carlo methods for non-contractive ergodic dynamics

Bruno Rodriguez Carrillo

Professor:
Fabio Nobile

Assistants:
Juan Pablo Madrigal Cianci
Sundar Ganesh

June 11th, 2021

Chapter 1

Preliminaries

1.1 Introduction

Let $\Omega \subseteq \mathbb{R}^m$ and denote by $\mathcal{F}$ its Borel σ-algebra. Computing the expected value of a function $\phi : \mathbb{R}^m \to \mathbb{R}$ with respect to an invariant measure π on a measurable space $(\Omega, \mathcal{F})$ is of great interest in different areas such as computational finance and physics. In particular, we are interested in computing:

$$\pi(\phi) := \int \phi(x) d\pi(x) = \lim_{t\to\infty} \mathbb{E}\left[\phi(X(t))\right] \ , \phi \in L^1(\pi), \tag{1.1}$$

where $X(t)$ is usually the solution to an m-dimensional Stochastic Differential Equation (SDE) driven by an m-dimensional Brownian motion $W(t)$ given by:

$$dX(t) = f(X(t))dt + dW(t), \tag{1.2}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ is called the *drift*. Throughout this work, we assume that f is Lipschitz continuous and satisfies the dissipativity condition given by:

$$\langle x, f(x)\rangle \leq -\alpha \left\|x\right\|^2 + \beta \ , \alpha, \beta > 0. \tag{1.3}$$

We say that the function f is globally Lipschitz if there exists $K > 0$ such that:

$$\|f(x) - f(y)\| \leq K\|x - y\|, \forall x, y \in \mathbb{R}^m. \tag{1.4}$$

It was shown in [5], Theorem 6.1 that SDEs satisfying Eq.(1.4) are ergodic and their solution converges exponentially fast to some invariant measure π.

More precisely, it has been shown in [5] that the exponential convergence to the invariant measure π is bounded by

$$\left|\mathbb{E}\left[\phi(X_T) - \pi(\phi)\right]\right| \leq \mu^\star e^{-\lambda^\star T}, \tag{1.5}$$

for some constants $\mu^\star$, $\lambda^\star > 0$. If we bound Eq.(1.5) by an accuracy ϵ, we have:

$$T \geq \frac{1}{\lambda^\star} \log\left(\epsilon^{-1}\right) + \frac{\log \mu^\star}{\lambda^\star}. \tag{1.6}$$

If equation (1.3) is fulfilled, Theorem 1 in [2] has shown that the p-th moments of the numerical solution to Eq.(1.2) with $p \geq 1$ are bounded uniformly with respect to T and so is the variance of the estimator.

An SDE is said to satisfy the *contractivity condition* if there exists $\lambda > 0$ such that

$$\langle x - y, f(x) - f(y)\rangle \leq -\lambda\|x - y\|^2 \ , \forall x, y \in \mathbb{R}^m. \tag{1.7}$$

In general, the exact solution to (1.2) is unknown and thus we need to make use of numerical methods to approximate Eq.(1.2) for a sufficiently long time T. To that end, one usually discretizes the time interval $[0, T]$ into N non-overlaping time sub-intervals of size h and uses the Euler-Maruyama (EM) method given as follows:

$$X_{t_{n+1}} = X_{t_n} + f\left(X_{t_n}\right) h + \Delta W_n. \tag{1.8}$$

where X_{t_n} denotes the approximate solution to Eq.(1.2) for $t_n \in [0, T]$ given an initial condition $X(0)$, $h = T/N$, $t_n = nh$, $X_{t_0} = X(0)$ and $\Delta W_n \sim N(0, h)$ for $n = 1, 2, ..., N-1$.

In the discretized setting, it can be shown that using Eq.(1.7), $\exists C, h > 0$ such that we have first order strong convergence and that the strong error is uniformly bounded by Ch^2 with $C > 0$ not depending on T and a discretization step $h > 0$, [2].

Generally, there are many SDEs that do not satisfy Eq.(1.7); instead, they satisfy the one-sided Lipschitz condition given by:

$$\langle x - y, f(x) - f(y)\rangle \leq \lambda \|x - y\|^2 \ , \ \lambda > 0. \tag{1.9}$$

One feature of this class of SDEs is that their solutions are highly sensitive to initial conditions [3]. An important consequence of contractivity Eq.(1.7) is that two solutions to the corresponding SDE starting from different initial data but driven by the same Brownian motion will converge exponentially fast in time, which means that the discretization error from previous time steps will decay exponentially. Thus, it was shown in [3] that the strong error is uniformly bounded. Whenever Eq.(1.7) is not satisfied, however, the strong error may increase exponentially fast with respect to T but is still bounded as given in Eq.(2.12).

Under the assumption that Eq.(1.3) is satisfied, we have that:

Weak order convergence is of order

$$\sup_{0 \leq t_n \leq T} \left(\mathbb{E}\left[X(t_n)\right] - \mathbb{E}\left[X_{t_n}\right]\right) = O(\Delta t).$$

Strong order convergence is of order

$$\mathbb{E}\left[\sup_{0 \leq t_n \leq T} |X(t_n) - X_{t_n}|\right] = O(\Delta t^{1/2}).$$

Typically, the expectation Eq.(1.1) is estimated using Monte Carlo methods, which are explained briefly in next section.

1.2 Monte Carlo method

Monte Carlo methods are a general and useful set of techniques to compute expectations arising from stochastic simulations with high-dimensional input uncertainties. However, given that their accuracy is controlled by the number of samples, they can rapidly become computationally expensive, particularly when the cost of generating individual stochastic samples is high. We can use MC methods to approximate Eq.(1.1) as follows.

We denote by X_t the exact solution to Eq.(1.2), and refer to Eq.(1.1) as $\mathbb{E}[Q]$ where $Q = \phi\left(X_t\right)$ in a time interval $t \in [0, T]$. We can estimate Eq.(1.1) by using a standard Monte Carlo estimator together with a numerical method to approximate Eq.(1.2) for a given accuracy $\epsilon > 0$. Moreover, we let $X_{t,h}$ represent the numerical solution to Eq.(1.2) and $Q_h = \phi\left(X_{t,h}\right)$ represent the approximation to Q for a given discretization step $h > 0$. We can approximate the expected value of our quantity of interest Q, which we denote by $\mathbb{E}[Q_h]$, using a Monte Carlo estimator in the following manner:

$$\mathbb{E}\left[Q\right] \approx \mathbb{E}^{MC}\left[Q_h\right] = \frac{1}{N}\sum_{n=1}^{N} Q_h^{(n)}, \tag{1.10}$$

where $Q_h^{(n)}$, $n = \{1, 1, ..., N\}$, are N independent and identically distributed realization of Q_h.

