Hans-Peter SCHREI

The Multilevel Monte Carlo Method and the Wiener-Hopf Factorisation

MASTERARBEIT

zur Erlangung des akademischen Grades eines Diplom-Ingenieur

Masterstudium Finanz- und Versicherungsmathematik

Technische Universität Graz

Betreuer:
O.Univ.-Prof. Dr.phil. Robert F. Tichy

Institut für Analysis und Computational Number Theory

Graz, im März 2014

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ....
(Unterschrift)

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

....
date (signature)

Contents

Introduction and Acknowledgements

The following thesis attempts to give a comprehensive treatment of the Multilevel Monte Carlo method introduced by [Giles, 2008a] and the Wiener-Hopf Multilevel Monte Carlo method introduced by [Ferreiro-Castilla et al., 2013]. In the last chapter, a short overview of Multilevel Quasi-Monte Carlo methods is given. All implementations were done in MATLAB 2013a.

I especially want to thank my advisors Professor Robert Tichy and Dr. Markus Hofer for their patience and their excellent guidance and supervision.

Lastly I want to thank my parents for their ubiquitous and never-ending support.

Graz, 16.03.2014

Chapter 1

Stochastic Processes

1.1 General Introduction

Randomness permeates financial markets. Shares in stock of a company are traded on markets, where buyers and sellers determine their price. Myriads of different factors like speculation, information flow and macroeconomic trends drive market sentiment in an unpredictable manner. Thus, there is inherent randomness present in the markets at all times, in the traded assets as well as in the derivatives based on them. The mathematical model for such random price movements in time are stochastic processes.

Definition 1. *A stochastic process is a collection of random variables $(X_t)_{t\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. That is, for each time instant $t \geq 0$, $\omega \to X_t(\omega)$ is a measurable function from Ω to a measure space $(E, \mathcal{E})$.*

In the following only the case $E = \mathbb{R}^d$ for $d \geq 1$ with the Borel sigma-algebra $\mathcal{E} = B(\mathbb{R}^d)$ is considered, as all processes of interest with regard to Monte Carlo simulation take their values in the n-dimensional real space $\mathbb{R}^d$.

Sometimes the interpretation of a stochastic process as a random variable of which the realisations are paths in $\mathbb{R}^d$ is of interest, that is, the stochastic process is understood as a measurable function $X : \Omega \to \left(\mathbb{R}^d\right)^{[0,\infty)}$ from the probability space Ω into the space $\left(\mathbb{R}^d\right)^{[0,\infty)}$ of all $\mathbb{R}^d$-valued functions on $[0, \infty)$.

The mathematical model for a stochastic process does not only include the process itself, but also a dynamic system of information in the context of which the process is observed. In the system, more and more information becomes accessible as time goes on. To strictly define the principles of causality and predictability in such a system the concept of a filtration is introduced. Essentially, a filtration encapsulates the information available at each time instant $t \geq 0$.

Definition 2. *A filtration $(\mathcal{F}_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of sub-sigma-algebras of $\mathcal{F}$ satisfying $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s \leq t$.*

Thus, a filtration allows us to draw a clear distinction between quantities for which the value can be derived from the available information at a specific time point t and quantities for which the outcome is still uncertain. Intuitively, the set $\mathcal{F}_t$ is a snapshot of the information available in a model at time $t \geq 0$. Conditioning on $\mathcal{F}_t$ for different $t \geq 0$ changes the underlying probabilities according to their relation to the information contained in $\mathcal{F}_t$. A process $(X_t)_{t\geq 0}$ for which each X_t is observable at time $t \geq 0$ is called adapted.

Definition 3. *A stochastic process process $(X_t)_{t\geq 0}$ is adapted if X_t is an $\mathcal{F}_t$-measurable random variable for each time instant $t \geq 0$.*

The set of functions measurable with respect to the sigma-algebra generated by a set of real-valued functions may be characterised as the smallest set of real-valued functions which contains the generating set and which is closed under taking linear combinations and increasing limits.

Definition 4. *The class of predictable processes is characterised by the smallest set which contains the adapted left-continuous processes and which is closed under taking linear combinations and increasing limits.*

Lemma 1. *Alternatively, the set of predictable processes is generated by the continuous and adapted processes or by sets of the form*

$$\{(s,t] \times A : t > s \geq 0, A \in \mathcal{F}_s\} \cup \{\{0\} \times A : A \in \mathcal{F}_0\}.$$

The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taken together with the filtration $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called a filtered probability space. It is assumed that the filtration under consideration fulfils the usual conditions.

Definition 5. *A filtration $\mathcal{F}$ is said to be right-continuous if*

$$\mathcal{F}_t = \mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$$

holds. It is said to be complete if $\mathcal{F}$ contains all subsets of zero probability elements of $\mathcal{F}$. It fulfils the usual conditions if it is right-continuous and complete.

In general, it is always possible to enlarge a filtration in a minimal way to one satisfying the usual conditions. In many instances modifications of stochastic processes that are right continuous and for which the left limit exists at every time instance are of interest.

Definition 6. *A process $(Y_t)_{t\geq 0}$ is a version or modification of $(X_t)_{t\geq 0}$ if $\mathbb{P}(X_t = Y_t) = 1$ holds for each time instant $t \geq 0$. A process $(Y_t)_{t\geq 0}$ is said to be indistinguishable from a process $(X_t)_{t\geq 0}$, if $\mathbb{P}(X_t = Y_t \ \forall t \geq 0) = 1$.*

Definition 7. *A stochastic process $(X_t)_{t\geq 0}$ taking values in the metric space $(\mathbb{R}^d, B(\mathbb{R}^d))$ is called a càdlàg process, short for continue à droite, limite à gauche, if $\lim_{s\uparrow t} X_s < \infty$ and if $\lim_{s\downarrow t} X_s = X_t$ for all time instances $t \geq 0$. The space of càdlàg functions $D([0,\infty), \mathbb{R}^d)$ is called the Skorokhod space.*[1]

If $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are right-continuous processes such that they are modifications of each other, they are additionally indistinguishable.

An important class of stochastic processes are martingales. A martingale is a process which stays the same on average. That is, the expected future value conditional on the present is equal to the current value.

Definition 8. *A martingale $(X_t)_{t\geq 0}$ is an adapted process with $\mathbb{E}(|X_t|) < \infty$ for all $t \geq 0$ satisfying*

$$X_s = \mathbb{E}(X_t | \mathcal{F}_s).$$

In many cases, however, processes under consideration are close to satisfying the martingale property, but are not actually martingales. This is for example the case when certain limiting processes are considered or when stochastic integration with respect to martingales is performed. It is therefore necessary to generalise the concept of a martingale to the concept of a locale martingale. Localisation is done via stopping times, which are positive random variables used to stop a stochastic process at a clearly defined, yet random time.

Definition 9. *A stopping time is a map $\tau : \Omega \to [0,\infty) \cup \{\infty\}$ such that $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \geq 0$.*

Definition 10. *Let P be a class of stochastic processes. Then, a process $(X_t)_{t\geq 0}$ is locally in P if there exists a sequence of stopping times $\tau_n \uparrow \infty$ such that the stopped processes*

$$\mathbb{1}_{\{\tau_n \geq 0\}} X^{\tau_n}$$

are in P. The sequence τ_n is called a localizing sequence for X with respect to P. A process is a local martingale if it is locally in the class of càdlàg martingales.

A further important class of processes are semimartingales which are the composition of locale martingales and finite variation processes. Semimartingales are good integrators in the sense that they are the largest class of processes with respect to which the Itō integral can be defined. The class of semimartingales is quite large and includes for example all continuously differentiable processes. In order to introduce semimartingales, the concept of variation is needed.

Definition 11. *The variation of a function $g : [a, b] \to \mathbb{R}$ on $[a, b]$ is given by*

$$V(g)_{[a,b]} = \sup_{\pi_n} \sum_{k=1}^{n} |g(t_k) - g(t_{k-1})|.$$

The supremum is taken over all finite partitions $\pi_n : a = t_0 < t_1 < \cdots < t_n = b$ of $[a, b]$.

A function $g : [0, \infty) \to \mathbb{R}$ is said to be of bounded variation on compact intervals if $V(g)_t = V(g)_{[0,t)} < \infty$ for all $t \geq 0$.

A stochastic process $V = (V_t)_{t\geq 0}$ is said to be a process of bounded variation if $V : [0, \infty) \to \mathbb{R}$ is of bounded variation almost surely.

Definition 12. *A process $(X_t)_{t\geq 0}$ defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called a semimartingale if it can be decomposed as*

$$X_t = M_t + V_t$$

where $(M_t)_{t\geq 0}$ is a local martingale and $(V_t)_{t\geq 0}$ is a càdlàg adapted process of locally bounded variation.

In particular, many financial models are built on one particular stochastic process, namely Brownian motion. Due to its tractable dynamics, it became an indispensable tool for the modelling of assets in financial markets. Apart from its ubiquitous use in mathematical finance, Brownian motion is an essential stochastic process for modelling in the natural sciences, where it also has its origins and from where it derives its name as a result of the seminal work of the botanist Robert Brown.

Definition 13. *A process $(B_t)_{t\geq 0}$ is a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ if it is an adapted process for which*

1. *$B_0 = 0$,*
2. *$B_t - B_s$ is independent of $\mathcal{F}_s$ for each $t > s \geq 0$, following a normal distribution with mean zero and variance $t - s$ and*
3. *for which the sample paths are continuous.*

1.2 Lévy Processes

Another special class of stochastic processes often used in the modelling of financial markets and used as a basis of risk models in insurance theory are Lévy processes. The diffusion processes analysed in Chapter 5 do not fully capture the dynamics observed in actual financial markets and are therefore of limited use in modelling the essential characteristics necessary for derivative pricing. In particular, the occurrence of sudden jumps are not taken into consideration when using diffusion models. Jumps have wide implications on the resulting properties of the model, a survey can be found in [Tankov and Cont, 2004]. Although some properties of real markets can be mimicked by diffusion models, the same properties already form an integral part of Lévy models.

Heavy tails in the distribution of asset price processes observed in the market are often used to justify the choice of Lévy processes in favour of diffusion processes. While heavy tails are certainly more natural to Lévy processes, they are also attainable in diffusion processes through the the use of highly non-stationary diffusion coefficients in local volatility models or through improbably high values for the parameters of the volatility process in stochastic volatility models.

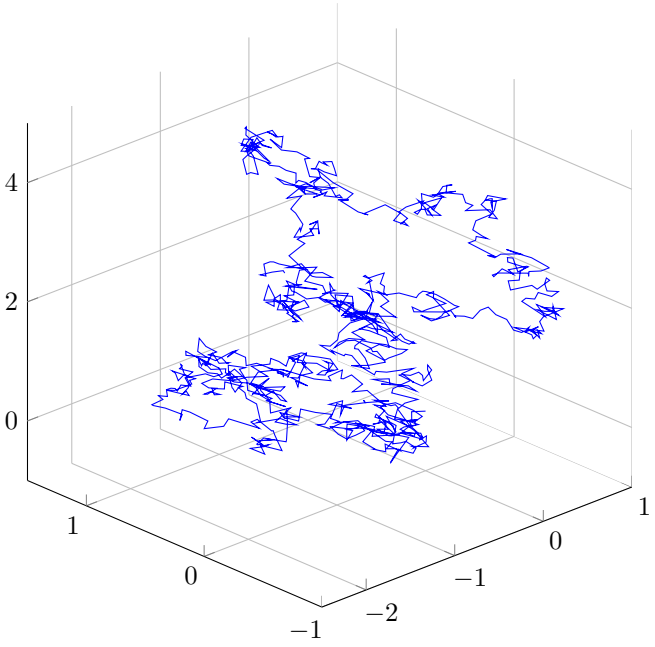

Figure 1.1: Three-dimensional Brownian motion

However, whereas stochastic volatility models can reasonably reproduce the implied volatility surface [Gatheral, 2006], their viability is not so evident when looking at the whole term structure of implied volatilities [Rebonato, 1999]. In contrast, Lévy models cannot only fully capture the term structure, they also offer a reasonable interpretation in the form of the risk averse behaviour of investors with regards the possibility of large, sudden jumps in the markets.

Furthermore, local volatility models and stochastic volatility are complete market models or can be easily extended to be complete, whereas Lévy models, which in general are incomplete market models, accurately reproduce the real world hedging issues that result from market incompleteness [Tankov and Cont, 2004].

Definition 14. *A stochastic process $(X_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^d$ such that $X_0 = 0$ is called a Lévy process if it possesses the following properties.*

1. ***Independent increments.*** *For every increasing sequence of time instances $t_0, \ldots, t_n$ the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.*
2. ***Stationary increments.*** *The law of $X_{t+\Delta t} - X_t$ does not depend on t.*
3. ***Stochastic continuity.*** *X is a stochastic continuous process, that is*

$$\lim_{\Delta t \to 0} \mathbb{P}(|X_{t+\Delta t} - X_t| \geq \epsilon) = 0$$

holds.

Due to the fact that a càdlàg version of a Lévy process always exists [Protter, 1992], it can be assumed that a càdlàg version is always used in the following.

Theorem 1. *Let $(X_t)_{t\geq 0}$ be a Lévy process. There exists a unique modification $(Y_t)_{t\geq 0}$ of $(X_t)_{t\geq 0}$ which is càdlàg and which is also a Lévy process.*

Since a Lévy process has stationary and independent increments, sampling of the process at regular time intervals leads to a random walk. Thus, any X_t at a specific time instance $t > 0$ can be decomposed into the sum of $n \in \mathbb{N}$ independent and identically distributed random variables, which have the same distribution as $X_{t/n}$. The distribution of a random variable which can be represented by the sum of n independent and identically distributed random variables is called an infinitely divisible distribution.

Definition 15. *A probability distribution F on $\mathbb{R}^d$ is said to be infinitely divisible if for any integer $n \geq 2$ there exist n independent and identically distributed random variables $Y_1, \ldots, Y_n$ such that $\sum_{i=1}^n Y_i$ has distribution F.*

Not only does every Lévy process lead to on an infinitely divisible distribution, the converse holds as well. Every infinitely divisible distribution F can be used to construct a Lévy process X such that the law of X_1 is given by F.

Proposition 1. *Let $(X_t)_{t\geq 0}$ be a Lévy process. Then on the one hand the random variable X_t has an infinitely divisible distribution for every $t > 0$. On the other hand, if F is an infinitely divisible distribution then there exists a Lévy process $(X_t)_{t\geq 0}$ such that the distribution of X_1 is given by F.*

A basic building block for characterizing Lévy processes is the Lévy measure, which gives the expected number of jumps $\Delta X_t = X_t - \lim_{s\downarrow t} X_s$ per unit time whose size belongs to $A \in B(\mathbb{R}^d)$.

Definition 16. *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$. The measure ν on $\mathbb{R}^d$ defined by*

$$\nu(A) = \mathbb{E}(|\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\}|)$$

with $A \in B(\mathbb{R}^d)$ is called the Lévy measure of $(X_t)_{t\geq 0}$.

A measure built on the Lévy measure is the jump measure, which extends the time interval in which the number of jumps is observed to an arbitrary length.

Definition 17. *Let $E \subseteq \mathbb{R}^d$. A Radon measure on (E, B) is a measure μ such that for every compact measurable set $A \in B$, $\mu(A) < \infty$.*

Definition 18. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $E \subseteq \mathbb{R}^d$ and μ a given positive Radon measure μ on (E, B). A Poisson random measure on E with intensity measure μ is an integer value random measure $M : \Omega \times B \to \mathbb{N}$ such that*

1. *For almost all $\omega \in \Omega$, $M(\omega, .)$ is an integer-valued Radon measure on E. For any bounded measurable $A \subseteq E$, $M(A) < \infty$ is an integer valued random variable.*
2. *For each measurable set $A \subseteq E$, $M(., A) = M(A)$ is a Poisson random variable with parameter $\mu(A)$,*

$$\mathbb{P}(M(A) = k) = e^{-\mu(A)} \frac{\mu(A)^k}{k!}$$

   for all $k \in \mathbb{N}$.

3. *For disjoint measurable sets $A_1, \ldots, A_n \in B$, the variables $M(A_1), \ldots, M(A_n)$ are independent.*

These jump measure J_X is a Poisson random measure on $[0, \infty) \times \mathbb{R}^d$ with intensity $\nu(dx)dt$.

Definition 19. *Let $(X_t)_{t\geq 0}$ be a Lévy process with Lévy measure ν. Its jump measure J_X is a Poisson random measure on $[0, \infty) \times \mathbb{R}^d$ with intensity measure $\mu(dx \times dt) = \nu(dx)dt$.*

The following theorem gives the basic stepping stone to the characterisation of Lévy processes, the Lévy-Itō-decomposition.

Proposition 2. *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ and ν its Lévy measure. Then ν is a Radon measure on $\mathbb{R}^d \setminus \{0\}$ that verifies*

$$\int_{|x|\leq 1} |x|^2 \nu(dx) < \infty \quad \text{and} \quad \int_{|x|\geq 1} \nu(dx) < \infty. \tag{1.1}$$

Additionally, there exist a vector $\gamma \in \mathbb{R}^d$ and a d-dimensional Brownian motion $(B_t)_{t\geq 0}$ and covariance matrix $A \in \mathbb{R}^{d\times d}$ such that

$$X_t = \gamma t + B_t + X_t^l + \lim_{\epsilon\downarrow 0} X_t^\epsilon$$

with

$$X_t^l = \int_{|x|\geq 1, s\in[0,t]} xJ_X(ds \times dx),$$
$$X_t^\epsilon = \int_{\epsilon\leq|x|<1, s\in[0,t]} x(J_X(ds \times dx) - \nu(dx)ds).$$

From the Lévy-Itō decomposition it is a short step to the Lévy-Khinchin representation.

Theorem 2. *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic triplet (A, ν, γ). Then*

$$\mathbb{E}\left(e^{izX_t}\right) = e^{t\psi(z)}$$

with $z \in \mathbb{R}^d$ and

$$\psi(z) = -\frac{1}{2}zAz + i\gamma z + \int_{\mathbb{R}^d}(e^{izx} - 1 - izx\mathbb{1}_{|x|\leq 1})\nu(dx).$$

The notion of a subordinator is a further concept used in the latter part of the text. Just as Lévy processes can be understood as an extension of discrete random walks into continuous time, subordinators can be understood as the continuous time version of discrete random walks with non-negative increments.

Definition 20. *A subordinator $X = (X_t)_{t\geq 0}$ is a real-valued, non-decreasing Lévy process starting from zero.*

Lemma 2. *If a Lévy process $X = (X_t)_{t\geq 0}$ is of bounded variation, it can be represented as the difference of two independent subordinators.*

Proof. Set $H_t^1 = \frac{1}{2}(V(X)_t + X_t)$ and $H_t^2 = \frac{1}{2}(V(X)_t - X_t)$. Both $H^1 = (H_t^1)_{t\geq 0}$ and $H^2 = (H_t^2)_{t\geq 0}$ are non-decreasing and positive. In general, H^1 and H^2 are not unique, another possible choice would be $H_t^1 = V(X)_t$ and $H_t^2 = X_t - V(X)_t$. □

Chapter 2

Risk Neutral Pricing

In this section, the use of an expectation to calculate the price of a derivative is justified. That an expectation is the right instrument to price a financial contract is a non-trivial result. Rather than using the expectation under a real world measure $\mathbb{P}$ of some presupposed asset dynamics to find prices, it is vital in financial markets to price derivatives correctly in relation to each other. The conventional expectation under the real world measure could be correct only by a fortunate streak of luck, as the market price would then coincide with its long-term average given by the expectation. That the price at any given point in time is identical to its historical average would be pure coincidence.

Rather, the fundamental principle enforcing the price of tradeable assets is market opportunism, that is, the idea that market participants take immediately advantage of any discrepancies in the relative valuation of incorrectly priced different instruments. An opportunity for riskless profit, also called arbitrage opportunity, thus instantly vanishes in an instant. While this might not be the case in real markets, this assumption is reasonable in mathematical models, for in the presence of arbitrage opportunities a potential for riskless unlimited loss, respectively gain, would emerge.

A change of measure to a so called risk neutral measure $\mathbb{Q}$ guarantees that the expected values of derivative payoffs produce a pricing framework that is free of arbitrage.

2.1 The Market Setup

The introduced market model is a generic, conventional market model used in financial theory based on the treatment presented in [Bingham and Kiesel, 2004].

In a first step, the market model is assumed to be very simple in order to ease the presentation. All these assumptions are in general not faithful representations of real markets.

Definition 21. *A market is called frictionless, if there are no transaction costs, no bid/ask spreads, no taxes, no margin requirements and no restrictions on short sales in place.*

As described in Chapter 1, the uncertainty present in a financial market is modelled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is assumed that $\mathcal{F}_0$ is generated only by null sets under $\mathbb{P}$, that is sets with $\mathbb{P}(A) = 0$, and that $\mathcal{F}_T = \mathcal{F}$.

It is assumed that the market consists of $d+1$ primary traded assets, for which the stochastic price processes are denoted by $S^0, \ldots, S^d$. The overall vector process $S = (S^0, \ldots, S^d)$ is furthermore assumed to follow an adapted, continuous and strictly positive semi-martingale on $(\Omega, \mathcal{F}, \mathbb{P})$.

The implicit reference asset used for transactions in a domestic market is chosen to be denoted by S^0. The asset is required to be a numéraire and it is assumed that it is a non-dividend paying asset [2].

Definition 22. *A numéraire is a price process $(S_t^0)_{t\geq 0}$ that is almost surely strictly positive for each $t \geq 0$.*

The central aim in introducing the market is the pricing and, to a lesser degree, hedging of derivatives. These derivatives or contingent claims are modelled as $\mathcal{F}_T$-measurable random variables, when $[0,T]$ is the time frame under consideration. For hedging purposes, the goal is to establish trading strategies that replicate the contingent claim.

Definition 23. *A trading strategy $\varphi = (\varphi_t)_{t\geq 0}$ is a $\mathbb{R}^{d+1}$-valued, predictable, locally bounded process $\varphi_t = (\varphi_t^0, \ldots, \varphi_t^d)$ with*

$$\int_0^T \mathbb{E}(\varphi_t^0)dt < \infty \quad \textit{and} \quad \sum_{i=0}^{d} \int_0^T \mathbb{E}\left((\varphi_t^i)^2\right) < \infty.$$

The $\{\varphi_t^i\}_{i\in\{0,\ldots,d\}}$ are interpreted to be the number shares of asset $i \in \{0,\ldots,d\}$ held in the portfolio at time $t \in [0,T]$. Predictability ensures that the holder of the portfolio constructs it on the basis of the information available in $\mathcal{F}_{t-}$.

Definition 24. *The value $V(\varphi) = (V_t(\varphi))_{t\geq 0}$ of the portfolio φ at time $t \in [0,T]$ is given by the scalar product*

$$V_t(\varphi) = \varphi_t \cdot S_t = \sum_{i=0}^{d} \varphi_t^i S_t^i.$$

The process V is called the value process of the trading strategy. A gains process $G_t(\varphi)$ is defined by

$$G_t(\varphi) = \int_0^t \varphi_u dS_u = \sum_{i=0}^{d} \int_0^t \varphi_u^i dS_u^i.$$

It is convenient to single out trading strategies where all changes in the portfolio value are due to movements in the assets, in contrast to trading strategies which allow for the withdrawal or injection of funds.

Definition 25. *A trading strategy φ is called self-financing if the value process $V_t(\varphi)$ satisfies*

$$V_t(\varphi) = V_0(\varphi) + G_t(\varphi)$$

for all $t \in [0,T]$.

Self-financing portfolios are invariant under change of the numéraire, which gives sufficient flexibility in the choice of the numéraire. The following proposition was proved in full generality in [Delbaen and Schachermayer, 1995].

Proposition 3. *Self-financing portfolios remain self-financing after a numéraire change.*

To simplify the notation, discounted processes are introduced. These processes are the asset and the value processes divided by the numéraire process and the gains process of the discounted assets. That is

$$\tilde{S}_t = \frac{S_t}{S_t^0}, \quad \tilde{V}_t(\varphi) = \frac{V_t(\varphi)}{S_t^0} \quad \text{and} \quad \tilde{G}_t(\varphi) = \sum_{i=1}^{d} \int_0^t \varphi_t^i d\tilde{S}_t^i.$$

The same definition of self-financing portfolios used for undiscounted processes holds in an analogous manner for discounted processes.

2.2 The Arbitrage Principle

As noted above, trading strategies of particular interest are strategies which achieve with positive probability a riskless profit. In the constructed mathematical model these strategies would allow with positive probability for unlimited risk-less profit.

Definition 26. *A self-financing trading strategy φ is called an arbitrage strategy if the value process $V(\varphi)$ satisfies the set of conditions*

$$V_0(\varphi) = 0, \quad \mathbb{P}(V_T(\varphi) \geq 0) = 1 \quad \textit{and} \quad \mathbb{P}(V_T(\varphi) > 0) > 0. \tag{2.1}$$

The main instrument in developing the risk neutral pricing framework are equivalent martingale measures.

Definition 27. *A probability measure $\mathbb{Q}$ defined on $(\Omega, \mathcal{F})$ is an equivalent martingale measure, respectively strong equivalent martingal measure if*

1. *$\mathbb{Q}$ is equivalent to $\mathbb{P}$,*
2. *and if the discounted price process $\tilde{S}$ is a $\mathbb{Q}$-local martingale, respectively a $\mathbb{Q}$-martingale.*

The set of all equivalent martingale measures is denoted by $\mathcal{P}$.

In order to eliminate certain arbitrage opportunities from the market, further restrictions are placed on the available set of self-financing strategies.

Definition 28. *A self-financing trading strategy φ is called C-admissible relative to the numéraire S^0 if $\tilde{V}_t(\varphi) \geq -C$ for a positive constant $C \in \mathbb{R}$ for all $t \in [0, T]$ and called admissible if it is C-admissible for some $C \in \mathbb{R}$. The set of all admissible strategies is denoted by Φ.*

It turns out that for admissible strategies the discounted value process is a supermartingale. A more general proposition holds.

Proposition 4. *Let S be a local martingale and let $\int \varphi dS < \infty$. If $\int \varphi dS$ is bounded below, then $\int \varphi dS$ is a local martingale. Every local martingale that is bounded below is a supermartingale.*

Proof. To show the first statement, observe that because $Z = \int_0^t X dS \geq -C$ for positive $C \in \mathbb{R}$, the negative part of the integral Z^- is bounded above by C. Due to a standard result in probability theory, Z is a local martingale if and only if the process $\sup_{0 \leq s \leq t} Z_s^-$ is locally integrable [Lowther, 2010].

For the second statement let M be a local martingale and $M_t \geq -C$ for all $t \geq 0$. Without loss of generality, it can be assumed that $C = 0$. Let $(T_n)_{n \in \mathbb{N}}$ be a localisation sequence, that is the stopped processes $(M_t^{T_n})_{t \geq 0}$ are martingales. For $0 \leq s \leq t$, $A \in \mathcal{F}_s$ and arbitrary $N \in \mathbb{N}$

$$\mathbb{E}\left(\mathbb{1}_{A \cap \{M_s^{T_n} \leq N\}} M_s^{T_n}\right) = \mathbb{E}\left(\mathbb{1}_{A \cap \{M_s^{T_n} \leq N\}} M_t^{T_n}\right)$$

holds. Applying dominated convergence on the left hand side and Fatou's Lemma on the right hand side yield together with taking the limit $n \to \infty$

$$\mathbb{E}\left(\mathbb{1}_{A \cap \{M_s \leq N\}} M_s\right) = \liminf_{n \to \infty} \mathbb{E}\left(\mathbb{1}_{A \cap \{M_s^{T_n} \leq N\}} M_t^{T_n}\right) \geq \mathbb{E}\left(\liminf_{n \to \infty} \mathbb{1}_{A \cap \{M_s^{T_n} \leq N\}} M_t^{T_n}\right) \geq \mathbb{E}\left(\mathbb{1}_{A \cap \{M_s \leq N\}} M_t\right).$$

Monotone convergence yields for $N \to \infty$ the result $\mathbb{E}(\mathbb{1}_A M_s) \geq \mathbb{E}(\mathbb{1}_A M_t)$ for $0 \leq s \leq t$, $A \in \mathcal{F}_s$. □

The main result in this section is that the existence of equivalent martingale measures guarantees that there are no arbitrage strategies in Φ.

Theorem 3. *Assume that $\mathcal{P} \neq \emptyset$. Then the market model contains no arbitrage opportunities in Φ.*

Proof. For any $\varphi \in \Phi$ under any $\mathbb{Q} \in \mathcal{P}$, the discounted value process $\tilde{V}_t(\varphi)$ is a supermartingale by proposition (4),

$$\mathbb{E}_{\mathbb{Q}}\left(\tilde{V}_t(\varphi) | \mathcal{F}_u\right) \leq \tilde{V}_u(\varphi)$$

for all $u \leq t \leq T$. In order for a $\varphi^* \in \Phi$ to be an arbitrage opportunity, it is necessary for $\tilde{V}(\varphi^*)$ to satisfy $\tilde{V}_0(\varphi^*) = 0$ according to the conditions in (2.1). Thus

$$\mathbb{E}_{\mathbb{Q}}\left(\tilde{V}_t(\varphi^*)\right) \leq 0$$

for all $t \in [0,T]$. Since $\varphi^* \in \Phi$, $\tilde{V}_t(\varphi^*) \geq -C$ with positive $C \in \mathbb{R}$ holds for all $t \in [0,T]$, implying in particular $\mathbb{E}_{\mathbb{Q}}\left(\tilde{V}_T(\varphi^*)\right) \leq 0$, which in turn implies $\mathbb{P}(V_T(\varphi) > 0) = 0$.

An arbitrage strategy φ^* additionally fulfils $\mathbb{P}\left(V_T(\varphi^*) \geq 0\right) = 1$, and, since $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, $\mathbb{Q}(V_T(\varphi^*) \geq 0) = 1$. In combination the aforementioned results yield

$$\mathbb{Q}(V_T(\varphi^*) > 0) = \mathbb{P}(V_T(\varphi^*) > 0) = 0$$

and thus the third condition in (2.1) is not fulfilled. The strategy φ^* cannot be an arbitrage strategy. □

The converse, that absence of an arbitrage strategy implies the existence of equivalent martingale measures, is more difficult to prove. Both directions taken together form a fundamental theorem in asset pricing theory, which in a general form was proven in [Delbaen and Schachermayer, 1994].

To formulate it, the notions of a simple predictable trading strategy and a free lunch with vanishing risk are required.

Definition 29. *A simple predictable trading strategy is a predictable process that can be represented as a finite linear combination of process of the form $\psi \mathbb{1}_{(\tau_1,\tau_2]}$, where τ_1 and τ_2 are stopping times and ψ is a $\mathcal{F}_{\tau_1}$-measurable random variable.*

Definition 30. *A price process S satisfies the condition of No Free Lunch with Vanishing Risk if for any $\{\varphi_n\}_{n\in\mathbb{N}}$ of simple predictable trading strategies such that φ_n is C_n-admissible with $\lim_{n\to\infty} C_n = 0$ satisfies $\lim_{n\to\infty} V_T(\varphi_n) = 0$ in probability.*

Fortunately, the condition allows for an evident economic interpretation. In the case S does not satisfy the condition there exists for every small enough $\epsilon > 0$ an admissible strategy φ^ϵ such that $V_T(\varphi^\epsilon) > -\epsilon$ and $\mathbb{P}(V_T(\varphi^\epsilon) > 0) > 0$. That is, if a free lunch with vanishing risk exists, the holder of the portfolio is willing to take an arbitrarily small, yet nonetheless positive loss, in order to achieve a positive probability of receiving a strictly positive portfolio value.

Theorem 4. *Let S be a locally bounded semimartingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Then S satisfies the condition of No Free Lunch with Vanishing Risk if and only if there exists an equivalent probability measure $\mathbb{Q}$ such that S is a local martingale with respect to it.*

The theorem can be relaxed even further using the concept of σ-martingales and σ-martingale measures [Delbaen and Schachermayer, 1998].

2.3 Risk Neutral Pricing

Let $\mathbb{Q}$ be a strong equivalent martingale measure, implying that there are no arbitrage opportunities with respect to Φ. In a market in which such a martingale measure exists, it is of interest to price derivatives, which are referred to as contingent claims in this setting.

Definition 31. *A contingent claim H with maturity T is an arbitrary $\mathcal{F}_T$-measurable random variable. The class of all contingent claims is denoted by $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$.*

Pricing in a risk neutral framework is done by replicating the payoff of a contingent claim by a trading strategy.

Definition 32. *A contingent claim H is called attainable if there exists at least on admissible trading strategy such that $V_T(\varphi) = H$. Such a trading strategy φ is called a replicating strategy for H.*

A financial market model is said to be complete if any contingent claim is attainable.

The concept of attainability does not depend on the numéraire used: If a contingent claim is attainable in a given numéraire, it is also attainable in any other numéraire and the replicating strategies are the same.

If a claim H is attainable trough the use of a strategy $\varphi \in \Phi$, a market participant is indifferent between holding the claim and holding the portfolio under the assumptions of the simplified market model that is used. Bar any possibility of arbitrage, the price process $\pi_t(H)$ of a contingent claim must therefore be equal to the value process of the trading strategy φ, $\pi_t(H) = V_t(\varphi)$. As can be expected, the price process is intimately connected with the equivalent martingale measure. The connection is characterised through the risk neutral valuation formula.

Theorem 5. *The arbitrage price process of any attainable claim is given by the risk neutral valuation formula*

$$\pi_t(H) = S_t^0 \mathbb{E}_{\mathbb{Q}} \left(\frac{H}{S_T^0} \Big| \mathcal{F}_t \right).$$

Proof. Because H is attainable, there exists replicating strategy $\varphi \in \Phi$ such that $V_T(\varphi) = H$ and $\pi_t(H) = V_t(\varphi)$ for all $t \in [0, T]$. Since $\varphi \in \Phi$ the discounted value process $(\tilde{V}_t(\varphi))_{t \geq 0}$ is a martingale, leading for $t \in [0, T]$ to

$$\pi_t(H) = V_t(\varphi) = S_t^0 \tilde{V}_t(\varphi) = S_t^0 \mathbb{E}_{\mathbb{Q}} \left(\tilde{V}_T(\varphi) \Big| \mathcal{F}_t \right) = S_t^0 \mathbb{E}_{\mathbb{Q}} \left(\frac{V_T(\varphi)}{S_T^0} \Big| \mathcal{F}_t \right) = S_t^0 \mathbb{E}_{\mathbb{Q}} \left(\frac{H}{S_T^0} \Big| \mathcal{F}_t \right).$$

□

Given a particular equivalent martingale measure $\mathbb{Q}$, the resulting price is unique.

Corollary 1. *For any two replicating portfolios $\varphi, \psi \in \Phi$, the relation $V_t(\varphi) = V_t(\psi)$ holds for all $t \in [0, T]$.*

The theorem and the corollary assume that a martingale measure $\mathbb{Q}$ is already given and does not mention how it should be chosen in the presence of other available equivalent martingale measures. However, it can be shown that for two equivalent martingale measures $\mathbb{Q}_1$ and $\mathbb{Q}_2$, the uniqueness result

$$\mathbb{E}_{\mathbb{Q}_1} \left(\frac{H}{S_T^0} \Big| \mathcal{F}_t \right) = \mathbb{E}_{\mathbb{Q}_2} \left(\frac{H}{S_T^0} | \mathcal{F}_t \right)$$

holds for replicating strategies $\varphi_1 \in \Phi(\mathbb{Q}_1)$ and $\varphi_2 \in \Phi(\mathbb{Q}_2)$. Even more generally, it is true that

$$\pi_0(H) = \sup_{\mathbb{Q} \in \mathcal{P}} S_0^0 \mathbb{E}_{\mathbb{Q}} \left(\frac{H}{S_T^0} | \mathcal{F}_t \right) = \inf_{\theta(H)} V_0(\varphi),$$

where $\theta(H)$ is the class of all admissible trading strategies replicating H.

When only the price of a given derivative is of interest, it is sufficient to find an equivalent martingale measure. If it is necessary to hedge the contingent claim under consideration, it is important to analyse the specific nature of the replicating portfolio.

Lemma 3. *Assume that the discounted contingent claim H/S^0 is $\mathbb{Q}$-integrable. If the $\mathbb{Q}$-martingale $(M_t)_{t \geq 0}$ defined by*

$$M_t = \mathbb{E}_{\mathbb{Q}} \left(\frac{H}{S_T^0} \Big| \mathcal{F}_t \right)$$

for $t \in [0, T]$ admits an integral representation of the form

$$M_t = M_0 + \sum_{i=1}^{d} \int_0^t \varphi_u^i d\tilde{S}_u^i,$$

where $\varphi^1, \ldots, \varphi^d$ are predictable and locally bounded, then H is attainable.

Proof. The goal is to find a trading strategy $\varphi \in \Phi$ such that $V_T(\varphi) = H$. Any set of predictable processes $\varphi^1, \ldots, \varphi^d$ such that the stochastic integrals $\int \varphi^i d\tilde{S}^i$, $i = 1, \ldots, d$ exist can be uniquely extended to a self-financing strategy φ with specified initial value $\tilde{V}_0(\varphi) = M_0$ by setting

$$\varphi_t^0 = v + \sum_{i=1}^{d} \int_0^t \varphi_u^i d\tilde{S}_u^i - \sum_{i=1}^{d} \varphi_t^i \tilde{S}_t^i$$

for $t \in [0, T]$. The strategy φ is then a strategy with

$$\tilde{V}_t(\varphi) = M_0 + \tilde{G}_t(\varphi) = M_0 + \sum_{i=1}^{d} \int_0^t \varphi_u^i d\tilde{S}_u^i = M_t.$$

M is a non-negative martingale, which means that $\varphi \in \Phi$, and by definition

$$V_T(\varphi) = S_T^0 \tilde{V}_T(\varphi) = S_T^0 M_T = S_T^0 \frac{H}{S_T^0} = H.$$

□

Chapter 3

Discretisation Schemes

Stochastic modelling in the sciences and finance is largely done by formulating continuous stochastic processes which serve as the mathematical description of some natural, non-deterministic phenomenon. However, in order to simulate a continuous stochastic process one has to restrict oneself again to a finite number of time point evaluations by the necessity imposed by finite computational resources. Prominent discretisation schemes for Itō stochastic differential equations are the Euler-Maruyama scheme, the Milstein scheme and Runge-Kutta schemes.

Definition 33. *Let $\Delta t = \frac{T}{n}$ be the size of an interval in a partition of $[0,T]$ into $n \in \mathbb{N}$ intervals of equal length.*

A process $\{Y_{i\Delta t}\}_{0\leq i\leq n}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called a time discrete approximation with step size Δt on $[0,T]$, if it is a process such that $Y_{j\Delta}$ is $\mathcal{F}_{j\Delta t}$-measurable for all $j \in \{1,\dots,n\}$ and expressible as a function of $Y_0, Y_{\Delta t}, \dots, Y_{j\Delta t}$ and a finite number l of $\mathcal{F}_{j\Delta t}$-measurable random variables Z_{jk}, $k \in \{1,\dots,l\}$ for all $j \in \{1,\dots,n\}$.

A continuous extension of $\{Y_{i\Delta t}\}_{0\leq k\leq n}$, denoted by $(\tilde{Y}_t)_{t\geq 0}$, is a càdlàg process that coincides with $\{Y_{i\Delta t}\}_{0\leq i\leq n}$ at the time instances $\{i\Delta t\}_{0\leq i\leq n}$.

