Multi-level Monte Carlo for High-Dimensional Parabolic Equations

Xuda Ye, 2001110036

June 24, 2021

1 Introduction

Solving high-dimensional partial differential equations (PDE) numerically is usually a tough question dut to the curse of dimensionality. If we apply traditional finite difference or finite element methods, then the number of grid points will be $\mathcal{O}(h^{-d})$, where h is the space step and d is the spatial dimensional. Such large complexity is unacceptable when d is large.

In this note we aim to access solution of high-dimensional parabolic equations via the Feynman-Kac formula [1], which represents the solution $u(x,t)$ of the parabolic equation as the expectation related to a stochastic process X_t. Theoretically, one needs to choose the time step sufficiently small and the number of independent samples sufficiently large to obtain the accurate result, which thus requires huge computation cost. In this note we show that, in order to ahieve $\mathcal{O}(\varepsilon^2)$ mean square error (MSE), the computational cost is at least $\mathcal{O}(\varepsilon^{-3})$.

For this reason, we employ the multilevel Monte Carlo [2] to accelerate the computation. Although multilevel Monte Carlo is usually applied in finance [3, 4], this method is also effective for high dimensional PDEs. We will show that multilevel Monte Carlo is able to reduce the computational cost from $\mathcal{O}(\varepsilon^{-3})$ to $\mathcal{O}(\varepsilon^{-2}(\log\varepsilon)^2)$. These results are also verified in numerical tests, where two specific parabolic equations are considered.

In Section 2, we briefly review the Feynman-Kac formula to represent the solutions of parabolic equations. In Section 3, we introduce the multilevel Monte Carlo with complexity analysis. In Section 4, we present the numerical experiments of multilevel Monte Carlo and compare the results with standard Monte Carlo.

2 Feynman-Kac Formula

The Feynman-Kac formula provides an approach to access high dimensional PDEs using stochastic processes. For simplicity, consider the stochastic process $\{X_t\}_{t\geqslant 0}$ in $\mathbb{R}^d$ governed by the following time-homogeneous stochastic differential equation (SDE):

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad t \geqslant 0 \tag{2.1}$$

with the following assumptions:

1. $b(x) : \mathbb{R}^d \mapsto \mathbb{R}^d$ is Lipschitz continuous and of linear growth, i.e.,

$$|b(x) - b(y)| \leqslant L|x-y|, \quad |b(x)| \leqslant K(1+|x|), \quad \forall x, y \in \mathbb{R}^d. \tag{2.2}$$

2. $\sigma(x) : \mathbb{R}^d \mapsto \mathbb{R}^{d\times n}$ is Lipschitz continuous and globally bounded, i.e.,

$$|\sigma(x) - \sigma(y)| \leqslant L|x - y|, \quad |\sigma(x)| \leqslant K, \quad \forall x, y \in \mathbb{R}^d. \tag{2.3}$$

3. W_t is the standard Brownian process in $\mathbb{R}^n$.

The assumptions above ensure that the stochastic process $\{X_t\}_{t\geqslant 0}$ exists and is unique for $t \in [0, +\infty)$. The (time-independent) backforward operator $\mathcal{L}^*$ for the stochastic process $\{X_t\}_{t\geqslant 0}$ is

$$\mathcal{L}^* u = b(x) \cdot \nabla u + \frac{1}{2}(\sigma\sigma^{\mathrm{T}}) : \nabla^2 u \tag{2.4}$$

for any $u \in C^2(\mathbb{R}^d)$.

2.1 Solution Representation of Parabolic Equations

Consider the linear parabolic equation for $u(x, t)$ in $\mathbb{R}^d \times \mathbb{R}^+$:

$$\partial_t u = b(x) \cdot \nabla u + \frac{1}{2}(\sigma\sigma^{\mathrm{T}}) : \nabla^2 u + q(x)u + f(x, t) \tag{2.5}$$

with initial data $u(x, 0) = u_0(x)$ and the following assumptions:

1. $q(x) : \mathbb{R}^d \mapsto \mathbb{R}$ is continuous and globally bounded;
2. $f(x, t) : \mathbb{R}^d \times \mathbb{R}^+ \mapsto \mathbb{R}$ is continuous in (x, t), globally bounded and Lipschitz continuous in t. Also, $f(x, t)$ vanishes as $x \to \infty$.
3. $u_0(x) : \mathbb{R}^d \mapsto \mathbb{R}$ is C^2 differentiable, globally bounded and vanishes as $x \to \infty$.