We can estimate the accuracy of our estimator by defining the Mean Squared Error (MSE), which we denote by e^2_{MC}, as:

$$e^2_{MC} = \mathbb{E}\left[\left(\mathbb{E}^{MC}\left[Q_h\right] - \mathbb{E}\left[Q\right]\right)^2\right] = \underbrace{\left(\mathbb{E}\left[Q_h - Q\right]\right)^2}_{\text{squared bias}} + \underbrace{N^{-1}\mathbb{V}[Q_h]}_{\text{variance}}, \tag{1.11}$$

where $\mathbb{V}[Q_h]$ denotes the variance of Q_h. Consequently, there are two sources of error:

- The first term on the right-hand side of Eq.(1.11) represents the discretization error or bias squared and depends on the step size h
- The second term on the right-hand side of Eq.(1.11) represents the variance of the estimator or statistical error and decays inversely with the number of samples N. This is the error due to finite sampling

As a result, if we require an MSE of order ϵ^2 for our estimator in Eq.(1.10), we need both terms on the right hand side of Eq.(1.11) to be less than $\epsilon^2/2$, that is:

$$\left(\mathbb{E}\left[Q_h - Q\right]\right)^2 < \epsilon^2/2 \text{ and } N^{-1}\mathbb{V}[Q_h] < \epsilon^2/2.$$

By requiring the bias and the variance to satisfy a tolerance $\epsilon^2/2$, it can be shown [6] that the total computational cost C_{MC} for the standard Monte Carlo estimator (1.10) given a tolerance $\epsilon > 0$ grows as:

$$C_{MC} \lesssim \epsilon^{-2-\gamma/\alpha},$$

where the values of $\alpha, \gamma \in \mathbb{R} > 0$ are related to the simulation cost and the convergence of the bias, and as such are problem specific. Typically, for the Euler-Maruyama method, $\alpha = 1$ (weak error order) and $\gamma = 1$ and as such, we have a total computational cost of order $O(\epsilon^{-3})$. Clearly, this computational cost can rapidly become prohibitively expensive for small tolerances ϵ. In next section, we describe a method aimed at improving this issue.

Chapter 2

Multi-level Monte Carlo (MLMC) method

2.1 Description

Multilevel Monte Carlo (MLMC) [4] is an approach that improves the complexity of traditional MC methods by performing most simulations with low accuracy, and hence, at a correspondingly low cost, with relatively few simulations being performed at high accuracy and at high cost. The key idea of the MLMC method is to compute samples on several approximations to our quantity of interest Q, which we call Q_{h_ℓ}, built on decreasing discretization steps h_ℓ: $h_0 > h_1 > h_2, ..., h_L$, where $h_\ell = h_0\delta^{-\ell}$, for $L > 0$, $\ell = 0, 1, ..., L$, $\delta \in \mathbb{R} > 1$ and $h_0 \in \mathbb{R} > 0$. That is, we have a sequence Q_{h_0}, Q_{h_1},.., Q_{h_L} which approximates Q with increasing accuracy and computational cost.

We let X_{t,h_ℓ} represent the numerical solution to Eq.(1.2) and $Q_{h_\ell} = \phi\left(X_{t,h_\ell}\right)$ represent the approximation to Q for a given discretization step $h_\ell > 0$. It is worth noting that when $h_L = O(\epsilon)$, we have the discretization step needed by the EM scheme to achieve a weak error of order $O(\epsilon)$. We can exploit the linearity of the expected value to write the expectation on the finest level as a telescopic sum of the expectation on the coarsest level plus a sum of corrections terms adding the difference in the expectation between consecutive levels:

$$\mathbb{E}\left[Q_{h_L}\right] = \mathbb{E}\left[Q_{h_0}\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[Q_{h_\ell} - Q_{h_{\ell-1}}\right] = \sum_{\ell=0}^{L} \mathbb{E}\left[Y_{h_\ell}\right], \tag{2.1}$$

with $Y_\ell := Q_{h_\ell} - Q_{h_{\ell-1}}$ and $Q_{h_{-1}} := 0$. The right-hand side of Eq.(2.1) can be approximated using a MC estimator for each level as follows:

$$\mathbb{E}\left[Y_{h_\ell}\right] \approx \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \left(Q_\ell^{(n)} - Q_{\ell-1}^{(n)}\right). \tag{2.2}$$

On level ℓ, we generate N_ℓ independent samples and the difference $Q_{h_\ell}^{(n)} - Q_{h_{\ell-1}}^{(n)}$ is computed using the same underlying random event on the mesh sizes h_ℓ and $h_{\ell-1}$.

Thus, the full estimator is:

$$\mathbb{E}\left[Q\right] \approx \mathbb{E}\left[Q_{h_L}\right] \approx \hat{\mu}_{MLMC} := \frac{1}{N_0} \sum_{n=1}^{N_0} Q_0^{(n)} + \sum_{\ell=1}^{L} \left\{ \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \left(Q_{h_\ell}^{(n)} - Q_{h_{\ell-1}}^{(n)}\right) \right\}. \tag{2.3}$$

Similarly, the MSE of the standard MLMC estimator is given by:

$$e_{MLMC}^2 := \mathbb{E}\left[\left(\hat{\mu}_{MLMC} - \mathbb{E}\left[Q\right]\right)\right)^2\Big] = \left(\mathbb{E}\left[Q_{h_L} - Q\right]\right)^2 + \sum_{\ell=0}^{L} N_\ell^{-1} V_\ell. \tag{2.4}$$

The second term on the right-hand side of Eq.(2.4) corresponds to the variance of the MLMC estimator and is computed using the following expression

$$\mathbb{V}\left[\hat{\mu}_{MLMC}\right] \approx \sum_{\ell=0}^{L} V_\ell N_\ell^{-1} \text{ and } V_\ell := \mathbb{V}\left[Q_{h_\ell} - Q_{h_{\ell-1}}\right], \tag{2.5}$$

where $\mathbb{V}\left[Q_{h_\ell} - Q_{h_{\ell-1}}\right]$ represents N_ℓ independent samples of the variance of $Q_{h_\ell} - Q_{h_{\ell-1}}$.

We state below the result from [4] regarding the computational complexity of the standard MLMC estimator.

Theorem 1 *Let Q denote a random variable and let Q_{h_ℓ} denote the corresponding level ℓ numerical approximation. If there exist independent estimators Y_{h_ℓ} based on N_ℓ Monte Carlo samples, each with expected cost C_ℓ and variance V_ℓ, and positive constants α, β, γ, c_1, c_2, c_3 and c_4, independent of ℓ, such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$, and*

$$|\mathbb{E}[Q_{h_\ell} - Q]| \leq c_1 2^{-\alpha\ell}. \tag{2.6}$$

$$\mathbb{E}[Y_{h_\ell}] = \begin{cases} \mathbb{E}[Q_{h_0}] & , \ell = 0 \\ \mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}] & , \ell > 0. \end{cases} \tag{2.7}$$

$$V_\ell \leq c_2 2^{-\beta\ell}. \tag{2.8}$$

$$C_\ell \leq c_3 2^{\gamma\ell}. \tag{2.9}$$

Then there exists a positive constant c_4 independent of ϵ such that for any $\epsilon < e^{-1}$, there exist cost optimal $L(\epsilon)$ and $N_\ell(\epsilon)$ such that the cost to satisfy a tolerance ϵ^2 on the MSE of the $\hat{\mu}_{MLMC}$ defined in Eq.(2.1) grows as:

$$C \leq \begin{cases} c_4\epsilon^{-2} & , \beta > \gamma \\ c_4\epsilon^{-2}\left(\log\epsilon\right)^2 & , \beta = \gamma \\ c_4\epsilon^{-2-(\gamma-\beta)/\alpha} & , \beta < \gamma. \end{cases}$$

It is important to note that the previous theorem ensures that MSE$\leq \epsilon^2$ as long as the values of the constants c_1 and c_2 are known. To ensure that MSE$\leq \epsilon^2$, we must make sure that both terms on the right-hand side of Eq.(2.4) satisfy $\mathbb{V}\left[\hat{\mu}_{MLMC}\right] \leq \epsilon^2/2$ and $\left(\mathbb{E}[Q_{h_L} - Q]\right)^2 \leq \epsilon^2/2$. To that end, the number of levels L is chosen such that:

$$|\mathbb{E}[Q_{h_L}] - Q| \leq \epsilon/\sqrt{2}. \tag{2.10}$$

Thus, one can use Eq.(2.6) to choose L such that Eq.(2.10) is satisfied. We now focus on bounding the variance in terms of N_ℓ. It was shown in [4] that by choosing the number N_ℓ on each level ℓ as:

$$N_\ell = \left\lceil \frac{2}{\epsilon^2} \cdot \sqrt{\frac{V_\ell}{C_\ell}} \left(\sum_{\ell=0}^{L} \sqrt{C_\ell V_\ell} \right) \right\rceil, \tag{2.11}$$

the overall cost of the MLMC algorithm is minimized and, at the same time, $\mathbb{V}\left[\hat{\mu}_{MLMC}\right] \leq \epsilon^2/2$.