A process $\{Y_{i\Delta t}\}_{0\leq i\leq n}$ may be extended by different variants of interpolation, which is usually done out of convenience by working in continuous time. The piecewise linear interpolant $(\tilde{Y}_t)_{t\geq 0}$ is defined by

$$\tilde{Y}_{i\Delta t+\theta\Delta t} = (1-\theta)Y_{i\Delta t} + \theta Y_{(i+1)\Delta t}$$

for $0 \leq \theta < 1$ and $i \in \{1,\dots,n\}$. If no explicit interpolation method is given in a particular context in the following text, it is implicitly assumed that piecewise linear interpolation is used.

Usually two notions of error are used to measure the quality of discretisation schemes. The weak error of a discretisation scheme measures the proximity of the discretised process to the continuous process based on their respective distributions whereas the strong error measures the proximity of the processes pathwise.

Formally, a discrete approximation scheme is said to strongly converge against a process $X = (X_t)_{t\geq 0}$ on $[0,T]$ if

$$\lim_{\Delta t\to 0} \mathbb{E}(|X_T - Y_{n\Delta t}|) = 0$$

holds. A strongly convergent scheme is said to have convergence rate γ if for some constant $C \in \mathbb{R}$

$$\mathbb{E}(|X_T - Y_{n\Delta t}|) \leq C(\Delta t)^\gamma$$

holds.

An approximation process Y is converging weakly against a process X on $[0,T]$ with respect to a class of measurable functions H, if

$$\lim_{\Delta t\to 0} |\mathbb{E}(P(X_T) - P(Y_{n\Delta t}))| = 0$$

for all $P \in H$. Intuitively, Y_T converges to X_T in distribution. A weakly convergent scheme with respect to the class H is said to have convergence rate α if for some constant $C \in \mathbb{R}$

$$|\mathbb{E}(P(X_T) - P(Y_{n\Delta t}))| \leq C(\Delta t)^\alpha \tag{3.1}$$

holds for all $P \in H$.

In the following a d-dimensional process $X = (X_t)_{t\geq 0}$ satisfying a stochastic differential equation of the form

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \tag{3.2}$$

is considered with $X_0 = x_0 \in \mathbb{R}^d$ fixed, a m-dimensional Brownian motion $(W_t)_{t\geq 0}$ and measurable coefficient functions $a : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$, $b : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ fulfilling the usual conditions ensuring the existence and uniqueness of a strong solution. [3]

To lighten the notation in some instances the frequently used operators

$$L^0 = \frac{\partial}{\partial t} + \sum_{k=1}^{d} a^{(k)} \frac{\partial}{\partial x^k} + \frac{1}{2} \sum_{k,l=1}^{d} \sum_{j=1}^{m} b^{(k,j)} b^{(l,j)} \frac{\partial^2}{\partial x^k \partial x^l}$$

$$\underline{L}^0 = \frac{\partial}{\partial t} + \sum_{k=1}^{d} \underline{a}^{(k)} \frac{\partial}{\partial x^k}$$

$$L^j = \sum_{k=1}^{d} b^{(k,j)} \frac{\partial}{\partial x^k}$$

are defined for $j \in \{1, \ldots, m\}$.

3.1 The Euler-Maruyama Scheme

The Euler-Maruyama Scheme on a given interval $[0, T]$ is derived by performing a stochastic Taylor approximation. Assuming an equidistant discretisation of the interval $[0, T]$ into n time steps of size $\Delta t = T/n$ applying Itō's Lemma to the two coefficient functions $a(t, X_t)$ and $b(t, X_t)$ leads to the relation

$$\begin{aligned}
X_{t+\Delta t}^{(i)} &= X_t^{(i)} + \int_t^{t+\Delta t} a^{(i)}(s, X_s)ds + \int_t^{t+\Delta t} \sum_{k=1}^{m} b^{(i,k)}(s, X_s)dW_s^{(k)} \\
&= X_t^{(i)} + \int_t^{t+\Delta t} \left[a^{(i)}(t, X_t) + \int_t^s L^0 a^{(i)}(u, X_u)du + \sum_{j=1}^{m} \int_t^s L^j a^{(i)}(u, X_u)dW_u^{(j)} \right] ds \\
&\quad + \sum_{k=1}^{m} \int_t^{t+\Delta t} \left[b^{(i,k)}(t, X_t) + \int_t^s L^0 b^{(i,k)}(u, X_u)du + \sum_{j=1}^{m} \int_t^s L^j b^{(i,k)}(u, X_u)dW_u^{(j)} \right] dW_s^{(k)} \\
&= X_t^{(i)} + \int_t^{t+\Delta t} a^{(i)}(t, X_t)ds + \sum_{k=1}^{m} \int_t^{t+\Delta t} b^{(i,k)}(t, X_t)dW_t^{(k)} + \psi^{(i)} \\
&= X_t^{(i)} + a^{(i)}(t, X_t)\Delta t + \sum_{k=1}^{m} b^{(i,k)}(t, X_t)\Delta W_t^{(k)} + \psi^{(i)},
\end{aligned} \tag{3.3}$$

where $\Delta W_t = W_{t+\Delta t} - W_t$ and where the time and Brownian integrals of multiplicity one are explicitly expressed and the remaining integrals are collected under the term ψ. By discarding the terms subsumed under ψ for small Δt the Euler-Maruyama approximation

$$\begin{aligned}
Y_{t+\Delta t}^{(i)} &= Y_t^{(i)} + a^{(i)}(t, Y_t)\Delta t + \sum_{k=1}^{m} b^{(i,k)}(t, X_t)\Delta W_t^{(k)} \\
&= Y_t^{(i)} + a^{(i)}(t, Y_t)\Delta t + \sum_{k=1}^{m} b^{(i,k)}(t, X_t)\sqrt{\Delta t} Z^{(k)}
\end{aligned} \tag{3.4}$$

is obtained, where $Z^{(k)}$ is a standard normally distributed random variable.

Algorithm 1 The Euler-Maruyama Scheme

Let $\Delta t = T/n$ for a given $n \in \mathbb{N}$.
The Euler-Maruyama approximation at the discretisation points is calculated via the following steps.

(1) **Set** $Y_0 = X_0 = x_0$.

(2) **For** $j = 0$ to $n-1$ **do**

Generate a sample of a m-dimensional standard normal variable $Z \sim N(0, I_d)$.

Calculate for every $i \in \{1, \dots, d\}$ of the d-dimensional process Y

$$Y^{(i)}_{(j+1)\Delta t} = Y^{(i)}_{j\Delta t} + a^{(i)}(j\Delta t, Y_{j\Delta t})\Delta t + \sum_{k=1}^{m} b^{(i,k)}(j\Delta t, Y_{j\Delta t})\sqrt{j\Delta t}Z^{(k)}.$$

The Euler-Maruyama method converges strongly with order 1/2 under Lipschitz and linear growth conditions on the coefficient functions [Kloeden and Platen, 1995]. However, in special cases the method may reach a higher order of convergence. In case the diffusion function $b(t,x)$ is additive, i.e., $b(t,x) = b(t)$, the Euler-Maruyama scheme has under further smoothness assumptions on the coefficient functions an order of strong convergence of 1.

Theorem 6. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(X_t)_{t\geq 0}$ be a stochastic process of the form (3.2) and $(Y_t)_{t\geq 0}$ its Euler-Maruyama approximation with step size Δt as given in (3.4). Requiring of the initial random variable*

$$\mathbb{E}(|X_0|^2) < \infty,$$
$$\mathbb{E}(|X_0 - Y_0|^2)^{\frac{1}{2}} \leq C_1(\Delta t)^{\frac{1}{2}}$$

for a constant $C_1 \in \mathbb{R}$ and requiring of the coefficient functions the Lipschitz and linear growth conditions for all $s, t \in [0,T]$ and $x, y \in \mathbb{R}^d$

$$|a(t,x) - a(t,y)| + |b(t,x) - b(t,y)| \leq C_2|x-y|,$$
$$|a(t,x)| + |b(t,x)| \leq C_3(1+|x|),$$
$$|a(s,x) - a(t,x)| + |b(s,x) - b(t,x)| \leq C_4(1+|x|)|s-t|^{\frac{1}{2}},$$

for constants $C_2, C_3, C_4 \in \mathbb{R}$ independent of Δt, the estimate

$$\mathbb{E}(|X_T - Y_T|^p)^{\frac{1}{p}} \leq C_5(\Delta t)^{\frac{1}{2}}$$

holds for the Euler-Maruyama approximation $(Y_t)_{t\geq 0}$ with the constant $C_5 \in \mathbb{R}$ independent of Δt and for $p \geq 2$. Additionally for the whole path over $[0,T]$ the result

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |X_t - Y_t|^p\right)^{\frac{1}{p}} \leq C_p(\Delta t)^{\frac{1}{2}}. \tag{3.5}$$

holds for some constant C_p and for $p \geq 2$, which includes the weaker statement that for any $\delta > 0$ and $p \geq 2$ a constant $C_p^\delta \in \mathbb{R}$ exists such that

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |X_T - Y_T|^p\right) \leq C_p^\Delta(\Delta t)^{\frac{p}{2}-\delta}. \tag{3.6}$$

Furthermore there exists a constant $C_p \in \mathbb{R}$ such that

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |X_T - Y_T|^p\right) \leq C_p|\Delta t \log \Delta t|^{\frac{p}{2}}.$$

Considering weak convergence, the Euler-Maruyama scheme has a convergence of order 1. Since the order of weak convergence of the Euler-Maruyama scheme is on par with the order of weak convergence of the Milstein scheme presented below, it is usually sensible to use the Euler-Maruyama scheme when weak convergence is sufficient.

Theorem 7. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(X_t)_{t\geq 0}$ be a stochastic process of the form (3.2) and $(Y_t)_{t\geq 0}$ its Euler-Maruyama approximation with step size Δt as given in (3.4).*

Suppose that the coefficient functions $a(t,x) = (a^{(i)}(t,x))_{1\leq i\leq d}$ and $b(t,x) = (b^{(i,k)}(t,x))_{1\leq i\leq d, 1\leq k\leq m}$ are Lipschitz continuous with components $a^{(i)}(t,x)$, $b^{(i,k)}(t,x)$ that are 4 times continuously differentiable, are polynomially bounded and have partial derivatives of up to and including order 4 that are polynomially bounded. Then there exists for each function $P \in C^4_{pol}(\mathbb{R}^d, \mathbb{R}^d)$ which has partial derivatives of up to and including order 4 that are polynomially bounded, a constant C_P independent of Δt such that

$$|\mathbb{E}(P(X_T) - P(Y_T))| \leq C_P \Delta t.$$

Thus, the Euler-Maruyama approximation Y converges weakly with order 1 to X.

3.2 The Milstein Scheme

The Milstein scheme is a refinement of the Euler-Maruyama scheme in that it uses an additional term in the Taylor expansion. Whereas the drift term in the Euler-Maruyama scheme (3.4) is of order $O(\Delta t)$, the diffusion term is only of order $O\left(\Delta t^{\frac{1}{2}}\right)$. The Milstein approximation makes use of the fact that a further expansion of the diffusion term increases the order of the term to $O(\Delta t)$ as well, thus increasing the overall order to $O(\Delta t)$.

Before the diffusion coefficient function in (3.3) is expanded further to derive the Milstein scheme, some definitions and relations are given.

For $j, k \in \{1, \ldots, m\}$ let I_{jk} denote the double integral against two independent Brownian motions $W^{(j)}$ and $W^{(k)}$ over the interval $[t, t+\Delta t]$, i.e.

$$I_{jk} = \int_t^{t+\Delta t} \int_t^s dW_u^{(j)} dW_s^{(k)}.$$

By Itō's Lemma and associativity of the integral

$$\begin{aligned}
I_{jk} + I_{kj} &= \int_t^{t+\Delta t} \int_t^s dW_u^{(j)} dW_s^{(k)} + \int_t^{t+\Delta t} \int_t^s dW_u^{(k)} dW_s^{(j)} \\
&= \int_t^{t+\Delta t} \int_t^s dW_u^{(j)} d\left(\int_t^s W_u^{(k)}\right) + \int_t^{t+\Delta t} \int_t^s dW_u^{(k)} d\left(\int_t^s W_u^{(j)}\right) \\
&\qquad + \left[\int_t^{t+\Delta t} dW_s^{(j)}, \int_t^{t+\Delta t} dW_s^{(j)}\right] - \left[\int_t^{t+\Delta t} dW_s^{(j)}, \int_t^{t+\Delta t} dW_s^{(j)}\right] \\
&= \int_t^{t+\Delta t} dW_s^{(j)} \int_t^{t+\Delta t} dW_s^{(k)} - \Xi_{jk} \Delta t \\
&= \Delta W_t^{(j)} \Delta W_t^{(k)} - \Xi_{jk} \Delta t, \qquad (3.7)
\end{aligned}$$

holds, where Ξ denotes the correlation matrix of the m Brownian motions. Usually, m-dimensional standard Brownian motions are considered, where the correlation matrix Ξ coincides with the m-dimensional identity matrix.

Furthermore the Lévy area [4] $A_t^{(j,k)}$ is defined by the difference of two double integrals,

$$\begin{aligned} A_t^{(j,k)} &= I_{jk} - I_{kj} \\ &= \int_t^{t+\Delta t} \left(W_s^{(j)} - W_t^{(j)}\right) dW_s^{(k)} - \int_t^{t+\Delta t} \left(W_s^{(k)} - W_t^{(k)}\right) dW_s^{(j)}. \end{aligned} \tag{3.8}$$

Combining (3.7) and (3.8) results in

$$I_{jk} = \frac{1}{2}\left(\Delta W_t^{(j)} \Delta W_t^{(k)} - \delta_{jk}\Delta t + A_t^{(j,k)}\right). \tag{3.9}$$

Expanding the diffusion coefficient function for an additional term in (3.3) using (3.9) yields

$$\begin{aligned} X_{t+\Delta t}^{(i)} &= X_t^{(i)} + \int_t^{t+\Delta t} \left[a^{(i)}(t, X_t) + \int_t^s L^0 a^{(i)}(u, X_u) du + \sum_{j=1}^m \int_t^s L^j a^{(i)}(u, X_u) dW_u^{(j)} \right] ds \\ &\quad + \sum_{k=1}^m \int_t^{t+\Delta t} \left[b^{(i,k)}(t, X_t) + \int_t^s L^0 b^{(i,k)}(u, X_u) du + \sum_{j=1}^m \int_t^s L^j b^{(i,k)}(u, X_u) dW_u^{(j)} \right] dW_s^{(k)} \\ &= X_t^{(i)} + a^{(i)}(t, X_t)\Delta t + \sum_{k=1}^m b^{(i,k)}(t, X_t)\Delta W_t^{(k)} + \sum_{j,k=1}^m L^j b^{(i,k)}(t, X_t) \int_t^{t+\Delta t} \int_t^s dW_u^{(j)} dW_s^{(k)} \qquad (3.10) \\ &\quad + \sum_{j,k=1}^m \int_t^{t+\Delta t} \int_t^s \left[\int_t^u L^0 L^j b^{(i,k)}(v, X_v) dv + \sum_{r=1}^m \int_t^u L^r L^j b^{(i,k)}(v, X_v) dW_v^{(r)} \right] dW_u^{(j)} dW_s^{(k)} + \eta^{(i)} \\ &= X_t^{(i)} + a^{(i)}(t, X_t)\Delta t + \sum_{k=1}^m b^{(i,k)}(t, X_t)\Delta W_t^{(k)} \qquad (3.11) \\ &\quad + \frac{1}{2}\sum_{j,k=1}^m \sum_{l=1}^d \frac{\partial b^{(i,k)}(t, X_t)}{\partial x^{(l)}} b^{(l,j)}(t, X_t) \left(\Delta W_t^{(j)} \Delta W_t^{(k)} - \Xi_{jk}\Delta t + A_t^{(j,k)}\right) + \psi^{(i)}. \end{aligned}$$

The Milstein approximation is accordingly given by

$$\begin{aligned} Y_{t+\Delta t}^{(i)} &= Y_t^{(i)} + a^{(i)}(t, Y_t)\Delta t + \sum_{k=1}^m b^{(i,k)}(t, Y_t)\Delta W_t^{(k)} \\ &\quad + \frac{1}{2}\sum_{j,k=1}^m \sum_{l=1}^d \frac{\partial b^{(i,k)}(t, Y_t)}{\partial y^{(l)}} b^{(l,j)}(t, Y_t) \left(\Delta W_t^{(j)} \Delta W_t^{(k)} - \Xi_{jk}\Delta t + A_t^{(j,k)}\right) \end{aligned} \tag{3.12}$$

and, as the informal reasoning above suggests, has an order of strong convergence of 1 [Kloeden and Platen, 1995]. The additional error bound was proved in [Müller-Gronbach, 2002].

Theorem 8. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(X_t)_{t\geq 0}$ be a stochastic process of the form (3.2) and $(Y_t)_{t\geq 0}$ its Milstein approximation with step size Δt as given in (3.16). Furthermore, define for $k \in \{1, \ldots, d\}$ the Stratonovich coefficient*

$$\underline{a}^{(k)} = a^{(k)} - \frac{1}{2}\sum_{j=1}^m L^{(j)} b^{(k,j)}.$$

Requiring of the initial random variable

$$\mathbb{E}(|X_0|^2) < \infty$$
$$\mathbb{E}(|X_0 - Y_0|^2)^{\frac{1}{2}} \leq C_1 (\Delta t)^{\frac{1}{2}}$$

for a constant $C_1 \in \mathbb{R}$ and requiring of the coefficient functions the Lipschitz and linear growth conditions for all $s,t \in [0,T]$, $x,y \in \mathbb{R}^d$ and $j \in \{0,\dots,m\}$, $j_1, j_2 \in \{1,\dots m\}$

$$\begin{aligned} |\underline{a}(t,x) - \underline{a}(t,y)| &\leq C_2|x-y| & |\underline{a}(t,x)| + |L^j\underline{a}(t,x)| &\leq C_3(1+|x|) \\ |b^{j_1}(t,x) - b^{j_1}(t,y)| &\leq C_2|x-y| & |b^{j_1}(t,x)| + |L^j b^{j_2}(t,x)| &\leq C_3(1+|x|) \\ |L^{j_1}b^{j_2}(t,x) - L^{j_1}b^{j_2}(t,y)| &\leq C_2|x-y| & |L^j L^{j_1}b^{j_2}(t,x)| &\leq C_3(1+|x|) \end{aligned} \tag{3.13}$$

$$\begin{aligned} |\underline{a}(s,x) - \underline{a}(t,x) &\leq C_4(1+|x|)|s-t|^{\frac{1}{2}} \\ |b^{j_1}(s,x) - b^{j_1}(t,x)| &\leq C_4(1+|x|)|s-t|^{\frac{1}{2}} \\ |L^{j_1}b^{j_2}(s,x) - L^{j_1}b^{j_2}(t,x)| &\leq C_4(1+|x|)|s-t|^{\frac{1}{2}}, \end{aligned}$$

where the constants $C_2, C_3, C_4 \in \mathbb{R}$ do not depend on Δt, the estimate

$$\mathbb{E}\left(|X_T - Y_T|\right) \leq C_5 \Delta t$$

holds for the Milstein approximation Y, where the constant $C_5 \in \mathbb{R}$ is independent of Δt. Furthermore there exists a constant $C_p \in \mathbb{R}$ such that

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |X_T - Y_T|^p\right) \leq C_p |\Delta t \log \Delta t|^{\frac{p}{2}}.$$

For $d = 1$, the Lévy area vanishes and the Milstein scheme (3.16) reduces to

$$Y_{t+\Delta t} = Y_t + a(t,Y_t)\Delta t + b(t,Y_t)\Delta W_t + \frac{1}{2}b'(t,Y_t)b(t,Y_t)(\Delta W_t^2 - \Delta t). \tag{3.14}$$

Kloeden and Platen define for $s \in [t, t+\Delta t]$ in [Kloeden and Platen, 1995] the interpolation scheme

$$Y_s^{KP} = Y_t + a(t,Y_t)(s-t) + b(t,Y_t)(W_s - W_t) + \frac{1}{2}b'(t,Y_t)b(t,Y_t)((W_s - W_t)^2 - (s-t)). \tag{3.15}$$

Theorem 9. *Under the conditions of theorem (8), there exists for all $m \in \mathbb{N}$ a constant C_m such that*

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |X_t - Y_t^{KP}|^m\right) < C_m(\Delta t)^m \quad \textit{and} \quad \mathbb{E}\left(\sup_{0\leq t\leq T} |Y_t^{KP}|^m\right) < C_m.$$

The order of weak convergence and the conditions ensuring weak convergence for the Milstein scheme are the same as in the Euler-Maruyama scheme, see Theorem 6.

However, the gain in accuracy obtained by expanding the Taylor expansion to the order of $O(\Delta t)$ in the Milstein scheme turns out to be spurious in the multidimensional case $d > 1$. Whereas the Lévy area vanishes in the case $d = m = 1$ [5], leaving explicit terms which can be readily simulated, the sampling of triples $\left(\Delta W^{(j)}, \Delta W^{(k)}, A^{(j,k)}\right)$ escapes an efficient approach. There is in general no method for sampling the Lévy area conditional on $\Delta W^{(j)}$ and $\Delta W^{(k)}$ exactly and although the characteristic function of the Lévy area may be determined in various ways[6], the Fourier inversion step is not analytically possible in the general case [Scheicher, 2007].

Methods approximating the Lévy area weakly to first order based on the decomposition of the covariance matrix of the finite-dimensional distribution of a Brownian motion are presented in [Scheicher, 2007]. Additionally, one can use the same extrapolation methods as in the deterministic case, for example the Romberg extrapolation method, to improve the accuracy of the approximation.

Kloeden and Platen proved that in order to accomplish a strong convergence rate of $O(\Delta t)$ the mean square error in the approximation of the Lévy areas must be neglegible in comparison to the discretisation error or, to be more precise, must be of order $O(\Delta t^3)$ [Kloeden and Platen, 1995, Corollary 10.6.5]. In [Kloeden et al.,

Algorithm 2 The Milstein Scheme

Let $\Delta t = T/n$ for a given $n \in \mathbb{N}$.
The Milstein approximation at the discretisation points is calculated via the following steps.

(1) **Set** $\hat{Y}_0 = X_0 = x_0$.

(2) **For** $s = 0$ to $n-1$ **do**

Generate a sample of a m-dimensional of a standard normal variable $Z \sim N(0, I_d)$.
Generate a sample of $(\Delta W^{(j)}_{s\Delta t}, \Delta W^{(k)}_{s\Delta t}, A^{(j,k)}_{s\Delta t})$ for every $j,k \in \{1,\dots,m\}$.
Calculate for every $i \in \{1,\dots,d\}$ of the d-dimensional process Y

$$Y^{(i)}_{(s+1)\Delta t} = Y^{(i)}_{s\Delta t} + a^{(i)}(s\Delta t, Y_{s\Delta t})\Delta t + \sum_{k=1}^{m} b^{(i,k)}(t, Y_{s\Delta t})\sqrt{s\Delta t}Z^{(k)} + \frac{1}{2}\sum_{j,k=1}^{m}\sum_{l=1}^{d} \frac{\partial b^{(i,k)}(s\Delta t, Y_{s\Delta t})}{\partial y^{(l)}} b^{(l,j)}(s\Delta t, Y_{s\Delta t})\left(\Delta W^{(j)}_{s\Delta t}\Delta W^{(k)}_{s\Delta t} - \Xi_{jk}\Delta t + A^{(j,k)}_{s\Delta t}\right) \quad (3.16)$$

1992] an algorithm based on truncating the Karhunen–Loève expansion of the Lévy area, which can be derived by looking at the expansion for a Brownian bridge process and which is given by

$$A^{(j,k)}_t = \frac{\Delta t}{\pi}\sum_{l=1}^{\infty}\frac{1}{l}\left(\Upsilon_l\left(\Phi_l - \sqrt{\frac{2}{\Delta t}}\Delta W^{(j)}_{\Delta t}\right) - X_l\left(\Psi_l - \sqrt{\frac{2}{\Delta t}}\Delta W^{(k)}_{\Delta t}\right)\right),$$

where Υ_l, Φ_l, X_l and Ψ_l are independent standard normal variables, was introduced. If the expansion is cut off after s terms the algorithm has a mean square error of $O\left(\frac{1}{s}\Delta t^2\right)$. Subsequently, Wiktorsson refined the algorithm by replacing the tail sum with an appropriate normal random variable $\frac{\sqrt{2}\Delta t}{\pi}\left(\frac{\pi^2}{6} - \sum_{l=1}^{s}\frac{1}{l^2}\right)^{\frac{1}{2}} Z$, where Z is a standard normal variable, in [Wiktorsson, 2001]. This reduces the mean square error to $O\left(\frac{1}{s^2}\Delta t^2\right)$, requiring s only to be proportional to $O\left(\Delta t^{-\frac{1}{2}}\right)$ to achieve the necessary mean square error. In fact the Milstein discretisation combined with the algorithm by Kloeden et al would, at a minimum, require the simulation of the same amount of Gaussian random variables as the Euler discretisation, and would therefore underperform with regards to the Euler discretisation provided that the evaluation of the coefficient functions is not too costly. To achieve strong convergence with a mean error of ϵ both methods require the simulation of $O\left(\epsilon^{-2}\right)$ random variables, whereas the method by Wiktorsson requires the simulation of only $O\left(\epsilon^{-\frac{3}{2}}\right)$ random variables.

Another approximation method for simulating the Lévy area by Wiktorsson is based on an expansion of the area into a series of Laplace random variables [Rydén and Wiktorsson, 2001]. The expansion rests on the idea that the characteristic function of the Lévy area is the product of a logistic random variable and a compound Poisson process where the summands are Laplace random variables. This decomposition allows the series representation

$$A^{(j,k)}_t = \frac{\Delta t}{\pi}\left(X + \sum_{l=1}^{\infty}\frac{1}{l}\sum_{r=1}^{P_l} Y_{rl}\right),$$

where X is a Logistic variable with location parameter 1, P_l are Poisson random variables with parameter

$$\lambda = \frac{\left(\left(W^{(j)}_{\Delta t}\right)^2 + \left(W^{(k)}_{\Delta t}\right)^2\right)}{\Delta t},$$

and where the Y_{rl} are Laplace random variables with parameter 1. All random variables are independent. Again, the series expansion is cut off after s terms and the tail sum replaced by $\frac{\Delta t}{\pi}\sqrt{\frac{2\lambda}{s}}Z$, where Z is a standard normal

variable. The complexity is the same as in the algorithm mentioned above, i.e. s must be of order $O\left(\Delta t^{-\frac{1}{2}}\right)$ to achieve an overall strong convergence rate of $O(\Delta t)$.

As already hinted at above, for a practical implementation the number of simulated random variables is a relevant measure of complexity. Two improvements on the algorithm of Wiktorsson for the case $d = 2$ with regards to this measure were proposed by [Malham and Wiese, 2011], which bring down computational complexity in terms of the number of random variables required for the simulation of the Lévy area to achieve an over all strong convergence with a mean error of ϵ from $O\left(\epsilon^{-\frac{1}{2}}\right)$ to $O\left(\log(\epsilon)\right)$.

An exact sampling approach for the case $d = 2$ was presented in [Gaines and Lyons, 1994]. It is based on the inversion of the joint characteristic function of λ and the Lévy area $A_t^{(j,k)}$ with $\Delta t = 2\pi$ through Marsaglia's rectangle-wedge-tail method. It is the fastest of the methods presented, as the simulation of the triple $\left(\Delta W^{(j)}, \Delta W^{(k)}, A^{(j,k)}\right)$ takes about the same time as the simulation of approximately 15 standard normal variables. A drawback of the algorithm is the rather forbidding complicated implementation. Subsequently, Gaines and Lyons developed an approximation method based on replacing the Lévy area by its conditional expectation on intervening Brownian path information, which in terms of the number of random variables required for the simulation of the Lévy area, has a computational complexity of $O\left(\epsilon^{-1}\right)$ [Gaines and Lyons, 1997].

Further methods to handle the difficulty of simulating the Lévy area in the Milstein discretisation are the elliptic operator approach presented in [Cruzeiro et al., 2004], which under certain conditions circumvents the sampling of the Lévy area altogether, and the combinatorial approach given in [Levin and Wildon, 2008].

For d-dimensional stochastic processes the approach presented in [Rydén and Wiktorsson, 2001] has the greatest relevance to Monte Carlo simulations. The approach is summarised in Algorithm 3.

Simulating the Lévy area efficiently is an important problem. Every discretisation scheme which exclusively relies on the discrete increments of the underlying Brownian motion such as the Euler-Maruyama scheme cannot exceed the order of strong convergence of that scheme, that is, every scheme depending on discrete increments necessarily has an order of strong convergence smaller than 1/2. This was shown to be the case for a particular stochastic differential equation in [Clark and Cameron, 1980] and extended to general stochastic differential equations in [Müller-Gronbach, 2002]. However, the Multilevel Monte Carlo method when used with the Euler-Maruyama scheme achieves in no case the optimal lower computational cost bound, not even when the payoff function is Lipschitz such as in the case of European, Asian and lookback options [Müller-Gronbach and Ritter, 2009]. Therefore application of the Milstein scheme in the context of the Multilevel Monte Carlo method combined with an efficient simulation of the Lévy area is desirable.

3.3 Runge-Kutta Schemes

A disadvantage of the Milstein method is the occurrence of the derivatives in (3.16), which can be difficult to handle in a computationally efficient manner. An alternative to the truncation of the Taylor expansion are Runge-Kutta methods. Just as in the case of Runge-Kutta methods for ordinary differential equation, the partial derivatives involved in Taylor methods are traded in for extra function evaluations resulting from an approximation through the underlying stochastic differential equation. However, although the derivative-free methods for stochastic differential equations are generically called Runge-Kutta methods they are not simple heuristical extensions of the Runge-Kutta methods for deterministic differential equations, as the extensions constructed by exchanging the derivative in a Taylor expansion for the corresponding difference ratios are not strongly consistent [Kloeden and Platen, 1995].

Definition 34. *A discrete approximation scheme to a solution of a stochastic differential equation* (3.2) *with step size* Δt *yielding discretisation values* $\{Y_{k\Delta t}\}_{k\in\{1,\dots,n\}}$ *is called strongly consistent if there exists a non-negative function* $c = c(\Delta t)$ *with* $\lim_{\Delta t\to 0} c(\Delta t) = 0$ *such that*

$$\mathbb{E}\left(\left|\mathbb{E}\left(\frac{Y_{(k+1)\Delta t} - Y_{k\Delta t}}{\Delta t}|\mathcal{F}_{k\Delta t}\right) - a(k\Delta t, Y_{k\Delta t})\right|^2\right) \leq c(\Delta t) \tag{3.17}$$

Algorithm 3 The Wiktorsson-Ryden Iterated Integral Algorithm

Let $\Delta t = T/n$ for a given $n \in \mathbb{N}$, let $s \in \{1, \ldots, n\}$ and $j, k \in \{1, \ldots, d\}$.
The Milstein approximation at the discretisation points is calculated via the following steps.

(1) **Fix** a constant C. Let p be the smallest number such that

$$p \geq \frac{1}{C\pi}\sqrt{\frac{m(m-1)}{24\Delta t}}\sqrt{m + 4\sum_{i=1}^{m}\left(W_{s\Delta t}^{(i)}\right)^2/\Delta t}.$$

(2) **Generate** samples a_{jr}, b_{jr} for $j \in \{1, \ldots, d\}$ and $r \in \{1, \ldots, p\}$ from $N(0, \Delta t/2\pi^2 r^2)$. Set $\eta_{jr}^* = \sqrt{\pi r}a_{jr}$ and $\zeta_{jr}^* = \sqrt{\pi r}b_{jr}$.

(3) **Calculate** the Lévy area and Brownian motion approximations

$$A_{s\Delta t}^{(j,k),p} = \frac{1}{\Delta t}\sum_{r=1}^{p}\zeta_{jr}^*\eta_{kr}^* - \eta_{jr}^*\zeta_{kr}^*$$

and

$$a_{j0}^p = -\sum_{r=1}^{p}\frac{2}{\sqrt{\pi r}}\zeta_{jr}^*.$$

(4) **Calculate** the approximated iterated integral

$$I_{jk}^p = \frac{1}{2}I_jI_k - \frac{1}{2}(a_{k0}^pI_j - a_{j0}^pI_k) + \Delta t A_{s\Delta t}^{(j,k),p}$$

with $I_j = \int_{s\Delta t}^{(s+1)\Delta t} dW_u^{(j)}$.

(5) **Calculate** the tail approximation $\mathcal{A}^{p,tail}$.

Let $x, y \in \mathbb{R}^m$ and $M = \frac{1}{2}m(m-1)$. Denote with e_j^m the j-th basis vector of $\mathbb{R}^m$. Let $\otimes$ denote the tensor product and define $P_m : \mathbb{R}^{m^2} \to \mathbb{R}^M$ with $P_m(x \otimes y) = y \otimes x$ and $K_m : \mathbb{R}^{m^2} \to \mathbb{R}^M$ with $K_m(e_j^m \otimes e_k^m) = e_{v(j,k)}^M$ and $K_m(e_k^m \otimes e_j^m) = 0 = K_m(e_k^m \otimes e_k^m)$, where $j < k$ and $v(j,k)$ is the position of (j,k) in the M term sequence

$$(1,2),(1,3),\ldots,(1,m),(2,3),\ldots,(2,m),\ldots,(m-1,m).$$

Calculate the quantity

$$\Sigma_\infty = 2E_M + \frac{2}{\Delta t}K_m(E_{m^2} - P_m)(E_m \otimes W_{s\Delta t}W_{s\Delta t}^t)(I_{m^2} - P_m)K_m^t,$$

with E_m the $m \times m$ identity matrix and $W_{s\Delta t} = (W_{s\Delta t}^{(1)}, \ldots, W_{s\Delta t}^{(m)})^t$.

Sample G_p from $N(0, E_M)$, calculate $a_p = \sum_{k=p+1}^{\infty}\frac{1}{k^2}$ and calculate

$$\mathcal{A}^{p,tail} = (E_{m^2} - P_m)K_m^t\frac{\Delta t}{2\pi}\sqrt{a_p\Sigma_\infty}G_p.$$

(6) **Calculate** $I_{jk}^{p+tail} = I_{jk}^p + \mathcal{A}_{jk}^{p,tail}$.

and

$$\mathbb{E}\left(\frac{1}{\Delta t}|Y_{(k+1)\Delta t} - Y_{k\Delta t} - \mathbb{E}(Y_{(k+1)\Delta t} - Y_{k\Delta t}|\mathcal{F}_{k\Delta t}) - b(k\Delta t, Y_{k\Delta t})\Delta W_{k\Delta t}|^2\right) \leq c(\Delta t) \tag{3.18}$$

for all fixed values $Y_{k\Delta t}$ *and* $k \in \{0, \ldots, n\}$.

The first condition (3.17) forces the mean of the increment in the discretisation to converge to that of the approximated process. In case no diffusion term appears in the underlying stochastic differential equation (3.2), the condition coincides with the definition of consistency of a one-step scheme of a deterministic differential equation. The second condition requires that the variance of the difference between the diffusion terms of the discretisation and the approximated process converges to zero, which implies strong convergence.

In general, Runge-Kutta schemes derived from Taylor schemes by exchanging the derivatives for the respective difference ratios are not strongly consistent. To show that the derivation of Runge-Kutta schemes is non-trivial the following short example [Kloeden and Platen, 1995] is given.

Definition 35. *A discrete approximation scheme to a solution of a stochastic differential equation* (3.2) *with step size* Δt *defined by*

$$Y_{t+\Delta t} = Y_t + \frac{1}{2}\left(a(t, \Xi_t) + a(t, Y_t)\right)\Delta t + \frac{1}{2}\left(b(t, \Xi_t) + b(t, Y_t)\right)\Delta W_t$$

with

$$\Xi_t = Y_t + a(t, Y_t)\Delta t + b(t, Y_t)\Delta W_t$$

is the Heun discretisation generalised heuristically from the Heun method for deterministic differential equations.

The Heun discretisation, which coincides with the explicit trapezoidal rule, is a two-stage Runge-Kutta method that extends the Euler method. In case of the simple stochastic differential equation

$$dX_t = 1.5X_t dt + 0.1X_t dW_t, \tag{3.19}$$

the Heun method does not converge on $[0, 1]$ for $Y_0 = 1$. Instead, for stochastic differential equation the following Heun method is useful.

Definition 36. *A discrete approximation scheme to a solution of a stochastic differential equation* (3.2) *with step size* Δt *defined by*

$$Y_{t+\Delta_t} = Y_t + \frac{1}{2}\left(a(t, \Xi_t) + a(t, Y_t)\right)\Delta t + \frac{1}{2}b(t, Y_t)\Delta W_t \tag{3.20}$$

with

$$\Xi_t = Y_t + a(t, Y_t)\Delta t + b(t, Y_t)\Delta W_t$$

is the Heun discretisation used for stochastic differential equations.

The discretisation in Definition 3.20 is indeed strongly consistent and converges for the example (3.19).

A Runge-Kutta scheme derived from the multidimensional Milstein scheme in definition (3.16) is given by

$$Y^{(i)}_{t+\Delta_t} = Y^{(i)}_t + a^{(i)}(t, Y_t) + \sum_{j=1}^{m} b^{(i,j)}\Delta W^{(j)}_t \tag{3.21}$$

$$+ \frac{1}{\sqrt{\Delta t}}\sum_{j_1, j_2=1}^{m}\left(b^{(i,j_2)}(t, \Xi^{(j_1)}_t) - b^{(i,j_2)}(t, Y_t)\right)\int_t^{t+\Delta t}\int_t^s dW^{(j_1)}_u dW^{(j_2)}_s \tag{3.22}$$

with

$$\Xi^{(j)}_t = Y_t + a(t, Y_t)\Delta t + b^{(j)}(t, Y_t)\sqrt{\Delta t}$$

for $j \in \{1, \ldots, m\}$. It should be noted that in (3.21) each component $b^{(i,j)}$ of the diffusion matrix b must be evaluated at the $m + 1$ vector valued points $Y_t, \Xi^{(1)}_t, \ldots, \Xi^{(m)}_t$. The higher number of evaluations is the price for replacing the derivative. [7]

Chapter 4

The Monte Carlo Method

Monte Carlo simulation fittingly has its roots in harnessing the computational power of electronic computers for solving mathematical and physical problems. In 1947 Stanislaw Ulam at the Los Alamos laboratories in New Mexico outlined the method in a research report, which soon became a research project on the first electronic computer ENIAC due to the recommendation of John von Neumann, who strongly encouraged research in that direction, seeing the value of the new statistical method in solving complex problems. Independently, Enrico Fermi had basically applied the same technique 15 years earlier, leaving the work unpublished.

The first problems solved with Monte Carlo simulation were differential equations describing the neutron diffusion in fissionable material. Soon the technique became an important research effort. Enormous batches of punch cards, which ENIAC took as input, were filled with generated random numbers produced for further use in Monte Carlo simulations [Metropolis, 1987].

Monte Carlo simulations were introduced to the problems of mathematical finance with the work of Boyle [Boyle, 1977] for the purpose of pricing options in a straightforward way. To this day, this approach is widely used by practitioners in the financial industry, which makes academic research in this area of vital interest to financial institutions such as banks and insurance companies.

4.1 The Classic Monte Carlo Method

The central idea which defines the Monte Carlo method is the approximation of the expected value $\mathbb{E}(f(X))$ of a random variable $f(X)$ given by

$$\mathbb{E}(f(X)) = \int_{\Omega} f(X(\omega))dP(\omega), \tag{4.1}$$

for a Lebesgue-integrable function $f \in L^2$, by an arithmetic average of the outcomes of a large number of independent experiments which are all equal in law to $f(X)$. Due to $f(X)$ being again a random variable for measurable $f \in L^2$, it is sufficient to restrict the attention to random variables denoted by X.