The assumptions above ensure that the solution $u(x, t)$ exists for $t \in [0, +\infty)$. The Feynman-Kac formula for the homogeneous parabolic equation, i.e., $f(x, t) \equiv 0$ is given by:

Theorem 1 (Feynman-Kac formula) *Let $\{X_t\}_{t\geqslant 0}$ be the stochastic process defined in* (2.1) *with the initial data $X_0 = x$. Define the stochastic process $\{Q_t\}_{t\geqslant 0}$ by*

$$Q_t = \int_0^t q(X_s)\mathrm{d}s, \qquad t \geqslant 0, \tag{2.6}$$

then the function

$$u(x, t) = \mathbb{E}^x\big(e^{Q_t} u_0(X_t)\big), \qquad x \in \mathbb{R}^d, \quad t \geqslant 0 \tag{2.7}$$

satisfies the parabolic equation (2.5) *with $f(x, t) \equiv 0$, i.e.,*

$$\partial_t u = b(x) \cdot \nabla u + \frac{1}{2}(\sigma\sigma^{\mathrm{T}}) : \nabla^2 u + q(x)u.$$

The proof of this result can be seen from [5] or Prof Li's notes[1]. For general $f(x, t)$, the solution representation can be derived from the Duhamel's principle.

[1] `http://dsec.pku.edu.cn/~tieli/notes/appl_stoch/lect14.pdf`

Theorem 2 (Feynman-Kac formula with source) *Let $\{X_t\}_{t\geqslant 0}, \{Q_t\}_{t\geqslant 0}$ be the stochastic processes defined in* (2.6)*, then the function*

$$u(x,t) = \mathbb{E}^x(e^{Q_t}u_0(X_t)) + \mathbb{E}^x\left(\int_0^t e^{Q_s}f(X_s, t-s)\mathrm{d}s\right) \tag{2.8}$$

satisfies the parabolic equation with source (2.5)*.*

The equation (2.5) can be written as $\partial_t u = \mathcal{L}^* u + f(x,t)$, whose solution is

$$\begin{aligned} u(\cdot,t) &= e^{t\mathcal{L}}u_0 + \int_0^t e^{(t-s)\mathcal{L}}f(\cdot,s)\mathrm{d}s \\ &= \mathbb{E}^x(e^{Q_t}u_0(X_t)) + \mathbb{E}^x\left(\int_0^t e^{Q_{t-s}}f(X_{t-s},s)\mathrm{d}s\right) \\ &= \mathbb{E}^x(e^{Q_t}u_0(X_t)) + \mathbb{E}^x\left(\int_0^t e^{Q_s}f(X_s,t-s)\mathrm{d}s\right) \end{aligned}$$

by the Duhamel's principle.

2.2 Numerical discretization

The discretization scheme for the Feynman-kac formula is given as follows. Fixing $x \in \mathbb{R}^d$, $t > 0$ and the number of iterations $M \in \mathbb{N}$, then the time step $h = t/M$. The stochastic process $\{X_t\}_{t\geqslant 0}$ is approximated by the Euler-Maruyama scheme

$$X_{k+1}^h = X_k^h + b(X_k^h)h + \sigma(X_k^h)\xi_k\sqrt{h}, \quad k = 0, 1, \cdots, M-1 \tag{2.9}$$

with initial data $X_0^h = x$, where $\xi_k \sim \mathcal{N}(0,1)$ are independent Gaussian random variables. The stochastic process $\{Q_t\}_{t\geqslant 0}$ is approximated by the numerical integral

$$Q_k^h = h\sum_{i=0}^{k-1} q(X_i^h), \quad k = 0, 1, \cdots, M \tag{2.10}$$

and the numerical solution is represented as

$$u^h(x,t) = \mathbb{E}^x\big(e^{Q_M^h}u_0(X_M^h)\big) + \mathbb{E}^x\left(h\sum_{k=1}^{M} e^{Q_k^h}f(X_k^h, (M-k)h)\right) \tag{2.11}$$

Now we analyze the error of the discretization (2.9)–(2.11). Define the random variables

$$W = e^{Q_t}u_0(X_t) + \int_0^t e^{Q_s}f(X_s, t-s)\mathrm{d}s \tag{2.12}$$

$$W_M = e^{Q_M^h}u_0(X_M^h) + h\sum_{k=1}^{M} e^{Q_k^h}f(X_k^h, (M-k)h) \tag{2.13}$$

then the solution of the parabolic equation is represented as

$$u(x,t) = \mathbb{E}^x[W], \quad u^h(x,t) = \mathbb{E}^x[W_M] \tag{2.14}$$

Therefore, the quality of the numerical solution $u^h(x,t)$ relies on how well the random variable W_M approximates W. Under the assumption $q(x) \equiv 0$, we have the following results for W and W_M:

Theorem 3 (Error analysis) *Assume $q(x) \equiv 0$ and let the random variables W, W_M be defined as in* (2.12)(2.13)*, then we have*

$$|\mathbb{E}^x[W - W_M]| \lesssim h, \qquad \mathbb{E}^x|W - W_M|^2 \lesssim h \tag{2.15}$$

The proof of the result is left in Appendix.