The parameters β and γ play an important point in how the overall computational cost of the MLMC estimator compares to the standard Monte Carlo approach. We have three cases:

1. If $\beta < \gamma$ gets larger with increasing level ℓ, the most expensive part comes from the finest levels. Then $C \sim \frac{1}{\epsilon^2} V_L C_L$. That is, the MLMC cost is reduced by a factor V_L/V_0, which is the ratio of the variances $\mathbb{V}\left[Q_{h_L} - Q_{h_{L-1}}\right]$ and $\mathbb{V}\left[Q_{h_L}\right]$. Because of condition (2.6), we have $2^{-\alpha L} \sim O(\epsilon)$, and hence $C_L = O(\epsilon^{-\gamma/\alpha})$. If $\beta = 2\alpha$, since $\left|\mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}]\right| \sim \left|\mathbb{E}\left[(Q_{h_\ell} - Q_{h_{\ell-1}})^2\right]\right| > (\mathbb{E}\left[Q_{h_\ell} - Q_{h_{\ell-1}}\right])^2$, we have $O(1)$ samples at the finest level

2. If $\beta > \gamma$ decreases with increasing level ℓ, the most expensive part comes from first level $\ell = 0$ where the standard MC is used. As a result, $C \sim \frac{1}{\epsilon^2} V_0 C_0$. Then the MLMC cost is reduced by a factor C_0/C_L which is the ratio of the costs of computing Q_0 and $Q_{h_L} - Q_{h_{L-1}}$. Moreover, we have that $C_\ell \sim O(1)$ and $O(\epsilon^{-2})$ samples are required to achieve the desired accuracy; this is the result when we use the standard MC approach

3. If $\beta = \gamma$ is constant from level to level, then the cost is $\epsilon^{-2} L^2 V_0 C_0 = \epsilon^{-2} L^2 V_L C_L$. Thus, both the computational effort and the contributions to the overall variance are equally spread across all levels

This implies that the total cost for the standard MC method for the application consider here, $\alpha = 1$, $\beta = 2$, $\gamma = 1$, goes as $O(\epsilon^{-3})$.

When we apply the MLMC method to non-contractive SDEs, we run into the problem that V_ℓ does not decay exponentially with levels. If the SDE does not satisfy contractivity Eq.(1.7), the difference between fine and coarse paths will increase exponentially in T until $\mathbb{V}\left[Q_{h_\ell} - Q_{h_{\ell-1}}\right] = \mathbb{V}\left[Q_{h_\ell}\right] + \mathbb{V}\left[Q_{h_{\ell-1}}\right]$, which increases the total computational cost to $O\left(\epsilon^{-2-\kappa/2\lambda^\star} |\log \epsilon|\right)$, where κ is the Lyapunov exponent of the system [3]. For instance, in some chaotic SDEs, κ can be such that $\frac{\kappa}{2\lambda^\star} \geq 1$. For this type of problems, the standard MLMC will have difficulties [1], [3] as shown in next section.

2.2 Model example

We consider the numerical results for a Lipschitz version of the stochastic Lorenz equation with additive noise as given in [3]:

$$f(x_1, x_2, x_3) = \left(10\left(B(x_2) - x_1\right), (28 - x_3)B(x_1) - x_2, B(x_1)x_2 - \frac{8}{3}x_3\right).$$

When $|x_1| > 65$ and $|x_2| > 65$, we use:

$$f(x_1, x_2, x_3) = \left(650\mathrm{sgn}(x_2) - 10x_1, 65\mathrm{sgn}(x_1)(28 - x_3) - x_2, 65\mathrm{sgn}(x_1)x_2 - \frac{8}{3}x_3\right).$$

with $B(x) = \dfrac{65x}{\max(65, |x|)}$

For all results shown herein, we consider the function ϕ as the L_2 norm at time T of the numerical approximation to Eq.(1.2); that is $\phi = \|X_T\|_2$. The initial value X_{t_0} is chosen at random but fixed over different realizations of the Brownian paths.

2.3 Numerical results for standard MLMC

When we apply the standard MLMC to our model problem we have the following plots for the variance V_ℓ behavior on each level ℓ.

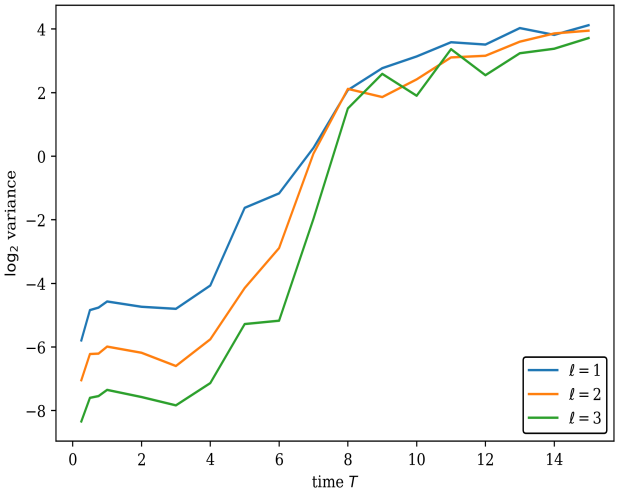

Figure 2.1: Variance for each level ℓ

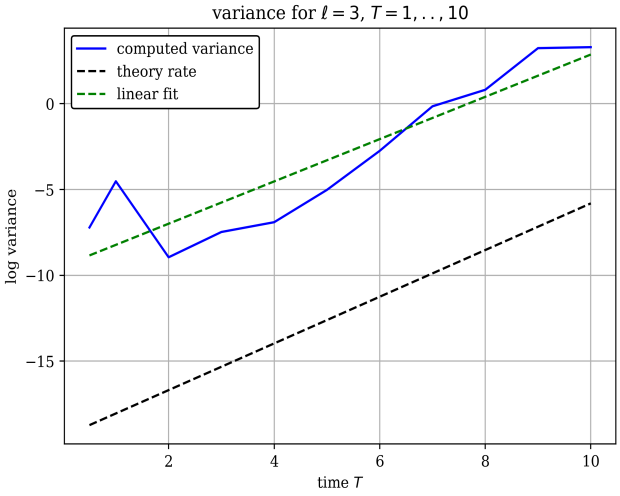

Figure 2.2: Linear increase of $\log_2$ variance

Both Figs.(2.1) and (2.2) show on the y axis $\mathbb{V}\left[\phi\left(X_{T,h_\ell}\right) - \phi\left(X_{T,h_{\ell-1}}\right)\right]$ for $N_\ell = 200$ samples on each level, $h_0 = 2^{-11}$ for $\ell = 1, .., 3$. In Fig.(2.1), as time T increases, V_ℓ achieves a constant value for all ℓ. That is, for values of $T > 10$, the standard MLMC does not results in a variance reduction with levels. The theoretical rate for variance V_ℓ on each level as shown in [1] is:

$$V_\ell \leq \min\left\{B_1, B_2 e^{2LT}\delta^{-2\ell}\right\}, \tag{2.12}$$

for suitable time independent constants B_1, B_2 and $\delta > 1$. As observed in Fig.(2.2), the variance on level $\ell = 3$, V_3, has an exponential increase for $T < 10$. It is, however, relevant to mention that standard MLMC does decrease the variance V_ℓ with level for small values of T. To show how the variance in the standard MLMC depends on the value of h_ℓ on each level ℓ and increasing time T, we have the Fig.(2.3), where on the y axis, we have $\mathbb{V}\left[\phi\left(X_{T,h_\ell}\right) - \phi\left(X_{T,h_{\ell-1}}\right)\right]$.