Convergence of the arithmetic mean to the expected value is assured by the Strong Law of Large Numbers. In its most basic form it stipulates that the arithmetic mean of a sequence of independent and identically distributed random variables $(X_i)_{i\in\mathbb{N}}$ with mean μ and variance σ^2 converges almost surely to the expected value $\mu = \mathbb{E}(X_1)$.

Theorem 10. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(X_i)_{i\in\mathbb{N}}$ be a sequence of real-valued, independent and identically distributed random variables with finite expected values $\mu = \mathbb{E}(X_1)$. Then*

$$\overline{X}_N \coloneqq \frac{1}{N}\sum_{i=1}^{N} X_i \xrightarrow{a.s.} \mu$$

holds for $N \to \infty$. [8]

The partial sums $\overline{X}_N$ are unbiased estimators of the expected value μ due to

$$\mathbb{E}\left(\overline{X}_N\right) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(X_i) = \frac{1}{N}\sum_{i=1}^{N}\mu = \mu. \tag{4.2}$$

It is convenient to use the root mean square error of the partial sums as a measure for the error in the estimation. For unbiased estimators like $\overline{X}_N$ it reduces to the standard deviation of the partial sums, that is

$$\begin{aligned} \mathrm{RMSE}(\overline{X}_N) &= \sqrt{\mathrm{MSE}(\overline{X}_N)} = \sqrt{\mathbb{E}((\overline{X}_N - X)^2)} && (4.3)\\ &= \sqrt{\left(\mathbb{V}ar(\overline{X}_N) + (\mathbb{E}(\overline{X}_N) - X)^2\right)} = \frac{\sigma}{\sqrt{N}}. && (4.4) \end{aligned}$$

As the the error is of order $O\left(\frac{1}{\sqrt{N}}\right)$, a significant increase in computational resources is needed to increase the accuracy of the method. Since the slow rate of convergence caused by the order of the error is a main drawback of the Monte Carlo method, mathematical research is focused on improving the order of the error by various extensions like control variates, antithetic sampling, stratification and Latin Hyper Cube sampling, for reference see [Glasserman, 2004] and [Korn et al., 2010].

For large N the use of the standard deviation as a measure for the error is justified by the Central Limit Theorem, as the standard deviation characterises uniquely the spread of the values around the mean of a normal distribution.

Theorem 11. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(X_i)_{i\in\mathbb{N}}$ a sequence of real-valued, independent and identically distributed random variables with finite expected values $\mu = \mathbb{E}(X_1)$ and finite variances $\sigma^2 = \mathbb{V}ar(X)$.*

Then the arithmetic mean of the random variables converges in distribution to a normally distributed random variable with expected value μ and variance $\frac{\sigma^2}{N}$,

$$\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i \xrightarrow{D} Z \sim N\left(\mu, \frac{\sigma^2}{N}\right).$$

Therefore, the Monte Carlo estimator is asymptotically normally distributed with mean μ and variance $\frac{\sigma^2}{N}$ for large values of N. A theoretical approximate confidence interval for the mean μ is given by

$$\left[\overline{X}_N - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}}, \overline{X}_N + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{N}}\right], \tag{4.5}$$

where z_q denotes the q-th quantile of the normal distribution. The unknown variance σ^2 renders the confidence interval of no practical use. However, σ^2 can be estimated by the sample variance

$$\overline{\sigma}^2_N = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X}_N)^2.$$

By Slutsky's Lemma, the limit in distribution in Theorem 11 does not change if the unknown variance σ^2 is exchanged for the calculable estimate $\overline{\sigma}^2_N$. Consequently, the calculable confidence interval

$$\left[\overline{X}_N - z_{1-\frac{\alpha}{2}}\frac{\overline{\sigma}^2_N}{\sqrt{N}}, \overline{X}_N + z_{1-\frac{\alpha}{2}}\frac{\overline{\sigma}^2_N}{\sqrt{N}}\right]$$

is obtained.

It should be noted that the confidence interval is only valid for sufficiently large N and that in particular cases, for example when the Monte Carlo estimator is applied to random variables that are non-zero only far away from the center of the underlying distribution, the confidence interval is particularly inaccurate [Korn et al., 2010].

From (4.5) it is clear that if a Monte Carlo estimator with a small error is of interest, the variance of the estimator should be reduced. All other things being equal, between two unbiased estimators the estimator with the lower variance should be chosen.

It is however seldomly the case that two estimators only differ in their variance. An estimator with a lower variance is usually computationally more expensive. The trade-off between the accuracy of an estimator and its computationally complexity was analysed by Glynn and Whitt [Glynn and Whitt, 1992]. The intuitive efficiency principle that

> [...] the efficiency of a Monte Carlo process may be taken as inversely proportional to the product of the sampling variance and the amount of labour expended in obtaining this estimate, [...]
>
> - *Monte Carlo Methods, Hammersley & Handscomb, 1964*

although applicable an most simulation situations, is shown to lead to inaccurate conclusions in some instances. Glynn and Whitt developed a framework encompassing very general situations in which to compare different variance reduction techniques in a rigorous manner. For the Multilevel Monte Carlo setting it is sufficient to note that in a general Monte Carlo setting the principle established by Hammersley and Handscomb is indeed applicable. It is therefore possible to fix a bound for the root mean square error and then compare variance reduction methods on the basis of computational complexity necessary to achieve the error bound.

A direct application of Theorem 10 is the Classic Monte Carlo algorithm for approximating the mean $\mathbb{E}(X)$ of a random variable X.

Algorithm 4 The Classic Monte Carlo Algorithm

Let X be a real-valued random variable and $N \in \mathbb{N}$ be fixed. The Monte Carlo approximation of the mean $\mathbb{E}(X)$ is calculated via the following steps.

(1) **For** $i = 1$ to N **do**

Sample an observation X_i from a random variable equal in law to X.

(2) **Calculate**

$$\overline{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i.$$

4.2 The Classic Monte Carlo Method for Stochastic Processes

In the following the main focus lies in simulating stochastic processes $X = (X_t)_{t \geq 0}$ in continuous time defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

The expectation of measurable functionals $P : \left(\mathbb{R}^d\right)^{[0,\infty)} \to \mathbb{R}$ of the stochastic process X is considered. It is assumed that the expectation $\mathbb{E}(P(X))$ is well-defined and finite. If independent realisations $X(\omega_i)$, $\omega_i \in \Omega$, for $i \in \{1, \ldots, N\}$, can be sampled, the random variable $P(X)$ can be sampled. However, because the realisations $X(\omega_i)$ are functions on a time continuum, one has to resort to a discretisation scheme in order to approximate the continuous stochastic process for sampling in manageable discrete finite time steps. Approximating the processes $X(\omega_i)$ through a time discretisation $Y(\omega_i)$ and subsequent sampling of the resulting processes with P applied to them results in the approximate sampling of $P(X)$.

Applying the Classic Monte Carlo Algorithm to $P(X)$ using the construction outlined above yields an estimator for $\mathbb{E}(P(X))$,

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} P\left(Y(\omega_i)\right),$$

Algorithm 5 The Classic Monte Carlo Algorithm for Stochastic Processes

Let X be a stochastic process, P a functional, and $N \in \mathbb{N}$ be fixed. The Monte Carlo approximation of the mean $\mathbb{E}(P(X))$ is calculated via the following steps.

(1) **For** $i = 1$ to N **do**

Approximate the process X with a discretisation Y and sample an observation $Y(\omega_i)$.

(2) Calculate

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} P\left(Y(\omega_i)\right).$$

and consequently a Classic Monte Carlo algorithm for stochastic processes.

Again the root mean square error

$$\text{RMSE} = \sqrt{\mathbb{E}([\hat{Y} - \mathbb{E}(P(X))]^2)}$$

should be kept small as efficiently as possible when estimating $P(X)$. That means minimizing the computational complexity to achieve a root mean square error RMSE $\leq \epsilon$, where ϵ is fixed in advance. Minimizing the root mean square error is equivalent to minimizing

$$\begin{aligned} \mathbb{E}([\hat{Y} - \mathbb{E}(P(X))]^2) &= \mathbb{E}([\hat{Y} - \mathbb{E}(\hat{Y}) + \mathbb{E}(\hat{Y}) - \mathbb{E}(P(X))]^2) \\ &= \underbrace{\mathbb{E}([\hat{Y} - \mathbb{E}(\hat{Y})]^2)}_{\text{variance of the Monte Carlo estimator}} + \underbrace{\Big[\mathbb{E}(\hat{Y}) - \mathbb{E}(P(X))\Big]^2}_{\text{discretisation error}} \end{aligned} \tag{4.6}$$

Due to the discretisation error the estimator has a bias in contrast to the conventional Monte Carlo estimator (4.2). However, the order of the variance is the same as in (4.3). Because of

$$\begin{aligned} \mathbb{V}ar(\hat{Y}) &= \frac{1}{N^2} \mathbb{V}ar\left(\sum_{i=1}^{N} P\left(Y_{i\Delta t}\right)\right) \\ &= \frac{1}{N} \mathbb{V}ar(P(Y_{\Delta t})) \end{aligned}$$

the variance is of order $O\left(\frac{1}{N}\right)$. For both the Euler-Maruyama and the Milstein discretisation the weak error is of order $O(\Delta t)$. In sum, for the MSE of $\hat{Y}$

$$\mathbb{E}([\hat{Y} - \mathbb{E}(P(X))]^2) = O\left(\frac{1}{N}\right) + O(\Delta t^2).$$

is obtained. In order to guarantee a root mean square error proportional to ϵ, it is therefore necessary to have $1/N = O(\epsilon^2)$ and $\Delta t^2 = O(\epsilon^2)$, which leads to $N = O(\epsilon^{-2})$ and $\Delta t = O(\epsilon)$. The computational complexity of a standard Monte Carlo simulation is proportional to the number of paths N multiplied by the cost of creating each path, which is sensibly measured by the number of time steps $1/\Delta t$ used for each path. Multiplication yields that the overall computational cost has to be of order $C = O(\epsilon^{-3})$.

Chapter 5

The Multilevel Monte Carlo Method

The Multilevel Monte Carlo method is a recent development introduced by Giles, generalizing on the two level approach developed by Kebaier in [Kebaier, 2005] and drawing on concepts of the multi-grid method for the iterative solution of linear systems of equations arising from the discretisation of elliptic partial differential equations. It is also similar to the multilevel method for parametric integration developed by Heinrich in [Heinrich, 2001]. The goal is to efficiently compute the expectation of the functional of a solution of a stochastic differential equation by using a sequence of discretisations with decreasing step size.

The method is based on the fact that the salient features or the long-time trends of a process can be captured by the use of a coarse time grid to a satisfying accuracy. Less important features are then added onto these coarse paths as corrections to the long-time trend by the use of a finer time grid. In course of the method, no additional bias appears and the successive corrections on the long-time trend are cheaper in terms of variance than the original variance of a process not using time grids of different granularity. Thus, the Multilevel Monte Carlo method is at heart an additional versatile variance reduction technique, adding to the many reduction techniques already employed frequently.

For a given error, composed of the bias introduced by the underlying discretisation scheme and the standard deviation, recursively applying the successive increase in fineness to the time grid used leads to an optimal computational complexity.

5.1 The Multilevel Monte Carlo Theorem

The goal is to calculate $\mathbb{E}(P(X))$ where $P : \mathbb{R}^d \to \mathbb{R}$ is some functional of $X : \omega \to \mathbb{R}^d$. In the Multilevel Monte Carlo approach, different levels of resolution for the approximation of a realisation $P(\omega) = P(X(\omega))$, $l = 0, 1, \ldots, L$ are considered with $l = 0$ being the coarsest and $l = L$ being the the finest level. For brevity, P_l denotes the approximation of $P(X)$ at the level l. By adding up the expectations of different approximation levels in a telescoping sum, the identity

$$\mathbb{E}(P_L) = \mathbb{E}(P_0) + \sum_{l=1}^{L} \mathbb{E}(P_l - P_{l-1}) \tag{5.1}$$

is obtained. The basic idea underpinning the Multilevel Monte Carlo Method is to now estimate each of the expectations $\mathbb{E}(P_l - P_{l-1})$ independently of each other in such a way that the overall variance for a fixed computational cost is minimised. Let $\hat{Y}_0$ be an estimator for $\mathbb{E}(P_0)$ using N_0 samples and let $\hat{Y}_l$, $l > 0$, be an estimator for $\mathbb{E}(P_l - P_{l-1})$ using N_l samples. A natural estimator for $\mathbb{E}(P_l - P_{l-1})$ is the mean of N_l independent samples, which for $l > 0$ is given by

$$\hat{Y}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} (P_l(\omega_i) - P_{l-1}(\omega_i)),$$

where $P_l(\omega_i)$ and $P_{l-1}(\omega_i)$ should come from two discrete approximations $Y^l(\omega_i)$ and $Y^{l-1}(\omega_i)$ of the same underlying stochastic sample $X(\omega_i)$. This ensures that the difference $P_l(\omega_i) - P_{l-1}(\omega_i)$ is small on finer levels of resolution because of strong convergence, resulting in a small variance $\mathbb{V}ar(P_l(\omega_i) - P_{l-1}(\omega_i))$ and consequently a small number of samples on finer levels to accurately estimate $\mathbb{E}(P_l - P_{l-1})$ via $\hat{Y}_l$. The overall Multilevel Monte Carlo estimator for $\mathbb{E}(P(X))$ is then given by

$$\begin{aligned}\hat{Y} &= \sum_{l=0}^{L} \hat{Y}_l \\ &= \frac{1}{N_0}\sum_{i=1}^{N_0} P_0(\omega_i) + \sum_{l=0}^{L}\frac{1}{N_l}\sum_{i=1}^{N_l}(P_l(\omega_i) - P_{l-1}(\omega_i))\end{aligned} \tag{5.2}$$

From the identity

$$\begin{aligned}\mathbb{E}(\hat{Y}) &= \mathbb{E}(\hat{Y}_0) + \sum_{l=1}^{L}\mathbb{E}(\hat{Y}_l) \\ &= \mathbb{E}(P_0) + \sum_{l=1}^{L}\frac{1}{N_l}\sum_{i=1}^{N_l}\mathbb{E}(P_l(\omega_i) - P_{l-1}(\omega_i)) \\ &= \mathbb{E}(P_0) + \sum_{l=1}^{L}\frac{1}{N_l}\mathbb{E}(P_l - P_{l-1}) = \mathbb{E}(P_L)\end{aligned} \tag{5.3}$$

it is clear that although different approximation levels with different approximation errors are used to estimate $\mathbb{E}(P(X))$ the accuracy of the overall estimation only depends on the accuracy of the finest level L. The complexity of the Multilevel Monte Carlo simulation was established by Giles in [Giles, 2008a] . The Theorem is formulated in great generality so that it not only applies to derivatives with Lipschitz payoffs, but also to derivatives for which the payoff function is a discontinuous function of an underlying asset at the terminal state T. It describes the computational complexity of the Multilevel Monte Carlo method in dependence on the underlying discretisation scheme used and the Multilevel estimators.

Theorem 12. *Let P denote a functional of the solution of a stochastic differential equation an let P_l denote the corresponding approximation at level l. If there exist independent estimators $\hat{Y}_l$ with computational complexity C_l based on N_l Monte Carlo samples and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and*

a) $|\mathbb{E}(P_l - P)| \leq \frac{c_1}{M^{\alpha l}}$

b) $\mathbb{E}(\hat{Y}_l) = \begin{cases}\mathbb{E}(P_0) & \text{if} \quad l = 0 \\ \mathbb{E}(P_l - P_{l-1}) & \text{if} \quad l > 0\end{cases}$

c) $\mathbb{V}ar(\hat{Y}_l) \leq \frac{c_2}{N_l M^{\beta l}}$

d) $C_l \leq c_3 N_l M^{\gamma l}$

then there exists a positive constant c_4 such that for any $\epsilon < \frac{1}{e}$ there are values L and N_l for which the Multilevel Monte Carlo estimator

$$\hat{Y} = \sum_{l=0}^{L}\hat{Y}_l$$

has a mean square error with bound

$$\mathbb{E}((\hat{Y} - \mathbb{E}(P))^2) < \epsilon^2$$

and computational complexity C with bound

$$C \leq \begin{cases}\frac{c_4}{\epsilon^2} & \text{if} \quad \beta > \gamma \\ \frac{c_4 \log^2 \epsilon}{\epsilon^2} & \text{if} \quad \beta = \gamma \\ \frac{c_4}{\epsilon^{2+\frac{\gamma-\beta}{\alpha}}} & \text{if} \quad 0 < \beta < \gamma.\end{cases} \tag{5.4}$$

Proof. Let L be given by

$$L = \left\lceil \frac{\log_M \frac{\sqrt{2}c_1}{\epsilon}}{\alpha} \right\rceil < \frac{\log_M \frac{\sqrt{2}c_1}{\epsilon}}{\alpha} + 1 \tag{5.5}$$

so that

$$\frac{\epsilon}{\sqrt{2}M^{-\alpha}} < c_1 M^{-\alpha L} \leq \frac{\epsilon}{\sqrt{2}}. \tag{5.6}$$

Combining (5.6) with conditions a) and b) of the theorem leads to

$$(\mathbb{E}(P_L - P))^2 \leq \frac{1}{2}\epsilon^2. \tag{5.7}$$

According to (4.6), an error bound of $\frac{1}{2}\epsilon^2$ in (5.7) and an error bound of $\frac{1}{2}\epsilon^2$ on the variance shown below, yield an overall error bound on the MSE of ϵ^2. Additionally, using (5.6) and extending the sum on the left hand side to a geometric series results in

$$\sum_{l=0}^{L} M^{\gamma l} < \frac{M^{\gamma L}}{1 - M^{-\gamma}} < \frac{M^{\gamma}(\sqrt{2}c_1)^{\frac{\gamma}{\alpha}}}{1 - M^{-\gamma}} \epsilon^{-\frac{\gamma}{\alpha}}. \tag{5.8}$$

Three different cases for β are possible.

i) $\beta = \gamma$
Setting

$$N_l = \left\lceil \frac{2(L+1)c_2}{\epsilon^2 M^{\beta l}} \right\rceil \tag{5.9}$$

achieves using condition c)

$$\mathbb{V}ar(P_L) = \sum_{l=0}^{L} \mathbb{V}ar(\hat{Y}_l) \leq \sum \frac{c_2}{N_l M^{-\beta l}} \leq \frac{1}{2}\epsilon^2,$$

which gives the required bound on the variance of the estimator such that the overall MSE is smaller than ϵ^2. Rearranging the terms in (5.9) shows that

$$C_L \leq c_3 \sum_{l=0}^{L} N_l M^{\gamma l} \leq c_3 \left(\frac{2(L+1)^2 c_2}{\epsilon^2} + \sum_{l=0}^{L} M^{\gamma l} \right)$$

If $\epsilon < \frac{1}{e} < 1$ it follows that $1 < \log \frac{1}{\epsilon}$ and because of the assumption $\alpha \leq \frac{1}{2}\gamma$ the bounds $\epsilon^{-\frac{\gamma}{\alpha}} \leq \epsilon^{-2} \leq \epsilon^{-2}(\log \epsilon)^2$ are obtained. Combining these inequalities with (5.5) and (5.8) leads to $C_L \leq \epsilon^{-2}(\log \epsilon)^2$, showing the desired result in (5.4).

ii) $\beta > \gamma$
The number of paths on the level $l \in \{1, \ldots, L\}$ is set to

$$N_l = \left\lceil \frac{2c_2}{\epsilon^2 (1 - M^{-\frac{\beta-\gamma}{2}}) M^{\frac{(\beta+\gamma)l}{2}}} \right\rceil$$

in order to get

$$\sum_{l=0}^{L} \mathbb{V}ar(\hat{Y}_l) \leq \sum \frac{c_2}{N_l M^{-\beta l}} \leq \frac{1}{2}\epsilon^2 \left(1 - M^{-\frac{(\beta-\gamma)}{2}}\right) \sum_{l=0}^{L} M^{-\frac{(\beta-\gamma)l}{2}} < \frac{1}{2}\epsilon^2.$$

Because

$$N_l \leq \frac{2c_2}{\epsilon^2 (1 - M^{-\frac{\beta-\gamma}{2}}) M^{\frac{(\beta+\gamma)l}{2}}} + 1$$

the computational complexity bound

$$C_L \leq c_3 \sum_{l=0}^{L} N_l M^{\gamma l} \leq c_3 \left(\frac{2c_2}{\epsilon^2 \left(1 - M^{-\frac{(\beta-\gamma)}{2}}\right)^2} + \sum_{l=0}^{L} M^{\gamma l} \right)$$

follows by using the geometric series as an upper bound on the sum on the left hand side. For $\epsilon < \frac{1}{e} < 1$ the inequality $\epsilon^{-\frac{\gamma}{\alpha}} \leq \epsilon^{-2}$ follows and from (5.8) the desired result $C_L \leq \epsilon^{-2}$ in (5.4) is achieved.

iii) $\beta < \gamma$
Setting

$$N_l = \left\lceil \frac{2c_2 M^{\frac{(\gamma-\beta)L}{2}}}{\epsilon^2 (1 - M^{-\frac{(\gamma-\beta)}{2}}) M^{\frac{(\beta+\gamma)l}{2}}} \right\rceil$$

leads to

$$\sum_{l=0}^{L} \mathbb{V}ar(\hat{Y}_l) \leq \sum \frac{c_2}{N_l M^{-\beta l}} < \frac{1}{2}\epsilon^2 \frac{\left(1 - M^{\frac{-(\gamma-\beta)}{2}}\right)}{M^{-\frac{(\gamma-\beta)L}{2}}} \sum_{l=0}^{L} M^{\frac{(\gamma-\beta)l}{2}} < \frac{1}{2}\epsilon^2.$$

Due to

$$N_l < \frac{2c_2 M^{\frac{(\gamma-\beta)L}{2}}}{\epsilon^2 (1 - M^{-\frac{\gamma-\beta}{2}}) M^{-\frac{(\beta+\gamma)l}{2}}} + 1$$

a bound on the computational complexity is given by

$$\begin{aligned} C_L &\leq c_3 \sum_{l=0}^{L} N_l M^{\gamma l} \\ &\leq c_3 \left(\frac{2c_2 M^{\frac{(\gamma-\beta)L}{2}}}{\epsilon^2 (1 - M^{-\frac{(\gamma-\beta)}{2}})} \sum_{l=0}^{L} M^{\frac{(\gamma-\beta)l}{2}} + \sum_{l=0}^{L} M^{\gamma l} \right) \\ &\leq c_3 \left(\frac{2c_2 M^{(\gamma-\beta)L}}{(1 - M^{-\frac{\gamma-\beta}{2}})^2} + \sum_{l=0}^{L} M^{\gamma l} \right). \end{aligned}$$

From (5.6) it follows that

$$M^{(\gamma-\beta)L} < (\sqrt{2}c_1)^{\frac{(\gamma-\beta)}{\alpha}} \frac{M^{(\gamma-\beta)}}{\epsilon^{\frac{\gamma-\beta}{\alpha}}}.$$

Furthermore, for $\epsilon < \frac{1}{e} < 1$ the inequality $\epsilon^{-\frac{\gamma}{\beta}} \leq \epsilon^{-2-\frac{(\gamma-\beta)}{\alpha}}$ follows because of $\alpha \geq \frac{1}{2}\beta$. Due to inequality (5.8) finally the result $C_L \leq \epsilon^{-2-\frac{\gamma-\beta}{\alpha}}$ is obtained.

□

In general $\mathbb{V}ar(\hat{Y}_l)$ is similar in magnitude to $\mathbb{E}([P_l - P_{l-1}]^2) \geq [\mathbb{E}(P_l - P_{l-1})]^2$ which implies $\beta \leq 2\alpha$ and thus $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$.

The different bounds for the computational complexity C result from balancing the computational costs over all levels in different ways.

In the case $\beta > \gamma$, the major part of the overall computational effort is spent on the coarsest levels close to the level $l = 0$, where $C_l = O(1)$ holds and where $O(\epsilon^{-2})$ paths are required to achieve the fixed accuracy level. This case coincides with the Classic Monte Carlo approach with only one level and one accordingly has to resort to standard variance reduction techniques or Quasi-Monte Carlo methods.

In the case $\beta < \gamma$, the major part of the computational effort is located at the finest levels close to the level $l = L$. As the MSE and thus the discretisation error are of order $O(\epsilon^2)$, combining $(\mathbb{E}(P_L - P))^2 = O(\epsilon^2)$ with condition i) from Theorem 31 results in $M^{-\alpha L} = O(\epsilon)$. Further combination of this relation with condition iv)

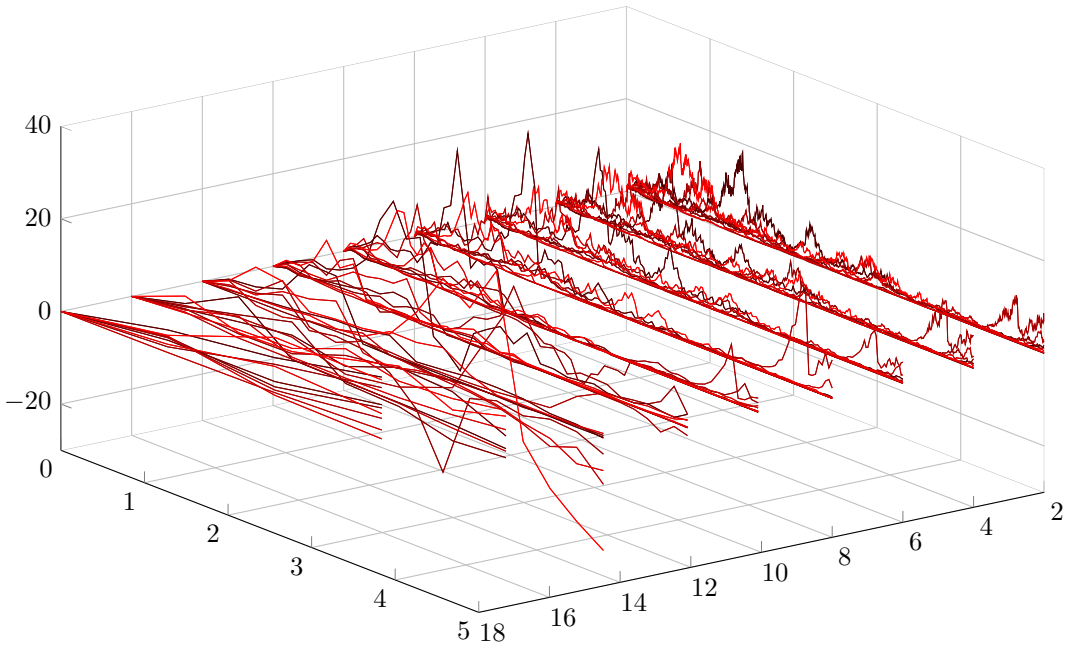

Figure 5.1: Realized paths of a geometric Brownian motion with $a(t, X_t) = 3X_t$ and $b(t, X_t) = X_t$ on different levels $l = 1, \ldots, 9$ with $M = 2$.

leads to $C_L = O(\epsilon^{-\gamma/\alpha})$. If $\beta = 2\alpha$, which for a fixed α is usually the maximum achievable for β, then the overall complexity is $O(C_L)$, which is the result of sampling $O(1)$ paths on the finest level $l = L$.

In the final case $\beta = \gamma$ the computational effort and the variance is balanced almost evenly over all levels. Each level contributes approximately the same computational cost and variance to the total tally of both cost and variance.

If V_0 and V_l are used to denote the variance of sampling one path of P_0 respectively $P_l - P_{l-1}$ then the overall variance of the multilevel estimator is given by $V = \sum_{l=0}^{L} V_l/N_l$. The overall computational cost of a multilevel estimator is given by $C = \sum_{l=0}^{L} C_l N_l$.

The variance reduction in the Multilevel Monte Carlo method is based on the mean square convergence of P_l to $P(X)$, which is corresponding to the mean square convergence of the Cauchy sequence $P_l - P_{l-1}$, leading to $\lim_{n_l \to \infty} V_l = \lim_{n_l \to \infty} \mathbb{V}ar(P_l - P_{l-1}) = 0$. This means that the contribution from the fine levels to the overall variance is falling with $n_l \to \infty$, which means that the use of a few computationally expensive samples on the fine level is sufficient to refine the use of computationally cheap samples on the coarse levels.

Under the constraint that the overall cost C is fixed, the variance V can be minimised by finding the solution of the optimisation problem

$$\sum_{l=0}^{L} \frac{V_l}{N_l} + \eta \left(\sum_{l=0}^{L} N_l C_l - C \right) \tag{5.10}$$

with the Lagrangian multiplier method. First order conditions reveal that the minimum is achieved at $N_l = \eta \sqrt{V_l/C_l}$, where η is the Lagrange multiplier which may attain some arbitrary value. If an overall variance of $\epsilon^2/2$ is required then $\eta = 2\epsilon^{-2} \sum_{l=0}^{L} \sqrt{V_l C_l}$ has to be chosen and the optimal number of sample paths for level l is then given by

$$N_l = \left\lceil 2\epsilon^{-2} \sqrt{V_l C_l} \sum_{k=0}^{L} \sqrt{V_k/C_k} \right\rceil . \tag{5.11}$$

The overall computational cost is then given by

$$C = \frac{1}{\epsilon^2}\left(\sum_{l=0}^{L}\sqrt{V_l C_l}\right)^2.$$

For different behaviours of the product $V_l C_l$, the dominant contribution comes from different parts of the level spectrum. If $V_l C_l$ increases with the level, the major part of the overall effort goes to the finest level L, which results in $C_L \approx \epsilon^{-2} V_L C_L$. Vice versa, $C_0 \approx \epsilon^{-2} V_0 C_0$ holds. In the classic Monte Carlo simulation on the other hand, the overall computational cost is $C_{CMC} \approx \epsilon - 2 V_0 C_L$ under the assumption that the cost of simulating P_L is close to the cost of $P_L - P_{L-1}$ and that $\mathbb{V}ar(P_L) \approx \mathbb{V}ar(P_0)$.

Comparing the improvement in overall computational complexity in the Multilevel Monte Carlo method to the classic Monte Carlo method shows that the improvement takes different forms, varying from case to case. In the case of increasing $V_l C_l$ the Multilevel Monte Carlo cost is reduced by the factor $C_L/C_{CMC} = V_L/V_0$, which means that the improvement stems from a reduction in variance, whereas in the case of decreasing $V_l C_l$ the multilevel cost is reduced by a factor of $C_0/C_{CMC} = C_0/C_L$, which means that the improvement occurs through speeding up the path calculation.

Under the assumption of an order of weak convergence of 1, the bias of the overall method is of order $O(\Delta t_L)$, since only the last level of fineness contributes to the discretisation bias because of condition b) in Theorem 31. This translates to $c\Delta t_L = cTM^{-L}$. To achieve a discretisation error and therefore squared bias proportional to $\epsilon^2/2$, it is useful to set

$$L_{\max} = \frac{\log(cT\sqrt{2}/\epsilon)}{\log M},$$

in order to calculate the overall computational complexity of the Multilevel Monte Carlo estimator via a short algorithm.

Algorithm 6 The Multilevel Monte Carlo Complexity Algorithm

Let L be the number of levels in a Multilevel Monte Carlo setting, V_l be the variance of $\hat{Y}_l$ and N_l be the number of sample paths at level l for $l = 1, \ldots, L$.

(1) Set $L = 0$.

(2) Determine a first estimate of V_L using 100 sample paths.

(3) Determine the optimal N_l using

$$N_l = \left\lceil 2\epsilon^{-2}\sqrt{V_l C_l}\sum_{k=0}^{L}\sqrt{V_k/C_k}\right\rceil.$$

(4) Sample as many additional paths as needed for the optimal N_l.

(5) **If** $L < L_{\max}$

Set $L = L + 1$.

Go to step 2.

5.2 The Multilevel Monte Carlo Method and Control Variates

The control variate method is based on the idea that given a random variable Y with known $\mathbb{E}(Y)$ that is correlated with X as much information as possible from Y should be used to calculate $\mathbb{E}(X)$. The identity

$$\mathbb{E}(X) = \mathbb{E}(X - \lambda Y) + \mathbb{E}(\lambda Y) \tag{5.12}$$

for $\lambda \in \mathbb{R}$ motivates the control variate estimator

$$\overline{X}_Y = \frac{1}{n}\sum_{i=1}^{n} X(\omega_i) - \lambda(Y(\omega_i) - \mathbb{E}(Y)) \tag{5.13}$$

for $\mathbb{E}(X)$ with $(X(\omega_i))_{1\leq i\leq n}, (Y(\omega_i))_{1\leq i\leq n}$ being independent samples of X and Y. From

$$\mathbb{V}ar(\overline{X}_Y) = \frac{1}{n}(\mathbb{V}ar(X) + \lambda^2\mathbb{V}ar(Y) - 2\lambda\mathbb{C}ov(X,Y))$$

it is clear that a variance reduction by use of a control variate is only achieved if $\mathbb{V}ar(X) \geq \lambda^2\mathbb{V}ar(X-Y)$ holds. The variance is minimised for $\lambda^* = \sigma_Y\rho_{XY}/\sigma_X$, where ρ_{XY} denotes the correlation between X and Y. Then the amount of variance reduction attained is

$$2\lambda^*\mathbb{C}ov(X,Y) - (\lambda^*)^2\mathbb{V}ar(Y) = \frac{\sigma_{XY}^2}{\sigma_Y^2} = 1 - \rho_{XY}^2.$$

Equations (5.12) and (5.13) bear a striking resemblance to equations (5.1) and (5.2), especially when one considers the parallels between (5.12) and the two-level Multilevel Monte Carlo estimator

$$\mathbb{E}(P_1) = \mathbb{E}(P_0) + \mathbb{E}(P_1 - P_0). \tag{5.14}$$

Whereas the goal is to estimate $\mathbb{E}(X)$ through the use of $\mathbb{E}(Y)$ because of our superior knowledge about $\mathbb{E}(Y)$ in comparison to $\mathbb{E}(X)$ in the case of the usual control variate approach, in the Multilevel Monte Carlo approach $\mathbb{E}(P_0)$ is used to estimate $\mathbb{E}(P_1)$ because of the reduced computational costs of estimating $\mathbb{E}(P_0)$ in comparison to estimating $\mathbb{E}(P_1)$. The control variate estimator in the case of 5.14,

$$\frac{1}{N_0}\sum_{i=1}^{N_0} P_0(\omega_i) + \frac{1}{N_1}\sum_{i=1}^{N_1}(P_1(\omega_i) - P_0(\omega_i)),$$

coincides with the Multilevel Monte Carlo estimator $\hat{Y}$ with $L = 1$. The two main differences between the control variate and the Multilevel approach are that the expectation $\mathbb{E}(P_0)$ is not taken as a given and therefore has to be estimated, and that $\lambda = 1$ is fixed ab ante.

5.3 Improved Multilevel Monte Carlo

5.3.1 Estimator Construction

The main idea behind the Multilevel Monte Carlo method is to exploit the improvement in total computational complexity by the reduction of variance through the introduction of different path approximation levels $l = 1, \ldots, L$. A further analysis of the variance $\mathbb{V}ar(\hat{Y}_l)$ allows for even greater improvements in overall computational cost. The payoff for constructing a more sophisticated estimator is a reduced variance $\mathbb{V}ar(\hat{Y}_l)$ and consequently a need for fewer samples paths.

Instead of relying on the telescoping property of the terms in (5.1), it is also possible to use different estimators for the finer and coarser levels under consideration [Giles, 2008b]. If P_l^f denotes the finer approximation of the path and P_l^c the coarser approximation, then

$$\mathbb{E}(P_l^f) = \mathbb{E}(P_l^c) \tag{5.15}$$

is required for $l = 1, \ldots, L$, that is, the P_l in $\mathbb{E}(P_l - P_{l-1})$ and $\mathbb{E}(P_{l+1} - P_l)$ must have the same expectation, so that again the pivotal relation in the Multilevel Monte Carlo approach

$$\mathbb{E}(P_L^f) = \mathbb{E}(P_0^f) + \sum_{l=1}^{L}\mathbb{E}(P_l^f - P_{l-1}^c) \tag{5.16}$$

is obtained. Theorem 31 is still valid for the estimators

$$\hat{Y}_0 = \frac{1}{N_0}\sum_{i=1}^{N_0} P_0^f(\omega_i),$$
$$\hat{Y}_l = \frac{1}{N_l}\sum_{i=1}^{N_l}(P_l^f(\omega_i) - P_{l-1}^c(\omega_i)),$$

for $l = 1, \ldots, L$. The advantage of this setting is that it allows for a greater range of possibilities to construct approximations for which

$$\mathbb{V}ar(P_l^f - P_{l-1}^c) < \mathbb{V}ar(P_l - P_{l-1}), \tag{5.17}$$

yielding larger values for the value β in condition c) of Theorem 31. The concrete construction of the approximation should be adapted to the specific problem under consideration. For derivatives more complex than European options such as Asian, lookback, barrier and Digital options it may be profitable to define P_l^c in such a way that it involves information from P_{l+1}^f which was not available during the previous iteration involving P_l^f. In the subsequent applications of this technique the improvement in variance (5.17) in combination with (5.15) is usually achieved by interpolating in between discretisation points with a Brownian interpolant.

5.3.2 Brownian Interpolation

In case the drift $a(t, X_t)$ and the volatility $b(t, X_t)$ are constant, the stochastic differential equation (3.2) has the solution

$$X_t = X_0 + at + bW_t.$$

Using constant diffusion coefficients in the intervals $[t_i, t_{i+1}]$, $i = 0, \ldots, n$, in between equally spaced discretisation points leads to the Brownian interpolation

$$X_t = X_{i\Delta t} + \lambda(X_{(i+1)\Delta t} - X_{i\Delta t}) + b(W_t - W_{i\Delta t} - \lambda(W_{(i+1)\Delta t} - W_{i\Delta t})), \tag{5.18}$$

with $\lambda = (t - t_i)/\Delta t$. The deviation of X_t from the piecewise linear interpolated values $X_{i\Delta t}$, $i = 0, \ldots, n$, is therefore proportional to the deviation of W_t from its piecewise linear interpolated values.

In [Glasserman, 2004] it is shown that the distribution of W_t is independent of the increment $W_{(i+1)\Delta t} - W_{i\Delta t}$ and that the following lemmas hold.

Lemma 4. *Conditional on $X_{i\Delta t}$ and $X_{(i+1)\Delta t}$ the distribution of the integral of X_t over the interval $[t_i, t_{i+1}]$, $i = 1, \ldots, n$ is given by*

$$\int_{t_i}^{t_{i+1}} X_t dt = \frac{1}{2}\Delta t(X_{i\Delta t} + X_{(i+1)\Delta t}) + bI_i,$$

where

$$I_i = \int_{t_i}^{t_{i+1}} (W_t - W_{i\Delta t} - \lambda(W_{(i+1)\Delta t} - W_{i\Delta t}))dt$$

is a random variable with $I_i \sim N\left(0, \frac{1}{12}\Delta t^3\right)$, independent of $W_{(i+1)\Delta t} - W_{i\Delta t}$.

Lemma 5. *Conditional on $X_{i\Delta t}$ and $X_{(i+1)\Delta t}$ the distributions for the minimum and maximum of X_t over the interval $[t_i, t_{i+1}]$ are given by*

$$X_{i\Delta t,min} = \min_{t\in[t_i,t_{i+1}]} \frac{1}{2}\left(X_{i\Delta t} + X_{(i+1)\Delta t} - \sqrt{(X_{i\Delta t} + X_{(i+1)\Delta t})^2 - 2b^2\Delta t \log U_i}\right),$$
$$X_{i\Delta t,max} = \max_{t\in[t_i,t_{i+1}]} \frac{1}{2}\left(X_{i\Delta t} + X_{(i+1)\Delta t} + \sqrt{(X_{i\Delta t} + X_{(i+1)\Delta t})^2 - 2b^2\Delta t \log V_i}\right),$$

where U_i and V_i, $i = 1, \ldots, n$ are uniformly distributed on the unit interval $(0, 1)$.