2.3 Numerical Examples

We present two parabolic equations for numerical treatment and verify the weak convergence order of the discretized Feynman-Kac formula (2.9)–(2.11) numerically.

Heat Equation with Source Consider the heat equation with source in $\mathbb{R}^d$

$$u_t = \Delta u + f(x,t) \tag{2.16}$$

with initial data $u_0(x) \equiv 0$, where the source term is

$$f(x,t) = e^{-t|x|^2}\Big((1 + t + 2dt^2 - |x|^2(t + 4t^3))\cos x_1 - 4t^2 x_1 \sin x_1\Big) \tag{2.17}$$

The exact solution is

$$u(x,t) = te^{-t|x|^2}\cos x_1 \tag{2.18}$$

Heat Equation with Linear Growth Consider the heat equation with linear growth in $\mathbb{R}^d$

$$u_t = \Delta u + (2 + \cos x_1)u + f(x,t) \tag{2.19}$$

with initial data $u_0(x) = e^{-|x|^2}$, where the source term is

$$f(x,t) = e^{-|x|^2 + t\cos x_1}(2d - 4|x|^2 + t\cos x_1 - 4tx_1 \sin x_1 - t^2 \sin^2 x_1) \tag{2.20}$$

The exact solution is

$$u(x,t) = e^{-|x|^2 + t\cos x_1} \tag{2.21}$$

We employ the discretized Feynman-Kac formula (2.9)–(2.11) to solve the equations above.

For the heat equation source with source (2.16), set $d = 8$ and evaluate the numerical solution at

$$x = (0.2303, 0.1940, 0.2095, 0.1305, 0.1156, 0.0629, 0.12890.0588) \in \mathbb{R}^8$$

and $t = 1, 2$.

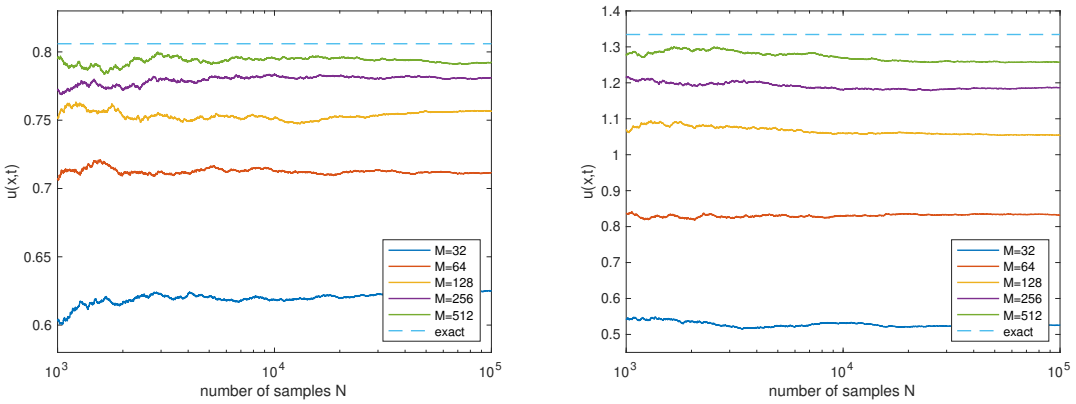

Figure 1: The numerical solution of the heat equation with source (2.16) with different number of iterations M. The left and right figures correspond to the evolution time $t = 1, 2$.

M	32	64	128	256	512
$t = 1$	22.43%	11.71%	6.11%	3.11%	1.68%
$t = 2$	60.64%	37.65%	21.03%	11.08%	5.80%

Table 1: Relative error of the numerical solution (converged result as $N \to \infty$) of the heat equation with source (2.16) with $t = 1, 2$.

For the heat equation with linear drift (2.19), set $d = 8$ and evaluate the numerical solution at

$$x = (0.2102, 0.1792, 0.2868, 0.1832, 0.2321, 0.07420.2222, 0.0102) \in \mathbb{R}^8$$

and $t = 1, 2$.