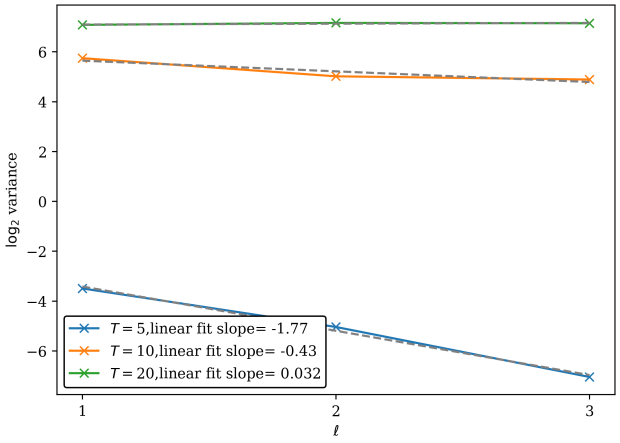

Figure 2.3: Variance for level ℓ

In Fig.(2.3), we can clearly observe that the variance is roughly constant no matter how small h_ℓ is for values $T = 10$ and $T = 20$. We use the same simulation parameters of Fig.(2.1).

Correspondingly, we illustrate how the coarse and fine levels behave for one sample using $h_0 = 2^{-11}$, $\ell = 5$ and $T = 20$:

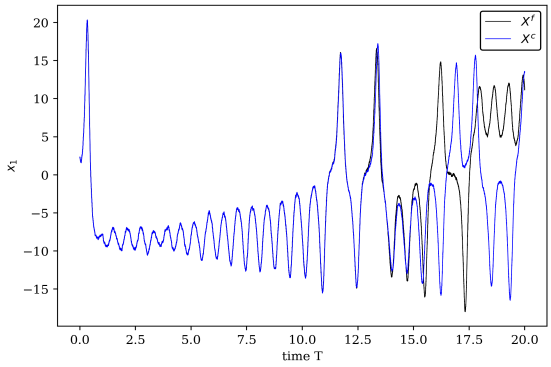

Figure 2.4: Path for x_1.

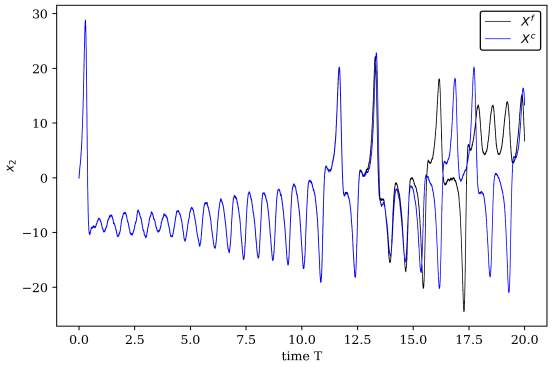

Figure 2.5: Path for x_2.

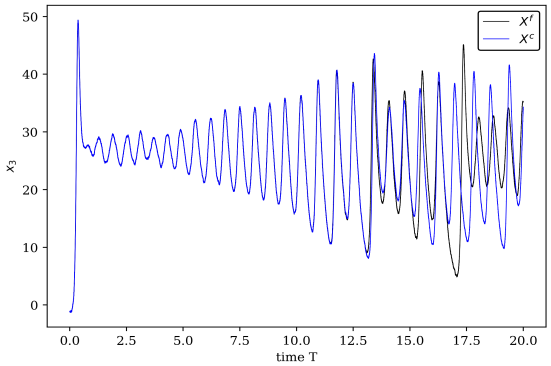

Figure 2.6: Path for x_3.

We can state that for values $T < 10$ the fine and coarse paths are well-coupled and then the standard MLMC works as predicted in theory as shown in Fig.(2.7). But as shown in the figure above, even for a discretization step $h_5 = 2^{-16}$, the paths start diverging for $T > 10$.

Now, we illustrate how the variance behaves for a fixed T when we optimally calibrate the hierarchy to achieve a tolerance ϵ^2 on the MSE.

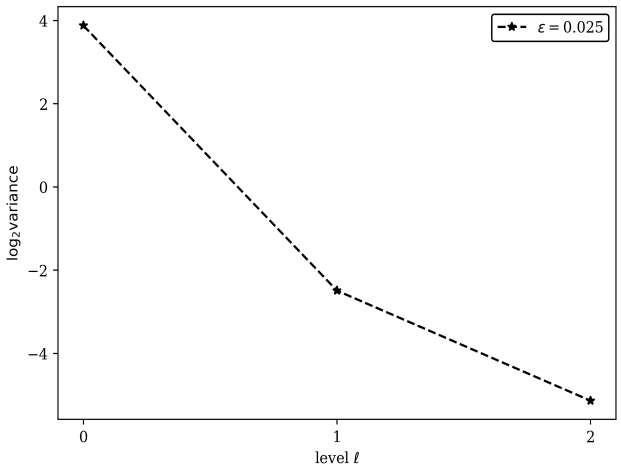

Figure 2.7: Variance for each level ℓ, $\epsilon = 0.025$.

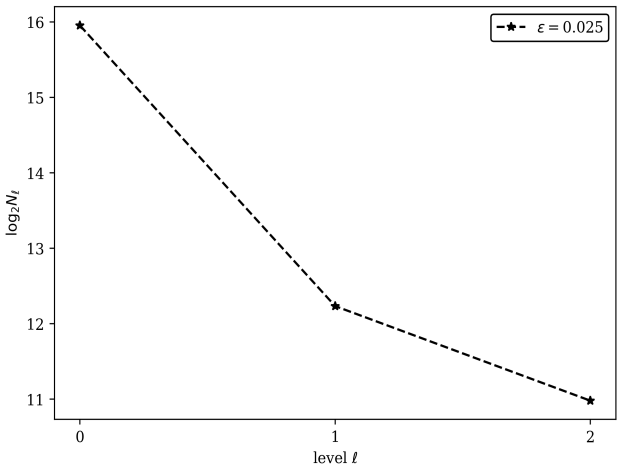

Figure 2.8: N_ℓ for each level ℓ.

For Figs.(2.7) and(2.8), we set $T = 5$ and $h_0 = 2^{-11}$. In Fig.(2.7), we observe that the variance decay as level ℓ increases. On the y axis we plot $\mathbb{V}\left[\phi\left(X_{T,h_\ell}\right) - \phi\left(X_{T,h_{\ell-1}}\right)\right]$ and in Fig.(2.8) we have the number of samples on each level ℓ. For this case, the computed values of the exponents in the Theorem 1 are $\alpha \approx 1.51$, $\beta \approx 2.65$ and $\gamma \approx 1.00$. When $\ell = 0$, we use the standard MC estimator. Based on the complexity analysis shown previously, we state that the most computationally expensive part comes from level 0.