Lemma 6. *Provided $b \neq 0$, conditional $X_{i\Delta t}$ and $X_{(i+1)\Delta t}$ the probability that the minimum, respectively the maximum, of X_t over the interval $[t_i, t(i+1)]$, $i = 1, \ldots, n$, is less than, respectively greater than, some value B is given by*

$$\mathbb{P}\left(\inf_{t\in[t_i,t_{i+1}]} X_t < B | X_{i\Delta t}, X_{(i+1)\Delta t}\right) = \exp\left(\frac{-2(X_{i\Delta t}-B)^+(X_{(i+1)\Delta t}-B)^+}{b^2\Delta t}\right),$$
$$\mathbb{P}\left(\sup_{t\in[t_i,t_{i+1}]} X_t > B | X_{i\Delta t}, X_{(i+1)\Delta t}\right) = \exp\left(\frac{-2(B-X_{i\Delta t})^+(B-X_{(i+1)\Delta t})^+}{b^2\Delta t}\right).$$

Corollary 2. *If $(W_t)_{t\geq 0}$ is a Brownian motion with $W_0 = W_1 = 0$ then for $x > 0$*

$$\mathbb{P}\left(\sup_{t\in[0,1]} W_t > x\right) = \mathbb{P}\left(\inf_{t\in[0,1]} W_t < -x\right) = \exp(-2x^2).$$

Thus, $\mathbb{E}\left(\sup_{t\in[0,1]} |W_t|^m\right)$ is finite for all positive integers m.

From extreme value theory [Shalizi, 2007] the following results that determine the limiting distribution of the maximum of a large set of independent and identically distributed random variables [Embrechts et al., 2008] are used.

Lemma 7. *If the U_i, $i = 1, \ldots, n$ are independent samples from a uniform distribution on the unit interval $[0, 1]$ then for any positive integer m*

$$\mathbb{E}(\max_{i\in\{1,\ldots,n\}} |\log U_i|^m) = O((\log n)^m)$$

as $n \to \infty$.

Lemma 8. *If Z_i, $i = 1, \ldots, n$ are independent samples from a standard normal distribution then for any positive integer m*

$$\mathbb{E}(\max_{i\in\{1,\ldots,n\}} |Z_i|^m) = O((\log n)^{m/2})$$

as $n \to \infty$.

Corollary 3. *If $(W_t^i)_{t\geq 0}$, $i = 1, \ldots, n$ are independent Brownian paths on $[0, 1]$ conditional on $W_0^i = W_1^i = 0$ then for any positive integer m*

$$\mathbb{E}\left(\max_{i\in\{1,\ldots,n\}} \sup_{t\in[0,1]} |W_t^i|^m\right) = O((\log n)^{m/2})$$

as $n \to \infty$.

Proof. This is a direct result from Corollary 2, since for sufficiently large x the tail probability for $|W_t^i|$ is less than that of a standard normal random variable,

$$\mathbb{P}\left(\sup_{t\in[0,1]} |W_t^i| > x\right) \leq 2e^{-2x^2} \leq \mathbb{P}\left(Z > x\right)$$

for x large enough and $Z \sim N(0, 1)$. □

5.3.3 Conditional Monte Carlo

To improve the convergence of the variance when considering barrier and digital options in the Multilevel Monte Carlo setting, the conditional Monte Carlo method is frequently used. The goal remains to calculate $\mathbb{E}(P)$. By the law of iterated expectation, it is possible to write $\mathbb{E}(P) = \mathbb{E}(\mathbb{E}(P|Z))$, where Z is a random

vector. The variance of P can be decomposed according to $\mathbb{V}ar(P) = \mathbb{E}(\mathbb{V}ar(P|Z)) + \mathbb{V}ar(E(P|Z))$. Both statements are employed when using the improved Multilevel Monte Carlo construction, where each level $l = 1, \ldots, L$ is once taken to be fine and once taken to be a coarse level in (5.16). For fine levels, the conditional expectation $\mathbb{E}(P^f|Z^f)$ is used with $Z^f = \{Y_{i\Delta t_l}\}_{0\leq i\leq M^l}$, whereas for coarse levels instead of the conditional expectation $\mathbb{E}(P^c|Z^c)$ with $Z^c = \{Y_{i\Delta t_{l-1}}\}_{0\leq i\leq M^{l-1}}$ the conditional expectation $\mathbb{E}(P^c|Z^c, \tilde{Z}^c)$ is employed with $\tilde{Z}^c = \{Y_{i\Delta t_{(l-1)}+j\Delta t_l}\}_{0\leq i\leq M^l, 1\leq j\leq M-1}$, where the respective points are obtained from the interpolation formula (5.18). The condition (5.15) holds by the law of iterated conditional expectation

$$\mathbb{E}(\mathbb{E}(P^c|Z^c)) = \mathbb{E}(P^c) = \mathbb{E}(\mathbb{E}(P^c|Z^c, \tilde{Z}^c)).$$

5.4 Extreme Paths

In the Multilevel Monte Carlo analysis for some path-dependent derivatives a specific theoretical argument is used to discard the contribution of certain extreme paths. Their impact on the expected value to be estimated is small enough to ignore them in the resulting estimate.

The methodology used to perform analysis on these paths is built on the following two lemmas. In the proof of the asymptotic error bounds of the respective derivatives, Lemma 9 is used to establish the conditions necessary for the application of Lemma 10, which then justifies discarding the extreme paths by showing that their overall contribution is negligibly small in contrast to the other paths.

Lemma 9. *If X_l is a random variable defined on level $l = 1, \ldots, L$ of the Multilevel Monte Carlo method and if for each positive integer $m \in \mathbb{N}$ the expectation $\mathbb{E}(|X_l|^m)$ is uniformly bounded, then for any $\delta > 0$*

$$\mathbb{P}\left(|X_l| > \Delta t_l^{-\delta}\right) = o(\Delta t_l^p)$$

for all $p > 0$.

Proof. This is an immediate consequence of Markov's inequality

$$\mathbb{P}\left(|X_l| \geq \Delta t_l^{-\delta}\right) = \mathbb{P}\left(|X_l|^m \geq \Delta t_l^{-m\delta}\right) \leq \Delta t_l^{m\delta}\mathbb{E}(|X_l|^m)$$

with $m > p/\delta$. □

Lemma 10. *If Y_l is a random variable defined on level $l = 1, \ldots, L$ of the Multilevel Monte Carlo method, if $\mathbb{E}\left((Y_l)^2\right)$ is uniformly bounded and if for each $p > 0$ the function $\mathbb{1}_{E_l}$ on level l, which takes the values 1 or 0 depending on whether or not a path lies within a set E_l, satisfies*

$$\mathbb{E}(\mathbb{1}_{E_l}) = o(\Delta t_l^p),$$

then for each $p > 0$

$$\mathbb{E}(|Y_l|\mathbb{1}_{E_l}) = o(\Delta t_l^p).$$

Proof. This is a consequence of Hölder's inequality, which yields

$$\mathbb{E}(|Y_l|\mathbb{1}_{E_l}) \leq \left(\mathbb{E}(Y_l^2)\right)^{\frac{1}{2}} \left(\mathbb{E}(\mathbb{1}_{E_l})\right)^{\frac{1}{2}}.$$

□

The following Lemma builds on the preceding results and is the Lemma directly used in the analysis of barrier and Digital options [Giles et al., 2013]. If $u > 0$, $u \prec h^\alpha$ denotes that there exists a constant $c > 0$ such that $u < ch^\alpha$ for h sufficiently small. The implications

$$u_1 \prec h^{\alpha_1},\ u_2 \prec h^{\alpha_2} \implies u_1 + u_2 \prec h^{\min(\alpha_1,\alpha_2)},\ u_1u_2 \prec h^{\alpha_1+\alpha_2}$$

are useful in establishing the following results.

Lemma 11. *For any $\gamma > 0$, the probability that a Brownian path $(W_t)_{t\geq 0}$, its increments $\{\Delta W_{i\Delta t}\}_{i\leq n}$ and the corresponding solutions $(X_t)_{t\geq 0}$ of the stochastic differential equation (5.23) and its fine as well as coarse path approximations $\{Y^f_{i\Delta t}\}_{0\leq i\leq n}$ and $\{Y^c_{i\Delta t}\}_{0\leq i\leq n}$ satisfy any of the following extreme conditions*

$$\begin{aligned}
\max_{0\leq i\leq n}\left(\max\left(|X_{i\Delta t}|, |Y^f_{i\Delta t}|, |Y^c_{i\Delta t}|\right)\right) &> \Delta t^{-\gamma},\\
\max_{0\leq i\leq n}\left(\max\left(X_{i\Delta t} - Y^c_{i\Delta t}|, |X_{i\Delta t} - Y^f_{i\Delta t}|, |Y^f_{i\Delta t} - Y^c_{i\Delta t}|\right)\right) &> \Delta t^{1-\gamma},\\
\max_{0\leq i\leq n}|\Delta W_{i\Delta}| &> \Delta t^{\frac{1}{2}-\gamma},\\
\sup_{t\in[0,T]}|Y^f_t - X_t| &> \Delta t^{1-\gamma},\\
\sup_{t\in[0,T]}|W_t - \overline{W}_t| &> \Delta t^{\frac{1}{2}-\gamma}
\end{aligned}$$

is $o(\Delta t^p)$ for all $p > 0$, where $\overline{W}_t$ is defined to be the piecewise linear interpolant of the discrete values $\{W_{i\Delta t}\}_{0\leq i\leq n}$.

In addition, if none of these extreme conditions is satisfied and if $\gamma < \frac{1}{2}$ then

$$\max_{0\leq i\leq n}|Y^f_{i\Delta t} - Y^f_{(i-1)\Delta t}| \prec \Delta t^{\frac{1}{2}-2\gamma}, \tag{5.19}$$

$$\max_{0\leq i\leq n}|b(i\Delta t, Y^f_{i\Delta t}) - b((i-1)\Delta t, Y^f_{(i-1)\Delta t})| \prec \Delta t^{\frac{1}{2}-2\gamma}, \tag{5.20}$$

$$\max_{0\leq i\leq n}\max\{|b(i\Delta t, Y^f_{i\Delta t})|, |b(i\Delta t, Y^c_{i\Delta t})|\} \prec \Delta t^{-\gamma}, \tag{5.21}$$

$$\max_{0\leq i\leq n}|b(i\Delta t, Y^f_{i\Delta t}) - b(i\Delta t, Y^c_{i\Delta t})| \prec \Delta t^{\frac{1}{2}-2\gamma}, \tag{5.22}$$

where $b(i\Delta t, Y^c_{i\Delta t})$ is defined to be equal to $b((i-1)\Delta t, Y^c_{(i-1)\Delta t})$ if $i \in \{1, \ldots, n\}$ is odd.

5.5 Antithetic Sampling

An alternative to the above construction is a variant based on antithetic sampling. Given an antithetic ω^a_i equal in distribution to ω_i, one can use with the same basic idea as in antithetic sampling the estimators

$$\hat{Y}_0 = \frac{1}{N_0}\sum_{i=1}^{N_0}\frac{1}{2}(P_0(\omega_i) + P_0(\omega^a_i)),$$

$$\hat{Y}_l = \frac{1}{N_l}\sum_{i=1}^{N_l}\frac{1}{2}(P_l(\omega_i) + P_l(\omega^a_i)) - P_{l-1}(\omega_i).$$

Because of $\mathbb{E}(P_l(\omega^a_i)) = \mathbb{E}(P_l(\omega_i))$ condition b) of Theorem 31 is again satisfied.

5.6 Derivative Pricing with the Multilevel Monte Carlo Method

In mathematical finance, one of the main applications of the Monte Carlo method is risk neutral derivative pricing. Application of the Monte Carlo method in the risk neutral valuation of a derivative reduces to the estimation of an expectation $\mathbb{E}(P(X))$ where the functional $P : \mathbb{R}^d \to \mathbb{R}$ describes the payoff function of the derivative including the appropriate discount factor and where $X = (X_t)_{t\geq 0}$ is the stochastic process underlying the mathematical model of the market under consideration.

In the following models based on a stochastic process $(X_t)_{t\geq 0}$ which is the solution of a one-dimensional stochastic differential equation of the form

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \tag{5.23}$$

are considered with $X_0 = x_0 \in \mathbb{R}$ fixed, a one-dimensional Brownian motion $(W_t)_{t\geq 0}$ and measurable coefficient functions $a : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $b : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ fulfilling the usual conditions ensuring the existence and uniqueness of a strong solution.

Although Lévy processes will be covered in later sections, the analysis of diffusion processes defined by (5.23) can be put to considerable practical use. Due to the simplicity of the majority of diffusion processes used in finance, a Monte Carlo simulation approach for a diffusion model seems hardly necessary. The ubiquitous Black-Scholes model based on geometric Brownian motion and the Vasicek model based on the Ornstein-Uhlenbeck process both have analytic solutions, which are quick to evaluate. For the Cox-Ingersoll-Ross model based on a square-root-diffusion process, a credit risk model introduced in [Brigo and Alfonsi, 2005] based on the same process, as well as the Black-Karasinski model based on a log-normal-diffusion process, the distribution of the process at a specific time instance is known. For these processes, it is thus advantageous to simulate the process at a time instance $t \geq 0$ directly, rather than to use a Monte Carlo method simulating entire paths.

However, local volatility models enjoy broad practical acceptance and both approaches outlined above are not available in their case. Fitting to a local volatility surface consists in calibrating $b(t, X_t)$ in (5.23) to market data in two dimensions, maturity and strike of the European options under consideration. Let $M_1 \in \mathbb{N}$ maturity dates and $M_2 \in \mathbb{N}$ strikes be available, yielding $M_1 \times M_2$ data points. A frequently used technique [Gatheral, 2006] consists in approximating the 2D-Fourier transform of a high resolution local volatility fit $\tilde{b}(t, x)$ by

$$b(t, x) = \frac{1}{M_1 M_2} \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} e^{i(\omega_{m_1} t + \omega_{m_2} X_t)}.$$

For a resulting stochastic differential equation of the type

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t$$

a Monte Carlo approach is preferable, as neither an analytic formula nor the distribution of X_t for $t \geq 0$ are likely to be available.

5.6.1 Derivatives with Lipschitz payoff functions

We are first considering European options which provide for only one possibility to exercise at the maturity of the option. Thus, the approximation of an expectation of the form $\mathbb{E}(P(X_T))$ is of interest. Specifically, globally Lipschitz continuous payoffs P are considered in this section, that is functionals $P : \mathbb{R}^d \to \mathbb{R}$ for which

$$|P(x) - P(y)| \leq D|x - y|.$$

holds for a positive constant $D \in \mathbb{R}$ and all $x, y \in \mathbb{R}^d$, such as European call or put options.

Euler-Maruyama and Milstein Scheme

The weak convergence rate of an approximation process $\{Y_{i\Delta t}\}_{0\leq i\leq n}$ to a process $(X_t)_{t\geq 0}$ with regards to a smooth payoff function $P : \mathbb{R}^d \to \mathbb{R}$ and a choice of time step approximations Δt_l depending on level l is given by α in

$$|\mathbb{E}(P(X_T) - P(Y_{n_l \Delta t_l}))| \leq C(\Delta t_l)^\alpha$$

The convergence rate α appears in condition a) of Theorem 31.

The variance in condition b) of Theorem 31 can be bound by

$$\begin{aligned}
\mathbb{V}ar(\hat{Y}_l) = \mathbb{V}ar(P_l - P_{l-1}) &\leq \mathbb{E}([P_l - P_{l-1}]^2) \\
&\leq 2\mathbb{E}([P_l - P]^2) + 2\mathbb{E}([P - P_{l-1}]^2) \\
&\leq 2D \left[|X_T - Y_{n_l \Delta t_l}|^2 + |X_T - Y_{n_{l-1} \Delta t_{l-1}}|^2 \right].
\end{aligned}$$

for Lipschitz continuous payoffs. Thus it becomes clear that one also has to analyse the strong convergence of the approximation process $\{Y_{i\Delta t}\}_{0\leq i\leq n}$, as the convergence of the Multilevel Monte Carlo variance in case of globally Lipschitz payoff functions is completely determined by the rate of strong convergence of the underlying

discretisation scheme used. For the Euler-Maruyama scheme a Multilevel Monte Carlo variance convergence of $\beta = 1/2$ is achieved, for the Milstein scheme accordingly a variance convergence rate of $\beta = 1$ [Giles and Szpruch, 2013].

Theorem 13. *The Multilevel Monte Carlo estimator for a derivative with a Lipschitz payoff function has variance $V_l = O(\Delta t_l^{1/2})$ using the Euler-Maruyama scheme and variance $V_l = O(\Delta t_l)$ using the Milstein scheme.*

Milstein Scheme

By allowing for an improved Multilevel Monte Carlo construction with Brownian interpolation (5.24) in between time steps the performance of the Multilevel Monte Carlo method can be improved. Generalizing from European derivatives to derivatives with payoff functions depending on multiple time instances leads to

$$P(X) = f(X_{T_1}, \dots, X_{T_m})$$

for $m \geq 1$, with the Lipschitz condition

$$|P(x^{(1)}, \dots, x^{(m)}) - P(y^{(1)}, \dots, y^{(m)})| \leq D \sum_{j=1}^{m} |y^{(j)} - x^{(j)}|$$

for $x, y \in \mathbb{R}^m$.

Before the convergence result is considered it is useful to contrast the interpolation scheme resulting from (5.18),

$$Y_s = Y_t + \lambda(Y_{(t+\Delta t} - Y_t) + b(t, Y_t)(W_s - W_t - \lambda(W_{t+\Delta t} - W_t)), \tag{5.24}$$

where $\lambda = (s-t)/\Delta t$, with the interpolation scheme (3.15) from [Kloeden and Platen, 1995].

Theorem 14. *If $(Y_t)_{t\geq 0}$ denotes a Milstein discretisation interpolated with Brownian motion according to (5.24) and if $(Y_t^{KP})_{t\geq 0}$ denotes a Milstein discretisation interpolated according to the Kloeden and Platen scheme (3.15), then for any $m \in \mathbb{N}$,*

$$\mathbb{E}\left(\sup_{0\leq t\leq T} |Y_t - Y_t^{KP}|^m\right) = O((\Delta t \log \Delta t)^m), \quad \sup_{0\leq t\leq T} \mathbb{E}\left(|Y_t - Y_t^{KP}|^m\right) = O(\Delta t^m),$$
$$\mathbb{E}\left(\left[\int_0^T Y_t - Y_t^{KP} dt\right]^2\right) = O(\Delta t^3). \tag{5.25}$$

Theorem 15. *For Lipschitz payoffs the single replication variance obtainable is $\beta = 2$, leading to an optimal overall complexity $O(\epsilon^{-2})$.*

Proof. Using Jensen's inequality in combination with the Lipschitz condition yields

$$\mathbb{V}ar(P_l^f - P_{l-1}^c) \leq \mathbb{E}((P_l^f - P_{l-1}^c)^2) \leq D^2 m \sum_{j=1}^{m} \mathbb{E}((Y_{T_j}^f - Y_{T_j}^c)^2).$$

By adding and subtracting the respective terms under the expectation on the right hand side for one summand the inequality

$$\frac{1}{4}\mathbb{E}((Y_{T_j}^f - Y_{T_j}^c)^2) \leq \mathbb{E}\left(\left(Y_{T_j}^f - Y^{f\,KP}_{T_j}\right)^2\right) + \mathbb{E}\left(\left(Y_{T_j}^c - Y^{c\,KP}_{T_j}\right)^2\right)$$
$$+ \mathbb{E}\left(\left(Y^{f\,KP}_{T_j} - Y_{T_j}\right)^2\right) + \mathbb{E}\left(\left(Y^{c\,KP}_{T_j} - Y_{T_j}\right)^2\right).$$

can be obtained.

Due to Theorem 14, the first two terms on the right hand side are therefore of order $O(\Delta t^2)$ and due to Theorem 9 the second two terms are of order $O(\Delta t^2)$, giving the desired overall order. □

5.6.2 Asian Options

Asian options which are continuously monitored have payoff functions which are uniform Lipschitz functions of the average of the underlying over the time interval $[0, T]$ and the value of the underlying at maturity,

$$P(X) = f\left(X_T, \frac{1}{T}\int_0^T X_t dt\right).$$

One possibility to treat Asian options is to directly integrate the Brownian interpolant (5.18).

Lemma 12. *Conditional on $X_{i\Delta t}$ and $X_{(i+1)\Delta t}$, the distribution of the integral of X_t over the interval $[i\Delta t, (i+1)\Delta t]$ is given by*

$$\int_{i\Delta t}^{(i+1)\Delta t} X_t dt = \frac{1}{2}\Delta t(X_{i\Delta t} + X_{(i+1)\Delta t}) + b(i\Delta, X_{i\Delta t})I_{i\Delta t} \tag{5.26}$$

with

$$I_{i\Delta t} = \int_{i\Delta t}^{(i+1)\Delta t} (W_t - W_{i\Delta t} - \frac{t - i\Delta t}{\Delta t}(W_{(i+1)\Delta t} - W_{i\Delta t}))dt$$

is a normally distributed random variable with mean zero and variance $\frac{1}{12}\Delta t^3$ independent of $W_{(i+1)\Delta t} - W_{i\Delta t}$.

Thus, integration of the approximation on the fine level yields

$$\int_0^T Y_t^f dt = \sum_{i=0}^{N-1} \frac{1}{2}\Delta t_l(Y^f_{i\Delta t_l} + Y^f_{(i+1)\Delta t_l}) + b(i\Delta t_l, Y^f_{i\Delta t_l})I^f_{i\Delta t_l}$$

where the $I^f_{i\Delta t_l}$ are independent normally distributed random variables with mean zero and variance $\frac{1}{12}\Delta t_l^3$. The coarse approximation may be derived from the fine approximation with

$$\begin{aligned} I^c_{i\Delta t_l} &= \int_{i\Delta t_l}^{(i+2)\Delta t_l} \left(W_t - W_{i\Delta t_l} - \frac{t - i\Delta t_l}{2\Delta t_l}(W_{(i+2)\Delta t_l} - W_{i\Delta t_l})\right) dt \\ &= I^f_{i\Delta t_l} + I^f_{(i+1)\Delta t_l} - \frac{1}{2}\Delta t_l(W_{(i+2)\Delta t_l} - 2W_{(i+1)\Delta t_l} + W_{i\Delta t_l}). \end{aligned}$$

Interpolation with Brownian motion leads to an optimal convergence result.

Theorem 16. *The approximation with Brownian interpolation for Asian payoffs leads to $V_l = O(\Delta t_l^2)$.*

Proof. Due to the Lipschitz continuity of the payoff function the proof of Theorem 15 for Lipschitz continuous functions can be mimicked. Integrating

$$Y_t^f - Y_t^c = (Y_t^f - {Y^f}_t^{KP}) - (Y_t^c - {Y^c}_t^{KP}) + ({Y^f}_t^{KP} - X_t) - ({Y^c}_t^{KP} - X_t)$$

in addition to squaring and taking the expectation leads to

$$\begin{aligned} \mathbb{E}((Y_t^f - Y_t^c)^2) \leq 4\Bigg(&\mathbb{E}\left(\left(Y_t^f - {Y^f}_t^{KP}\right)^2\right) + \mathbb{E}\left(\left(Y_t^c - {Y^c}_t^{KP}\right)^2\right) \\ &+ \mathbb{E}\left(\left({Y^f}_t^{KP} - X_t\right)^2\right) + \mathbb{E}\left(\left({Y^c}_t^{KP} - X_t\right)^2\right)\Bigg). \end{aligned}$$

□

An alternative is to leave off the integrals I^f, respectively I^c, in equation (5.26). The integral of the interpolants are thus approximated via trapezoidal integration or, from a different viewpoint, are the result of an averaging process over piecewise linear interpolants. The order of the variance does not change when compared to the case of Brownian interpolation.

Theorem 17. *The approximation with piecewise linear interpolation for Asian payoffs leads to $V_l = O(\Delta t_l^2)$.*

Proof. A compact proof can be found in [Giles et al., 2013]. □

5.6.3 Lookback Options

In this section floating strike lookback call options are considered. The payoff function of such an option is a uniform Lipschitz function of the underlying at maturity and depending on the contract either the infimum or the supremum of the underlying over the time interval until maturity. In the following the analysis for the case of dependence of the payoff on the infimum is performed, the supremum case is treated analogously. The undiscounted lookback payoff is then given by

$$P(X) = X_T - \inf_{0\leq t\leq T} X_t \tag{5.27}$$

and a possible numerical approximation of the payoff by setting for every level $l = 1, \ldots, L$

$$P_l = Y_{n_l \Delta t_l} - \inf_{0\leq m\leq T/\Delta t_l} Y_{\Delta t_l m}. \tag{5.28}$$

Bar any further treatment, the weak convergence rate of the approximated payoff in (5.28) to the exact payoff in (5.27) is of order $O\left(\sqrt{\Delta t_l}\right)$ [Gobet and Menozzi, 2004].

A standard approximation result [Broadie et al., 1999], which relates the price of a continuously monitored lookback option to the price of discretely monitored one by way of adjusting the resulting minimum in (5.28), can be used to numerically approximate the lookback option payoff. In order to compensate for the time instances in between discretisation points at which the minimum attained could be have been lower, the minimum is lowered by a multiple of the constant $\beta = -\zeta(1/2)/2\pi \approx 0.5826$, where ζ is the Riemann-Zeta function,

$$Y_{\min} = \min_{0\leq i\leq n_l} \left(Y_{i\Delta t_l} - \beta b(i\Delta t_l, Y_{i\Delta t_l})\sqrt{\Delta t_l} \right). \tag{5.29}$$

The correction term results from a limiting theorem proven in [Asmussen et al., 1995] and improves the weak convergence of order $O(\sqrt{\Delta t_l})$ to $O(\Delta t_l)$ [Giles, 2008b].

Euler-Maruyama Scheme

For simplicity, the case $M = 2$ is considered.

Theorem 18. *The Multilevel Monte Carlo estimator for a lookback option has variance $V_l = O(\Delta t_l^{1-\delta})$ for any $\delta > 0$ using the Euler-Maruyama scheme.*

Proof. By considering (3.5) the Euler-Maruyama scheme with linear interpolation $(Y_t^{\Delta t_l})_{t\geq 0}$ yields

$$\begin{aligned}\mathbb{E}(|P - P_l|^2) &\leq 2\mathbb{E}(|X_T - Y_T^{\Delta t_l}|^2) + 2\mathbb{E}\left(|\inf_{0\leq t\leq T} X_t - \inf_{0\leq t\leq T} Y_t^{\Delta t_l}|^2 \right) \\ &= O(\Delta t) + 2\mathbb{E}\left(|\inf_{0\leq t\leq T} X_t - \inf_{0\leq t\leq T} Y_t^{\Delta t_l}|^2 \right). \end{aligned} \tag{5.30}$$

From the inequality

$$|\inf_{0\leq t\leq T} X_t - \inf_{0\leq t\leq T} Y_t^{\Delta t_l}| \leq \sup_{0\leq t\leq T} |X_t - Y_t^{\Delta t_l}|$$

it follows from (5.30) and (3.6) that given any $\delta > 0$ the order relation

$$\mathbb{E}(|P - P_l|^2) \leq O(\Delta t) + 2\mathbb{E}\left(\sup_{0\leq t\leq T} |X_t - Y_t^{\Delta t_l}|^2 \right) = O(\Delta t^{1-\delta})$$

holds, resulting in $\beta = 1 - \delta$ for the floating strike lookback call. □

An analogous result holds for the floating strike lookback put with payoff function $P = \sup_{0\leq t\leq T} X_t - X_T$.

Milstein Scheme

The most recent analysis of lookback options in the Multilevel Monte Carlo setting has been done in [Giles et al., 2013]. Using the Milstein scheme, the convergence rate of the variance can be increased to $O(\Delta t_l^2 (\log \Delta t_l)^2)$ by

applying the improved Multilevel Monte Carlo method. The idea is to compute the minimum for both the fine and the coarse paths.

The Brownian interpolant for the fine path simulation in the time interval $[i\Delta t_l, (i+1)\Delta t_l]$ for $i \in \{1, \ldots, n_l - 1\}$ is according to (5.18) defined to be

$$Y_t^f = Y_{i\Delta t_l}^f + \lambda(Y_{(i+1)\Delta t_l}^f - Y_{i\Delta t_l}^f) + b(t, Y_{i\Delta t_l}^f)(W_t - W_{i\Delta t_l} - \lambda(W_{(i+1)\Delta t_l} - W_{i\Delta t_l}))$$

with $\lambda = (t - i\Delta t_l)/\Delta t_l$. Applying Lemma 5 the minimum in the time interval $[i\Delta t_l, (i+1)\Delta t_l]$ can be simulated by

$$Y_{i\Delta t_l,\min}^f = \frac{1}{2}\left(Y_{i\Delta t_l}^f + Y_{(i+1)\Delta t_l}^f - \sqrt{\left(Y_{(i+1)\Delta t_l}^f - Y_{i\Delta t_l}^f\right)^2 - 2b(i\Delta t_l, Y_{i\Delta t_l}^f)^2 \Delta t_l \log U_i}\right),$$

where U_i is a uniform random variable on the unit interval. The global minimum is attained by taking the minimum over all time instances. The global minimum is then used to compute the value for the fine path P_l^f.

The treatment for the coarse path P_{l-1}^c differs slightly. For even $i \in \{1, \ldots, n_l - 1\}$ the same underlying Brownian interpolant (5.18) is used to define $Y_{(i+1)\Delta t_l}^c$. The minimum value over the interval $[i\Delta t_l, (i+1)\Delta t_l]$ can then be determined by choosing the smaller of the minima for the two intervals $[i\Delta t_l, (i+1)\Delta t_l]$ and $[(i+1)\Delta t_l, (i+2)\Delta t_l]$,

$$Y_{i\Delta t_l,\min}^c = \frac{1}{2}\left(Y_{i\Delta t_l}^c + Y_{(i+1)\Delta t_l}^c - \sqrt{\left(Y_{(i+1)\Delta t_l}^c - Y_{i\Delta t_l}^c\right)^2 - 2b(i\Delta t_l, Y_{i\Delta t_l}^c)^2 \Delta t_l \log U_i}\right),$$

$$Y_{(i+1)\Delta t_l,\min}^c = \frac{1}{2}\left(Y_{(i+1)\Delta t_l}^c + Y_{(i+2)\Delta t_l}^c - \sqrt{\left(Y_{(i+2)\Delta t_l}^c - Y_{(i+1)\Delta t_l}^c\right)^2 - 2b(i\Delta t_l, Y_{i\Delta t_l}^c)^2 \Delta t_l \log U_{i+1}}\right).$$

The uniform random numbers U_i and U_{i+1} from the fine path simulation are reused. Consequently, the identity (5.15) necessary in order for the improved Multilevel Monte Carlo method to work holds due to

$$\min(Y_{i\Delta t_l,\min}^c, Y_{(i+1)\Delta t_l,\min}^c) \stackrel{D}{=} Y_{i/2\Delta t_{l-1},\min}^f,$$

which in turn holds because of the fact that both the left hand and the right hand side are based on the same Brownian interpolant.

Theorem 19. *For the improved Multilevel Monte Carlo method using the Milstein scheme with Brownian interpolation, $V_l = O(\Delta t_l^2 (\log \Delta t_l)^2)$ holds.*

Proof. If $Y_{\min}^f$ and $Y_{\min}^c$ are the minima for the fine and the coarse part, then

$$\begin{aligned} |Y_{\min}^f - Y_{\min}^c| &\leq \max_{0 \leq i \leq n_l} |Y_{i\Delta t_l,\min}^f - Y_{i\Delta t_l,\min}^c| \\ &\leq \max_{0 \leq i \leq n_l} |Y_{i\Delta t_l}^f - Y_{i\Delta t_l}^c| + \max_{0 \leq i \leq n_l} |R_{i\Delta t_l}^f - R_{i\Delta t_l}^c| \end{aligned}$$

with

$$R_{i\Delta t_l}^f = \frac{1}{2}\sqrt{(Y_{(i+1)\Delta t_l}^f - Y_{i\Delta t_l}^f)^2 - 2b(i\Delta t_l, Y_{i\Delta t_l}^f)^2 \Delta t_l \log U_i}.$$

If $R_{i\Delta t_l}^f = R_{i\Delta t_l}^c = 0$, then $|R_{i\Delta t_l}^f - R_{i\Delta t_l}^c| = 0$, otherwise the summands $R_{i\Delta t_l}^f$ and $R_{i\Delta t_l}^c$ yield a strictly positive

sum and via the inequality $\Big||x|-|y|\Big| \leq |x-y|$ for $x, y \in \mathbb{R}$ it follows that

$$\begin{aligned}
|R^f_{i\Delta t_l} - R^c_{i\Delta t_l}| &= \frac{|{R^f_{i\Delta t_l}}^2 - {R^c_{i\Delta t_l}}^2|}{R^f_{i\Delta t_l} - R^c_{i\Delta t_l}} \\
&\leq \frac{|(Y^f_{(i+1)\Delta t_l} - Y^f_{i\Delta t_l})^2 - (Y^c_{(i+1)\Delta t_l} - Y^c_{i\Delta t_l})^2|}{4(R^f_{i\Delta t_l} + R^c_{i\Delta t_l})} + \frac{|b(i\Delta t_l, Y^f_{i\Delta t_l})^2 - b(i\Delta t_l, Y^c_{i\Delta t_l})^2|\Delta t_l|\log U_i|}{2(R^f_{i\Delta t_l} + R^c_{i\Delta t_l})} \\
&\leq \frac{1}{2}\Big||Y^f_{(i+1)\Delta t_l} - Y^f_{i\Delta t_l}| - |Y^c_{(i+1)\Delta t_l} - Y^c_{i\Delta t_l}|\Big| + \frac{1}{\sqrt{2}}\Big||b(i\Delta t_l, Y^f_{i\Delta t_l})| - |b(i\Delta t_l, Y^c_{i\Delta t_l})|\Big|\sqrt{\Delta t_l|\log U_i|} \\
&\leq \frac{1}{2}\left(|Y^f_{(i+1)\Delta t_l} - Y^c_{(i+1)\Delta t_l}| + |Y^f_{i\Delta t_l} - Y^c_{i\Delta t_l}|\right) + \frac{1}{2}|b(i\Delta t_l, Y^f_{i\Delta t_l}) - b(i\Delta t_l, Y^c_{i\Delta t_l})|\sqrt{\Delta t_l|\log U_i|}
\end{aligned}$$

from which

$$(Y^f_{\min} - Y^c_{\min})^2 \leq 8 \max_{0\leq i\leq n_l}(Y^f_{i\Delta t_l} - Y^c_{i\Delta t_l})^2 + \Delta t_l \left(\max_{0\leq i\leq n_l}(b(i\Delta t_l, Y^f_{i\Delta t_l}) - b(i\Delta t_l, Y^c_{i\Delta t_l})^2\right)\left(\max_{0\leq i\leq n_l}|\log U_i|\right) \quad (5.31)$$

follows. From the uniform Lipschitz condition in the main Theorem 8 on the strong convergence rate of the Milstein scheme, for even $i \in \{1, \ldots, n_l\}$

$$(b(i\Delta t_l, Y^f_{i\Delta t_l}) - b(i\Delta t_l, Y^c_{i\Delta t_l})^2 \leq C_2^2 (Y^f_{i\Delta t_l} - Y^c_{i\Delta t_l})^2$$

follows, whereas for odd $i \in \{1, \ldots, n\}$

$$\begin{aligned}
\Big(b(i\Delta t_l, Y^f_{i\Delta t_l}) - b(i\Delta t_l, Y^c_{i\Delta t_l})\Big)^2 &= \Big((b(i\Delta t_l, Y^f_{i\Delta t_l}) - b((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})) + (b((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})\Big)^2 \\
&\leq 2C_2^2 (Y^f_{i\Delta t_l} - Y^f_{(i-1)\Delta t_l})^2 + 2C_2^2 (Y^f_{(i-1)\Delta t_l} - Y^c_{(i-1)\Delta t_l})^2 \qquad (5.32)
\end{aligned}$$

holds. The fine Milstein approximation on level l is according to (3.14) given by

$$\begin{aligned}
Y^f_{i\Delta t_l} - Y^f_{(i-1)\Delta t_l} &= a((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})\Delta t_l + b((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})\Delta W_{(i-1)\Delta t_l} \\
&\quad + \frac{1}{2} b'((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l}) b((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})(\Delta W^2_{(i-1)\Delta t_l} - \Delta t_l). \qquad (5.33)
\end{aligned}$$

The second summand on the right hand side of (5.33) is of order $O\left(\sqrt{\Delta t_l}\right)$ whereas the other summands are of order $O(\Delta t_l)$. This observation in combination with Jensen's inequality, Hölder's inequality and Lemma 8 yields

$$\begin{aligned}
\mathbb{E}\left(\max_{0\leq i\leq n_l}(Y^f_{i\Delta t_l} - Y^f_{(i-1)\Delta t_l})^2\right) &\leq \left(\mathbb{E}(\max_{0\leq i\leq n}|Y^f_{i\Delta t_l} - Y^f_{(i-1)\Delta t_l}|)\right)^2 \\
&= \left(\mathbb{E}(\max_{0\leq i\leq n_l}|b((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})\Delta W_{(i-1)\Delta t_l}|)\right)^2 \\
&\quad + o\left(\left(\mathbb{E}(\max_{0\leq i\leq n_l}|b((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})\Delta W_{(i-1)\Delta t_l}|)\right)^2\right) \\
&\leq \mathbb{E}\left(\max_{0\leq i\leq n_l} b((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})^2\right)\mathbb{E}\left(\max_{0\leq i\leq n_l}\left(\Delta W_{(i-1)\Delta t_l}\right)^2\right) \\
&\quad + o\left(\left(\mathbb{E}(\max_{0\leq i\leq n_l}|b((i-1)\Delta t_l, Y^f_{(i-1)\Delta t_l})\Delta W_{(i-1)\Delta t_l}|)\right)^2\right) \\
&= O(\Delta t_l|\log \Delta_l|) \qquad (5.34)
\end{aligned}$$

From the definition of Y^c, inequality (5.32) and statement (5.34) it follows that

$$\mathbb{E}\left(\max_{0\leq i\leq n_l}(b^f(i\Delta t_l, Y^f_{i\Delta t_l}) - b^c(i\Delta t_l, Y^c_{i\Delta t_l}))^2\right) = O(\Delta t_l|\log \Delta t_l|).$$

Combining Lemma 7 with inequality (5.31) yields

$$\mathbb{E}\left((Y^f_{\min} - Y^c_{\min})^2\right) = O(\Delta t_l^2 (\log \Delta t_l)^2).$$

Since the payoff function is uniform Lipschitz and since

$$\max_{0\leq i\leq n_l} \mathbb{E}\left((Y^f_{i\Delta t_l} - Y^c_{i\Delta t_l})^2\right) = O(\Delta t_l^2), \tag{5.35}$$

the variance of a single sample is of the desired order

$$V_l = O(\Delta t_l^2 (\log \Delta t_l)^2).$$

□

5.6.4 Barrier Options

This section contains the analysis of barrier options. The type of barrier options considered are down-and-out options for which the payoff is a Lipschitz function $f : \mathbb{R} \to \mathbb{R}$ of the value of the underlying asset X at maturity, provided the asset has never dropped below a fixed value B,

$$P(X) = f(X_T)\mathbb{1}_{\tau_B^- > T}, \tag{5.36}$$

where the first passage time τ_B^- is defined as

$$\tau_B^- = \inf_{t>0}\{X_t < B\}. \tag{5.37}$$

Direct approximation of the stopping time on the level l via

$$\hat{\tau}_B^- = \inf_{0\leq i\leq n_l}\{Y_{i\Delta t_l} < B\}$$

overestimates the exact stopping time as it is possible for X to cross the barrier at some time instance $t > 0$ in between the discretisation points $\{i\Delta t_l\}_{0\leq i\leq n_l}$ but never to be above the barrier at any of the discretisation points themselves. For the resulting weak error

$$\mathbb{E}(P(X) - P(Y)) \leq C\sqrt{\Delta t_l} + o(\sqrt{\Delta t_l}) \tag{5.38}$$

holds [Gobet and Menozzi, 2010].