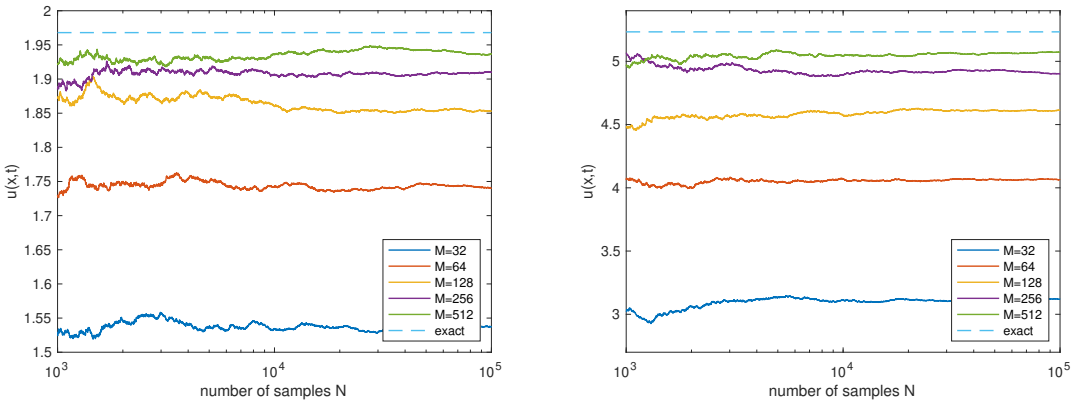

Figure 2: The numerical solution of the heat equation with linear drift (2.19) with different number of iterations M. The left and right figures correspond to the evolution time $t = 1, 2$.

M	32	64	128	256	512
$t=1$	21.88%	11.57%	5.82%	2.94%	1.60%
$t=2$	40.41%	22.35%	11.86%	6.32%	3.10%

Table 2: Relative error of the numerical solution (converged result as $N \to \infty$) of the heat equation with linear drift (2.19) with $t = 1, 2$.

These numerical results verify the weak convergence order in (2.15). Approximately, one has

$$\mathbb{E}^x[W - W_M] \approx Ch \tag{2.22}$$

for some constant C. To obtain the accurate estimation of $u(x,t) = \mathbb{E}^x[W]$, one thus needs to choose the number of iterations M sufficiently large.

3 Multilevel Monte Carlo

In the Feynman-Kac formula, the solution $u(x,t)$ to the parabolic equation is represented as the expectation $\mathbb{E}[W]$ (the superscript x is omitted), and can be approximated by the limit of $\mathbb{E}[W_M]$,

$$\mathbb{E}[W] = \lim_{M\to\infty} \mathbb{E}[W_M] \tag{3.1}$$

In a general setting, we assume

Assumption 1 *For the random variables W, W_M, there exists constants $\alpha, \beta, \gamma > 0$ and a constant C (independent of the number of iterations M) such that*

1. (weak error) $|\mathbb{E}[W - W_M]| \lesssim M^{-\alpha}$;
2. (strong error) $\mathbb{E}|W - W_M|^2 \lesssim M^{-\beta}$;
3. (complexity) $\mathcal{C}(W_M) \lesssim M^{\gamma}$.

In the case of Feynman-Kac formula and its discretization (2.9)–(2.11), $\alpha = \beta = \gamma = 1$.

To compute $\mathbb{E}[W]$, we employ a random variable $\hat{W}_M$ to estiamte $\mathbb{E}[W_M]$ with no bias, i.e., $\mathbb{E}[\hat{W}_M] = \mathbb{E}[W_M]$. The mean square error (MSE) of the estimator $\hat{W}_M$ is thus

$$\text{MSE} = |\mathbb{E}[W - W_M]|^2 + \text{Var}(\hat{W}_M) \tag{3.2}$$

Now we discuss various approaches to construct the estimator $\hat{W}_M$. Using multilevel Monte Carlo technique, we are able to reduce the computational cost from $\mathcal{O}(\varepsilon^{-3})$ (standard Monte Carlo) to $\mathcal{O}(\varepsilon^{-2}(\log\varepsilon)^2)$ (multilevel Monte Carlo).

3.1 Standard Monte Carlo

Standard Monte Carlo uses the empirical average of independent samples to construct $\hat{W}_M$, i.e.,

$$\hat{W}_M = \frac{1}{N}\sum_{i=1}^{N} W_M^{(i)} \tag{3.3}$$

where $\{W_M^{(i)}\}_{i=1}^N$ are N independent samples of the random variable W_M. Standard Monte Carlo is the simplest choice to construct $\hat{W}_M$, and is employed in the numerical tests in Section 1. The MSE of standard Monte Carlo is

$$\text{MSE} = |\mathbb{E}[W - W_M]|^2 + \frac{1}{N}\text{Var}(W_M) \lesssim h^{2\alpha} + \frac{1}{N} \tag{3.4}$$

To achieve $\mathcal{O}(\varepsilon^2)$ MSE, one need to choose

$$M \gtrsim \varepsilon^{-\frac{1}{\alpha}}, \quad N \gtrsim \varepsilon^{-2} \tag{3.5}$$

and the computational cost is thus

$$\mathcal{C}_\varepsilon(\hat{Q}_M) \gtrsim \varepsilon^{-2-\gamma/\alpha} \tag{3.6}$$

In the case of the Feynman-Kac formula, the complexity is at least $\mathcal{O}(\varepsilon^{-3})$.