We have the following when we fix $T = 5$ and compute the time complexity in $\log_2$ scale of the standard MLMC estimator as a function of the accuracy ϵ:

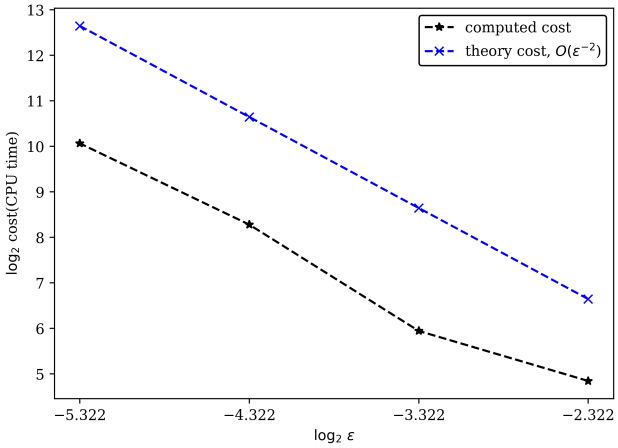

Figure 2.9: Cost as a function of ϵ

As observed, for $T = 5$, the standard MLMC method works for our model problem. However, as seen in Fig.(2.3), for larger values of T, there is no variance reduction with levels. The model problem motivates the following numerical scheme presented in [3].

Chapter 3

Multi-level Monte Carlo with change of measure

3.1 Description

When the SDE does not satisfy Eq.(1.7), standard MLMC might work for small values of T, but such a small T is usually not of practical interest. To overcome this issue, [3] proposes the use of an MLMC estimator with change of measure. As with the standard MLMC approach, this method uses the idea of computing samples on several approximations to our quantity of interest built on decreasing discretization steps but adds a spring term that keeps fine and coarse levels close to each other.

In continuous time, in standard MLMC we have that measures for the fine and coarse paths are different and given by:

$$\begin{aligned} \mathbb{Q}^f &: dX^f(t) = f\left(X^f(t)\right) dt + dW(t), \\ \mathbb{Q}^c &: dX^c(t) = f\left(X^c(t)\right) dt + dW(t), \end{aligned} \tag{3.1}$$

where $X^f(t)$ and $X^c(t)$ correspond to the fine and coarse levels, respectively.

The change-of-measure technique proposed in [3] adds a spring term $S > \lambda/2$ for the coarse and fine level for $\ell \geq 1$, and then computes the fine and coarse paths in the same probability measure $\mathbb{P}$; the new dynamics for both paths is now defined by:

$$dX^f(t) = S\left(X^c(t) - X^f(t)\right) dt + f\left(X^f(t)\right) dt + dW(t), \tag{3.2}$$

$$dX^c(t) = S\left(X^f(t) - X^c(t)\right) dt + f\left(X^c(t)\right) dt + dW(t). \tag{3.3}$$

Notice that since we introduced the spring term S, we have to recover the expected value for both paths in the original SDE. Thus, from the Girsanov's theorem, we obtain that

$$\mathbb{E}^{\mathbb{Q}^f}\left[X^f(t)\right] - \mathbb{E}^{\mathbb{Q}^c}\left[X^c(t)\right] = \mathbb{E}^{\mathbb{P}}\left[X^f(t)\frac{d\mathbb{Q}^f}{d\mathbb{P}} - X^c(t)\frac{d\mathbb{Q}^c}{d\mathbb{P}}\right], \tag{3.4}$$

where $\frac{d\mathbb{Q}^f}{d\mathbb{P}}$ and $\frac{d\mathbb{Q}^c}{d\mathbb{P}}$ are the Radon-Nikodym derivatives for the fine and coarse levels, respectively.

The advantage of this change of measure is that the difference between both paths satisfies:

$$d\left(X^f(t) - X^c(t)\right) = 2S\left(X^c(t) - X^f(t)\right) dt + \left(f(X^f(t)) - f(X^c(t))\right) dt.$$

Ito's formula and Eq.(1.9) give:

$$d\|X^f(t) - X^c(t)\|^2 \leq 2\left(\lambda - 2S\right) \|X^f(t) - X^c(t)\|^2 dt,$$

which recovers contractivity Eq.(1.7) between coarse and fine paths provided $S > \lambda/2$, i.e., the chains are contractive for sufficiently large S.

It was shown in [3] that due to contractivity the difference between fine and coarse levels is uniformly bounded with respect to T and that, together with the Radon-Nikodym derivatives, the variance of the new estimator increases linearly in T, which corresponds to Theorem 5 shown below.

It is important to mention that the above equations are in continuous time and in practice, we use their discretized versions, which are discussed in next section.

3.1.1 Proposed numerical scheme

The following numerical scheme is proposed in [3]. We consider the discretized Radon–Nikodym derivatives, which we denote by $\frac{d\widehat{\mathbb{Q}}^f}{d\mathbb{P}} = R_T^f$ and $\frac{d\widehat{\mathbb{Q}}^c}{d\mathbb{P}} = R_T^c$, for fine and coarse levels, respectively.

Similarly, we denote by $\hat{X}_{t_n}^f$ and $\hat{X}_{t_n}^c$ the discretized solution to Eq.(1.2) in the fine and coarse paths, respectively. Here we have introduced a notation where the superscripts f and c correspond to level ℓ and $\ell - 1$ in (2.3), respectively.

1. At level $\ell = 0$, the numerical estimator is the one computed with standard MC
2. At t_0, $\hat{X}_{t_0}^f = \hat{X}_{t_0}^c = x_0$ and $R_{t_0}^f$=$R_{t_0}^f = 1$
3. We update both paths and the fine Radon–Nikodym derivative at odd time-steps $t_{2n+1} = t_{2n} + h$, $n \geq 0$

$$\begin{aligned}
\hat{X}_{t_{2n+1}}^c &= \hat{X}_{t_{2n}}^c + S\left(\hat{X}_{t_{2n}}^f - \hat{X}_{t_{2n}}^c\right)h + f\left(\hat{X}_{t_{2n}}^c\right)h + \Delta W_{2n} \\
\hat{X}_{t_{2n+1}}^f &= \hat{X}_{t_{2n}}^f + S\left(\hat{X}_{t_{2n}}^c - \hat{X}_{t_{2n}}^f\right)h + f\left(\hat{X}_{t_{2n}}^f\right)h + \Delta W_{2n} \\
R_{t_{2n+1}}^f &= R_{t_{2n}}^f R\left(\hat{X}_{t_{2n+1}}^f, \hat{X}_{t_{2n}}^f, S\left(\hat{X}_{t_{2n}}^c - \hat{X}_{t_{2n}}^f\right), h\right)
\end{aligned}$$

4. We update spring and drift terms of the fine path, keep the coarse level unchanged and update coarse and fine Radon–Nikodym derivatives at even time-steps $t_{2n+2} = t_{2n+1} + h$, $n \geq 0$

$$\begin{aligned}
\hat{X}_{t_{2n+2}}^c &= \hat{X}_{t_{2n+1}}^c + S\left(\hat{X}_{t_{2n}}^f - \hat{X}_{t_{2n}}^c\right)h + f\left(\hat{X}_{t_{2n}}^c\right)h + \Delta W_{2n+1} \\
\hat{X}_{t_{2n+2}}^f &= \hat{X}_{t_{2n+1}}^f + S\left(\hat{X}_{t_{2n+1}}^c - \hat{X}_{t_{2n+1}}^f\right)h + f\left(\hat{X}_{t_{2n+1}}^f\right)h + \Delta W_{2n+1} \\
R_{t_{2n+2}}^f &= R_{t_{2n+1}}^f R\left(\hat{X}_{t_{2n+2}}^f, \hat{X}_{t_{2n+1}}^f, S\left(\hat{X}_{t_{2n+1}}^c - \hat{X}_{t_{2n+1}}^f\right), h\right) \\
R_{t_{2n+2}}^c &= R_{t_{2n}}^c R\left(\hat{X}_{t_{2n+2}}^c, \hat{X}_{t_{2n}}^c, S\left(\hat{X}_{t_{2n}}^f - \hat{X}_{t_{2n}}^c\right), 2h\right)
\end{aligned}$$