The overshoot problem in simulating the payoff of barrier option has been extensively analysed by Gobet in [Gobet and Menozzi, 2010], [Gobet and Menozzi, 2004], [Gobet, 2000] and [Gobet and Temam, 2001] and an array of methods to handle this particular problem has been developed.

1. A standard approach taken is shifting the barrier by a constant to account for the overestimation of the stopping time. The shifted barrier

   $$B^+ = B + \beta b(i\Delta t_l, Y_{i\Delta t_l})\sqrt{\Delta t_l}$$

   with $\beta = -\zeta(1/2)/2\pi \approx 0.5826$, where ζ is the Riemann-Zeta function, corrects the leading error in (5.38) resulting in an $o(\sqrt{\Delta t_l})$ error and $O(\Delta t_l)$ weak convergence.

2. Another approach consists of interpolating in between discretisation times with Brownian bridges and directly sampling the infimum of the bridges.

3. Instead of simulating the Brownian bridges it can be advantageous to only compute the probabilities of bridges crossing the barrier in a conditional Monte Carlo approach [Boyle et al., 1997].

The first method may seem to be the equivalent of the similar construction in (5.29), however, the correction term for lookback option payoff is not a straightforward extension of the arguments used to justify the correction for barrier payoffs [Broadie et al., 1999]. The other two methods are presented below.

Milstein Scheme

In contrast to the lookback option, the barrier option is not a Lipschitz continuous function of the infimum of the underlying asset price over the time interval $[0, T]$. Therefore, the chain of argumentation, specifically including the argument using the Lipschitz continuity of the payoff function (5.35) and Lemma 5 in the proof of the lookback option cannot be employed in the case of a barrier option.

The route taken in [Giles, 2008b], which follows an argument presented in [Glasserman, 2004], uses the unique structure of the barrier payoff function and samples the indicator $\mathbb{1}_{Y_{i\Delta t,\min}<B}$ for $i \in \{1, \ldots, n\}$ from Lemma 5 directly by calculating in each timestep the probability that the Brownian bridge used to interpolate in between discretisation points crosses the barrier. Conditional on $Y_{i\Delta t}$ and $Y_{(i+1)\Delta t}$, which are assumed to be greater than B, the indicator takes the value 1 with probability

$$\hat{p}_i = \mathbb{P}(Y_{i\Delta t,\min} \leq B | Y_{i\Delta t}, Y_{(i+1)\Delta t}) = \exp\left(-\frac{2(Y_{i\Delta t} - B)(Y_{(i+1)\Delta t} - B)}{b(i\Delta t, Y_{i\Delta t})^2 \Delta t}\right).$$

Thus, $\mathbb{1}_{\tau_B^- > T}$ can be approximated by $\prod_{i=0}^{n-1} \mathbb{1}_{U_i \leq \hat{p}_i}$ with U_i, $i \in \{0, \ldots, n-1\}$, uniformly distributed on the unit interval $(0, 1)$ and consequently the exact payoff (5.36) can be approximated by

$$f(Y_n \Delta t) \prod_{i=0}^{n-1} \mathbb{1}_{Y_{i\Delta t,\max}<B}. \tag{5.39}$$

In case the discretisation $\{Y_{i\Delta t}\}_{0 \leq i \leq n}$ is a Markov process, as is the case for the Euler-Maruyama and Milstein scheme, the indicators in (5.39) do not have to be sampled, as the conditional expectation of the payoff is then given by

$$\begin{aligned}\mathbb{E}\left(f(Y_n\Delta t) \prod_{i=0}^{n-1} \mathbb{1}_{Y_{i\Delta t,\max}<B} | Y_0, \ldots, Y_{n\Delta t}\right) &= f(Y_n\Delta t) \prod_{i=0}^{n-1} \mathbb{E}\left(\mathbb{1}_{Y_{i\Delta t,\max}<B} | Y_{i\Delta t}, Y_{(i+1)\Delta t}\right) \\ &= f(Y_n\Delta t) \prod_{i=0}^{n-1} \hat{p}_i. \end{aligned} \tag{5.40}$$

Using the conditional expectation as an estimator is an example for a more general method in Monte Carlo variance reduction [Boyle et al., 1997]. By the law of iterated expectations, the estimators (5.39) and (5.40) have the same expectation and discretisation bias, but due to the law of total variance, the variance is reduced by using the estimator (5.40) instead of (5.39). However, the computational effort in computing the estimator (5.40) is higher than the effort involved in computing the estimator (5.39), as the computation of (5.39) can be aborted as soon as $Y_{i\Delta t,\min}$ for one $i \in \{1, \ldots, n\}$ exceeds the barrier, whereas the whole path has to be simulated for the estimator (5.40) if for no $i \in \{1, \ldots, n\}$ the value $Y_{i\Delta t,\max}$ is below the barrier.

In the Multilevel Monte Carlo setting the paths on the fine level are set to

$$P_l^f = f(Y^f_{n_l \Delta t_l}) \prod_{i=0}^{n_l-1} (1 - \hat{p}_i^f),$$

with

$$\hat{p}_i^f = \exp\left(-\frac{2(Y^f_{i\Delta t_l} - B)(Y^f_{(i+1)\Delta t_l} - B)}{b(i\Delta t_l, Y^f_{i\Delta t_l})^2 \Delta t_l}\right).$$

Analogously the coarse path is defined by

$$P_l^c = f(Y^c_{n\Delta t_l}) \prod_{i=0}^{n_l-1} (1 - \hat{p}_i^c)$$

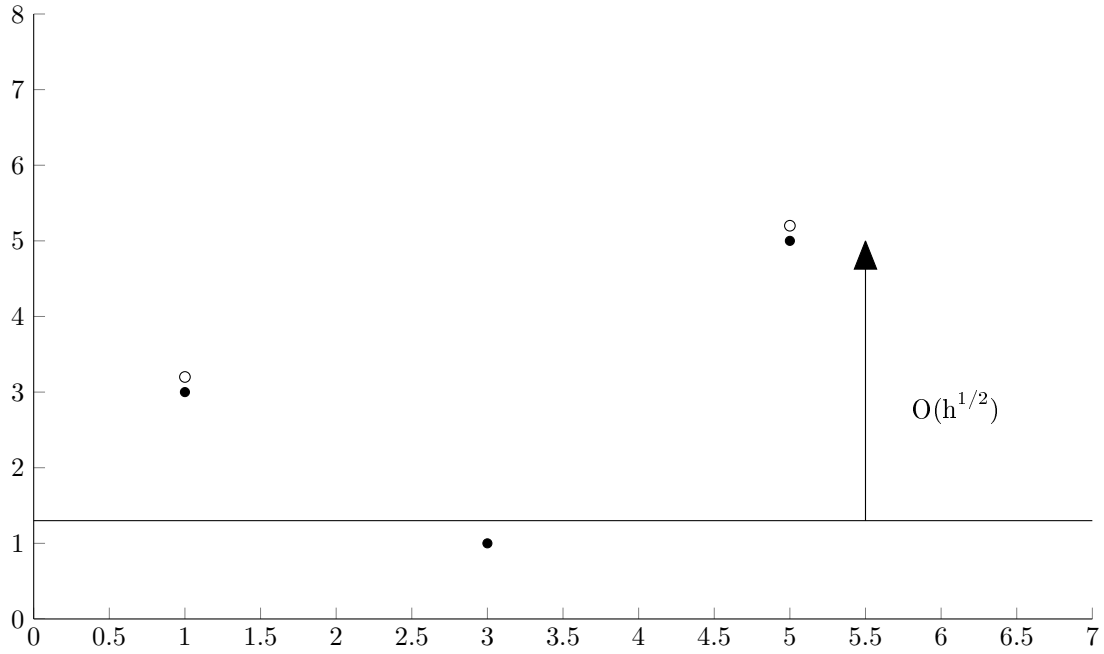

Figure 5.2: An illustration of a situation for which the improved Multilevel Monte Carlo method yields an improvement over the Classic Monte Carlo method. The conventional Multilevel Monte Carlo method would not yield an improvement, as the case of the fine path (filled circles) passing the barrier with the coarse path (empty circles) not passing the barrier still gives $\mathbb{V}ar(\Delta t_l) = O(\Delta t_l)$.

with

$$\hat{p}_i^c = \exp\left(-\frac{2(Y^c_{i\Delta t_l} - B)(Y^c_{(i+1)\Delta t_l} - B)}{b(i\Delta t_l, Y^c_{i\Delta t_l})^2 \Delta t_l}\right),$$

where for odd values of $i \in \{1, \ldots, n_l\}$ the value $Y^f_{i\Delta t_l}$ is given by the Brownian interpolant and where $b(i\Delta t_l, Y^c_{i\Delta t_l}) = b((i-1)\Delta t_l, Y^c_{(i-1)\Delta t_l})$. The identity (5.15) necessary for the improved Multilevel Monte Carlo setting to work is satisfied due to the law of iterated expectations leading to

$$\mathbb{P}\left(\inf_{t\in[i\Delta t_l,(i+2)\Delta t_l]} Y_t^c > B | Y^c_{i\Delta t_l}, Y^c_{(i+2)\Delta t_l}\right) = \mathbb{E}\left(\mathbb{P}\left(\inf_{t\in[i\Delta t_l,(i+2)\Delta t_l]} Y_t^c > B | Y^c_{i\Delta t_l}, Y^c_{(i+1)\Delta t_l} Y^c_{(i+2)\Delta t_l}\right)\right).$$

Theorem 20. *Assuming that* $\inf_{t\in[0,T]} |b(t,B)| > 0$ *and that* $\inf_{t\in[0,T]} X_t$ *has a bounded density in the neighbourhood of* B*, the Multilevel Monte Carlo estimator for a down-and-out barrier option has variance* $V_l = o(\Delta t_l^{3/2-\delta})$ *for any* $\delta > 0$ *using the Milstein scheme.*

A detailed proof of the theorem can be found in [Giles et al., 2013]. That the variance V_l is of order $o(\Delta t_l^{3/2-\delta})$ is a result of the strong convergence of the discretisation schemes used [Giles and Szpruch, 2013]. Due to Markov's inequality

$$\mathbb{P}\left(\sup_{0\le i\le n_l} |X_{i\Delta t_l} - Y_{i\Delta t_l}| \ge \Delta t_l^{1-\epsilon}\right) \le \Delta t_l^{-p+p\epsilon} \mathbb{E}\left(\sup_{0\le i\le n} |X_{i\Delta t_l} - Y_{i\Delta t_l}|^p\right) = O(\Delta t_l^{p\epsilon})$$

holds, which means that the probability that the discretisation $\{Y_{i\Delta t_l}\}_{0\le i\le n}$ is outside a $\Delta t_l^{1-\delta}$ neighbourhood of the exact underlying process $(X_t)_{t\ge 0}$ is arbitarily small. Consequently, if the infimum of $(X_t)_{t\ge 0}$ over the interval $[0,T]$ is outside of a $\Delta t_l^{1/2}$-neighbourhood of the barrier B then so is the discretisation. In this case the barrier becomes irrelevant, as asymptotically the probability of either process crossing the barrier is either 0 or 1 depending on which side of the barrier the infimum of the exact process $(X_t)_{t\ge 0}$ has realised. Therefore this situation reduces back to the case of a Lipschitz payoff function.

If on the other hand the infimum of $(X_t)_{t\geq 0}$ has realised within a $\Delta t_l^{1/2}$-neighbourhood of the barrier, then the same holds for the discretisation $\{Y_{i\Delta t_l}\}_{0\leq i\leq n_l}$. In this case it can be shown that $\mathbb{E}(P_l^f - P_{l-1}^c)^2 = O(\Delta t_l^{1-\delta})$.

In the proof it is shown that due to the boundedness assumption on the density, the probability that the infimum of $(X_t)_{t\geq 0}$ is within a $\Delta t_l^{1/2}$-neighbourhood the barrier is of order $O(\Delta t_l^{1/2-\delta})$. All put together the overall variance V_l is thus of order $o(\Delta t_l^{3/2-\delta})$ for any $\delta > 0$.

5.6.5 Digital Options

A digital call option, which is a derivative that pays out one unit if the value of the asset is above the fixed strike price K and pays out nothing otherwise, has a payoff function of the form

$$P = \mathbb{1}_{\{X_T > K\}}$$

with the numerical approximation

$$P_l = \mathbb{1}_{\{Y_{n_l \Delta t_l} > K\}}.$$

Euler-Maruyama Scheme

Theorem 21. *Under the additional assumptions that*

$$|a(0,x)| \leq C, \quad |a'(0,x)| \leq C, \quad |b(0,x)| \leq C, \quad |b'(0,x)| \leq C,$$

for some constant $C \in \mathbb{R}_+$ and

$$\mathbb{E}\left(\left|\int_0^t b(s,X_s)^2 ds\right|^{-\frac{p_0}{2}}\right) < \infty \tag{5.41}$$

for some $p_0 > 2$ and for all $t \in [0,T]$ [9] the Multilevel Monte Carlo estimator for a digital option has variance $V_l = O(\Delta t_l^{1/2-\delta})$ for any $\delta > 0$ using the Euler-Maruyama scheme.

Proof. Because of the simple structure of the payoff function, the error due the numerical approximation is readily calculated. An error in the payoff only occurs if either the exact solution is below and the approximation is above the strike price and vice versa.

$$\mathbb{E}(|P - P_l|^2) = \mathbb{P}(\{X_T > K\} \cap \{Y_T \leq K\}) + \mathbb{P}(\{X_T \leq K\} \cap \{Y_T > K\}) \tag{5.42}$$

For a given $\delta \in (0, \frac{1}{2})$, a number $m \in \mathbb{N}$ large enough may be chosen such that $1/(2m+2) < \delta$. It is then convenient to set

$$\hat{\beta} = \frac{1}{2} - \frac{1}{2m+2} > \frac{1}{2} - \delta. \tag{5.43}$$

For the first summand on the right hand side of (5.42), which handles the case that the exact solution is above the strike but the discretisation is below the strike, the inequality

$$\begin{aligned}\mathbb{P}(\{X_T > K\} \cap \{Y_T \leq K\}) &= \mathbb{P}(\{K + \Delta t_l^{\hat{\beta}} > X_T > K\} \cap \{Y_T \leq K\}) + \mathbb{P}(\{X_T > K + \Delta t_l^{\hat{\beta}}\} \cap \{Y_T \leq K\}) \\ &= \mathbb{P}(\{K + \Delta t_l^{\hat{\beta}} > X_T > K\} \cap \{Y_T \leq K\}) + \mathbb{P}(\{X_T - Y_T > \Delta t_l^{\hat{\beta}}\})\end{aligned} \tag{5.44}$$

holds for the threshold value $\Delta t_l^{\hat{\beta}}$, where $\hat{\beta}$ in (5.43) has been chosen in such a way that it balances two opposing aims. On the one hand, it is desirable to keep $\Delta t_l^{\hat{\beta}}$ small in order to keep $\mathbb{P}(\{K + \Delta t_l^{\hat{\beta}} > X_T > K\})$ small. On the other hand, $\Delta t_l^{\hat{\beta}}$ should be large so that by strong convergence the probability $\mathbb{P}(\{X_T - Y_T > \Delta t_l^{\hat{\beta}}\})$ is small. Due to assumption (5.41) the density of X_T is bounded [Caballero et al., 1998] resulting in

$$\mathbb{P}(\{K + \Delta t_l^{\hat{\beta}} > X_T > K\} = O(\Delta t_l^{\hat{\beta}}). \tag{5.45}$$

On the other hand the Markov inequality together with the bound (3.5) yields

$$\mathbb{P}(\{X_T - Y_T > \Delta t_l^{\hat{\beta}}\}) \leq \frac{\mathbb{E}(|X_T - Y_T|^m)}{\Delta t_l^{\hat{\beta}m}} \leq \frac{C_m \Delta t_l^{m/2}}{\Delta t_l^{\hat{\beta}m}} = O\left(\Delta t_l^{\frac{1}{2}-\frac{1}{2m+2}}\right) = O(\Delta t_l^{\hat{\beta}}). \tag{5.46}$$

Thus, plugging in (5.46) and (5.45) into (5.44) results in the desired

$$\mathbb{E}(|P - P_l|^2) = O(\Delta t_l^{\hat{\beta}}) = O(\Delta t_l^{\frac{1}{2}-\delta}).$$

□

Milstein Scheme

In [Giles et al., 2013] the Multilevel Monte Carlo estimator variance for digital options was analysed for the Milstein scheme after numerical results were obtained in [Giles, 2008b]. Again a conditional expectation approach is used.

Using the same approach as for global Lipschitz functions would lead to an $O(\Delta t_l)$ fraction of all S_T of fine and coarse paths to end up on different sides of the strike K, leading to an error of $|P_l^f - P_{l-1}^c| = 1$. This leads to an overall estimator variance of $V_l = O(\Delta t_l)$ in the absence of any advanced treatment.

The basic idea is to observe the behaviour of a path one time step before T, that is at time $(n_l - 1)\Delta t_l$, where the value of the fine path approximation is $Y^f_{(n_l-1)\Delta t_l}$. Approximating the path in between $\Delta t_l(n_l - 1)$ and $n_l \Delta t_l$ with a Brownian motion that has constant drift $a^f_{(n_l-1)\Delta t_l} = a(Y^f_{(n_l-1)\Delta t_l}, T - \Delta t_l)$ and constant volatility $b^f_{(n_l-1)\Delta t_l} = b(Y^f_{(n_l-1)\Delta t_l}, T - \Delta t_l)$, results in a conditional expectation for the payoff at time T that is given by $\mathbb{P}(Y^f_{n_l \Delta t_l} > K)$ after one additional time step or

$$P_l^f = \mathbb{E}(\mathbb{1}_{\{Y^f_{n_l\Delta t_l} > K\}} | Y^f_{(n_l-1)\Delta t_l}) = \Phi\left(\frac{Y^f_{(n_l-1)\Delta t_l} + a^f_{(n_l-1)\Delta t_l}\Delta t_l - K}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}}\right) \tag{5.47}$$

with the standard normal cumulative distribution function Φ, due to the fact that an arithmetic Brownian motion at a time instance $t \in \mathbb{N}$ follows a normal distribution. Given the Brownian increment $\Delta W_{(n_{l-1}-1)\Delta t_{l-1}+\Delta t_l}$, which is the same increment as used in the fine path simulation, the probability $\mathbb{P}(Y^c_{n_{l-1}\Delta t_{l-1}} > K)$ is given by

$$\begin{aligned} P_{l-1}^c &= \mathbb{E}(\mathbb{1}_{\{Y^c_{n_{l-1}\Delta t_{l-1}} > K\}} | Y^c_{(n_{l-1}-1)\Delta t_{l-1}}, \Delta W_{(n_{l-1}-1)\Delta t_{l-1}+\Delta t_l}) \\ &= \Phi\left(\frac{Y^c_{(n_{l-1}-1)\Delta t_{l-1}} + a^c_{(n_{l-1}-1)\Delta t_{l-1}}\Delta t_{l-1} + b^c_{(n_{l-1}-1)\Delta t_{l-1}}\Delta W_{(n_{l-1}-1)\Delta t_{l-1}+\Delta t_l} - K}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|\sqrt{\Delta t_l}}\right) \end{aligned} \tag{5.48}$$

The conditional expectation of P_{l-1}^c is equal to the conditional expectation of P_{l-1}^f, satisfying equality (5.15).

Theorem 22. *Assuming that $b(T, K) \neq 0$ and that X_t has a bounded density in the neighbourhood of K, the Multilevel Monte Carlo estimator for a digital option has variance $V_l = o(\Delta t_l^{3/2-\delta})$ for any $\delta > 0$ using the Milstein scheme.*

Proof. The proof is is split into three parts. In the first part, **extreme paths** are considered, in the second part, **paths which are not extreme with** $|X_T - K| > \Delta t_l^{1/2-4\gamma}$ and in the third part the **the rest** of the remaining paths.

a.) Paths that are defined to be **extreme** satisfy any of the conditions of Lemma 11 for $0 < \gamma < \frac{1}{4}$. The fourth moments $\mathbb{E}\left((P^f)^4\right)$ and $\mathbb{E}\left((P^c)^4\right)$ are both finite. Therefore, the extreme paths contribute $o(\Delta t_l^p)$ for all $p > 0$ to the overall variance.

b.) **Paths which are not extreme with** $|X_T - K| > \Delta t_l^{1/2-4\gamma}$ need their own analysis. Let $Y^f_{n_l\Delta t_l}$ and $Y^c_{n_{l-1}\Delta t_{l-1}}$ be the values that are obtained in the fine, respectively coarse, path simulations after the final time step, then by plugging in the Milstein approximation (3.14)

$$\begin{aligned}&\frac{Y^f_{n_l\Delta t_l} + a^f_{(n_l-1)\Delta t_l}\Delta t_l - K}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}} \\ &\quad = \frac{Y^f_{n_l\Delta t_l} - K}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}} - \frac{b^f_{(n_l-1)\Delta t_l}}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}}\left(\Delta W_{(n_l-1)\Delta t_l} + \frac{1}{2}(b')^f_{(n_l-1)\Delta t_l}\left((\Delta W_{(n_l-1)\Delta t_l})^2 - \Delta t_l\right)\right)\end{aligned} \tag{5.49}$$

and

$$\begin{aligned}&\frac{Y^c_{(n_{l-1}-1)\Delta t_{l-1}} + a^c_{(n_{l-1}-1)\Delta t_{l-1}}\Delta t_{l-1} + b^c_{(n_{l-1}-1)\Delta t_{l-1}}\Delta W_{(n_{l-1}-1)\Delta t_{l-1}} - K}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|\sqrt{\Delta t_l}} \\ &\quad = \frac{Y^c_{n_{l-1}\Delta t_{l-1}} - K}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|\sqrt{\Delta t_l}} - \frac{b^c_{(n_{l-1}-1)\Delta t_{l-1}}}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|\sqrt{\Delta t_l}} \\ &\qquad \cdot\left(\Delta W_{(n_l-1)\Delta t_l} + \frac{1}{2}(b')^c_{(n_{l-1}-1)\Delta t_{l-1}}\left((\Delta W_{(n_{l-1}-1)\Delta t_{l-1}} + \Delta W_{(n_l-1)\Delta t_l})^2 - \Delta t_{l-1}\right)\right).\end{aligned} \tag{5.50}$$

Because the paths are not extreme $|\Delta W_{i\Delta t_l}| \leq \Delta t_l^{1/2-\gamma}$ for all $i \in \{1, \ldots, n_l\}$ and $|X_T - Y^f_{n_l\Delta t_l}| \leq \Delta t_l^{1-\gamma}$. Due to Lemma 11, $|b^f_{(n_l-1)\Delta t_l}| \prec \Delta t_l^{-\gamma}$. Combining these statements leads in the case of $X_T > K + \Delta t_l^{1/2-3\gamma}$ for sufficiently small Δt_l to

$$\frac{Y^f_{(n_l-1)\Delta t_l} + a^f_{(n_l-1)\Delta t_l}\Delta t_l - K}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}} > C\Delta t_l^{-2\gamma}$$

for an appropriate constant $C \in \mathbb{R}$. An analogue result follows for the corresponding coarse path, which results in $P_l^f - P_{l-1}^c = o(\Delta t_l^p)$ for all $p > 0$.

An analogue argument applies to the paths for which $X_T < K - \Delta t_l^{1/2-3\gamma}$. Combining both cases leads to $\mathbb{E}((P_l^f - P_{l-1}^c)^2 \mathbb{1}_{b.)}) = o(\Delta t_l^p)$ for all $p > 0$.

c.) **The rest** of the paths satisfy $|X_T - K| \leq \Delta t_l^{1/2-2\gamma}$. The triangle inequality yields

$$|b^f_{(n_l-1)\Delta t_l} - b(T,K)| \leq |b^f_{(n_l-1)\Delta t_l} - b^f_{n_l\Delta t_l}| + |b^f_{n_l\Delta t_l} - b(T,K)|. \tag{5.51}$$

The first summand on the right hand side in (5.51) is bound by $\Delta t^{1/2-2\gamma}$ with $\gamma < \frac{1}{4}$ due to the relation (5.20),

$$|\max_{0\leq i\leq n} |b(i\Delta t, Y^f_{i\Delta t}) - b((i-1)\Delta t, Y^f_{(i-1)\Delta t})| \prec \Delta t^{1/2-2\gamma}$$

in Lemma 11. For the second summand on the right hand side in (5.51) the Lipschitz condition imposed on $b(t,x)$ in (3.13) in Theorem 8 provides the upper bound

$$|b^f_{n_l\Delta t_l} - b(T,K)| \leq C|Y^f_{n_l\Delta t_l} - X_T + X_T - K| \leq C(|Y^f_{n_l\Delta t_l} - X_T| + |X_T - K|) \leq C(\Delta t_l^{1-\gamma} + \Delta t_l^{1/2-2\gamma}).$$

For small enough Δt_l the left hand side in (5.51) can thus be bound by $|b^f_{(n_l-1)\Delta t_l} - b(T,K)| < \frac{1}{2}b(T,K)$. This shows that $b^f_{(n_l-1)\Delta t_l}$ is not zero and of the same sign as $b(T,K)$. The same chain of argumentation can be applied to $b^c_{(n_{l-1}-1)\Delta t_{l-1}}$.

Applying the fact that the cumulative distribution function of the normal distribution Φ is Lipschitz continuous with Lipschitz constant $C = 1$, $|\Phi(x) - \Phi(y)| \leq |x - y|$, to the fine payoff approximation P_l^f in

(5.47) and the coarse payoff approximation P^c_{l-1} in (5.48) and using the identities (5.49) and (5.50) leads to

$$|P^f_l - P^c_{l-1}| \leq \left| \frac{Y^f_{(n_l-1)\Delta t_l} + a^f_{(n_l-1)\Delta t_l}\Delta t_l - K}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}} \right. \tag{5.52}$$
$$\left. - \frac{Y^c_{(n_{l-1}-1)\Delta t_{l-1}} + a^c_{(n_{l-1}-1)\Delta t_{l-1}}\Delta t_{l-1} + b^c_{(n_{l-1}-1)\Delta t_{l-1}}\Delta W_{(n_{l-1}-1)\Delta t_{l-1}+\Delta t_l} - K}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|\sqrt{\Delta t_l}} \right|$$
$$\leq \left| \frac{Y^f_{n_l\Delta t_l} - K}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}} - \frac{Y^c_{n_{l-1}\Delta t_{l-1}} - K}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|\sqrt{\Delta t_l}} \right|$$
$$+ \frac{1}{2\sqrt{\Delta t_l}} K \left((\Delta W_{(n_l-1)\Delta t_l})^2 + (\Delta W_{(n_{l-1}-1)\Delta t_{l-1}} + \Delta W_{(n_l-1)\Delta t_l})^2 + 3\Delta t_l \right)$$

In the next step, the identity

$$f_1 g_1 - f_2 g_2 = \frac{1}{2}(f_1 - f_2)(g_1 + g_2) + \frac{1}{2}(f_1 + f_2)(g_1 - g_2)$$

is used to arrive at

$$\frac{Y^f_{n_l\Delta t_l} - K}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}} - \frac{Y^c_{n_{l-1}\Delta t_{l-1}} - K}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|\sqrt{\Delta t_l}} = \frac{1}{2\sqrt{\Delta t_l}}(Y^f_{n_l\Delta t_l} - Y^c_{n_{l-1}\Delta t_{l-1}}) \left(\frac{1}{|b^f_{(n_l-1)\Delta t_l}|} + \frac{1}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|} \right)$$
$$+ \frac{1}{2\sqrt{\Delta t_l}}(Y^f_{n_l\Delta t_l} + Y^c_{n_{l-1}\Delta t_{l-1}} - 2K) \left(\frac{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}| - |b^f_{(n_l-1)\Delta t_l}|}{|b^f_{(n_l-1)\Delta t_l}| \cdot |b^c_{(n_{l-1}-1)\Delta t_{l-1}}|} \right).$$

Using the identities (5.19), (5.21) and (5.22) it follows that

$$\frac{Y^f_{n_l\Delta t_l} - K}{|b^f_{(n_l-1)\Delta t_l}|\sqrt{\Delta t_l}} - \frac{Y^c_{n_{l-1}\Delta t_{l-1}} - K}{|b^c_{(n_{l-1}-1)\Delta t_{l-1}}|\sqrt{\Delta t_l}} = O(\Delta t_l^{1/2-5\gamma})$$

and in consequence due to (5.52) that $P^f_l - P^c_{l-1} = O(\Delta t_l^{1/2-5\gamma})$. Due to assumption that the probability density of X_T is bounded, the probability that a path belongs to the rest of the paths treated in section c) is given by the expectation $\mathbb{E}(\mathbb{1}_{c.}) = O(\Delta t_l^{1/2-3\gamma})$. Combining these results yields $\mathbb{E}((P^f_l - P^c_{l-1})^2 \mathbb{1}_{c.}) = O(\Delta t_l^{3/2-13\gamma})$. Setting $\gamma < \min(\frac{1}{4}, \delta/13)$ leads to the desired result.

□

5.6.6 Basket Options

Basket options are derivatives whose payoff is dependent on the weighted average of J underlying assets, $X^A_t = \sum_{j=1}^J \mu_j X^j_t$. Each of these underlying asset price processes X^j is assumed to be the solution of a stochastic differential equation of the form given in (3.2),

$$dX^j_t = a^j(t, X^j_t)dt + b^j(t, X^j_t)dW^j_t,$$

where Σ is the correlation matrix of the multidimensional Brownian motion $W_t = (W^1_t, \ldots, W^J_t)$. As in the case of the other derivatives treated above, Brownian interpolation is used in between discretisation points. It can be shown that the resulting averaged Brownian interpolation has the same form as (5.18) [Giles, 2009], rendering the construction of the Multilevel Monte Carlo estimators exactly the same as in the case of the non-Basket derivatives. Let $\{Y^A_{i\Delta t}\}_{1\leq i\leq n}$ denote the averaged discretisation values of the underlying assets, that is $Y^A_{i\Delta t} = \sum_{j=1}^J \mu_j Y^j_{i\Delta t}$. Then the Brownian interpolant in between discretisation points is given by

$$Y^A_t = Y^A_{i\Delta t} + \lambda(Y^A_{(i+1)\Delta t} - Y^A_{i\Delta t}) + \sum_{j=1}^J \mu_j b^j(i\Delta t, Y^j_{i\Delta t}) \left(W^j_t - W^j_{i\Delta t} - \lambda(W^j_{(i+1)\Delta t} - W^j_{i\Delta t} \right)$$
$$= Y^A_{i\Delta t} + \lambda(Y^A_{(i+1)\Delta t} - Y^A_{i\Delta t}) + \tilde{b}_{i\Delta t}(\tilde{W}_t - \tilde{W}_{i\Delta t} - \lambda(\tilde{W}_{(i+1)\Delta t} - \tilde{W}_{i\Delta t}),$$

where $\tilde{W}_t = \sum_{j=1}^{J} \mu_j W_t^j$ and where

$$\tilde{b}_{i\Delta t}^2 = \sum_{j,k} \mu_j b^j(i\Delta t, Y_{i\Delta t}^j)\Sigma^{(j,k)}\mu_k b^k(i\Delta t, Y_{i\Delta t}^k).$$

Y_t^A thus has the same form as the interpolant defined in (5.18).

No numerical analysis of Basket options for the Multilevel Monte Carlo method exists as of yet. However, the construction of the estimators suggests that convergence rates remain in line with the rates derived for the non-Basket derivatives and thus optimal. Giles has found as much for the case of Black-Scholes model in [Giles, 2009] via numerical simulation. Improvements in terms of computational cost over the Classic Monte Carlo simulation range from an 100-fold improvement in case of the Asian option, 150-fold in case of the lookback option, 100-fold in case of the barrier option to 500-fold in case of the Digital option, always considered with a target accuracy level of $\epsilon = 0.01$.

5.6.7 Summary

A look at the table of convergence rates of V_l for the derivatives presented above shows that β, determined by V_l, always satisfies $\beta > 1$ in the case of the Milstein scheme. Under the assumption $\gamma = 1$ Theorem (31) guarantees that in every case an optimal computational complexity of $O(\epsilon^{-2})$ is achieved. Thus, the Multilevel Monte Carlo Method effectively brings down the computational complexity of $O(\epsilon^{-3})$ under the Classic Monte Carlo method to $O(\epsilon^{-2})$.

It should be explicitly said that the Multilevel Monte Carlo method achieves optimal results asymptotically, but not in every case that may arise in industry applications. If, for example ϵ is relatively large, the savings achieved by coupled levels of grid granularity are trumped by the additional computational overhead that is necessary to estimate the variances V_l for every $l \in \{1, \ldots, L\}$. Roughly, the savings outweigh the overhead obtained through the Multilevel construction for an $\epsilon \approx 2^{-10}$. Additionally, if the estimator variance on the coarse levels is relatively small in comparison to the finer levels, the computational cost is spread more evenly across all levels, defeating the purpose of the Multilevel approach.

Nonetheless, the Multilevel Monte Carlo method features enormous savings in the case of the financial derivatives discussed in the preceding sections, as in these cases the variance of the estimators is relatively uneven going from coarse to fine levels. If a high accuracy in form of a very small ϵ is required, the method offers vast computational savings. Its usefulness can be increased even more by performing the computations on a processor which allows parallelisation, as the independent sampling of the estimators on different levels of the Multilevel Monte Carlo technique is predilected to be done in parallel.

	Euler Scheme		Milstein Scheme	
Option Type	Numerical	Analytical	Numerical	Analytical
Lipschitz Payoff	$O(\Delta t_l)$	$O(\Delta t_l)$	$O\left(\Delta t_l^2\right)$	$O\left(\Delta t_l^2\right)$
Asian Option	$O(\Delta t_l)$	$O(\Delta t_l)$	$O\left(\Delta t_l^2\right)$	$O\left(\Delta t_l^2\right)$
Lookback Option	$O(\Delta t_l)$	$O(\Delta t_l)$	$O\left(\Delta t_l^2\right)$	$O\left(\Delta t_l^2(\log \Delta t_l)^2\right)$
Barrier Option	$O\left(\Delta t_l^{1/2}\right)$	$o\left(\Delta t_l^{1/2-\delta}\right)$	$O\left(\Delta t_l^{3/2}\right)$	$o\left(\Delta t_l^{3/2-\delta}\right)$
Digital Option	$O\left(\Delta t_l^{1/2}\right)$	$O\left(\Delta t_l^{1/2} \log \Delta t_l\right)$	$O\left(\Delta t_l^{3/2}\right)$	$o\left(\Delta t_l^{3/2-\delta}\right)$

Table 5.1: Convergence rates for the variance V_l, proven analytically and shown numerically using the Euler scheme respectively the Milstein scheme [Giles et al., 2013]. $\delta > 0$ is an arbitrary constant.

Chapter 6

The Multilevel Wiener-Hopf Monte Carlo Simulation

6.1 The Wiener-Hopf Factorisation

A fundamental property common to all Lévy processes is the Wiener-Hopf factorisation, which is a highly efficient method to study vital functionals of Lévy processes like the extrema of the underlying process, first passage times and the resulting overshoot. The factorisation is heavily used in theoretical research [Bertoin, 1996], [Sato, 1999], but due to the nature of many problems in mathematical finance, such as the pricing of derivatives involving the extrema of the underlying asset process like lookback or barrier options and the modelling of default processes credit risk management [Schoutens, 2003], [Cariboni and Schoutens, 2009], it is of increasing interest in areas connected to finance. In insurance, ruin processes are frequently modelled by Lévy processes, where the problems are also strongly related to quantities which can be treated with the Wiener-Hopf factorisation [Albrecher et al., 2008].

6.1.1 History of the Wiener-Hopf factorisation

The factorisation has its basis in a method developed by Norbert Wiener [Wiener and Payley, 1934] and Eberhard Hopf [Hopf, 1934] to solve for $f(x)$ linear integral equations of the first kind[10], which take the form

$$\int_0^\infty k(x-y)f(y)dy = g(x) \tag{6.1}$$

with $x \in (0,\infty)$, $k(x-y)$ a known function depending only on the difference $x-y$, the difference kernel [11], and $g(x)$ a given function defined for $x \in (0,\infty)$. In a first step the method constructs a continuation of the integral equation to values on the negative axis by setting

$$\int_0^\infty k(x-y)f(y)dy = \begin{cases} g(x), & x \in (0,\infty) \\ h(x), & x \in (-\infty,0) \end{cases} \tag{6.2}$$

where $h(x)$ is an unknown function. This allows to perform a Fourier transformation of (6.2) resulting in

$$G_+(\alpha) + H_-(\alpha) = F_+(\alpha)K(\alpha), \tag{6.3}$$

where $H_-(\alpha)$ and $F_-(\alpha)$ are half-range Fourier transforms respectively defined over $(-\infty,0)$ and $(0,\infty)$ of the unknown functions $f(x)$ and $h(x)$, whereas $G_+(\alpha)$ is the known half-range Fourier transform of $g(x)$ and $K(\alpha)$ is the full-range Fourier transform of $k(x)$. The subscripts show whether the respective functions are analytic in the upper or in the lower half regions of the complex plane.
The Wiener-Hopf method exploits the fact that a factorisation, the Wiener-Hopf factorisation, of the Fourier-transformed kernel can be found, that is

$$K(\alpha) = K_+(\alpha)K_-(\alpha), \tag{6.4}$$

where both factors on the right hand side are zero-free in their respective half-planes of analyticity. From this decomposition it follows that (6.3) can be written as

$$\frac{G+(\alpha)}{K_-(\alpha)} + \frac{H_-(\alpha)}{K_-(\alpha)} = F_+(\alpha)K_+(\alpha).$$

Factorizing the first summand further [12] into

$$\frac{G_+(\alpha)}{K_-(\alpha)} = L_+(\alpha) + L_-(\alpha)$$

makes it possible to rewrite (6.3) as

$$L_-(\alpha) + \frac{H_-(\alpha)}{K_-(\alpha)} = F_+(\alpha)K_+(\alpha) - L_+(\alpha).$$

The left hand side contains only functions analytic in the lower half-plane and the right hand side contains only functions analytic in the upper half region. The next step consists of employing analytic continuation arguments to express both sides as an entire function $E(\alpha)$ and using constraints on the asymptotic behaviour of $f(x), g(x)$ and $k(x)$ for $x \to 0$ and respectively on their Fourier-transforms for $\alpha \to \infty$ to determine $E(\alpha)$, thereby working out the unique $H_-(\alpha)$ and $F_+(\alpha)$ fixed by $E(\alpha)$. In the last step, Fourier inversion gives the desired function $f(x)$ [Lawrie and Abrahams, 2007].