3.2 Multilevel Monte Carlo

Multilevel Monte Carlo employs consecutive levels of estimators to approximate $\mathbb{E}[W_M]$. Introduce an increasing sequence $M_0 < M_1 < \cdots < M_L := M$ such that

$$M_0 = 1, \quad M_l = sM_{l-1}, \quad l = 1, \cdots, L \tag{3.7}$$

where $s \in \mathbb{N}$ is a given integer ($s = 2$ in numerical tests), and L denotes the number of levels. The idea of multilevel Monte Carlo is to write $\mathbb{E}[W_M]$ as the summation of consecutive levels,

$$\mathbb{E}[W_M] = \mathbb{E}[W_{M_0}] + \sum_{l=1}^{M} \mathbb{E}[W_{M_l} - W_{M_{l-1}}] = \sum_{l=0}^{L} \mathbb{E}[Y_l] \tag{3.8}$$

where the random variables $\{Y_l\}_{l=0}^L$ of different levels are defined by

$$Y_0 = W_{M_0}, \quad Y_l = W_{M_l} - W_{M_{l-1}}, \quad l = 1, \cdots, L \tag{3.9}$$

Multilevel Monte Carlo does not approximate $\mathbb{E}[W_M]$ as a whole, but approximates different levels $\{\mathbb{E}[Y_l]\}_{l=0}^L$ separately. At the l-th level, $\mathbb{E}[Y_l]$ is approximated by $\hat{Y}_l$, the empirical average of N_l independent samples, i.e.,

$$\hat{Y}_l = \frac{1}{N_l}\sum_{i=1}^{N_l} Y_l^{(i)}, \quad l = 0, 1, \cdots, L \tag{3.10}$$

and the overall estimator of multilevel Monte Carlo is

$$\hat{W}_M = \sum_{l=0}^{L} \hat{Y}_l \tag{3.11}$$

Let $\mathcal{C}_l = \mathcal{C}(Y_l)$ be the complexity at the l-th level, and $V_l = \text{Var}(Y_l)$ be the variance of Y_l. Then $\mathcal{C}_l \lesssim M_l^\gamma$, and Y_l has the upper bound

$$V_l \leqslant \mathbb{E}|Y_l|^2 \leqslant 2\Big(\mathbb{E}|W_{M_l} - W|^2 + \mathbb{E}|W - W_{M_{l-1}}|^2\Big) \leqslant Ch^\beta \tag{3.12}$$

The MSE of this estimator is

$$\mathrm{MSE} = |\mathbb{E}[W - W_M]|^2 + \sum_{l=0}^{L} N_l^{-1} V_l, \tag{3.13}$$

and the total computational cost is

$$\mathcal{C}(\hat{W}_M) = \sum_{l=0}^{L} N_l \mathcal{C}_l \tag{3.14}$$

To achieve $\mathcal{O}(\varepsilon^2)$ MSE, it is required that

(A) $M = s^L$ is sufficiently large so that $|\mathbb{E}[W - W_M]| \leqslant \dfrac{\varepsilon}{\sqrt{2}}$;

(B) The numbers of samples $\{N_l\}_{l=0}^{L}$ satisfies $\displaystyle\sum_{l=0}^{L} N_l^{-1} V_l \leqslant \frac{\varepsilon^2}{2}$.

and we aim to minimize the total computational cost $\mathcal{C}(\hat{W}_M)$. Using Cauchy's inequality, $\{N_l\}_{l=0}^{L}$ can be chosen as

$$N_l = \left\lceil 2\varepsilon^{-2} \sqrt{V_l/\mathcal{C}_l} \left(\sum_{l=0}^{L} \sqrt{\mathcal{C}_l V_l} \right) \right\rceil \propto s^{-(\beta+\gamma)l/2} \tag{3.15}$$

Based on the discussion above, the complexity of multilevel Monte Carlo method is concluded in the following theorem [6]:

Theorem 4 (complexity) *Suppose there are constants $\alpha, \beta, \gamma > 0$ such that $\alpha \geqslant \frac{1}{2}\min(\beta, \gamma)$ and*

1. $|\mathbb{E}[W - W_{M_l}]| \lesssim M_l^{-\alpha}$;
2. $V_l := \mathrm{Var}(Y_l) \lesssim M_l^{-\beta}$;
3. $\mathcal{C}_l := \mathcal{C}(W_{M_l}) \lesssim M_l^{\gamma}$.