After N steps, the Radon Nikodym derivatives for both paths are:

$$\begin{aligned}
\frac{d\hat{\mathbb{Q}}^f}{d\mathbb{P}} = R_T^f &= \prod_{n=0}^{N-1} R\left(\hat{X}_{t_{n+1}}^f, \hat{X}_{t_n}^f, S\left(\hat{X}_{t_n}^c - \hat{X}_{t_n}^f\right), h\right) \\
\frac{d\hat{\mathbb{Q}}^c}{d\mathbb{P}} = R_T^c &= \prod_{n=0}^{N/2-1} R\left(\hat{X}_{t_{2n+2}}^c, \hat{X}_{t_{2n}}^c, S\left(\hat{X}_{t_{2n}}^f - \hat{X}_{t_{2n}}^c\right), 2h\right).
\end{aligned}$$

with

$$R\left(\hat{X}_{t_{n+1}}, \hat{X}_{t_n}, S\left(\hat{X}_{t_{n+1}} - \hat{X}_{t_n}\right), h\right) = \exp\left(-\langle \Delta W_n, S\left(\hat{X}_{t_{n+1}} - \hat{X}_{t_n}\right)\rangle - |S|\|\hat{X}_{t_{n+1}} - \hat{X}_{t_n}\|^2 h/2\right).$$

In this numerical scheme, the spring term S is needed when the inner product $\langle x-y, f(x)-f(y)\rangle$ is positive in order to prevent the exponential divergence of the fine and coarse paths. We can also choose S to be a function of the current state $\hat{X}_{tn}$ and as a consequence it can be minimized to reduce the size of the Radon-Nikodym derivative.

Regarding the coarse path updates, we note that they can be combined to give:

$$X^c_{t_{2n+2}} = \hat{X}^c_{t_{2n}} + S\left(\hat{X}^f_{t_{2n}} - \hat{X}^c_{t_{2n}}\right)2h + f\left(\hat{X}^c_{t_{2n}}\right)2h + \Delta W_{2n} + \Delta W_{2n+1}.$$

That means that for the coarse path the terms $S\left(\hat{X}^f_{t_{2n}} - \hat{X}^c_{t_{2n}}\right)$ and $f\left(\hat{X}^c_{t_{2n}}\right)$ are computed once, which results in fewer computations with respect to the fine path where the same terms are computed for both odd and even time steps.

The new multilevel estimator is thus:

$$\mathbb{E}\left[Q_L\right] \approx \mathbb{E}^{\mathbb{P}}\left[\phi\left(\hat{X}^L_T\right)\right] \approx \mathbb{E}^{\mathbb{P}}\left[\phi\left(\hat{X}^0_T\right)\right] + \sum_{\ell=1}^{L}\mathbb{E}^{\mathbb{P}}\left[\phi\left(\hat{X}^f_T\right)R^f_T - \phi\left(\hat{X}^c_T\right)R^c_T\right].$$

And

$$\mathbb{E}\left[Q_L\right] \approx \frac{1}{N_0}\sum_{n=1}^{N_0}\phi\left(\hat{X}^{0,(n)}_T\right) + \sum_{\ell=1}^{L}\left\{\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(\phi\left(\hat{X}^{f,(n)}_T\right)R^{f,(n)}_T - \phi\left(\hat{X}^{c,(n)}_T\right)R^{c,(n)}_T\right)\right\}. \tag{3.5}$$

Such a numerical scheme is supported by Theorems 1,2,3 and 4 in [3], which we state below. For the following results in [3], we make use of the following assumptions.

Assumption 1. The function f is globally Lipschitz Eq.(1.4) and satisfies the dissipativity condition Eq.(1.3).

Assumption 2. The function f is one-sided Lipschitz Eq.(1.9), is differentiable and $\nabla f(x)$ is globally Lipschitz, that is $\|\nabla f(x) - \nabla f(y)\| \leq K\|x-y\|\ \forall x, y \in \mathbb{R}^m$ for some $K > 0$.

Theorem 2 (stability) *Suppose that Eq.(1.2) satisfies* Assumption 1*, then using the new change-of-measure algorithm with $S > 0$, there exist constants $C_{(1)}$, $C_{(2)} > 0$ such that for any $T > 0$ and $p \geq 1$, and for all $0 < h < C_{(1)}$:*

$$\sup_{0\leq n\leq N}\mathbb{E}\left[\left\|\hat{X}^f_{t_n}\right\|^p\right]^{1/p} \leq C_{(2)}p^{1/2}$$
$$\sup_{0\leq n\leq N}\mathbb{E}\left[\left\|\hat{X}^c_{t_n}\right\|^p\right]^{1/p} \leq C_{(2)}p^{1/2}.$$

Theorem 3 (difference bewtween fine and coarse paths) *Suppose that Eq.(1.2) satisfies* Assumption 1 *and* Assumption 2*, then using the new change-of-measure algorithm with $S > \lambda/2$, there exist constants $C_{(1)}$, $C_{(2)} > 0$ such that for any $T > 0$ and $p \geq 1$, and for all $0 < h < C_{(1)}$:*

$$\sup_{0\leq n\leq N}\mathbb{E}\left[\left\|\hat{X}^f_{t_n} - \hat{X}^c_{t_n}\right\|^p\right]^{1/p} \leq C_{(2)}\min\left(p^{1/2}h^{1/2}, ph\right).$$

Theorem 4 (Radon–Nikodym moments) *Suppose Eq.(1.2) satisfies* Assumption 1 *and* Assumption 2*, using the new change-of-measure algorithm with $S > \lambda/2$, there exist constants $C_{(1)}$, $C_{(2)} > 0$ such that, for any $T > 0$ and $p \geq 1$, and for all $0 < h < min(C_{(1)}, C_{(2)}/(Tp^2))$,*

$$\mathbb{E}\left[\left|\frac{d\hat{\mathbb{Q}}^c}{d\mathbb{P}}\right|^p\right]^{1/p} \leq 2$$
$$\mathbb{E}\left[\left|\frac{d\hat{\mathbb{Q}}^f}{d\mathbb{P}}\right|^p\right]^{1/p} \leq 2.$$

Theorem 5 (MLMC moments) *Suppose Eq.(1.2) satisfies* Assumption 1 *and* Assumption 2*, and* $\phi : \mathbb{R}^m \longrightarrow \mathbb{R}$ *is globally Lipschitz, then using the new change-of-measure algorithm with* $S > \lambda/2$*, for any* $T > 0$ *and* $p \geq 1$*, there exist constants* $C_{(1)}, C_{(2)}, C_{(3)} > 0$ *such that for all* $0 < h < \min(C_{(1)}, C_{(1)}/(Tp^2))$*:*

$$\mathbb{E}\left[\left|\phi\left(\hat{X}_T^f\right)\frac{d\hat{\mathbb{Q}}^f}{d\mathbb{P}} - \phi\left(\hat{X}_T^c\right)\frac{d\hat{\mathbb{Q}}^c}{d\mathbb{P}}\right|^p\right]^{1/p} \leq C_{(3)}p^2\sqrt{T}h.$$

We will numerically verify Theorems 4 and 5 for our model problem.