6.1.2 The Wiener-Hopf Factorisation for Lévy Processes

The Wiener-Hopf factorisation of a Lévy process $X = (X_t)_{t\geq 0}$ is given by a decomposition of the same type as in (6.4). Let a compound Poisson process X with the property that $\limsup_{t\to\infty} X_t < \infty$ be given. It can be shown that the Strong Markov property [13] yields that $\sup_{s\leq\infty} X_s$ is equal in distribution to $\max(\zeta + \sup_{s\leq\infty} X_s, 0)$ where ζ is not dependent on $\sup_{s\leq\infty} X_s$ and equal in distribution to the jump distribution of X [Kyprianou, 2006]. Under the assumption that the jump distribution of X has density k it directly follows that

$$\mathbb{P}\left(\sup_{s\leq\infty} X_s \leq x\right) = \int_{-\infty}^{x} \mathbb{P}\left(\sup_{s\leq\infty} X_s \leq x - y\right) k(y)dy = \int_0^{\infty} k(x-y)\mathbb{P}\left(\sup_{s\leq\infty} X_s \leq y\right) dy.$$

Setting $g(x) = f(x) = \mathbb{P}\left(\sup_{s\leq\infty} X_s \leq x\right)$ gives equation (6.1).
To switch to a more general setting, let $\varsigma(q)$ for any $q > 0$ denote an exponentially distributed random variable with mean $\mathbb{E}(\varsigma(q)) = 1/q$ and which is independent of X. Additionally, the notation

$$\overline{X}_t = \sup_{s\leq t} X_s \quad \text{and} \quad \underline{X}_t = \inf_{s\leq t} X_s$$

is introduced. Furthermore let $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im(z) > 0\}$ denote the open upper complex half plane and $\overline{\mathbb{C}}^- = \{z \in \mathbb{C} : \Im(z) \geq 0\}$ the closed upper complex half plane and let the open/closed lower complex half plane be defined analogously.

Theorem 23. *Assume that $q > 0$. Let the Wiener-Hopf factors be defined by*

$$\phi_q^{\pm}(u) = \exp\left(\int_{\mathbb{R}^+} \frac{1}{te^{qt}} \int_{\mathbb{R}^{\pm}} (e^{iux} - 1)d\mathbb{P}_{X_t}dt\right),\ u \in \overline{\mathbb{C}}^{\pm}$$

1. *For all $u \in \mathbb{R}$ the Wiener-Hopf factorisation*

$$\frac{q}{q+\psi(u)} = \phi_q^+(u)\phi_q^-(u)$$

   holds, where $\psi(u)$ is the characteristic exponent of the Lévy process X.
2. *The Wiener-Hopf factors $\phi_q^+(u)$ and $\phi_q^-(u)$ may be also expressed by*

$$\phi_q^{\pm}(u) = \exp\left(\pm\frac{u}{2\pi i}\int_{\mathbb{R}} \ln\left(\frac{q}{q+\psi(z)}\right)\frac{dz}{z(z-u)}\right),\ u \in \mathbb{C}^{\pm}$$

3. *The identities*

$$\phi_q^+(u) = \mathbb{E}\left(e^{iu\overline{X}_{\varsigma(q)}}\right),\ u \in \mathbb{C}^+ \quad \text{and} \quad \phi_q^-(u) = \mathbb{E}\left(e^{iu\underline{X}_{\varsigma(q)}}\right),\ u \in \mathbb{C}^-$$

hold.

4. $\overline{X}_{\varsigma(q)}$ *and* $\underline{X}_{\varsigma(q)}$ *are independent positive, respectively negative, infinitely divisible random variables with zero linear drift. Additionally,* $\underline{X}_{\varsigma(q)} \stackrel{D}{=} X_{\varsigma(q)} - \overline{X}_{\varsigma(q)}$.
5. *Assume that there exist a positive infinitely divisible random variable* X_1 *and a negative infinitely divisible random variable* X_2 *with zero linear drift such that for all* $u \in \mathbb{R}$

$$\frac{q}{q+\psi(u)} = \mathbb{E}\left(e^{iuX_1}\right)\mathbb{E}\left(e^{iuX_2}\right).$$

Then $\overline{X}_{\varsigma(q)} \stackrel{D}{=} X_1$ *and* $\underline{X}_{\varsigma(q)} \stackrel{D}{=} X_2$.

6. *Assume that there exist two functions* $f^+(u)$ *and* $f^-(u)$ *such that* $f^\pm(0) = 1$, $f^\pm(u)$ *are analytic in* $\mathbb{C}^\pm$, *continuous and have no roots in* $\overline{\mathbb{C}}^\pm$ *and* $u^{-1}\ln(f^\pm(u)) \to 0$ *as* $u \to \infty$, $u \in \overline{\mathbb{C}}^\pm$. *if*

$$\frac{q}{q+\psi(u)} = f^+(u)f^-(u), \quad u \in \mathbb{R}$$

then $f^\pm(u) = \phi_q^\pm(u)$ *for all* $u \in \overline{\mathbb{C}}^\pm$.

Exit problems, such as an asset price leaving an area described by a barrier, are closely connected to the Wiener-Hopf factorisation. For instance, if the positive Wiener-Hopf factor is known, then through the Pecherskii-Rogozin identity [Pecherskii and Rogozin, 1969] the joint Laplace transform of the first passage time and the overshoot is known.

Definition 37. *Let the first passage times* τ_h^+ *and* τ_h^- *of* X *be defined by*

$$\tau_h^+ = \inf\{t > 0 : X_t > h\} \quad \text{and} \quad \tau_h^- = \inf\{t > 0 : X_t < h\}$$

and let the respective overshoots be defined to be equal to $X_{\tau_h^+} - h$ *and* $h - X_{\tau_h^+}$.

Theorem 24. *Assume that* $q > 0$, $h \geq 0$, $\Re(w) > 0$ *and* $\Re(z) > 0$ *and* $w \neq z$. *Then*

$$\int_{\mathbb{R}^+} e^{-wh}\mathbb{E}\left(e^{-qT_h^+ - z(X_{\tau_h^+} - h)}\right)dh = \frac{1}{w-z}\left(1 - \frac{\phi_q^+(iw)}{\phi_q^-(iz)}\right).$$

Since $\tau_h^-(X) = \tau_{-h}^+(-X)$, an analogue statement holds for the joint law of $(\tau_h^-(X), X_{\tau_h^-(X)})$ for $h < 0$. The joint law of the first passage time and the overshoot is of natural interest in finance and insurance problems. For example, the pricing of an up-and-in barrier option requires evaluating an expectation of the form $\mathbb{E}(f(X_T)\mathbb{1}_{\overline{X}_T > B})$, the valuation of a credit default swap involves knowledge of the distribution function $\mathbb{P}(\overline{X}_T < x)$

Unfortunately, although the Wiener-Hopf factorisation holds for all Lévy processe, the factors $\phi_q^\pm$ in general do not allow for an explicit form [Kuznetsov and Peng, 2012]. Therefore, the search for closed-form Wiener-Hopf factors is reduced to finding rich enough families of Lévy processes with properties that make it possible to derive an explicit Wiener-Hopf factorisation.

For some classes of Lévy processes it was shown that the Wiener-Hopf factors can be identified and consequently that the distribution of the variables $\overline{X}_{\varsigma(q)}$ and $\underline{X}_{\varsigma(q)}$ can be explicitly determined. The sub-classes of Lévy processes are distinguished by the form their Lévy measure takes. More precisely, for the Lévy processes in the different sub-classes the Lévy density $f(x)$ exists, which is the density of the Lévy measure with respect to the Lebesgue measure, $v(dx) = f(x)dx$, and the processes are categorised according to their Lévy density.

1. Processes with jumps of rational transform. A Lévy process X is said to have jumps of rational transform, if its characteristic function is a rational function, that is the quotient of two polynomials. It can be shown that the Lévy density of such a process necessarily takes the form

$$f(x) = \sum_{k=1}^{m}\sum_{j=1}^{n_k} c_{kj}\alpha_k^j \frac{x^{j-1}}{(j-1)!}e^{-\alpha_k x}$$

with parameters $m, n_k \in \mathbb{N}$ and $c_{kj}, \alpha_k \in \mathbb{C}$. The $\alpha_k \in \mathbb{C}$ must furthermore satisfy

$$0 < \alpha_1 < \Re(\alpha_2) \leq \cdots \leq \Re(\alpha_m).$$

Processes with positive jumps of rational transform, which are defined by a Lévy measure

$$\nu(dx) = \begin{cases} \nu^+(dx) = \lambda p(x)dx, & x > 0 \\ \nu^-(dx), & x < 0 \end{cases}$$

where the behaviour of the negative jumps is determined by an arbitrary Lévy measure $\nu^-(dx)$ concentrated on the set $(-\infty, 0)$ and where the behaviour of the positive jumps is described by an intensity $\lambda > 0$ and a Lévy density $f(x)$, were studied in [Lewis and Mordecki, 2008].
Sub-classes of processes with jumps of rational transform are the following.

(a) Processes with phase-type jumps. A Lévy process X is said to have jumps of phase-type if its Lèvy density has the form

$$f(x) = ve^{Rx}r,$$

where $v = (v_1, \ldots, v_d)$, $d \in \mathbb{N}$, is the initial probability distribution of a finite state continuous time Markov process with one absorbing state and the remaining $d-1$ states transient, $R \in \mathbb{R}^{d\times d}$ is its intensity matrix and where $r \in \mathbb{R}^d$ is the associated exit rates vector, which together satisfy $r_j + \sum_{k=1}^d R_{jk} = 0$ for $j \in \{1, \ldots, d\}$. Phase-type distributions defined by the density $f(x)$ include and generalise the exponential distribution, which is exactly the phase-type distribution with $d = 1$, as well as the Erlang distribution and the hyper-exponential distribution explicitly listed below. Phase-type distributions form a dense subset of the set of all distributions on $(0, \infty)$ [Asmussen et al., 2000]. In [Pistorius, 2006] it is shown that a general Lévy process can be approximated with arbitrary precision by processes where the positive jumps follow phase type distributions.

(b) Processes with hyper-exponential jumps. A Lévy process X is said to have hyper-exponential jumps if its Lévy density has the form

$$f(x) = \mathbb{1}_{\{x>0\}} \sum_{i=1}^m p_i\eta_i e^{-\eta_i x} + \mathbb{1}_{\{x<0\}} \sum_{j=1}^n q_j\theta_j e^{\theta_j x}$$

with $p_i, \eta_i, q_j, \theta_j \in \mathbb{R}^+$ and $\eta_i > 1$ as well as $\sum_{i=1}^m p_i + \sum_{j=1}^n q_j = 1$. The sub-class of Lévy processes with hyper-exponential jumps is large enough to approximate processes with heavy-tailed distributed jumps with arbitrary precision [Cai, 2009].

2. Meromorphic processes. A Lévy process X is said to be meromorphic if its density has the form

$$f(x) = \mathbb{1}_{\{x>0\}} \sum_{i=1}^\infty p_i\eta_i e^{-\eta_i x} + \mathbb{1}_{\{x<0\}} \sum_{j=1}^\infty q_j\theta_j e^{\theta_j x}$$

with $p_i, \eta_i, q_j, \theta_j \in \mathbb{R}^+$ and strictly increasing sequences $\{\eta_i\}_{i\in\mathbb{N}}$ and $\{\theta_j\}_{j\in\mathbb{N}}$ that satisfy $\lim_{i\to\infty} \eta_i = \lim_{j\to\infty} \theta_j = \infty$. To ensure that the convergence condition

$$\int_{|x|\le 1} |x|^2 \nu(dx) = \int_{|x|\le 1} |x|^2 f(x)dx < \infty$$

in (1.1) is satisfied, the additional constraint

$$\sum_{i=1}^\infty p_i\eta_i^{-2} + \sum_{j=1}^\infty q_j\theta_j^{-2} < \infty$$

has to be imposed. The characteristic exponent $\psi(x)$ of the Lévy process X is a meromorphic function which has poles at the points $\{-\mathrm{i}\eta_i, \mathrm{i}\theta_j\}_{i,j\in\mathbb{N}}$ [Kuznetsov et al., 2012]. Clearly, Lévy processes with hyper-exponential jumps are a sub-class.

(a) Processes in the β-family. A Lévy process X is said to belong to the β-family, if it has the Lévy density

$$f(x) = C_1 \frac{e^{-\alpha_1\beta_1 x}}{(1-e^{-\beta_1 x})^{\lambda_1}} \mathbb{1}_{\{x>0\}} + C_2 \frac{e^{\alpha_2\beta_2 x}}{(1-e^{\beta_2 x})^{\lambda_2}} \mathbb{1}_{\{x<0\}},$$

where $C_i, \alpha_i, \beta_i \in \mathbb{R}^+$ for $i \in \{1, 2\}$ and $\lambda_1, \lambda_2 \in (0, 3) \setminus \{1, 2\}$. This family of processes allows for an arbitrary behaviour of small jumps and includes processes similar to CGMY processes. The processes in the β-family are processes where the meromorphic characteristic exponent consists of beta and digamma functions [Kuznetsov, 2010].

(b) Hyper-geometric Lévy processes. A Lévy process X is said to be a hyper-geometric Lévy process, if it has the Lévy density

$$f(x) = \begin{cases} -\frac{\Gamma(\eta)}{\Gamma(\eta-\tilde{\gamma})\Gamma(-\gamma)} e^{-(1-\beta+\gamma)x} {}_2F_1(1+\gamma, \eta; \eta-\tilde{\gamma}; e^{-x}), & x > 0 \\ -\frac{\Gamma(\eta)}{\Gamma(\eta-\gamma)\Gamma(-\tilde{\gamma})} e^{(\tilde{\beta}+\tilde{\gamma})x} {}_2F_1(1+\tilde{\gamma}, \eta; \eta-\gamma; e^{-x}) & x < 0 \end{cases}$$

where $\eta = 1 - \beta + \gamma + \tilde{\beta} + \tilde{\gamma}$ and ${}_2F_1$ is the respective hyper-geometric function. In the class of Lévy processes with double-sided jumps the distribution of the exponential functional $\left(e^{X_t}\right)_{t\geq 0}$ is known in closed form only in the case of Lévy processes with hyper-exponential jumps [Cai and Kou, 2010].

3. Processes with bounded positive jumps. Lévy processes with bounded positive jumps have a Lévy measure ν that has support on $(-\infty, k]$, where k is the right boundary of the support of ν, that is $k = \inf_{x>0}\{\nu((x,\infty)) = 0\}$. This class is of interest as it is very large and with a certain understanding dense in the class of all Lévy processes, as any Lévy measure can be approximated with arbitrary precision by cutting it off at a large positive number. Therefore, they are of practical interest, as any general Lévy process can be approximated with a given precision. Secondly, in insurance applications a natural cap on the positive jumps is already a given, as when a reinsurance contract is used to hand off any claims above a certain level to an insurer.

   Meromorphic processes and processes with bounded positive jumps are very similar to each other. In both cases the positive Wiener-Hopf factor ϕ^+ is the product including the solutions to the equation $\psi(z) = q$ in the half-plane $\Re(z) > 0$. For meromorphic processes these solutions are real, whereas for processes with bounded positive jumps they are complex.

All these processes have in common that their characteristic exponent $\psi(x)$ has an analytical structure which makes it possible to represent it as the product of two functions, which are analytic in the left, respectively right, complex half plane, and in consequence to retrieve the Wiener-Hopf factors. In general, an explicit Wiener-Hopf factorisation can be obtained if and only if $\psi(x)$ can be extended to a meromorphic function in the left or right complex plane, which underlines the importance of the meromorphic Lévy processes introduced above. This is the reason why no explicit Wiener-Hopf factorisations have been discovered for the more widely used processes usually found in financial applications, like the Variance Gamma and the CGMY process. The respective characteristic exponents $\psi(x)$ of theses processes have at least one branch point and therefore lack a meromorphic extension [Kuznetsov and Peng, 2012].

6.2 The Wiener-Hopf Monte Carlo Algorithm

The subsequent treatment based on the exposition in [Ferreiro-Castilla et al., 2013] assumes that the following conditions are satisfied.

A1. $\int_{|x|\geq 1} x^2 \nu(dx) < \infty$
A2. The payoff function $P : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ is a Lipschitz function with a Lipschitz constant assumed to be 1.
A3. The computational time to sample $\underline{X}_{\varsigma(q)}$ and $\overline{X}_{\varsigma(q)}$ is independent of the value of q.

A natural way to use Monte Carlo simulation when calculating the expectation of an unknown quantity involving the joint law of $(X_t, \overline{X}_t)_{t\geq 0}$ is given by using stationarity and independence of the increments of Lévy processes to approximate the continuous time processes by discrete random walks, simulating multiple path instances and by keeping record of the maximum in each instance. This approach requires that the distribution of the Lévy process is known, either through the availability of an analytical formula or alternatively through the use of numerical inversion of the characteristic function. However, although one samples examples exactly from the distribution of the underlying Lévy process X, the law of the recorded maximum of the sampled random walk will not be the same as the law of $(\overline{X}_t)_{t\geq 0}$, indeed, the incurred error will be significant.

Another method to simulate a Lévy process is by using an appropriate series representation. In [Rosiński, 2001] a Lévy process X at time $t \in [0, T]$ is developed into an almost surely and uniformly converging series according to

$$X_t = \gamma t + \sum_{i=1}^{\infty} (H(\Gamma_i/T, V_i)\mathbb{1}_{U_i \leq t} - tc_i), \tag{6.5}$$

where $\{\Gamma_i\}_{i\in\mathbb{N}}$ denotes a sequence of partial sums of independent and identically distributed standard exponential random variables, $\{U_i\}_{i\in\mathbb{N}}$ a sequence of independent and identically distributed random variables uniformly

distributed on $[0,T]$, independent of $\{\Gamma_i\}_{i\in\mathbb{N}}$, $\{V_i\}_{i\in\mathbb{N}}$ a sequence of independent and identically distributed random variables with common probability distribution F and where the $\{c_i\}_{i\in\mathbb{N}}$ are given by $c_i = A(i) - A(i-1)$ with

$$A(s) = \int_0^s \int_{|x|\leq 1} x\mathbb{P}(H(r,V_i)\in dx)dr.$$

The function $H : (0,\infty)\times\mathbb{R}^d$ is defined by $\mathbb{P}(H(r,V_i)\in A) = f(r,A)$ with $\nu(A) = \int_0^\infty f(r,A)dr$ for $A\in B(\mathbb{R}^d)$. $H(\Gamma_i,V_i)$ can be interpreted as the effect at time zero of a shot V_i that happened Γ_i time units ago, when the respective series in (6.5) is interpreted as a shot noise process. However, more direct and much faster methods which are adapted to the specific law of the Lévy process under consideration exist. For example, [Janicki and Weron, 1994] gives an overview for α-stable Lévy processes.
A further approach commonly used is to approximate the Lévy process X with a jump-diffusion process. Jump-diffusion processes are Lévy processes that are the linear combination of a Brownian motion and an independent compound Poisson process. The method is based on the idea of cutting off the Lévy measure by discarding the small jumps below an arbitrarily chosen threshold. To compensate for the removal of the small jumps the linear component or the Brownian motion are adjusted in an appropriate fashion. If the jump component of the Lévy process is of finite variation, it is for example possible to replace the small jumps with an adjustment in the linear trend, whereas if the jump component is of unbounded variation, it is more appropriate to use an adjustment in the Brownian motion component [Dereich and Heidenreich, 2011]. In general, the removal of the small jumps guarantees that the process has the structure of a Brownian motion with linear drift combined with a compound process. Simulating the paths of the Brownian motion interspersed with jumps injected by the Poisson process is relatively straightforward. Limitations were noted by [Asmussen and Rosiński, 2001].
In [Kuznetsov et al., 2011] an accurate Monte Carlo simulation technique for expectations involving the joint law of $(X_t,\overline{X}_t)_{t\geq 0}$ was introduced. It allows for the exact sampling from the law of $(X_\tau,\overline{X}_\tau)$ with τ being a random time whose distribution can be centred arbitrarily close around $t\geq 0$, that is it does not work with a fixed, deterministic grid, but rather uses an underlying grid composed of random times. The accuracy of the Monte Carlo simulation heavily depends on a parameter controlling how much the the distribution of the random times is concentrated around a given $t\geq 0$ and thus how high the resolution of the resulting grid is. The intervals in between those random times are the arrival times of a compound Poisson process, that is independent and identically distributed exponential variables.
More formally, assume that $\{\mathbf{e}_i(1)\}_{i\in\mathbb{N}}$ are a sequence of independent and identically distributed random variables with mean 1. By the strong law of large numbers, it follows that

$$\sum_{i=1}^{n}\frac{t}{n}\mathbf{e}_i(1)\xrightarrow{a.s.} t. \tag{6.6}$$

By the property of exponential variables that their sum is again a scaled exponential variable, the random variable on the left hand side of (6.6) is an exponentially distributed random variable with mean t/n and by the relationship between Gamma distributed random variables and exponentially distributed random variables, it is in addition a Gamma random variable with parameters n and n/t, denoted by $\mathbf{g}(n,n/t)$. Due to the strong law of large numbers and independence of $\mathbf{g}(n,n/t)$ from X,

$$(\overline{X}_{\mathbf{g}(n,n/t)}, X_{\mathbf{g}(n,n/t)})\xrightarrow{a.s.}(\overline{X}_t, X_t).$$

The main theorem in [Kuznetsov et al., 2011] establishes a procedure on sampling from $(\overline{X}_{\mathbf{g}(n,n/t)}, X_{\mathbf{g}(n,n/t)})$.

Theorem 25. *Let $\{S^j_{n/t}\}_{j\in\mathbb{N}}$ and $\{I^j_{n/t}\}_{j\in\mathbb{N}}$ be sequences of independent and identically distributed random variables with common distribution equal to that of $\overline{X}_{\mathbf{e}(n/t)}$ and $\underline{X}_{\mathbf{e}(n/t)}$, respectively. Then for all $n\in\mathbb{N}$*

$$(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)})\overset{D}{=}(V(n,n/t), J(n,n/t))$$

with $V(0,n/t) = 0$ and where for any $k\in\mathbb{N}$

$$V(k,n/t) = V(k-1,n/t) + (S^k_{n/t} + I^k_{n/t}), \tag{6.7}$$

$$J(k,n/t) = \max\{J(k-1,n/t), V(k-1,n/t) + S^k_{n/t}\}. \tag{6.8}$$

Proof. Let $n \geq 1$ be fixed and define $\overline{X}_{s,t} = \sup_{s \leq u \leq t} X_u$. Telescoping the sum in the first component and splitting up the supremum in the second results in

$$(X_{\mathbf{g}(k,n/t)}, \overline{X}_{\mathbf{g}(k,n/t)}) = (X_{\mathbf{g}(k-1,n/t)} + (X_{\mathbf{g}(k,n/t)} - X_{\mathbf{g}(k-1,n/t)}), \max(\overline{X}_{\mathbf{g}(k-1,n/t)}, \overline{X}_{\mathbf{g}(k-1,n/t),\mathbf{g}(k,n/t)})),$$

where $\mathbf{g}(0,\lambda) = 0$. Furthermore define $X_t^{(k)} = X_{\mathbf{g}(k-1,n/t)+t} - X_{\mathbf{g}(k-1,n/t)}$ and $\overline{X}^{(k)}_{\mathbf{e}_k t/n} = \sup_{s \leq \mathbf{e}_k t/n} X_s^{(k)}$. Using this notation it then follows that

$$(X_{\mathbf{g}(k,n/t)}, \overline{X}_{\mathbf{g}(k,n/t)}) = (X_{\mathbf{g}(k-1,n/t)} + \overline{X}^{(k)}_{\mathbf{e}_k t/n}, \max(\overline{X}_{\mathbf{g}(k-1,n/t)}, X_{\mathbf{g}(k-1,n/t)} + \overline{X}^{(k)}_{\mathbf{e}_k t/n})).$$

Due to the construction of $X^{(k)}$, it is independent of $\{X_s : s \leq \mathbf{g}(k-1,n/t)\}$. Additionally, from 4. in Theorem 26 the Wiener-Hopf factorisation it follows that $\overline{X}_{\mathbf{e}_1 t/n}$ and $X_{\mathbf{e}_1 t/n} - \overline{X}_{\mathbf{e}_1 t/n}$ are independent and that $X_{\mathbf{e}_1 t/n} - \overline{X}_{\mathbf{e}_1 t/n}$ is equal in distribution to $\underline{X}_{\mathbf{e}_1 t/n}$. Using the factorisation finally results in

$$(X_{\mathbf{g}(k,n/t)}, \overline{X}_{\mathbf{g}(k,n/t)}) \stackrel{D}{=} (X_{\mathbf{g}(k-1,n/t)} + S^k_{n/t} + I^k_{n/t}, \max(\overline{X}_{\mathbf{g}(k-1,n/t)}, X_{\mathbf{g}(k-1,n/t)} + S^k_{n/t})).$$

□

Given that sampling from the distributions of $S_{n/t}$ and $I_{n/t}$ is feasible, the theorem thus affords us the ability to simulate $(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)})$ exactly. Due to the Strong Law of Large Numbers, the conventional Monte Carlo estimator

$$\hat{Y}^n_N = \frac{1}{N} \sum_{i=1}^{N} P^n(\omega_i)$$

where $P^n = P(V(n,n/t), J(n,n/t))$, can be used to estimate the expected value $\mathbb{E}(P(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)}))$. Indeed, successively taking limits leads to almost sure convergence to $\mathbb{E}(P(X_t, \overline{X}_t))$,

$$\lim_{n\to\infty} \lim_{N\to\infty} \hat{Y}^n_N = \lim_{n\to\infty} \mathbb{E}(P(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)})) = \mathbb{E}(P(X_t, \overline{X}_t)).$$

The feasibility of the Monte Carlo simulation rests on the availability of samples from the distributions of $S_{n/t}$ and $I_{n/t}$ in Theorem (26). The discovery of Lévy processes for which these distributions are known is quite a recent one. Indeed, processes for which samples from $S_{n/t}$ and $I_{n/t}$ can be readily drawn are the rich class of meromorphic processes. The aforementioned closeness of a sub-class of meromorphic processes, namely the β-processes, to widely used Lévy processes in financial mathematics like the CGMY process, variance gamma processes and processes give the class practical importance.

6.3 The Wiener-Hopf Multilevel Monte Carlo Algorithm

As shown in section 5.6.7, the introduction of multiple levels of different fineness into a Monte Carlo simulation vastly improves asymptotic convergence rates which even tend to be optimal. As in the case of (4.6) a decomposition of the mean square error between the estimator $\hat{Y}^n_N$ and $\mathbb{E}(P(X_t, \overline{X}_t))$, $\mathbb{E}((\hat{Y}^n_N - \mathbb{E}(P(X_t, \overline{X}_t))^2)$, is due to $\mathbb{E}(\hat{Y}^n_N) = E(P^n)$ and $\mathbb{V}ar(\hat{Y}^n_N) = \frac{1}{N}\mathbb{V}ar(P^n)$ given by

$$\mathrm{MSE}(\hat{Y}^n_N) = \frac{1}{N}\mathbb{V}ar(P^n) + [\mathbb{E}(P^n - P(X_t, \overline{X}_t))]^2. \tag{6.9}$$

In contrast to (4.6), the decomposition breaks down the mean square error into the the variance of the Monte Carlo simulation and the bias introduced by the randomised time horizon. The decomposition of the expectation $E(P_l)$ into a telescoping sum and the subsequent construction of the Multilevel Monte Carlo estimator $\hat{Y}_l$ are the same as given by equations (5.1) and (5.2).
As noted in the introductory section on the Multilevel Monte Carlo Method it is vital for the construction of the Multilevel Monte Carlo estimator that the samples for P_l and P_{l-1} when simulation $P_l - P_{l-1}$ should come from the same draw, necessitating a mechanism to calculate P_{l-1} from P_l. Different mechanism were presented in section 5.6 on derivative pricing. When using Lévy processes on a random grid, the question of constructing such an adequate mechanism has to be revisited.

The technique used in [Ferreiro-Castilla et al., 2013] is Poisson thinning. When generating a sample on level $l \in \{1, \dots, L\}$ of $P_l - P_{l-1}$ a Poisson random grid with parameter $\lambda_l = n_l/t$ is constructed, that is, the time distances between arrival times in the grid are independent exponentially distributed random variables with mean $1/\lambda_l$. Let $(\overline{N}^l_s)_{s\geq 0}$ denote a Poisson process with arrival rate λ_l and let the respective arrival times of the process $\overline{N}^l$ be denoted by $\{T_k^{\overline{N}^l}\}_{k\geq 0}$ with $T_0^{\overline{N}^l} = 0$. The central property used in Poisson thinning is the thinning property of Poisson processes [Tankov and Cont, 2004, p. 65], which guarantees that discarding arrivals with probability p creates out of a Poisson process with parameter λ a Poisson process with parameter λp. Applied to the situation at hand, discarding the arrivals of a Poisson process $\overline{N}^l$, which has intensity $\lambda_l = n_l/t$, with probability $1/M$, where $n_l/n_{l-1} = M$, yields a Poisson process that has intensity $\lambda_{l-1} = n_{l-1}/t$ and thus follows the same law as $\overline{N}^{l-1}$. Again, the waiting times on this coarser Poisson grid are independent and exponentially distributed with mean $1/\lambda_l$.
In [Ferreiro-Castilla et al., 2013] only the case $M = n_l/n_{l-1} = 2$ is considered. Using the previously introduced notation, the identity $\mathbf{g}(n_l, n_l/t) = T_{n_l}^{\overline{N}^l}$ holds. Following Theorem 26 the sequences of independent random variables $\{S^j_{n_l/t}\}_{j\geq 1}$ and $\{I^j_{n_l/t}\}_{j\geq 1}$, which are identically distributed with the respective distributions of $\overline{X}_{\mathbf{e}(n_l/t)}$ and $\underline{X}_{\mathbf{e}(n_l/t)}$ with the exponential waiting times generated by the Poisson process $\overline{N}^l$, are considered. For $i \in \mathbb{N}$ let κ_i be the index of the i-th arrival in the process $\overline{N}^l$ that has not been discarded after using the Poisson thinning technique and set $\kappa_0 = 0$. In particular, it should be noted that the sequence $\{\kappa_i - \kappa_{i-1}\}_{i\in\mathbb{N}}$ is a sequence of independent and identically, geometrically distributed random variables with parameter $1/2$ and that for a $i \in \mathbb{N}$ the waiting time in the thinned process $\overline{N}^l$ in between the $(i-1)$-th arrival and the i-th arrival is equal in distribution to $\sum_{j=\kappa_{i-1}+1}^{\kappa_i} \mathbf{e}_j(n_l/t)$.
The technique of Poisson thinning can be consequently used to build the independent sequences $\{S^i_{n_{l-1}/t}\}_{i\in\mathbb{N}}$ and $\{I^i_{n_{l-1}/t}\}_{i\in\mathbb{N}}\}$, which are subsequently used to sample from P_{l-1} according to Theorem 26, from the sequences $\{S^i_{n_l/t}\}_{i\in\mathbb{N}}$ and $\{I^i_{n_l/t}\}_{i\in\mathbb{N}}\}$ by way of a deterministic transformation. The arguments above lead to the construction scheme

$$S^i_{n_{l-1}/t} = \bigvee_{k=1}^{\kappa_i-\kappa_{i-1}} \left\{ \sum_{j=1}^{k-1} \left(S^{\kappa_{i-1}+j}_{n_l/t} + I^{\kappa_{i-1}+j}_{n_l/t} \right) + S^{\kappa_{i-1}+k}_{n_l/t} \right\},$$

$$I^i_{n_{l-1}/t} = \sum_{j=\kappa_{i-1}+1}^{\kappa_i} (S^j_{n_l/t} + I^j_{n_l/t}) - S^i_{n_{l-1}/t}.$$

The construction scheme offers a straightforward way to construct the infinite sequence of pairs $\{(S^i_{n_{l-1}/t}, I^i_{n_{l-1}/t})\}_{i\in\mathbb{N}}$ from $\{(S^i_{n_l/t}, I^i_{n_l/t})\}_{i\in\mathbb{N}}$, however, the practical situation of finite sequences still has to be considered. The construction of the Poisson random grid is cut off at the time $\mathbf{g}(n_l, n_l/t)$. To guarantee that the number of arrivals that were not discarded from $\overline{N}^l$ to construct $\overline{N}^{l-1}$ sums up to n_{l-1} it is necessary to prophylactically extend the number of arrivals in $\overline{N}^l$ beyond $\mathbf{g}(n_l, n_l/t)$ before applying the Poisson thinning technique.
Another point to consider arises from that the algorithm is based on a grid that is entirely random. Let $\hat{P}_{l-1}$ be a sample produced from P_l by the Poisson thinning technique. As in the conventional Multilevel Monte Carlo algorithm it is necessary to ensure that the expectation of P_{l-1} is the same as the expectation of $\hat{P}_{l-1}$. That this is indeed the case, is ensured by the construction of the Poisson thinning technique, as thinning the Poisson process $\overline{N}^l$ yields in distribution a Poisson process with parameter n_{l-1}/t, which is therefore equal in distribution to $\overline{N}^{l-1}$. Consequently, the random time $\mathbf{g}(n_{l-1}, n_{l-1}/t)$ is of the same distribution as the thinned version $\hat{g}(n_{l-1}, n_{l-1}/t)$ produced from $\mathbf{g}(n_l, n_l/t)$. Considering this in connection with the fact that $\mathbf{g}(n_{l-1}, n_{l-1}/t)$ and $\hat{g}(n_{l-1}, n_{l-1}/t)$ are independent of the Lévy process on top of which they are constructed gives that $\hat{P}_{l-1}$ and P_{l-1} are indeed of the same law.

6.4 Numerical Analysis of the Wiener-Hopf Multilevel Monte Carlo Algorithm

In course of the numerical analysis of the Wiener-Hopf Multilevel Monte Carlo Algorithm in [Ferreiro-Castilla et al., 2013] the following notation is employed. The short hand $a \lesssim b$ is used to signify that for positive a, b the quantity a/b is uniformly bounded independent of any parameters. Accordingly, $a \eqsim b$ means that both $a \lesssim b$ and $b \lesssim a$ hold.
The basic methodology adopted to measure the efficiency of an algorithm in [Ferreiro-Castilla et al., 2013] is different from the methodology used in [Giles, 2008b] and [Giles et al., 2013], where an accuracy ϵ is fixed in advance and the cost needed to reach an error level below that accuracy is estimated subsequently. Instead, the convergence error of the analysed Wiener-Hopf Multilevel Monte Carlo algorithm is taken to be a function of the computational cost involved, measured in floating point operations. That is, the expectation of the computational complexity $C(\hat{P})$ is considered for a given estimator $\hat{P}$ of the expected value $\mathbb{E}(P(X_t, \overline{X}_t))$ and it is assumed that it is uniformly bounded by some value ν, $\mathbb{E}(C(\hat{P})) \lesssim \nu$. Given that bound on the estimated computational cost the goal consists in consequently bounding the root mean square error $\mathrm{RMSE}(\hat{P})$ by a bound dependent on ν.
A basic convergence result is the convergence of the Wiener-Hopf Monte Carlo method on a single level.
Theorem 26. *Let $t \geq 0$ be a given time instance and assume that Y converges in mean square to $P(X_t, \overline{X}_t)$. Furthermore suppose that there exist positive constants $\alpha, \gamma > 0$ such that*

$$a) \quad |\mathbb{E}(P^n - P(X_t, \overline{X}_t))| \lesssim \frac{1}{n^\alpha} \qquad \text{and} \qquad b) \quad \mathbb{E}(C^n) \lesssim n^\gamma,$$

where C^n is the cost of computing a single sample of P^n on a Poisson grid with rate n/t. Then for every $\nu \in \mathbb{N}$ there exist $n, N \in \mathbb{N}$ such that

$$\mathbb{E}(C(\hat{Y}_N^n)) \lesssim \nu \qquad \text{and} \qquad RMSE(\hat{Y}_N^n) \lesssim \frac{1}{\nu^{\frac{1}{2+\gamma/\alpha}}}.$$

Proof. The first step consists in showing that $\mathbb{V}ar(P^n)$ is bounded. Due to $\mathbb{V}ar(P^n) \leq \mathbb{E}\left((P^n)^2\right)$ the inequality

$$\frac{1}{2}\mathbb{V}ar(P^n) \leq \mathbb{E}((P^n - P(X_t, \overline{X}_t))^2) + \mathbb{E}((P(X_t, \overline{X}_t)^2)$$

holds. Since P^n converges in mean square to $P(X_t, \overline{X}_t)$ and P is assumed to be Lipschitz continuous, the first summand is bounded above independently of n. The second summand is bound again due to the Lipschitz continuity of P and assumption A1 that guarantees that X_t possesses second finite moments for all $t \geq 0$.
Under the assumption that $\mathbb{V}ar(P^n)$ is bounded by a constant independent of n, it becomes clear that when the two summands on the right hand side in (6.9) should be balanced, then assumption a) yields

$$\frac{1}{N}\mathbb{V}ar(P^n) \eqsim (\mathbb{E}(P^n - P(X_t, \overline{X}_t))^2 \eqsim n^{-2\alpha}$$

and consequently $N \eqsim n^{2\alpha}$. Due to $C(\hat{Y}_N^n) = NC_n$ and assumption b) the expected value of the overall computational cost of the estimator is going to be have the upper bound ν, if $n \eqsim \nu^{\frac{1}{2\alpha+\gamma}}$ because of

$$C(\hat{Y}_N^n) = NC_n \eqsim n^{2\alpha+\gamma} \eqsim \nu.$$

Consequently, $N \eqsim \nu^{\frac{2\alpha}{2\alpha+\gamma}}$, yielding the desired bound on $\mathrm{RMSE}(\hat{Y}_N^n)$. □

In an analogue fashion to the Multilevel Monte Carlo Theorem, Theorem 31, a multilevel variant of the Wiener-Hopf Monte Carlo method can be formulated. A proof of the theorem can be found in [Cliffe et al., 2011].
Theorem 27. *Let $t \geq 0$ be a given time instance and $n_l = n_0 M^l$ for some level $l \in \mathbb{N}$ and assume that there are positive constants $\alpha, \beta, \gamma > 0$ with $\alpha \geq \frac{1}{2}\max(\beta, \gamma)$ such that*

$$a) \quad |E(P_l - P(X_t, \overline{X}_t))| \lesssim \frac{1}{n_l^\alpha}, \qquad b) \quad \mathbb{V}ar(P_l - P_{l-1}) \lesssim \frac{1}{n_l^\beta}, \qquad c) \quad \mathbb{E}(C_l) \lesssim n_l^\gamma,$$

where C_l denotes the computational cost of computing a single sample of P_l on a Poisson grid with rate n_l/t. Then for every $\nu \in \mathbb{N}$ there exists a value L and a sequence of Multilevel Monte Carlo estimators $\hat{Y}_l$, $l \in \{1, \dots, L\}$, such that for the Multilevel Monte Carlo estimator $\hat{Y}$ the inequalities $\mathbb{E}(C(\hat{Y})) \lesssim \nu$ and

$$RMSE(\hat{Y}) \lesssim \begin{cases} \nu^{-\frac{1}{2}} & \text{if} \quad \beta > \gamma, \\ \nu^{-\frac{1}{2}} \log^2 \nu & \text{if} \quad \beta = \gamma, \\ \nu^{-\frac{1}{2+(\gamma-\beta)/\alpha)}} & \text{if} \quad \beta < \gamma \end{cases}$$

hold.
For Theorem 26 and Theorem 27 to be useful, constants α, β and γ have to be found such that the respective assumptions in the theorems are indeed satisfied. Since the computational time to sample $\overline{X}_{\mathbf{e}(q)}$ and $\underline{X}_{\mathbf{e}(q)}$ is independent of q because of assumption A3, $\gamma = 1$ always holds, whereas the values of α and β are dependent on the convergence rate of $\mathbf{g}(n, n/t)$ to t.
To determine suitable constants α and β, let τ and τ' be non-negative random variables that are independent of the Lévy process X. They nonetheless could be correlated. Additionally, let $\mathbf{X}_t = (X_t, \overline{X}_t)$. From the Lipschitz continuity assumption on P the inequalities

$$|\mathbb{E}(P(\mathbf{X}_\tau) - P(\mathbf{X}_t))| \leq \mathbb{E}(|\mathbf{X}_\tau - \mathbf{X}_t|) \leq \mathbb{E}(|X_\tau - X_t|) + E(|\overline{X}_\tau - \overline{X}_t|) \tag{6.10}$$

and

$$\mathbb{V}ar(P(\mathbf{X}_\tau - P(\mathbf{X}_{\tau'})) \leq \mathbb{E}((\mathbf{X}_\tau - \mathbf{X}_{\tau'})^2) \leq 2(\mathbb{E}((X_\tau - X_{\tau'})^2) + \mathbb{E}((\overline{X}_\tau - \overline{X}_{\tau'})^2)) \tag{6.11}$$

follow. The following Lemma helps in establishing estimates on the summands on the right hand sides of the respective inequalities.
Lemma 13. *Let X be a Lévy process that satisfies assumptions A1 to A3 and let τ be an arbitrary non-negative random variable not dependent on X. Then for a given time instance $t \geq 0$ the identities*

$$\begin{aligned} \mathbb{E}((X_\tau - X_t)^2) &= \mathbb{V}ar(X_1)\mathbb{E}(|\tau - t|) + \mathbb{E}(X_1)^2\mathbb{E}((\tau - t)^2), \\ \mathbb{E}((\overline{X}_\tau - \overline{X}_t)^2) &\leq 16\mathbb{V}ar(X_1)\mathbb{E}(|\tau - t|) + 2(\max(\mathbb{E}(X_1), 0))^2\mathbb{E}((\tau - t)^2) \end{aligned}$$

hold.