Then for any $\varepsilon < e^{-1}$, there exists $L \in \mathbb{N}$ and a sequence $\{N_l\}_{l=0}^{L}$ such that

$$\mathrm{MSE} = \mathbb{E}[\hat{W}_M - \mathbb{E}[W]]^2 \leqslant \varepsilon^2 \tag{3.16}$$

and the total computational cost of generating a sample $\hat{W}_M$ is

$$\mathcal{C}_\varepsilon(\hat{W}_M) \lesssim \begin{cases} \varepsilon^{-2}, & \beta > \gamma \\ \varepsilon^{-2}(\log \varepsilon)^2, & \beta = \gamma \\ \varepsilon^{-2-(\gamma-\beta)/\alpha}, & \beta < \gamma \end{cases} \tag{3.17}$$

We perform a simple discussion on the complexity of multilevel Monte Carlo:

1. As long as $\beta > 0$, the complexity of multilevel Monte Carlo is less than $\mathcal{O}(\varepsilon^{-2-\gamma/\alpha})$, the complexity of standard Monte Carlo.
2. At the l-th level, the number of samples $N_l \propto s^{-(\beta+\gamma)l/2}$. That is, the number of samples N_l decreases as the level number l increases. Intuitively, when l is large, the random variable $Y_l = W_{M_l} - W_{M_{l-1}}$ has small variance, thus fewer samples are needed to estimate $\mathbb{E}[Y_l]$.
3. In the case of Feynman-Kac formula, $\alpha = \beta = \gamma = 1$, hence the complexity of multilevel Monte Carlo is $\mathcal{O}(\varepsilon^{-2}(\log \varepsilon)^2)$, which is much more efficient than standard Monte Carlo.

3.3 Algorithm Design

The complexity theorem above provides a theoretical result to show that multilevel Monte Carlo is more efficient. To implement multilevel Monte Carlo in practice, one should choose the number of levels L and the numbers of samples $\{Y_l\}_{l=0}^L$ according to the requirements (A)(B).

In (A), the number of iterations $M = s^L$ should satisfy $|\mathbb{E}[W_M - W]| \leqslant \varepsilon/\sqrt{2}$. However, $\mathbb{E}[W]$ itself is unknown, hence (A) cannot be used directly. Inspired from the numerical results of standard Monte Carlo, we may assume

$$\mathbb{E}[W_{M_l} - W] \approx \frac{C}{M_l}, \quad l = 0, 1, 2, \cdots \tag{3.18}$$

then

$$\mathbb{E}[W_{M_{l-1}} - W] \approx \frac{C}{M_{l-1}} = \frac{sC}{M_l} \tag{3.19}$$

Taking the differece, we obtain

$$\mathbb{E}[Y_l] \approx \frac{(1-s)C}{M_l}, \quad l = 1, 2, \cdots \tag{3.20}$$

hence (A) can be replaced by

$$|\mathbb{E}[Y_L]| \leqslant \frac{(s-1)\varepsilon}{\sqrt{2}} \tag{3.21}$$

Since Y_L has small variance, the expectation $\mathbb{E}[Y_L]$ can be computed conveniently.

In (B), we may compute the variance $V_l = \mathrm{Var}(Y_l)$ numerically and choose $\{N_l\}_{l=0}^L$ as in (3.15). In summary, the multilevel Monte Carlo can be implemented as follows:

Algorithm 1: Multilevel Monte Carlo

Input: error tolerance ε, initial number of samples $\bar{N}$
Output: estimator $\hat{W}_M$ with $\mathcal{O}(\varepsilon^2)$ MSE
for $L = 0, 1, 2, \cdots$ **do**
  Compute $E_L := \mathbb{E}[Y_L]$ and $V_L := \mathrm{Var}(Y_L)$ using $\bar{N}$ independent samples.
  if $L \geqslant 1$ *and* $|E_L| \leqslant (s-1)\varepsilon/\sqrt{2}$ **then**
    Calculate the optimal numbers of samples $\{Y_l\}_{l=0}^L$ by

$$N_l = \left\lceil 2\varepsilon^{-2}\sqrt{V_l/\mathcal{C}_l}\left(\sum_{l=0}^{L}\sqrt{\mathcal{C}_l V_l}\right)\right\rceil \propto s^{-(\beta+\gamma)l/2}$$