3.2 Numerical results for MLMC with change of measure

As we did with the standard MLMC, we apply the change-of-measure MLMC to our model problem. For all figures below, the level $\ell = 0$ corresponds to the standard MC estimator.

3.2.1 Effect of Radon Nikodym derivative

We illustrate how the coarse and fine levels and Radon-Nikodym derivatives behave when $h_0 = 2^{-11}$, $\ell = 5$, $T = 20$ and $S = 5$; we have the following plots:

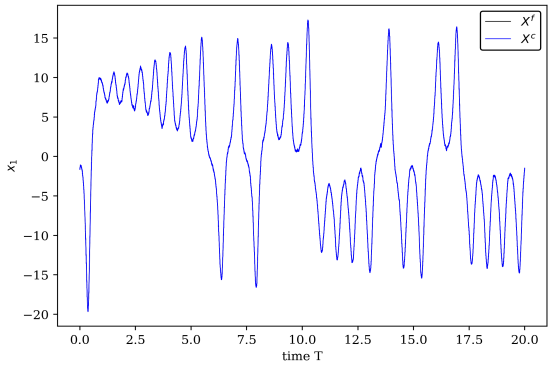

Figure 3.1: Path for x_1.

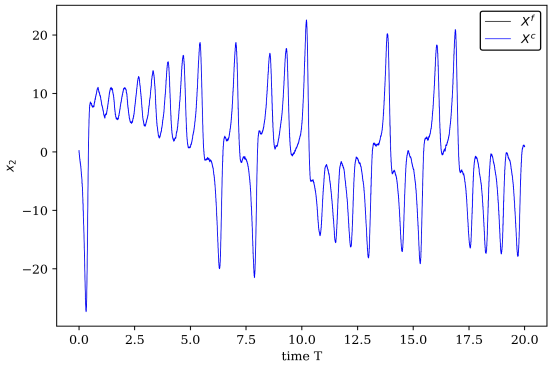

Figure 3.2: Path for x_2.

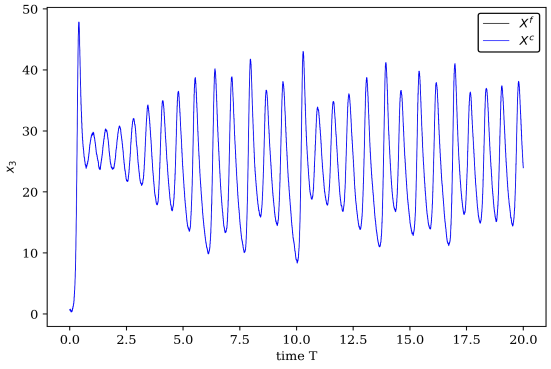

Figure 3.3: Path for x_3.

In the above figures, we clearly see that the coarse and fine paths are indistinguishable for $T = 20$, which means both paths stay together as time elapses. Nevertheless, the Radon-Nikodym derivatives vary in time as follows:

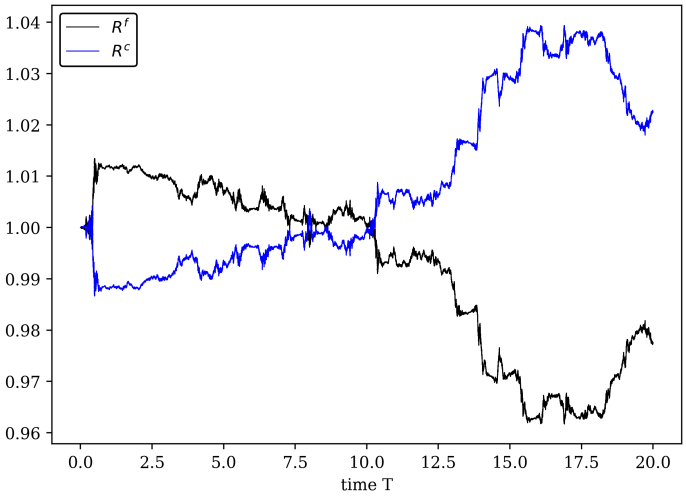

Figure 3.4: Evolution in time of the Radon-Nikodym derivative.

In Fig.(3.4), we observe that the difference between the Radon-Nikodym derivatives in the coarse and fine level is small. In fact, such a difference in the same figure does not exceed 0.08.

As stated in Theorem 4, for $p = 1$

$$\mathbb{E}\left[\left|\frac{d\hat{\mathbb{Q}}^c}{d\mathbb{P}}\right|\right] \approx 1.0 < 2 \text{ and } \mathbb{E}\left[\left|\frac{d\hat{\mathbb{Q}}^f}{d\mathbb{P}}\right|\right] \approx 1.0 < 2.$$

From Fig.3.4, we observe that the longer time is, the larger the magnitude of the Radon-Nikodym derivatives are. This could mean that as time increases, the paths tend to diverge from each other; then such derivatives, along with the spring term S, play the role of bringing the paths back together; this is observe in Figs.(3.1), (3.2) and (3.2) where there is not visible distinction between the paths for the three coordinates, that is they 'stay together' in the time interval $[0, T]$.

Moreover, it is interesting to mention that for values of $T > 10$, the magnitude of the Radon-Nikodym derivatives increases slightly. In the standard MLMC, it is for such values of T that we observe no variance reduction as level ℓ increases.

3.2.2 Variance and complexity results

We have the following plots for the variance V_ℓ behavior on each level ℓ for decreasing h_ℓ.

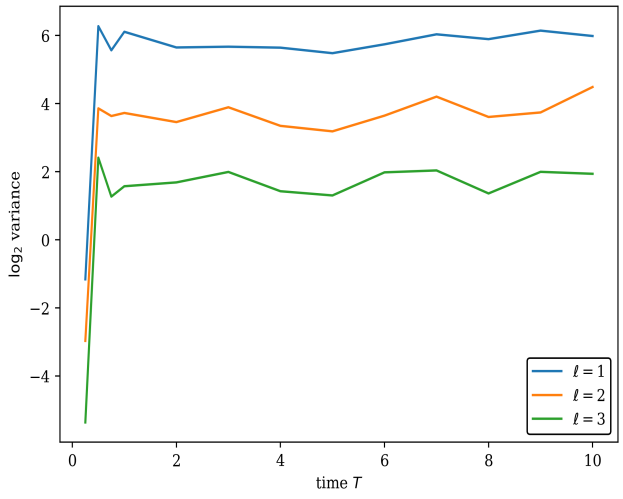

Figure 3.5: Variance for each level ℓ.

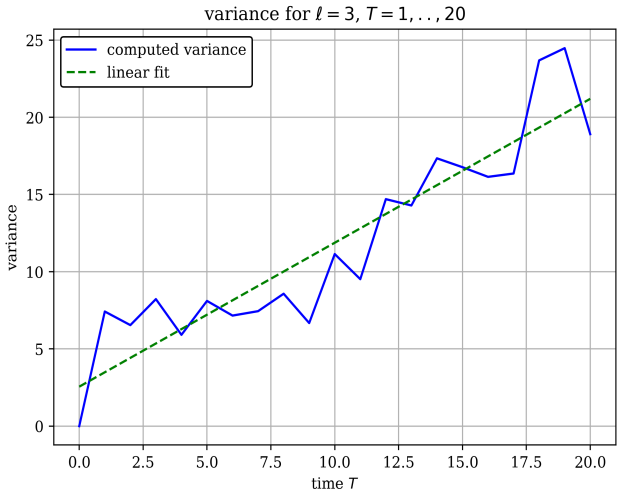

Figure 3.6: Linear increase of variance.