Proof. Since the increments of a Lévy process are independent and identically distributed, they can be written in the form

$$X_\tau - X_t \overset{D}{=} \begin{cases} X_{\tau - t}, & \text{if} \quad \tau \geq t, \\ -X_{t-\tau} & \text{if} \quad \tau < t. \end{cases}$$

Because τ is assumed to not depend on X and because of

$$\mathbb{E}(X_s^2) = \mathbb{V}ar(X_1)s + \mathbb{E}(X_1)s^2$$

for all $s \geq 0$, the identity

$$\mathbb{E}((X_\tau - X_t)^2) = \mathbb{V}ar(X_1)\mathbb{E}(|\tau - t|) + \mathbb{E}(X_1)^2\mathbb{E}((\tau - t)^2)$$

follows.
To show the second result, the identity

$$\overline{X}_t \overset{D}{=} \max(\overline{X}_s, X_s + \overline{X}'_{t-s})$$

with $s < t$ and $\overline{X}'_{t-s}$ independent of $\{X_{s'}\}_{s' \leq s}$ and identically distributed to $\overline{X}_{t-s}$. Considering the duality property of Lévy processes, which states that $\underline{X}_s$ is equal in distribution to $X_s - \overline{X}_s$, the result

$$\overline{X}_t - \overline{X}_s \overset{D}{=} \max(\underline{X}''_s + \overline{X}'_{t-s}, 0) \leq \overline{X}'_{t-s}$$

follows, with $\underline{X}''_s$ independent of $\overline{X}'_{t-s}$ and equal in distribution to $\underline{X}_s$. This identity yields the inequality

$$\mathbb{E}((\overline{X}_\tau - \overline{X}_t)^2) \leq \mathbb{E}(\overline{X}^2_{|\tau - t|}). \tag{6.12}$$

Assumption A1 affords us the ability to decompose X into a martingale part $X^* = \{X_t^*\}_{t\geq 0}$ and a drift such that it has the form $X_t = X_t^* + t\mathbb{E}(X_1)$ for $t \geq 0$. It then follows that

$$\begin{aligned}\mathbb{E}(\overline{X}_t^2) &= \mathbb{E}(\sup_{s\leq t}(X_s^* + \mathbb{E}(X_1)s)^2) \\ &\leq \mathbb{E}((\overline{X}_t^* + t(\max(\mathbb{E}(X_1), 0))^2) \\ &\leq 2\left(\mathbb{E}((\overline{X}_t^*)^2) + (\max(\mathbb{E}(X_1), 0)^2 t^2\right), \end{aligned} \tag{6.13}$$

with $\overline{X}_t^* = \sup_{s\leq t} X_t^*$. Doob's martingale inequality [Sato, 1999, p. 167] then leads to

$$\mathbb{E}(\overline{X}_t^2) \leq 2\left(8\mathbb{E}((X_t^*)^2) + (\max(\mathbb{E}(X_1), 0)^2 t^2\right). \tag{6.14}$$

□

Plugging in $|\tau - t|$ results in

$$\mathbb{E}(\overline{X}_{|\tau-t|}^2) \leq 16\mathbb{E}((X_{|\tau-t|}^*)^2) + 2(\max(\mathbb{E}(X_1), 0)))^2\mathbb{E}((\tau - t)^2).$$

But because of $\mathbb{E}((X_{|\tau-t|}^*)^2) = \mathbb{E}((X_\tau^* - X_t^*)^2)$ equation (6.12) can be used to get the result

$$\begin{aligned}\mathbb{E}(\overline{X}_{|\tau-t|}^2) &\leq 16\mathbb{E}((X_\tau^* - X_t^*)^2) + 2(\max(\mathbb{E}(X_1), 0))^2\mathbb{E}((\tau - t)^2) \\ &= 16\mathbb{V}ar(X_1^*)\mathbb{E}(|\tau - t|) + 2(\max(\mathbb{E}(X_1), 0))^2\mathbb{E}((\tau - t)^2).\end{aligned}$$

Combining this inequality with $\mathbb{V}ar(X_1^*) = \mathbb{V}ar(X_1)$ leads to the desired result.

Lemma 14. *Let $t \geq 0$ be a given time instance and X a Lévy process satisfying assumptions A1 to A3. Then for any $n \in \mathbb{N}$ the bounds*

$$\mathbb{E}(((X_{\boldsymbol{g}(n,n/t)} - X_t)^2) \underset{\sim}{=} n^{-\frac{1}{2}} \quad \textit{and} \quad \mathbb{E}(|X_{\boldsymbol{g}(n,n/t)} - X_t|) \underset{\sim}{\leq} n^{-\frac{1}{4}}, \tag{6.15}$$

$$\mathbb{E}(((\overline{X}_{\boldsymbol{g}(n,n/t)} - \overline{X}_t)^2) \underset{\sim}{=} n^{-\frac{1}{2}} \quad \textit{and} \quad \mathbb{E}(|\overline{X}_{\boldsymbol{g}(n,n/t)} - \overline{X}_t|) \underset{\sim}{\leq} n^{-\frac{1}{4}} \tag{6.16}$$

hold. If additionally X has paths of bounded variation then the improved bounds

$$\mathbb{E}(|X_{\boldsymbol{g}(n,n/t)} - X_t|) \lesssim n^{-\frac{1}{2}} \quad \textit{and} \quad \mathbb{E}(|\overline{X}_{\boldsymbol{g}(n,n/t)} - \overline{X}_t|) \lesssim n^{-\frac{1}{2}} \tag{6.17}$$

hold.

Proof. To show the bounds (6.15) and (6.16) it is sufficient to show the respective bounds for the second moments, as the bounds for the first moments then follow from an application of Jensen's inequality. With the goal of using Lemma 13, it will be shown that for all $n \in \mathbb{N}$ the identities

$$\mathbb{E}((\mathbf{g}(n, n/t) - t)^2) = \frac{t^2}{n} \tag{6.18}$$

$$\mathbb{E}(|\mathbf{g}(n, n/t) - t|) = 2te^{-n}\frac{n^n}{n!} \underset{\sim}{=} n^{-\frac{1}{2}}. \tag{6.19}$$

The last bound is a direct consequence of Stirling's formula $n! \approx \sqrt{2n\pi}n^n e^{-n}$. Because $\mathbb{E}(\mathbf{g}(n, n/t)) = t$ and $\mathbf{g}(n, n/t)$ has a Gamma distribution, equation (6.18) is just an application of the formula for the variance of the Gamma distribution. Equation (6.19) can be written as

$$\begin{aligned}\mathbb{E}(|\mathbf{g}(n, n/t) - t|) &= \int_0^t (t - s)\mathbf{g}(n, n/t, s)ds + \int_t^\infty (s - t)\mathbf{g}(n, n/t, s)ds \\ &= 2\int_0^t (t - s)\mathbf{g}(n, n/t, s)ds, \end{aligned} \tag{6.20}$$

where $\mathbf{g}(k,\theta,s) = \Gamma(k)^{-1}\theta^k s^{k-1}e^{-\theta s}$ is the density of a Gamma distributed random variable. Let $\gamma(k,u) = \int_0^u x^{k-1}e^{-x}dx$ denote the incomplete gamma function for which the identities

$$\int_0^t \mathbf{g}(k,\theta,s)ds = \frac{\gamma(k,t\theta)}{\Gamma(k)} \quad \text{and} \quad \int_0^t s\mathbf{g}(k,\theta,s)ds = \frac{\gamma(k+1,t\theta)}{\Gamma(k)\theta}$$

hold. With the incomplete gamma function equation (6.20) can be written as

$$\int_0^t (t-s)\mathbf{g}(n,n/t,s)ds = t\int_0^t \mathbf{g}(n,n/t,s)ds - \int_0^t s\mathbf{g}(n,n/t,s)ds = t\left(\frac{\gamma(n,n)}{\Gamma(n)} - \frac{\gamma(n+1,n)}{\Gamma(n+1)}\right). \tag{6.21}$$

The incomplete gamma function also allows for the representation

$$\frac{\gamma(k,u)}{\Gamma(k)} = \frac{1}{\Gamma(k)}\int_0^u x^{k-1}e^{-x}dx = 1 - \left(1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots + \frac{u^{k-1}}{(k-1)!}\right)e^{-u},$$

which can be found in [Abramowitz and Stegun, 1970, sec. 6.5.13]. Plugging in this representation into the right hand side of equation (6.21) yields

$$\int_0^t (t-s)\mathbf{g}(n,n/t,s)ds = t\frac{n^n}{n!}e^{-n}.$$

Substituting the right hand side into equation (6.20) yields the desired result

$$\mathbb{E}(|\mathbf{g}(n,n/t) - t|) = 2te^{-n}\frac{n^n}{n!}.$$

In case the Lévy process X is of bounded variation, the Doob's martingale inequality for absolute moments and the same decomposition into martingale and drift as used in equations (6.13) and (6.14) affords us the ability to derive that

$$\begin{aligned}\mathbb{E}(|\overline{X}_{\mathbf{g}(n,n/t)} - \overline{X}_t|) &\leq \mathbb{E}(\overline{X}^*_{|\mathbf{g}(n,n/t)-t|}) + \max(\mathbb{E}(X_1),0)\mathbb{E}(|\mathbf{g}(n,n/t) - t|) \\ &\leq 8\mathbb{E}(|X^*_{|\mathbf{g}(n,n/t)-t|}|) + \max(E(X_1),0)\mathbb{E}(|\mathbf{g}(n,n/t) - t|).\end{aligned}$$

Additionally, $\mathbb{E}(|X_{\mathbf{g}(n,n/t)} - X_t|) = \mathbb{E}(|X_{|\mathbf{g}(n,n/t)-t|}|)$ is thus also bounded by a weighted sum consisting of $\mathbb{E}(|X^*_{|\mathbf{g}(n,n/t)-t|}|)$ and $\mathbb{E}(|\mathbf{g}(n,n/t) - t|)$. The missing link to a complete proof is the bound in (6.19) on the expectation of the martingale part, $\mathbb{E}(|X^*_{|\mathbf{g}(n,n/t)-t|}|) \lesssim n^{-\frac{1}{2}}$. Given this bound, the proof is finished.
Due to Lemma 2, the Lévy process X^*, which is of bounded variation, is the difference of two independent subordinators, $X^*_t = X'_t - X''_t$. Again assumption A1 guarantees that the subordinators X' and X'' have finite second moments. Using (6.19) yields

$$\mathbb{E}(|X^*_{|\mathbf{g}(n,n/t)-t|}|) \leq E(X'_{|\mathbf{g}(n,n/t)-t|}) + \mathbb{E}(X''_{|\mathbf{g}(n,n/t)-t|}) = \mathbb{E}(|\mathbf{g}(n,n/t) - t|)\mathbb{E}(|X'_1| + |X''_1|) \eqsim n^{-\frac{1}{2}}.$$

□

The main theorem in [Ferreiro-Castilla et al., 2013], which gives the particular convergence rates needed to confirm the validity of the assumptions in Theorems 26 and 27, directly follows by combining the previous lemmas and theorems.

Theorem 28. *If the assumptions A1 to A3 are satisfied, the assumptions in Theorems 26 and 27 hold with $\alpha = 1/4, \beta = 1/2$ and $\gamma = 1$. If the Lévy process X has paths of bounded variation, $\alpha = 1/2$ holds. The root mean square errors for the single Wiener-Hopf Monte Carlo method are under the restriction that the expectation of the overall computational complexity is $O(\nu)$ for any $\nu \in \mathbb{N}$ given by*

$$RMSE(\hat{Y}^n_N) \lesssim \begin{cases}\nu^{-\frac{1}{6}}, & \text{if } X \text{ is not of bounded variation,} \\ \nu^{-\frac{1}{4}}, & \text{if } X \text{ is of bounded variation.}\end{cases}$$

For the Wiener-Hopf Multilevel Monte Carlo method the root mean square errors are given by

$$RMSE(\hat{Y}) \lesssim \begin{cases}\nu^{-\frac{1}{4}}, & \text{if } X \text{ is not of bounded variation,} \\ \nu^{-\frac{1}{3}}, & \text{if } X \text{ is of bounded variation.}\end{cases}$$

Proof. Since $\gamma = 1$ is ensured by assumption A3, this result follows immediately from Theorems 26 and 27 as well as Lemma 14 and equations (6.10) and (6.11). □

6.5 Numerical Analysis of the Random Grid

In [Ferreiro-Castilla et al., 2013] a suggestion for improvement on the random grid presented above is discussed. Unfortunately, the improvement suffers from the defect that it turns out to be of primarily theoretical value, as it requires the ability to simultaneously sample from three distributions in connection with the Wiener-Hopf factorisation. Sub-classes of Lévy processes which are practically important and for which this is possible are not known.
The basic insight of the improvement consists in the recognition that there is indeed nothing particular about the random times $\mathbf{g}(n, n/t)$ used in Theorem 26. Since by definition $\mathbf{g}(n, n/t) = \sum_{i=1}^{n} \frac{t}{n}\mathbf{e}_i(1)$, the random time $\mathbf{g}(n, n/t)$ is just the sum of n independent random variables exponentially distributed with mean t/n, converging to t as $n \to \infty$. An alternative, more efficient way to construct a sum of independent exponentially distributed random variables that converges to t is presented in the following.
Let $\overline{N} = \{\overline{N}_s\}_{s\geq 0}$ be the Poisson process with parameter t/n that is used as the basis for sampling the random walk $V(k, n/t)$, $k \in \mathbb{N}$. Furthermore, define

$$T_j^{\overline{N}} = \inf_{s>0}\{\overline{N}_s = j\} \quad \text{and} \quad T(n,t) = T^{\overline{N}}_{\overline{N}_t+1}.$$

From the lack of memory property of the exponential distribution it directly follows that the exponentially distributed waiting times of $\overline{N}$ imply that $T(n,t) - t$ is again exponentially distributed with mean t/n. Additionally,

$$\lim_{n\to\infty} T(n,t) \overset{a.s.}{=} t \quad \text{and} \quad \lim_{n\to\infty} (X_{T(n,t)}, \overline{X}_{T(n,t)}) \overset{a.s.}{=} (X_t, \overline{X}_t).$$

Theorem 26 indicates that it may be useful to approximate $(X_{T(n,t)}, \overline{X}_{T(n,t)})$ by $(V(\overline{N}_t + 1, n/t), J(\overline{N}_t + 1, n/t))$. The Wiener-Hopf factorisation, however, only offers knowledge about $(\overline{X}_{\mathbf{e}(q)}, X_{\mathbf{e}(q)} - \overline{X}_{\mathbf{e}(q)})$, but not $(\overline{X}_{\mathbf{e}(q)}, X_{\mathbf{e}(q)} - \overline{X}_{\mathbf{e}(q)}, \mathbf{e}(q))$, which is a salient point exploited in Theorem 26 and which makes it impossible to sample exactly and simultaneously from $(V(\overline{N}_t + 1, n/t), J(\overline{N}_t + 1, n/t), \overline{N}_t + 1)$.
To consider the theoretical improvement of the time $T(n,t)$ over $\mathbf{g}(n, n/t)$ let $P^n(\omega_i)$ to be understood as the $i-th$ realisation of the random variable

$$P^n = P(V(\overline{N}_t + 1, n/t), J(\overline{N}_t + 1, n/t)).$$

The Poisson thinning technique results again in$\{S^i_{n_{l-1}/t}\}_{i\in\mathbb{N}}$ and $\{I^i_{n_{l-1}/t}\}_{i\in\mathbb{N}}\}$ constructed through

$$S^i_{n_{l-1}/t} = \bigvee_{k=1}^{\kappa_i-\kappa_{i-1}} \left\{ \sum_{j=1}^{k-1} \left(S^{\kappa_{i-1}+j}_{n_l/t} + I^{\kappa_{i-1}+j}_{n_l/t} \right) + S^{\kappa_{i-1}+k}_{n_l/t} \right\},$$

$$I^i_{n_{l-1}/t} = \sum_{j=\kappa_{i-1}+1}^{\kappa_i} (S^j_{n_l/t} + I^j_{n_l/t}) - S^i_{n_{l-1}/t}.$$

However, it still has to be shown that the Poisson thinning procedure works on the random $T(n,t)$. Let $l \in \{1, \ldots, L\}$ and let $\tilde{T}(n_{l-1}, t)$ be the random time constructed from $\overline{N}^l$ by Poisson thinning. What remains to be seen is that $\tilde{T}(n_{l-1}, t)$ is in distribution equal to $T(n_{l-1}, t)$. Since Poisson thinning removes arrivals from the Poisson process $\overline{N}^l$, $T(n_l, t) \leq \tilde{T}(n_{l-1}, t)$ holds. Because $T(n_l, t) - t$ is equal in distribution to $\mathbf{e}(n_l/t)$, the relations

$$\tilde{T}(n_{l-1}, t) - t = T(n_l, t) - t + \sum_{i=1}^{\kappa} \mathbf{e}_i(n_l/t) \overset{D}{=} \sum_{i=0}^{\kappa} \mathbf{e}_i(n_l/t) \overset{D}{=} \mathbf{e}(n_{l-1}/t) \overset{D}{=} T(n_{l-1}, t) - t,$$

with κ being an independent geometrically distributed random variable on $\mathbb{N}_0$ with parameter $1/2$. Thus $\tilde{P}_{l-1}$, constructed via $\tilde{T}(n_{l-1}, t)$, is equal in distribution to P_{l-1}, constructed via $\tilde{T}(n_{l-1}, t)$.
Even more interestingly, the random time $T(n,t)$ allows for an additional simplification of the Multilevel Monte Carlo algorithm. Let $(X_{T(n_l,t)}, \overline{X}_{T(n_l,t)})$ be known. To construct a sample on the coarse level, only remove one of the arrival times via Poisson thinning with probability $1/2$. If it is not removed, then

$$(X_{T(n_{l-1},t)}(\omega_i), \overline{X}_{T(n_{l-1},t)}(\omega_i)) = (X_{T(n_l,t)}(\omega_i), \overline{X}_{T(n_l,t)}(\omega_i)) \quad \text{and} \quad P_l(\omega_i) - P_{l-1}(\omega_i) = 0.$$

If it is removed, then a sample $(S_{n_{l-1}/t}(\omega_i), I_{n_{l-1}/t}(\omega_i))$ of $(\overline{X}_{\mathbf{e}(n_{l-1}/t)}, \underline{X}_{\mathbf{e}(n_{l-1}/t)})$ is generated and the new pair $(\overline{X}_{T(n_{l-1},t)}(\omega_i), X_{T(n_{l-1},t)}(\omega_i))$ is constructed via

$$\begin{aligned}\overline{X}_{T(n_{l-1},t)}(\omega_i) &= \max(\overline{X}_{T(n_l,t)}(\omega_i), X_{T(n_l,t)}(\omega_i) + S^i_{n_{l-1}/t}),\\ X_{T(n_{l-1},t)}(\omega_i) &= X_{T(n_l,t)}(\omega_i) + S^i_{n_{l-1}/t} + I^i_{n_{l-1}/t}.\end{aligned}$$

Proposition 5. *Let $t \geq 0$ be a given time instance. Furthermore suppose that X satisfies the assumptions A1 to A3. Then for any $n \in \mathbb{N}$*

$$\begin{aligned}\mathbb{E}((X_{T(n,t)} - X_t)^2) &\underset{\sim}{=} n^{-1} \quad \textit{and} \quad \mathbb{E}(|X_{T(n,t)} - X_t|) \underset{\sim}{\leq} n^{-\frac{1}{2}},\\ \mathbb{E}((\overline{X}_{T(n,t)} - \overline{X}_t)^2) &\underset{\sim}{\leq} n^{-1} \quad \textit{and} \quad \mathbb{E}(|\overline{X}_{T(n,t)} - \overline{X}_t|) \underset{\sim}{\leq} n^{-\frac{1}{2}}.\end{aligned}$$

Proof. Due to Jensen's inequality it is sufficient to prove the inequalities for the second moments, as the inequalities for the first moments are a direct consequence. The random variable $T(n,t) - t$ is exponentially distributed with parameter n/t due to the lack of memory property. Thus,

$$\mathbb{E}((T(n,t) - t)^2) = 2\frac{t^2}{n^2} \quad \text{and} \quad \mathbb{E}(|T(n,t) - t|) = \mathbb{E}(T(n,t) - t) = \frac{t}{n}.$$

An application of Lemma 13 is then sufficient to show that the inequalities hold. □

Together with Theorems 26 and 27 and $\gamma = 1$ the next Theorem directly follows from Proposition 5.
Theorem 29. *Suppose that the assumptions A1 to A3 are satisfied. Let the time horizon be given by $T(n,t)$. Then for any $\nu \in \mathbb{N}$, the root mean square error for the Wiener-Hopf Monte Carlo method under the constraint that the expected value of the overall computational cost is limited to $O(\nu)$ operations, is $RMSE(\hat{Y}_n^N) \lesssim \nu^{-\frac{1}{4}}$. Correspondingly the root mean square error for the Wiener-Hopf Multilevel Monte Carlo method satisfies $RMSE(\hat{Y}) \lesssim \nu^{-\frac{1}{2}} \log^2 \nu$.*

6.6 Derivative Pricing with the Multilevel Monte Carlo Method

6.6.1 Barrier Options

In [Ferreiro-Castilla et al., 2013] the sub-class of β-processes in the class of meromorphic processes is used as the underlying process modelling the financial market. As remarked in section 6.1.2, this class of Lévy processes is one of the sub-classes that possesses an explicit Wiener-Hopf factorisation and which, under addition to a compound Poisson process, can be used as an approximation to a larger class of Lévy processes. Additionally, the class contains processes that allow for considerable freedom in the modelling of financial markets and includes important special cases of practical use.
The extrema distributions are given by

$$\begin{aligned}\mathbb{E}\left(e^{iu\overline{X}_{\mathbf{e}(q)}}\right) &= \frac{1}{1 + \frac{iu}{\zeta_0^-(q)}} \prod_{n \leq -1} \frac{1 + \frac{iu}{\beta_1(n+1-\alpha_1)}}{1 + \frac{iu}{\zeta_n(q)}}\\ \mathbb{E}\left(e^{iu\underline{X}_{\mathbf{e}(q)}}\right) &= \frac{1}{1 + \frac{iu}{\zeta_0^+(q)}} \prod_{n \geq 1} \frac{1 + \frac{iu}{\beta_2(n-1+\alpha_2)}}{1 + \frac{iu}{\zeta_n(q)}}\end{aligned}$$

with $\zeta_0^-(q)$ and $\{\zeta_n(q)\}_{n \leq -1}$ corresponding to the negative zero points and $\zeta_0^+(q)$ and $\{\zeta_n(q)\}_{n \geq 1}$ corresponding to the positive zeros of $\psi(-iu) = -q$. Simulation of the extrema may be accomplished by truncating the infinite series, which has the favourable property that the cost of the simulation is independent of q. This was assumed to be a condition for the analysis presented in the previous sections, namely assumption A3. The assumption A1 also holds for β-processes.
The payoff function of barrier call options with strike $K > 0$ and barrier $B > 0$ in a Lévy model has the form

$$\mathbb{E}\left(\left(e^{x_0 + X_t} - K\right)^+ \mathbb{1}_{x_0 + \overline{X}_t > B}\right).$$

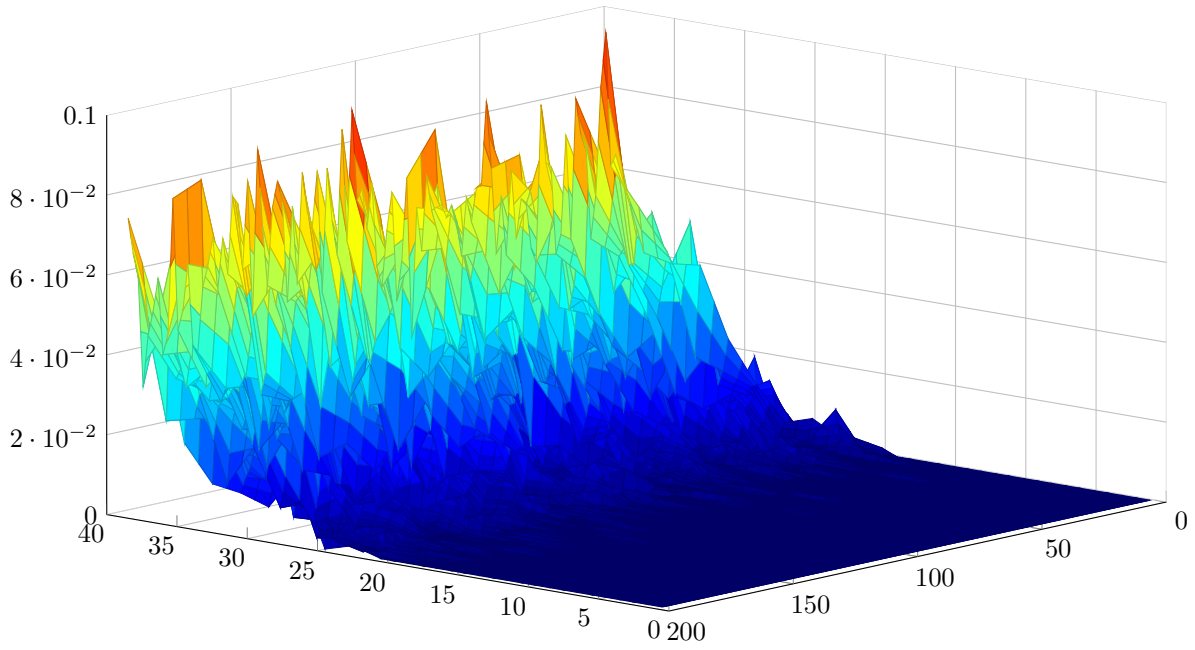

Figure 6.1: Price/Time grid of a β-process simulated with the Wiener-Hopf Multilevel Monte Carlo method with $M = 2$ and $L = 7$. The β-process with parameters $\mu = 1$, $\sigma = 0$, $C_1 = 1.5$, $C_2 = 1$, $\alpha_1 = 1$, $\alpha_2 = 1.5$, $\beta_1 = 1.5$, $\beta_2 = 1.5$, $\lambda_1 = 2$, $\lambda_2 = 2$ was simulated. The option parameters were set to $K = 5$ and $B = 10$. 40 initial prices $S_0 \in [0, 10]$ and 200 time instances $t \in [0, 1]$ were used as the underlying grid. The behaviour is as expected, with rising barrier option prices for increasing asset prices.

It should be noted that assumption A2 is not satisfied. Nonetheless, the performance analytically shown for payoff functions satisfying the assumption is achieved.
It should be noted that β-processes are adequate processes for modelling financial asset price processes. The class of β-processes servers as an excellent approximation to Meixner-processes [Ferreiro-Castilla and Schoutens, 2010]. In [Schoutens, 2001] it is shown that a near-perfect fit to financial time series data is achievable, in particular, the Nikkei-225 Index and the Standard and Poors 500 Index were fitted with the process using time series data spanning three years. Stochastic volatility effects such as autocorrelation may be incorporated using stochastic time.

6.6.2 Parisian Options

Parisian options are exotic options which can be used in a wide variety of financial applications. Essentially, they are barrier options for which the knock in/knock out event is not given by the crossing a pre-specified barrier, but rather by an excursion above or below that barrier, see Figure 6.4. The option yields a conditional payoff at maturity if the asset price stays above or below a pre-specified barrier for a pre-specified amount of time.
In contrast to barrier options and much like Asian options, the excursion time makes Parisian options impervious to manipulations of the underlying stock price. Whereas barrier options can be knocked in by pushing the price above the barrier for only a moment, Parisian options can be only knocked in if the stock price has stayed above a certain level for a longer time period. Applications are for example compensatory options for board members of a corporation, where barrier options could give an incentive to push the price above a bonus barrier for only a moment in time. Parisian options would incentivise long-term performance. In the insurance industry, the ruin of an insurance portfolio in conventional risk theory could be fitted with a Parisian option-style window, taking account of grace periods afforded under bankruptcy law.
Formally, let $g_t^B(X)$ be the last time before $t > 0$ at which an asset process $X = (X_t)_{t \geq 0}$ was, without loss of generalisation, below the barrier B,

$$g_t^B(X) = \sup_{s \leq t} \{X_s \leq B\}.$$

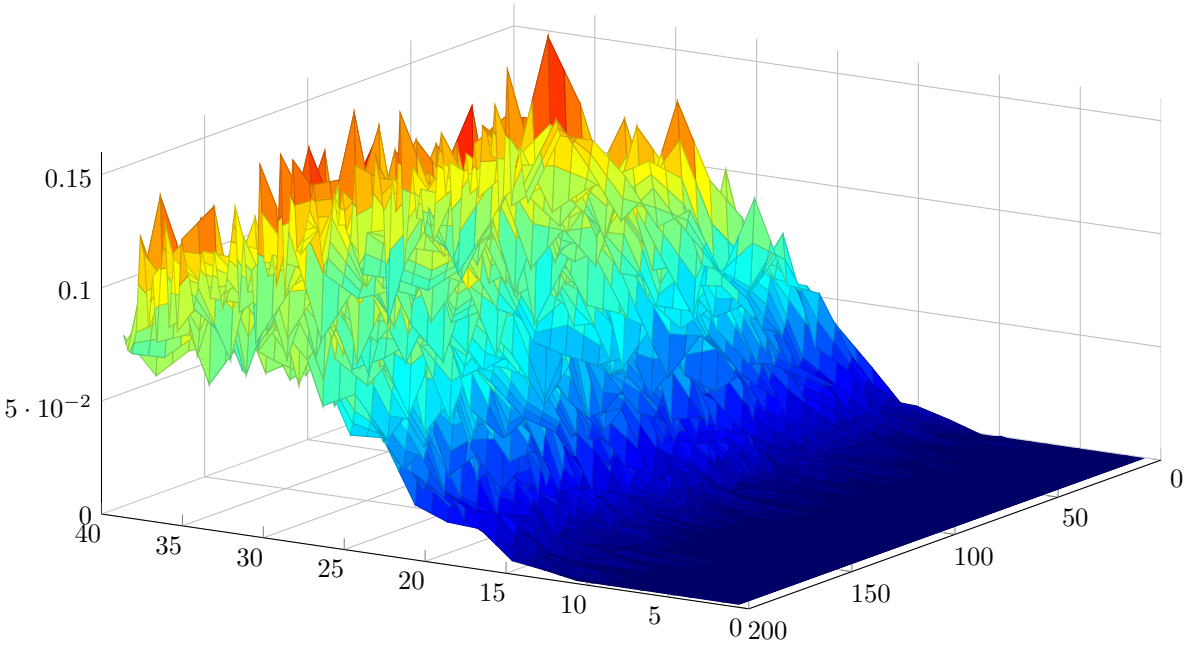

Figure 6.2: Price/Time grid of a β-process simulated with the Wiener-Hopf Multilevel Monte Carlo method with $M = 2$ and $L = 7$. The β-process with parameters $\mu = 1$, $\sigma = 1$, $C_1 = 1.5$, $C_2 = 1$, $\alpha_1 = 1$, $\alpha_2 = 1.5$, $\beta_1 = 1.5$, $\beta_2 = 1.5$, $\lambda_1 = 2$, $\lambda_2 = 2$ was simulated. The option parameters were set to $K = 5$ and $B = 10$. 40 initial prices $S_0 \in [0, 10]$ and 200 time instances $t \in [0, 1]$ were used as the underlying grid. Clearly, setting $\sigma = 1$ in comparison to Figure 6.1 has shifted the behaviour of the process markedly.

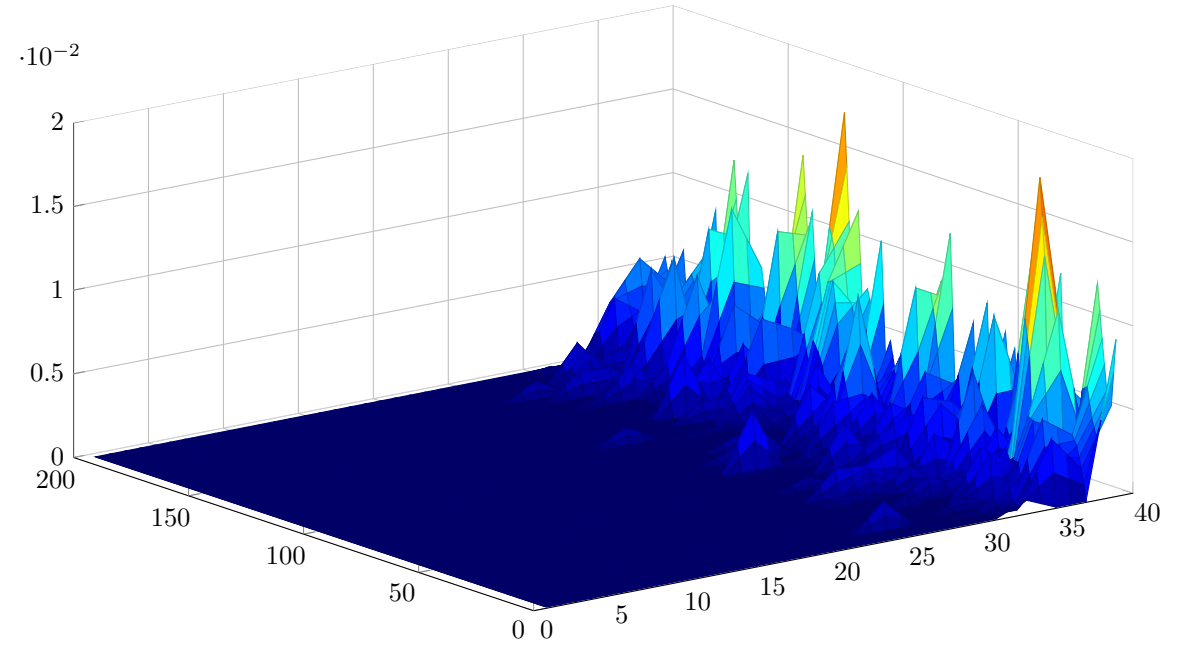

Figure 6.3: Price/Time grid of a β-process simulated with the Wiener-Hopf Multilevel Monte Carlo method with $M = 2$ and $L = 7$. The β-process with parameters $\mu = 1$, $\sigma = 1$, $C_1 = 0.8$, $C_2 = 1.3$, $\alpha_1 = 1.2$, $\alpha_2 = 1.7$, $\beta_1 = 1$, $\beta_2 = 1.6$, $\lambda_1 = 1.4$, $\lambda_2 = 0.5$ was simulated. The option parameters were set to $K = 5$ and $B = 10$. 40 initial prices $S_0 \in [0, 10]$ and 200 time instances $t \in [0, 1]$ were used as the underlying grid.

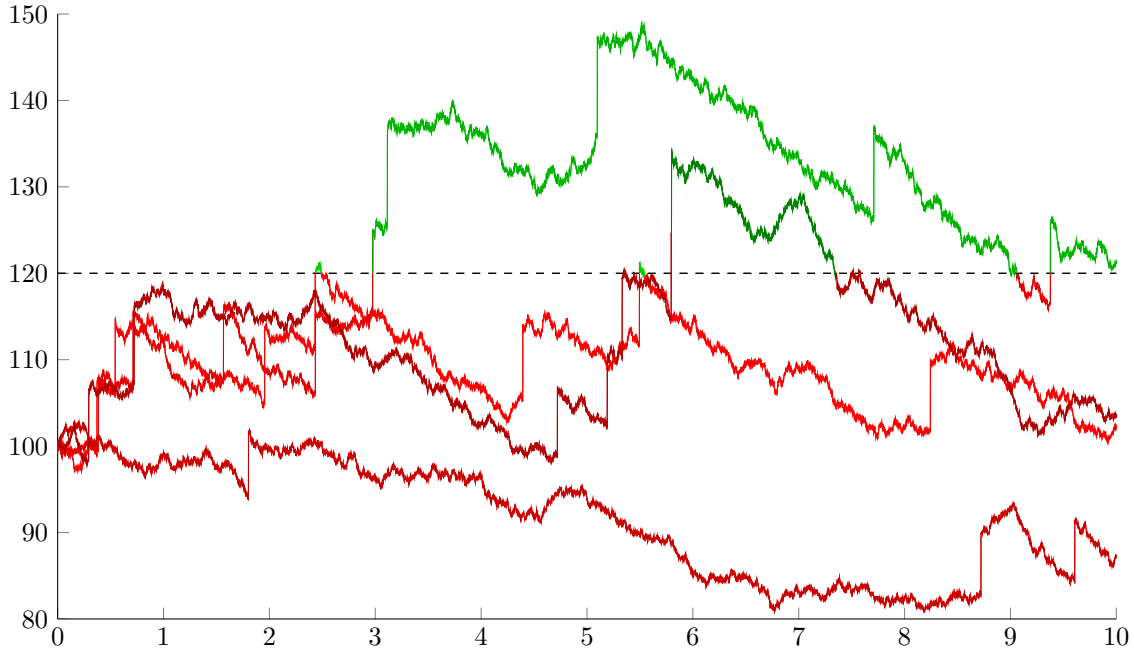

Figure 6.4: Lévy processes in red, with their excursions above a barrier set at $B = 120$ marked in green. If these excursions are longer than a prespecified time, a Parisian option knocks in respectively out. For cumulative Parisian options the overall excursion above the barrier is of interest.

The stopping time depending on a given time window $D \in \mathbb{R}$ of interest for the particular case of Parisian options is given by. If this stopping time occurs before the maturity date T, the option knocks in, respectively knocks out.

$$G_D^+ = \inf_{t>0}\{(t - g_t^B(X))\mathbb{1}_{S_t>B} \geq D\}.$$

Future research could implement a Wiener-Hopf Multilevel Monte Carlo method which would not only sample the supremum or the first hitting time of a Lévy process but also the respective excursions above a given barrier.