    Output the multilevel estimator

$$\hat{W}_M = \sum_{l=0}^{L}\left(\frac{1}{N_l}\sum_{i=1}^{N_l} Y_l^{(i)}\right)$$

  end
end

As a comparison, the standard Monte Carlo is implemented by

Algorithm 2: Standard Monte Carlo

Input: error tolerance ε, initial number of samples $\bar{N}$
Output: estimator $\hat{W}_M$ with $\mathcal{O}(\varepsilon^2)$ MSE
for $L = 0, 1, 2, \cdots$ **do**
  Compute $E_L := \mathbb{E}[Y_L]$ and $V_L' := \mathrm{Var}(W_{M_L})$ using $\bar{N}$ independent samples.
  if $L \geqslant 1$ *and* $|E_L| \leqslant (s-1)\varepsilon/\sqrt{2}$ **then**
    Choose the numbers of samples as

$$N_L = \lceil 2\varepsilon^{-2} V_L' \rceil$$

    Output the standard estimator

$$\hat{W}_M = \frac{1}{N_L} \sum_{i=1}^{N_l} W_M^{(i)}$$

  end
end

Choices of the number of samples $\{N_l\}_{l=0}^L$ (multilevel Monte Carlo) and N_L (standard Monte Carlo) ensure that the variance of the estimator $\hat{W}_M$ is less than $\varepsilon^2/2$.

4 Numerical Experiments

We compare the performance of standard Monte Carlo and multilevel Monte Carlo in the two parabolic equations (2.16)(2.19). The initial sample size $\bar{N}$ is fixed at 100. Different error tolerance ε are employed.

For the heat equation with source (2.16) with $d = 8$ and $t = 1$, we report the CPU time of two Monte Carlo methods and the variance sequence $\{V_l\}_{l=0}^L$ in multilevel Monte Carlo.

CPU time/s	$\varepsilon = 1/32$	$\varepsilon = 1/64$	$\varepsilon = 1/128$	$\varepsilon = 1/256$
standard MC	0.293	3.61	23.8	149
multilevel MC	0.213	0.857	2.78	11.66

Table 3: CPU time of standard Monte Carlo and multilevel Monte Carlo with different error tolerance ε for the heat equation with source (2.16).

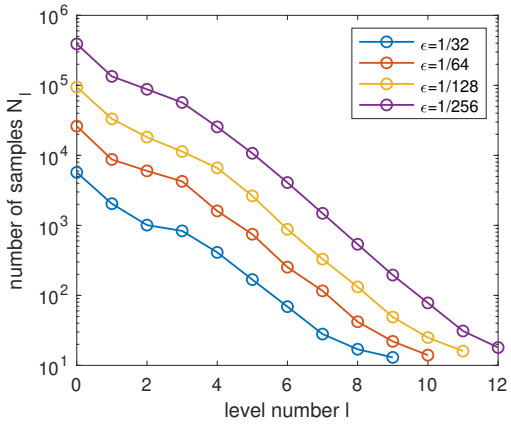

Figure 3: Number of samples $\{N_l\}_{l=0}^{L}$ of multilevel Monte Carlo with different error tolerance ε for the heat equation with source (2.16).

For the heat equation with linear drift (2.19) with $d = 8$ and $t = 1$, we report the CPU time of two Monte Carlo methods and the variance sequence $\{V_l\}_{l=0}^{L}$ in multilevel Monte Carlo.

CPU time/s	$\varepsilon = 1/32$	$\varepsilon = 1/64$	$\varepsilon = 1/128$	$\varepsilon = 1/256$
standard MC	5.39	45.5	578	>3600
multilevel MC	0.96	4.28	14.4	55.6

Table 4: CPU time of standard Monte Carlo and multilevel Monte Carlo with different error tolerance ε for the heat equation with linear drift (2.19).

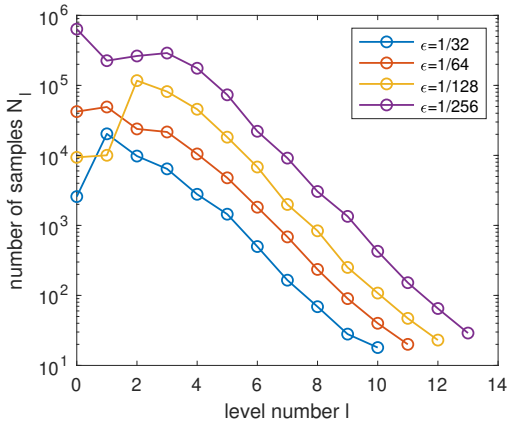

Figure 4: Number of samples $\{N_l\}_{l=0}^{L}$ of multilevel Monte Carlo with different error tolerance ε for the heat equation with linear drift (2.19).