For the above figures, we used $N_\ell = 200$ samples on each level, $h_0 = 2^{-11}$, $\ell = 1, .., 3$ and $S = 5$.

Furthermore, in figure (3.6), we clearly observe that the variance for level $\ell = 3$ has a linear increase with T, as expected.

We show in Fig.(3.7) how the variance behaves for decreasing time steps h_ℓ for $N_\ell = 200$ and $h_0 = 2^{-11}$ for $\ell = 1, 2, 3$. It can be observed that for fix T, $\mathbb{V}\text{ar} \sim h_\ell^2$. In Figs.(3.6) and (3.7), the linear fit is computed via the least squares method.

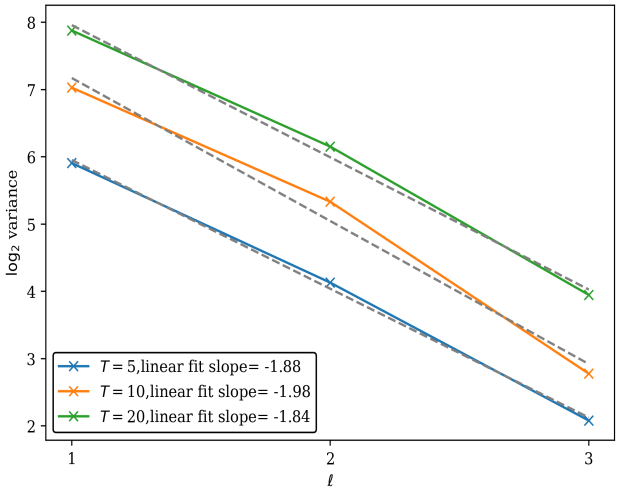

Figure 3.7: Variance decay with level ℓ.

As expected, we observe a variance reduction as level increases. However, we should remark that the variance in the new MLMC method highly depends on the value of the largest timestep, namely, h_0 and on the specifics of the problem.

We now examine the practical implementation of the MLMC with change of measure to achieve a tolerance ϵ on the MSE. Specifically, we require $\mathbb{V}[Q_{h_0}] > 2\mathbb{V}[Q_{h_1} - Q_{h_0}]$ because then the majority of the samples, and consequently, the cost are on level $\ell = 0$ and because it results in a good coupling between levels $\ell = 0$ and $\ell = 1$. However, the choice of h_0 impacts this requirement significantly. Then we adjust the variance by using the upper bound given in Eq.(3.6) and with this, we found numerically that $h_0 = 2^{-11}$ is appropriate.

$$\mathbb{V}[\phi(X_{T,h_0})] = \mathbb{E}\left[(\phi(X_{T,h_0}))^2\right] - (\mathbb{E}[\phi(X_{T,h_0})])^2 \geq 0 \rightarrow \mathbb{E}\left[(\phi(X_{T,h_0}))^2\right] \geq (\mathbb{E}[\phi(X_{T,h_0})])^2. \tag{3.6}$$

In other words, in order to observe a variance reduction from level $\ell = 0$ to $\ell = 1$ using the new MLMC estimator, we use a milder bound on the variance for level $\ell = 0$.

Thus, we have for $T = 12$, 500 initial samples on the first three levels, $h_0 = 2^{-11}$ and $\epsilon \in \{0.6, 0.5, 0.4, 0.3, 0.2\}$:

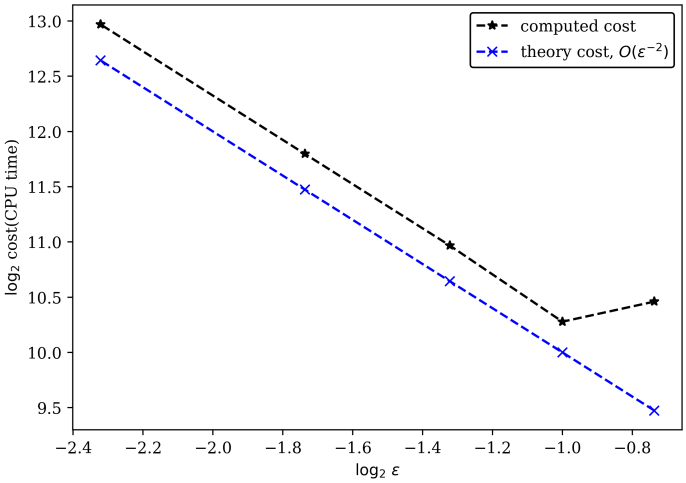

Figure 3.8: Complexity as function of ϵ.

Let us recall that the total cost C for each ϵ is computed by $C = \sum_{\ell=0}^{L} N_\ell C_\ell$, where N_ℓ and C_ℓ correspond to the number of samples and cost per sample on each level ℓ, respectively. Furthermore, for all values of ϵ shown above, we have $\beta > \gamma$; then, based on Theorem 1, the theoretical rate of the complexity cost should be $c_4\epsilon^{-2}$. This is observed in Fig.(3.8).

3.3 Conclusions

In summary, the standard MC method can be computational expensive for small tolerances ϵ. As explained, the main goal of using the MLMC method lies in reducing such a computational cost to achieve a given tolerance $\epsilon > 0$. However, for a class of SDEs, such an approach cannot result in cost savings. Hence, we can use the MLMC with change of measure to recover the computational complexity of the MLMC.

For this project, we implemented the standard MLMC method for our model problem and observed that such a technique does not result in any improvement with respect to the standard MC. Hence, we implemented the MLMC with change of measure proposed in [3] and applied it to our model problem. The numerical results show that the MLMC method with change of measure works for our analyzed problem.

It is also important to mention that for the new MLMC method the Radon-Nikodym derivatives have a large impact on the multilevel estimator. If these take values too distant from each other, the differences in the multilevel estimator, between levels ℓ and $\ell - 1$, could be large even if there is not visible difference between the trajectories in the fine and coarse paths, which results in large values of the variance. Then, such values of the derivatives should be as close to each one as possible for this new approach to work.

Likewise, the chosen value of h_0 for our experiments results in expensive computational simulations. Regarding future work, we can find the optimal values of the spring term S and h_0 such that the difference between the values of the Radon-Nikodym derivatives in the fine and coarse levels is small and, at the same time, we still have a good coupling between level 0 and level 1, in such a way that $V_0 > 2V_1$, [3].

Bibliography

[1] Quentin Ayoul-Guilmard et al. “D5. 4 Report on MLMC for time dependent problems”. In: (2021).

[2] Wei Fang and Michael B Giles. “Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift: Part II, infinite time interval”. In: *arXiv preprint arXiv:1703.06743* (2017).

[3] Wei Fang and Michael B Giles. “Multilevel Monte Carlo method for ergodic SDEs without contractivity”. In: *Journal of Mathematical Analysis and Applications* 476.1 (2019), pp. 149–176.

[4] Michael B Giles. “Multilevel Monte Carlo methods.” In: *Acta Numer.* 24 (2015), pp. 259–328.

[5] Sean P Meyn and Richard L Tweedie. “Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes”. In: *Advances in Applied Probability* (1993), pp. 518–548.

[6] Michele Pisaroni, Fabio Nobile, and Pénélope Leyland. “A Continuation Multi Level Monte Carlo (C-MLMC) method for uncertainty quantification in compressible inviscid aerodynamics”. In: *Computer Methods in Applied Mechanics and Engineering* 326 (2017), pp. 20–50.