Chapter 7

The Multilevel Quasi-Monte Carlo Method

The Multilevel Monte Carlo method may be extended in the direction of Quasi-Monte Carlo simulation in different ways. The Quasi-Monte Carlo method computes a desired integral on a d-dimensional hypercube like the expectation in (4.1) by way of a quadrature rule consisting of N equally weighted points of the form

$$\int_{[0,1]^d} f(x)dx \approx \frac{1}{N}\sum_{i=0}^{N-1} f(x_i). \tag{7.1}$$

In contrast to the Classic Monte Carlo method the d-dimensional points $\{x_i\}_{1\leq i\leq N}$ are not sampled uniformly from the unit cube, but are constructed by an algorithm in a deterministic manner. The Classic Monte Carlo method requires in general computational costs of order $O(\epsilon^{-3})$, whereas in the best case the Quasi-Monte Carlo method requires only costs of order $O(\epsilon^{-2})$ [Glasserman, 2004].
In general, the fractional part of low-discrepancy sequences $\{z_i\}_{1\leq i\leq N}$ is considered, that is $x_i = z_i - \lfloor z_i \rfloor$ for $i \in \{1,\dots,N\}$. In order to mimic salient features of the Classic Monte Carlo method, namely the provision of an unbiased estimate of a target quantity and of a confidence interval for that unbiased estimate, the Quasi-Monte Carlo method may be randomised [Tuffin, 2004]. This is done by setting $x_i = z_i + \Delta - \lfloor z_i + \Delta \rfloor$ with an offset vector $\Delta \in [0,1)^d$. Generating $q \in \mathbb{N}$ random offset vectors $\{\Delta_j\}_{1\leq j\leq q}$ gives one the ability to treat the $\{x_i\}_{1\leq i\leq N}$ as random variables with samples $\{x_i^{(j)}\}_{1\leq i\leq N, 1\leq j\leq q}$, $x_i^{(j)} = z_i + \Delta_j - \lfloor z_i + \Delta_j \rfloor$, and

$$\hat{Y} = \frac{1}{N}\frac{1}{q}\sum_{j=1}^{q}\sum_{i=1}^{N-1} f\left(x_i^{(j)}\right)$$

becomes an unbiased estimate of the integral in (7.1) with confidence intervals constructed in the fashion of the interval (4.5). Let $f_l = P_0$ for $l = 0$ and $f_l = P_l - P_{l-1}$ for $l \in \{1,\dots,L\}$ and write $\{x_i^{l,(j)}\}_{1\leq i\leq N_l, 1\leq j\leq q, 0\leq l\leq L}$, $x_i^{l,(j)} = z_i^l + \Delta_j^l - \lfloor z_i^l + \Delta_j^l \rfloor$. On a level $l \in \{1,\dots,L\}$ in a Multilevel Monte Carlo setting the quantity

$$\hat{Y}_0 = \frac{1}{N_0}\frac{1}{q}\sum_{j=1}^{q}\sum_{i=1}^{N_0} f_0\left(x_i^{0,(j)}\right)$$

can thus be used as an estimate for $\mathbb{E}(P_0)$ and

$$\hat{Y}_l = \frac{1}{N_l}\frac{1}{q}\sum_{j=1}^{q}\sum_{i=1}^{N_l-1} f_l\left(x_i^{l,(j)}\right)$$

can be used as an estimate for $\mathbb{E}(P_l - P_{l-1})$ with $l \in \{1,\dots,L\}$.
Considering the Milstein discretisation with step size $\Delta t = T/n$ of a scalar stochastic differential equation on the time horizon $[0,T]$ of the form

$$dX_t = a(t,X_t)dt + b(t,X_t)dW_t,$$

the dimensionality of the problem is the number of discretisation points $d = n$. When using the Milstein scheme (3.14) to approximate the stochastic differential equation, the question of how the covariance matrix of the Brownian increments should be factorised comes up. The expectation of a derivative with payoff function P based on the value of the solution $(X_t)_{t\geq 0}$ of the stochastic differential equation is given by

$$\int_{\mathbb{R}^d} P(x) \frac{e^{-\frac{1}{2}x^t \Sigma^{-1} x}}{(2\pi)^{d/2}\sqrt{\det \Sigma}} dx, \tag{7.2}$$

where Σ is the d-dimensional covariance matrix of the elements of x which are the values of the driving Brownian motion $\{W_{i\Delta t}\}_{1\leq i\leq n}$. Factorising the covariance matrix into $\Sigma = AA^t$ and substituting in (7.2) with $x = Ay$ and $y = \Phi^{-1}(z)$ leads to

$$\int_{\mathbb{R}^d} P(Ay) \frac{e^{-\frac{1}{2}y^t y}}{(2\pi)^{d/2}} dy = \int_{[0,1]^d} P(A\Phi^{-1}(z)) dz.$$

Thus, the transformed integral has the desired form (7.1). Whereas the factorisation of Σ is of no importance when using Monte Carlo simulation, it is important when considering Quasi-Monte Carlo simulation. Roughly, three approaches are frequently used.

1. **The Cholesky factorisation.** Σ is factorised into the Cholesky factorisation $\Sigma = AA^t$. The Brownian increments are then given by $\Delta W_{j\Delta t} = \sqrt{\Delta t}\Phi^{-1}(x_i^{(j)})$. This is usually called the standard construction.
2. **The Brownian Bridge construction.** The component $x^{(1)}$ is used to define W_T, the component $x^{(2)}$ is used to define $W_{T/2}$ conditional on $x^{(1)}$, the components $x^{(3)}$ and $x^{(4)}$ are used to define $W_{T/4}$ and $W_{3T/4}$ conditional on the components $x(1)$ and $x^{(2)}$, and so on and so forth. The factorisation matrix A is only implicitly defined.
3. **The Principal Components Analysis Method.** A is constructed column-wise. The k-the column of A is set to be equal to $\sqrt{\lambda_k}v_k$ where λ_k is the k-th largest eigenvalue of Σ and where v_k is the associated eigenvector.

In [Giles and Waterhouse, 2009] the second method is chosen, as it systematically outperforms the first and third methods in numerical simulations. However, in general there are problems in financial mathematics for which the standard construction and the PCA method fare better.
The Quasi-Monte Carlo method pendant to the Multilevel Monte Carlo complexity algorithm, Algorithm 4, is a straightforward variation of the original algorithm.

Algorithm 7 The Multilevel Quasi-Monte Carlo Complexity Algorithm

Let L be the number of levels in a Multilevel Monte Carlo setting, N_l be the number of Quasi-Monte Carlo points at level l for $l = 0, \ldots, L$, $\hat{Y}_l$ be the average of P_0 or $P_l - P_{l-1}$ for $l = 1, \ldots, L$ over $q = 32$ sets of N_l Quasi-Monte Carlo lattice points with each set based on a different random offset Δ and V_l be the variance of $\hat{Y}_l$.

(1) Set $L = 0$.
(2) Determine a first estimate of V_L using $q = 32$ random offsets and $N_L = 1$.
(3) **While** $\sum_{l=0}^{L} V_l > \epsilon^2/2$ double the N_l on the level with the largest $V_l/(2^l N_l)$.
(4) **If** $L < 2$ or **if** the bias estimate is greater than $\epsilon/\sqrt{2}$
    Set $L = L + 1$.
    Go to step 2.

The low-discrepancy sequence used in [Giles and Waterhouse, 2009] is a sequence generated by a rank-1 lattice rule, that is $z_i = \frac{i}{N}z$ with an integer vector z. To generate a suitable z the component-by-component sieve algorithm introduced in [Dick et al., 2008] in connection with a Fast Fourier Transformation developed in [Nuyens and Cools, 2006] can be used.
Numerical simulation of the different derivatives presented in section 5.6 is performed for the case of the Black-Scholes model in [Giles and Waterhouse, 2009]. In general, they observe that the introduction of the Quasi-Monte Carlo method has the greatest impact on the rough levels of the Multilevel Monte Carlo method, with

improvements flailing as the levels become finer. For European options, a 20 to 100-fold increase in performance is found when using the Multilevel Quasi-Monte Carlo method over the conventional Multilevel Monte Carlo method, with computational costs being of order $O(\epsilon^{-1})$. Similar results are found for the case of Asian and lookback options, whereas improvements in case of barrier and Digital options are less pronounced, with savings ranging around 5 to 10-fold improvements in computational efficiency.
In a more recent paper [Gerstner and Noll, 2013], a Sobol sequence is used instead of a sequence determined by a rank-1 lattice rule. The authors also perform a numerical analysis of the Multilevel Quasi-Monte Carlo method. In order to present the analysis, some definitions common to Quasi-Monte Carlo methods are introduced [Glasserman, 2004, p. 288f].
First, the concept of discrepancy of a sequence is introduced. Intuitively, discrepancy describes the deviation of a given sequence from the uniform distribution. The higher the discrepancy, the higher the irregularity of the given sequence. Let $P = \{x_1, \dots, x_N\}$ with $N \in \mathbb{N}$ and $x_1, \dots, x_N \in [0,1]^d$. For a given set $B \subset [0,1]^d$ the counting function $A(B,P)$ is defined by

$$A(B,P) = \sum_{i=1}^{N} \mathbb{1}_B(x_i),$$

indicating the number of $i \in \{1, \dots, N\}$ for which $x_i \in B$.
Definition 38. *Let $\mathcal{B}$ be a non-empty set of Lebesgue-measurable subsets of $[0,1]^d$ and let λ_d denote the d-dimensional Lebesgue measure. Then the general discrepancy of a point set P is defined by*

$$D_N(\mathcal{B}, P) = \sup_{B \in \mathcal{B}} \left| \frac{A(B,P)}{N} - \lambda_d(B) \right|.$$

By placing restrictions on the family $\mathcal{B}$, particular concepts of discrepancy are defined. For the Multilevel Quasi-Monte Carlo analysis especially one concept of discrepancy is of interest.
Definition 39. *The star discrepancy $D_N^*(P) = D_N^*(x_1, \dots, x_N)$ of the point set P is defined by $D_N^*(P) = D_N(\mathcal{J}^*, P)$, where $\mathcal{J}^*$ is the family of all subintervals of $[0,1]^d$ of the form $\prod_{i=1}^d [0, u_i)$.*
Next, two notions of measuring the variation of a function are introduced. Let f be a function on $[0,1]^d$ and let $J = [u_1^-, u_1^+] \times [u_2^-, u_2^+] \times \cdots \times [u_d^-, u_d^+]$ with $0 \le u_i^- \le u_i^+$ for $i \in \{1, \dots, d\}$. Furthermore, let $\mathcal{E}(J)$ be the set of all vertices for which the number of $+$ superscripts is even and let $\mathcal{O}(J)$ be the set of all vertices for which the number of $+$ superscripts is odd. Denote the sum of f over the vertices of J with opposite signs given to function values at vertices lying next to each other with $\Delta(f, J)$,

$$\Delta(f,J) = \sum_{u \in \mathcal{E}(J)} f(u) - \sum_{u \in \mathcal{O}(J)} f(u).$$

The hypercube $[0,1]^d$ may be partitioned into rectangles taking the form of J, together forming a set of rectangles denoted by $\mathcal{P}$.
Definition 40. *Let J, $\mathcal{P}$ and $\Delta(f,J)$ be defined as above. With $\mathcal{P}$ ranging over all possible partitions of $[0,1]^d$ into rectangles, the measure of the variation of f*

$$V^d(f) = \sup_{\mathcal{P}} \sum_{J \in \mathcal{P}} |\Delta(f,J)|$$

is defined to be the measure of variation in the sense of Vitali.
In particular, $V^d(f)$ allows for the representation

$$V^d(f) = \int_0^1 \dots \int_0^1 \left| \frac{\partial^d f}{\partial u_1 \dots \partial u_d} \right| du_1 \dots du_d$$

in case the partial derivative is continuous on the unit hypercube [Niederreiter, 1992, p. 19]. A measure of the variation of f employed in the analysis of the Multilevel Quasi-Monte Carlo Method is the measure of the variation in the sense of Hardy and Krause.
Definition 41. *For $1 \le k \le d$ and $1 \le i_1 < i_2 < \cdots < i_k \le d$, let $V^k(f, i_1, \dots, i_k)$ be the variation in the sense of Vitali of the restriction of f to the k-dimensional set $\{(u_1, \dots, u_d) \in [0,1]^d \mid u_j = 1 \, for \, j \notin \{i_1, \dots, i_k\}\}$. Then*

$$V(f) = \sum_{k=1}^{d} \sum_{1 \le i_1 < \cdots < i_k \le d} V^k(f, i_1, \dots, i_k)$$

is defined to be the measure of variation in the sense of Hardy and Krause.
Using the star discrepancy and the Hardy-Krause variation, the Koksma-Hlawka inequality can be formulated [Hlawka, 1961].
Theorem 30 (Koksma-Hlawka Inequality)**.** *If f has a bounded variation $V(f)$ on $[0,1]^d$ in the sense of Hardy and Krause, then for any $x_1, \ldots, x_N \in [0,1]^d$*

$$\left| \frac{1}{N} \sum_{i=1}^{N} f(x_i) - \int_{[0,1]^d} f(u)du \right| \leq V(f) D_N^*(x_1, \ldots, x_N).$$

Using the Koksma-Hlawka inequality, and setting $Q_{x_l} = \{z_i^l + \Delta_j^l\}_{1 \leq i \leq N_l, 1 \leq j \leq q}$ as well as $\mathbb{P} = \{z_i^l\}_{1 \leq i \leq N_l}$ the variance of the estimator $\hat{Y}_l$ can be bounded above by

$$\begin{aligned} \mathbb{V}ar(\hat{Y}_l) &= \frac{1}{q} \mathbb{V}ar\left(\frac{1}{N_l} \sum_{j=1}^{N_l} f_l\left(x_i^{l,(j)}\right) \right) = \frac{1}{q} \int_{[0,1]^d} \left(\frac{1}{N_l} \sum_{j=1}^{N_l} f_l\left(x_i^{l,(j)}\right) - \mathbb{E}(f_l) \right)^2 dx_l \\ &\leq \frac{1}{q} \int_{[0,1]^d} \left(V(f_l) D_{N_l}^*(Q_{x_l}) \right)^2 dx_l. \end{aligned}$$

Due to $D_{N_l}^*(Q_{x_l}) \leq 4^d D_{N_l}^*(\mathbb{P})$, the bound

$$\mathbb{V}ar(\hat{Y}_l) \leq \left(4^d V(f_l) D_{N_l}^*(\mathbb{P}) \right)^2$$

follows. For low-discrepancy sequences, $D_{N_l}^*(\mathbb{P}) = O(N_l^{-1}(\log N_l)^d)$ holds and because $4^d \leq (\log N)^d$ is true for all $N \geq 55$ the factor 4^d can be discarded by doubling the dimension d. In [Niederreiter, 1992] it is proposed that a star discrepancy of $D_{N_l}^*(\mathbb{P}) = O(N_l^{-1+\epsilon})$ for all $\epsilon > 0$ is sufficient in most cases. Thus, a variance of order $\mathbb{V}ar(\hat{Y}_l) = O\left((V(f_l))^2 N_l^{-2+\epsilon} \right)$ is obtained.
It remains to determine the behaviour of $V(f_l)$. The authors of [Gerstner and Noll, 2013] assume that $V(f_l)$ has the same behaviour as $\mathbb{V}ar(f_l)$ via the following chain of argumentation. In the Multilevel Monte Carlo setting the function $f_l = P_l - P_{l-1}$ is the difference of two functions at different time steps. As the difference of the time instances becomes increasingly smaller with successively higher levels, the difference of the two functions evaluated at a certain point successively decreases. With the additional assumption that the variation is connected to the strong convergence behaviour of the underlying discretisation method used, the authors arrive at the conclusion that $V(f_l)$ is similar in its asymptotic behaviour to $\mathbb{V}ar(f_l)$, leading to

$$\mathbb{V}ar(\hat{Y}_l) = O\left(\Delta_l^{2\beta} N_l^{-2+\epsilon} \right) \tag{7.3}$$

with the β from the Multilevel Monte Carlo Theorem, Theorem 31.
Under these assumptions a numerical analysis of the Multilevel Quasi-Monte Carlo method based on the initial work in [Giles, 2008a] is given in [Gerstner and Noll, 2013]. The underlying discretisation scheme is chosen to be the Euler scheme.
Since the order of weak convergence for the Euler scheme is 1, $\mathbb{E}(P_L - P) = O(\Delta t_L)$ holds. By fixing $L = \frac{\log \epsilon^{-1}}{\log M} + O(1)$ the order of the finest time step is $\Delta t_L = O(\epsilon)$ in the limit $\epsilon \to 0$. Due to $\hat{Y}$ having no bias as an estimate for $\mathbb{E}(P_L)$ as evident from equation (5.3), the weak error $\mathbb{E}(\hat{Y} - \mathbb{E}(P))$ is also of order $O(\epsilon)$. Discarding the ϵ in (7.3) leads to

$$\mathbb{V}ar(\hat{Y}_l) = O\left(\Delta t_l^{2\beta} N_l^{-2} \right).$$

Minimizing $\mathbb{V}ar(\hat{Y}) = \sum_{l=0}^{L} \mathbb{V}ar(\hat{Y}_l)$ in N_l leads to an optimisation problem similar to (5.10) and can accordingly be dealt with in the same way. Formulating it as an Lagrangian multiplier problem and solving it in combination with using the Euler scheme, which implies $\beta = 1/2$, gives the optimal solution

$$N_l = O\left(\frac{(L+1)^{1/2} \Delta t_l^{2/3}}{\epsilon^{7/6}} \right).$$

Plugging in this result in to $\mathbb{V}ar(\hat{Y})$ gives

$$\mathbb{V}ar(\hat{Y}) = \sum_{l=0}^{L} c_2 N_l^{-2} \Delta t_l = \sum_{l=0}^{L} c_1 \Delta t_l \left(\frac{(L+1)^{1/2} \Delta t_l^{2/3}}{\epsilon^{7/6}} \right)^{-2} = \sum_{l=0}^{L} c_2 \Delta t_l \left(\frac{\epsilon^{7/3}}{(L+1)\Delta t_l^{4/3}} \right) \leq c_2 \epsilon^{7/3} \epsilon^{-1/3} \leq c_2 \epsilon^2$$

leading to the desired target mean square error MSE $= O(\epsilon^2)$. Plugging in N_l into the overall computational cost gives

$$C = c_3 \sum_{l=0}^{L} N_l \Delta t_l^{-1} = c_3 \sum_{l=0}^{L} \left(\frac{(L+1)^{1/2} \Delta t_l^{2/3}}{\epsilon^{7/6}} \right) \Delta t_l^{-1} = c_3 \sum_{l=0}^{L} \frac{(L+1)^{1/2}}{\epsilon^{7/6} \Delta t_l^{1/3}} \leq c_3 \frac{(L+1)^{3/2}}{\epsilon^{7/6} \Delta t_L^{1/3}} \leq c_3 \epsilon^{-9/6} (L+1)^{3/2}. \tag{7.4}$$

Thus, substituting $L = \frac{\log \epsilon^{-1}}{\log M} + O(1)$ leads to $C = O(\epsilon^{-3/2} (\log \epsilon^{-1})^{3/2})$.
Generalizing upon the Multilevel Monte Carlo Theorem proven in [Giles, 2008a], an analogous Multilevel Quasi-Monte Carlo Theorem is presented in [Gerstner and Noll, 2013]. The Theorem is extended through the use of a parameter δ that changes according to the chosen simulation method and has to be determined numerically. Using Quasi-Monte Carlo methods $\delta = 2 - \epsilon$ with $\epsilon > 0$ applies; the conventional Multilevel Monte Carlo method can be obtained with $\delta = 1$.

Theorem 31. *Let P denote a functional of the solution of a stochastic differential equation an let P_l denote the corresponding approximation at level l with $\Delta t_l = M^{-l} T$. If there exist independent estimators $\hat{Y}_l$ with computational complexity C_l based on N_l Monte Carlo samples and positive constants $\alpha, \beta, \delta, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2}$, $\delta \geq 1$ and*

a) $|\mathbb{E}(P_l - P)| \leq \frac{c_1}{M^{\alpha l}}$

b) $\mathbb{E}(\hat{Y}_l) = \begin{cases} \mathbb{E}(P_0) & \text{if} \quad l = 0 \\ \mathbb{E}(P_l - P_{l-1}) & \text{if} \quad l > 0 \end{cases}$

c) $\mathbb{V}ar(\hat{Y}_l) \leq \frac{c_2}{N_l^{-\delta} M^{\beta l}}$

d) $C_l \leq c_3 N_l M^{\gamma l}$

then there exists a positive constant c_4 such that for any $\epsilon < \frac{1}{e}$ there are values L and N_l for which the Multilevel Monte Carlo estimator

$$\hat{Y} = \sum_{l=0}^{L} \hat{Y}_l$$

has a mean square error with bound

$$\mathbb{E}((\hat{Y} - \mathbb{E}(P))^2) < \epsilon^2$$

and computational complexity C with bound

$$C \leq \begin{cases} c_4 \max(\epsilon^{-\frac{2}{\delta}} (\log \epsilon^{-1})^{1+\frac{1}{\delta}}, \epsilon^{-\frac{1}{\alpha}}) & \text{if} \quad \beta \geq \delta \\ c_4 \max(\epsilon^{-\frac{2}{\delta} - \frac{1}{\alpha} + \frac{\beta}{\alpha\delta}} (\log \epsilon^{-1})^{1+\frac{1}{\delta}}, \epsilon^{-\frac{1}{\alpha}}) & \text{if} \quad \beta < \delta \end{cases} \tag{7.5}$$

As in the case of the conventional Multilevel Monte Carlo Theorem, Theorem (31), α is determined by the discretisation method via its weak convergence, whereas β is determined by the payoff function P. The newly introduced parameter δ depends on the payoff function and possibly on the dimension of the Quasi-Monte Carlo method used.
For Lipschitz bounded payoff functions and the Euler scheme, $\alpha = 1$ and $\beta = 1$ follows. With $\delta = 2$, the computational complexity (7.4) is obtained.

Notes

[1] Usually the Skorokhod space is equipped with a metrizable topology such that the Skorokhod space is Polish, which is called the Skorokhod topology. A Polish space is a separable, completely metrizable topological space. A topological space is called separable if it contains a countable dense set. There exists a metrizable topology on $D([0,\infty), \mathbb{R}^d)$ called the Skorokhod topology for which the Skorokhod space is Polish.

Let Λ be the set of all continuous functions $\lambda : [0,\infty) \to [0,\infty)$ that are strictly increasing with $\lambda(0) = 0$ and $\lim_{t\to\infty} \lambda(t) = \infty$. Furthermore, define for each $N \in \mathbb{N}$ the function k_N by

$$k_N(t) = \begin{cases} 1 & \text{if} \quad t \leq N \\ N+1-t & \text{if} \quad N < t < N+1 \\ 0 & \text{if} \quad t \geq N+1. \end{cases}$$

Then for all $\lambda \in \Lambda$ we set

$$||\lambda||_\Lambda = \sup_{s<t} \left| \log \frac{\lambda(t) - \lambda(s)}{t-s} \right|.$$

and for $\alpha, \beta \in D([0,\infty), \mathbb{R}^d)$ we set

$$\delta_N(\alpha,\beta) = \inf_{\lambda\in\Lambda} (||\lambda||_\Lambda + ||(k_N\alpha)\circ\lambda - k_N\beta||_\infty).$$

Then

$$\delta(\alpha,\beta) = \sum_{n=1}^{\infty} \frac{1}{2^N} \min(1, \delta_N(\alpha,\beta))$$

is a metric on $D([0,\infty), \mathbb{R}^d)$, which yields the Skorokhod topology [Jacod and Shiryaev, 2002]. The Borel sigma-algebra induced by the metric is the Skorokhod sigma-algebra.

[2]Historically, the reference asset was usually chosen to be the money market account given by $S_t^0 = e^{r_t}$ with $(r_t)_{t\geq 0}$ being a positive deterministic process.

[3]A strong solution $X = (X_t)_{t\geq 0}$ of the stochastic differential equation (3.2) is a stochastic process satisfying

1. X is adapted to the filtration generated by the Brownian motion $(W_t)_{t\geq 0}$ and by X_0,
2. X is continuous,
3. $\mathbb{P}(X_0 = x_0) = 1$,
4. $\mathbb{P}(\int_0^t |a(s,X_s)| + |b(s,X_s)|^2 ds < \infty) = 1$ for all $t > 0$ and
5. $X_t = X_0 + \int_0^t a(s,X_s)ds + \int_0^t b(s,X_s)dW_s$ for all $t \geq 0$ almost surely.

If the any two strong solutions X and Y with initial condition $X_0 = Y_0$ are indistinguishable, i.e. $\mathbb{P}(X_t = Y_t \forall t \geq 0) = 1$, the strong solution is said to be unique.

If the coefficient functions of (3.2) are locally Lipschiz continuous in x and satisfy for some $C \geq 0$

$$\langle x, a(t,x)\rangle + \text{trace}(b(t,x)b(t,x)^t) \leq C(1+|x|^2)$$

for all $x \in \mathbb{R}^d$ and $t \geq 0$, then the stochastic differential equation has a unique strong solution for any initial random variable X_0 satisfying $\mathbb{E}(|X_0|^2) < \infty$ [Durrett, 1996].

In the one-dimensional case strong uniqueness already holds for a Hölder-continuous coefficient function $b(t,x)$ of order $\frac{1}{2}$ [Karatzas and Shreve, 1991].

[4]The Lévy area derives its name from the fact that it describes the area enclosed by the paths of two Brownian motions and the chord joining their respective end points. This is a consequence of Stoke's Theorem.

[5]The Lévy area also vanishes in more general situations. If the diffusion function $b(t,x)$ is additive, i.e., $b(t,x) = b(t)$, the Milstein scheme reduces to the Euler scheme. If in the case $d = m$ the component $i \in \{1,\ldots,d\}$ of the d-dimensional process is only dependent on the component $i \in \{1,\ldots,m\}$ of the m-dimensional Brownian motion, the multi-dimensional Milstein scheme reduces to one-dimensional Milstein schemes. The most general case, which includes the previous two, is the case of commutative noise [Kloeden and Platen, 1995].

[6]Short treatments of the Lévy area which include the calculation of the characteristic function and the density can be found in [Protter, 1992] and [Kloeden and Platen, 1995].

[7]In special cases the number of evaluations of the diffusion matrix in a Runge-Kutta scheme may be lower. In the case of additive noise the Runge-Kutta scheme reduces to the Euler scheme, in which the diffusion matrix is only evaluated at Y_n. For diagonal noise the number of evaluations at each $b^{(i,i)}$ is reduced to two.

[8]Many variants relaxing the assumptions of the Strong Law of Large Numbers exist. An elementary proof of a Strong Law of Large Numbers requiring only pairwise independence was given by Etemadi [Etemadi, 1981]. A version weakening both the independence and the identical distribution down to a relatively loose correlation structure is presented in a report by Ninness [Ninness, 2000], which also showcases practical examples. Furthermore, the Strong Law is a special case of Birkhoff's ergodic theorem [Shalizi, 2007].

[9]The condition guarantees the boundedness of the density ofX_T. Another sufficient condition is that the coefficient functions $a(t,x)$ and $b(t,x)$ are both bounded and arbitrarily often continuously differentiable on $[0,T] \times \mathbb{R}$ and that $b(t,x)$ satisfies the uniform ellipticity condition, that is there exists a constant $C \in \mathbb{R}$ such that

$$\underline{(t,x)} \geq C \geq 0$$

for all $(t,x) \in [0,T] \times \mathbb{R}$ [Friedman, 1964].

[10]The same method works for integral equations of the second kind, which occur more often in problems, but involve a more complicated treatment.

[11]A general function $k(x,y)$ in (6.1) is called a kernel. If it only depends on the difference $x - y$, it is called a difference kernel.

[12]Such factorisations can usually be achieved by using Cauchy's integral formula.

[13]A stochastic process X on $(\Omega, \mathcal{F}, \mathbb{P})$ with the natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is said to have the Strong Markov Property if for each stopping time τ conditioned on $\{\tau < \infty\}$ the random variable $X_{\tau+t}$ is independent of $\mathcal{F}_{\tau+}$ given X_τ for all ≥ 0.

Bibliography

M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1970.

H. Albrecher, J. Renaud, and X. Zhou. A Lévy insurance risk process with tax. *Journal of Applied Probability*, 45(2):363–375, 2008.

S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *Journal of Applied Probability*, 38(2):482–493, 2001.

S. Asmussen, P. Glynn, and J. Pitmann. Discretization error in simulation of one-dimensional reflecting Brownian motion. *Annals of Applied Probability*, 5:875–896, 1995.

S. Asmussen, F. Avram, and M. Pistorius. Russian and American put options under exponential phase-type lévy models. *Stochastic Processes and their Applications*, 109(1):79–111, 2000.

J. Bertoin. *Lévy Processes*. Cambridge University Press, 1996.

N. Bingham and R. Kiesel. *Risk Neutral Valuation - Pricing and Hedging of Financial Derivatives*. Springer, 2004.

P. Boyle. Options: A Monte Carlo approach. *Journal of Financial Economics*, 4(3):323–338, 1977.

P. Boyle, M. Broadie, and P. Glasserman. Monte Carlo methods for security pricing. *Journal of Economic Dynmaics and Control*, 21:1267–1321, 1997.

D. Brigo and A. Alfonsi. Credit Default Swaps Calibration and Option Pricing with the SSRD Stochastic Intensity and Interest-Rate Model. *Finance and Stochastics*, 9(1):29–42, 2005.

M. Broadie, P. Glasserman, and S. Kou. Connecting discrete and continuous path-dependent options. *Finance and Stochastics*, 3:55–82, 1999.

M. Caballero, B. Fernández, and D. Nualart. Estimation of Densities and Applications. *Journal of Theoretical Probability*, 11(3):831–851, 1998.

N. Cai. On first passage times of a hyper-exponential jump diffusion process. *Operations Research Letters*, 37 (2):127–134, 2009.

N. Cai and S. Kou. Pricing Asian options under a general jump diffusion model. *Operations Research*, 60(1): 64–77, 2010.

J. Cariboni and W. Schoutens. *Lévy Processes in Credit Risk*. Wiley, 2009.

J. Clark and R. Cameron. The maximum rate of convergence of discrete approximations for stochastic differential equations. In B. Grigelionis, editor, *Lecture Notes in Control and Information Sciences*. Springer, 1980.

K. Cliffe, M. Giles, R. Scheichl, and A. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.

A. Cruzeiro, P. Malliavin, and A. Thalmaier. Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation. *Comptes Rendus de l'Académie des Sciences*, 338:481–486, 2004.

F. Delbaen and W. Schachermayer. A General Version of the Fundamental Theorem of Asset Pricing. *Mathematische Annalen*, 300:520–463, 1994.

F. Delbaen and W. Schachermayer. The no-arbitrage property under a change numéraire. *Stochastics and Stochastics Reports*, 53:213–226, 1995.

F. Delbaen and W. Schachermayer. The Fundamental Theorem of Asset Pricing for Unbounded Stochastic Processes. *Mathematische Annalen*, 312:212–250, 1998.

S. Dereich and F. Heidenreich. A multilevel Monte Carlo algorithm for lévy-driven stochastic differential equations. *Stochastic Processes and their Applications*, 121(7):1565–1587, 2011.

J. Dick, F. Pillichshammer, and B. Waterhouse. The construction of good extensible rank-1 lattices. *Mathematics of Computation*, 77(264):1345–1373, 2008.

R. Durrett. *Stochastic Calculus: A Practical Introduction*. Probability and Stochastics Series. CRC Press, 1996.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, 2008.

N. Etemadi. An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55:119–122, 1981.

A. Ferreiro-Castilla and W. Schoutens. The Beta-Meixner model. *Journal of Computational and Applied Mathematics*, 236(9):2466–2476, 2010.

A. Ferreiro-Castilla, A.E. Kyprianou, R. Scheichl, and G. Suryanarayana. Multi-level Monte-Carlo simulation for Lévy processes based on the Wiener-Hopf factorisation. *pre-print*, 2013. arXiv:1210.5868.

A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, 1964.

J. Gaines and T. Lyons. Random generation of stochastic area integrals. *Journal of Applied Mathematics*, 54: 1132–1146, 1994.

J. Gaines and T. Lyons. Variable step size control in the numerical solution of stochastic differential equations. *Journal of Applied Mathematics*, 57:1455–1484, 1997.

J. Gatheral. *The Volatility Surface*. Wiley Finance, 2006.

T. Gerstner and M. Noll. Randomized multilevel quasi-Monte Carlo path simulation. In T. Gerstner and P. Kloeden, editors, *Recent Developments in Computational Finance*. World Scientific Publishing, 2013.

M. Giles. Multilevel Monte Carlo Path Simulation. *Operations Research*, 56(3):607–617, 2008a.

M. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In Alexander Keller, Stefan Heinrich, and Harald Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*. Springer, 2008b.

M. Giles. Multilevel Monte Carlo for Basket Options. In M. Rossetti, R. Hill, B. Johansson, A. Dunkin, and R. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference*. WSC 2009, 2009.

M. Giles and L. Szpruch. Multilevel Monte Carlo methods for applications in finance. In T. Gerstner and P. Kloeden, editors, *Recent Developments in Computational Finance*. World Scientific Publishing, 2013.

M. Giles and B. Waterhouse. Multilevel quasi-Monte Carlo path simulation. *Radon Series in Computational and Applied Mathematics*, 8:161–181, 2009.

M. Giles, K. Debrabant, and A. Rößler. Numerical analysis of multilevel Monte Carlo path simultion using the Milstein discretisation. *pre-print*, 2013. arXiv:1302.4676.

P. Glasserman. *Monte Carlo Methods in Financial Engineering*, volume 53 of *Applications of Mathematics*. Springer-Verlag New York, 2004.

P. Glynn and W. Whitt. The Efficiency of Simulation Estimators. *Operations Research*, 40:505–520, 1992.

E. Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic Processes and their Applications*, 87:167–197, 2000.

E. Gobet and S. Menozzi. Exact approximation rate of killed hypoelliptic diffusions using the disrete Euler scheme. *Stochastic Processes and their Applications*, 112:210–223, 2004.

E. Gobet and S. Menozzi. Stopped diffusion processes: boundary corrections and overshoot. *Stochastic Processes and their Applications*, 120:130–162, 2010.

E. Gobet and E. Temam. Discrete time hedging errors for options with irregular pay-offs. *Finance and Stochastics*, 5:357–367, 2001.

S. Heinrich. Multilevel Monte Carlo methods. *Lecture Notes in Computer Science*, 2179:58–67, 2001.

E. Hlawka. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Annali di Matematica Pura ed Applicata*, 54:325–333, 1961.

E. Hopf. Mathematical problems of radiative equilibrium. *Cambridge tracts*, 31, 1934.

J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, 2002.

A. Janicki and A. Weron. *Simulation and Chaotic Behavior of alpha-Stable Stochastic Processes*. Marcel Dekker, 1994.

I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, second edition, 1991.

A. Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to options pricing. *Annals of Applied Probability*, 14(4):2681–2704, 2005.

P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics*. Springer-Verlag Heidelberg Berlin, 1995.

P. Kloeden, E. Platen, and W. Wright. The Approximation of Multiple Stochastic Integrals. *Stochastic Analysis and Applications*, 10:431–441, 1992.

R. Korn, E. Korn, and G. Kroisandt. *Monte Carlo Methods and Models in Finance and Insurance*. Chapman and Hall/CRC Financial Mathematics Series. CRC Press, Taylor and Francis Group, 2010.

A. Kuznetsov. Wiener-Hopf factorization and distribution of extrema for a family of Lévy processes. *Annals of Applied Probability*, 20(5):1801–1830, 2010.

A. Kuznetsov and X. Peng. On the Wiener-Hopf factorization for Levy processes with bounded positive jumps. *Stochastic Processes and their Applications*, 122(7):2610–2638, 2012.

A. Kuznetsov, A. Kyprianou, J. Pardo, and K. van Schaik. A Wiener-Hopf Monte Carlo Simulation Technique for Lévy Processes. *The Annals of Applie Probability*, 21(6):2171–2190, 2011.

A. Kuznetsov, A. Kyprianou, and J. Pardo. Meromorphic Lévy processes and their fluctuation identities. *Annals of Applied Probabilities*, 22(3):881–1300, 2012.

A. Kyprianou. *Introductory Lectures on fluctuations of Lévy processes with applications*. Springer, 2006.

J. Lawrie and I. Abrahams. A brief historical perspective of the Wiener-Hopf technique. *Journal of Engineering Mathematics*, 59(4):351–358, 2007.

D. Levin and M. Wildon. A combinatorial method for calculating the moments of the Lévy area. *Transactions of the American Mathematical Society*, 360:6695–6709, 2008.

A. Lewis and E. Mordecki. Wiener-Hopf factorization for Lévy processes having positive jumps with rational transforms. *Journal of Applied Probability*, 45:118–134, 2008.

G. Lowther. Preservation of the Local Martingale Property. `http://almostsure.wordpress.com/2010/03/25/preservation-of-the-local-martingale-property/`, 2010. Accessed: 2014-01-19.

S. Malham and A. Wiese. Lévy area logistic expansion and simulation. arXiv:1107.0151, 2011.

N. Metropolis. The Beginning of the Monte Carlo Method. *Los Alamos Science Special Issue*, pages 125–130, 1987.

T. Müller-Gronbach. The optimal uniform approximation of systems of stochastic differential equations. *The Annals of Applied Probability*, 12(2):664–690, 2002.

T. Müller-Gronbach and K. Ritter. Variable subspace sampling and multi-level algorithms. In P. L'Ecuyer and A. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2008*. Springer, 2009.

H. Niederreiter. *Random Number Generation and Qasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, 1992.

B. Ninness. Strong Laws of Large Numbers Under Weak Assumptions with Application. Technical Report EE9881, Department of Electrical and Computer Engineering, University of Newcastle, 2000.

D. Nuyens and R. Cools. Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces. *Mathematics of Computation*, 75:903–920, 2006.

E. Pecherskii and B. Rogozin. On joint distribution of random variables associated with fluctuations of a process with independent increments. *Theory of Probability and Its Applications*, 14:410–423, 1969.

M. Pistorius. On maxima and ladder processes for a dense class of lévy processes. *Journal of Applied Probability*, 43(1):208–220, 2006.

P. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Applications of Mathematics*. Springer-Verlag Heidelberg Berlin, 1992.

R. Rebonato. *Volatility and Correlation in the Pricing of Equity, FX and Interest Rate Options*. Wiley, 1999.

J. Rosiński. Series representation of Lévy processes from the perspective of point processes. In O. Barndorff-Nielsen, T. Mikosch, and S. Resnick, editors, *Lévy Processes - Theory and Applications*. Birkhauser, 2001.

T. Rydén and M. Wiktorsson. On the simulation of iterated Itô integrals. *Stochastic processes and their Applications*, 91:151–168, 2001.

K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Number 68 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

K. Scheicher. Complexity and effective dimension of discrete Lévy areas. *Journal of Complexity*, 23:152–168, April 2007.

W. Schoutens. *Lévy Processe in Finance: Pricing Financial Derivatives*. Wiley, 2003.

Wim Schoutens. Meixner processes in Finance. Technical report, K.U. Leuven, 2001.

C. Shalizi. Almost None of the Theory of Stochastic Processes. Lecture Notes, Department of Statistics, Carnegie Mellon University, 2007, 2007. URL `http://www.stat.cmu.edu/~cshalizi/almost-none/`.

P. Tankov and R. Cont. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC financial mathematics series. Chapman & Hall/CRC, 2004.

B. Tuffin. Randomization of Qasi-Monte Carlo Methods for Error Estimation: Survey and Normal Approximation. *Monte Carlo Methods and Applications*, 10(4):617–628, 2004.

N. Wiener and R. Payley. Fourier transforms in the complex domain. *American Mathematical Society Colloquium Publications*, 19, 1934.

M. Wiktorsson. Joint Characteristic Function and Simultaneous Sumulation of Iterated Itô Integrals for Multiple Independent Brownian Motions. *The Annals of Applied Probability*, 11:470–487, 2001.