5 Conclusion

The multilevel Monte Carlo provides a simple technique to accelerate the stochastic simulation in the Feynman-Kac formula, and can be applied to solve high-dimensional parabolic equations at a given point $(x,t) \in \mathbb{R}^d \times \mathbb{R}^+$. Numerical results show that the multilevel Monte Carlo is able to reduce the computational cost from $\mathcal{O}(\varepsilon^{-3})$ to $\mathcal{O}(\varepsilon^{-2}(\log \varepsilon)^2)$.

A Proof of Error Analysis

Under the assumption $q(x) \equiv 0$, the random variables W, W_M are given by

$$W = u_0(X_t) + \int_0^t f(X_s, t-s)\mathrm{d}s, \qquad W_M = u_0(X_M^h) + h\sum_{k=1}^{M} f(X_k^h, (M-k)h) \tag{A.1}$$

Since the weak convergence order of the Euler-Maruyama scheme is 1, we immediately obtain

$$|\mathbb{E}u_0(X_t) - \mathbb{E}u_0(X_M^h)| \leqslant Ch \tag{A.2}$$

In order to show $|\mathbb{E}W - \mathbb{E}W_M| \leqslant Ch$, we only need to verify

$$\left| h\mathbb{E}[f(X_k^h, (M-k)h)] - \mathbb{E}\int_{(k-1)h}^{kh} f(X_s, t-s)\mathrm{d}s \right| \leqslant Ch^2, \quad k = 1, \cdots, M \tag{A.3}$$

Again from the weak order of convergence, (A.3) is equivalent to

$$\left| h\mathbb{E}[f(X_{kh}, (M-k)h)] - \mathbb{E}\int_{(k-1)h}^{kh} f(X_s, t-s)\mathrm{d}s \right| \leqslant Ch^2 \tag{A.4}$$

Let $v(s) = \mathbb{E}f(X_s, t-s)$, then (A.4) reduces to

$$\left| v(kh) - \int_{(k-1)h}^{kh} v(s)\mathrm{d}s \right| \leqslant Ch^2 \tag{A.5}$$

Since $v(s)$ is differentiable in s, we have

$$\left| v(kh) - \int_{(k-1)h}^{kh} v(s)\mathrm{d}s \right| \leqslant \int_{(k-1)h}^{kh} |v(kh) - v(s)|\mathrm{d}s \leqslant \int_{(k-1)h}^{kh} Ch\mathrm{d}s = Ch^2, \tag{A.6}$$

and we finally obtain $|\mathbb{E}W - \mathbb{E}W_M| \leqslant Ch$. Since $u_0(x)$ is Lipschitz continuous,

$$\mathbb{E}|u_0(X_t) - u_0(X_M^h)|^2 \leqslant L^2\mathbb{E}|X_t - X_M^h|^2 \leqslant Ch \tag{A.7}$$

In order to show $\mathbb{E}|W - W_M|^2 \leqslant Ch$, we only need to verify

$$\mathbb{E}\left| \sum_{k=1}^{M} \left(hf(X_k^h, (M-k)h) - \int_{(k-1)h}^{kh} f(X_s, t-s)\mathrm{d}s \right) \right|^2 \leqslant Ch \tag{A.8}$$

Using Cauchy's inequality, we only need to prove

$$\mathbb{E}\left|hf(X_k^h,(M-k)h)-\int_{(k-1)h}^{kh}f(X_s,t-s)\mathrm{d}s\right|^2\leqslant Ch^2,\quad k=1,\cdots,M \tag{A.9}$$

Again using Cauchy's inequality, we only need to prove

$$\mathbb{E}\left|f(X_k^h,(M-k)h)-f(X_s,t-s)\right|^2\leqslant Ch \tag{A.10}$$

which is obvious from the strong convergence order $1/2$.

References

[1] Mark Kac. *Enigmas of chance: An autobiography*. Univ of California Press, 1987.

[2] Michael B Giles. Multilevel monte carlo methods. *Acta Numer.*, 24:259–328, 2015.

[3] Michael B Giles and Christoph Reisinger. Stochastic finite differences and multilevel monte carlo for a class of spdes in finance. *SIAM Journal on Financial Mathematics*, 3(1):572–592, 2012.

[4] Michael B Giles and Lukasz Szpruch. Multilevel monte carlo methods for applications in finance. *High-Performance Computing in Finance*, pages 197–247, 2018.

[5] E Weinan, Tiejun Li, and Eric Vanden-Eijnden. *Applied stochastic analysis*, volume 199. American Mathematical Soc., 2019.

[6] K Andrew Cliffe, Mike B Giles, Robert Scheichl, and Aretha L Teckentrup. Multilevel monte carlo methods and applications to elliptic pdes with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.