A multilevel Monte Carlo algorithm for SDEs with jumps

Johanna Rantala

Matematiikan pro gradu

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Kevät 2019

Tiivistelmä: Johanna Rantala, *Monitasoinen Monte Carlo -menetelmä ja stokastiset differentiaaliyhtälöt Lévy-prosesseilla*, matematiikan pro gradu -tutkielma, 49 s. Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, helmikuu 2019.

Monitasoinen Monte Carlo -menetelmä on perinteisen Monte Carlo -algoritmin laajennos. Se on numeerinen metodi, jonka avulla voidaan approksimoida satunnaismuuttujan X odotusarvoa. Menetelmä hyödyntää diskretisointimetodia, jolla muodostetaan approksimaatiot $\widehat{X}_1, \widehat{X}_2, \ldots, \widehat{X}_L$ siten, että jokainen approksimaatio on edellistä tarkempi. Approksimaation tarkkuuden kasvaessa myös aikavaativuus kasvaa. Monitasoinen Monte Carlo -menetelmä hyödyntää approksimointia useilla eri tasoilla. Tällöin osa laskennasta voidaan tehdä epätarkemmilla mutta aikavaativuudeltaan pienemmillä tasoilla silti saavuttaen kalliimpien approksimaatioiden tarkkuus. Toisin sanoen, eri tarkkuuksisien approksimaatioiden yhdistäminen laskee menetelmän aikavaativuutta.

Tässä tutkielmassa monitasoista Monte Carlo -menetelmää käytetään stokastisen differentiaaliyhtälön ratkaisun odotusarvon arvioimiseen. Diskretisointimenetelmänä käytetään Euler-Maruyama -metodia. Approksimaation paikkansapitävyyttä mitataan keskineliövirheellä suhteutettuna menetelmän aikavaativuuteen. Tutkielmassa todistetaan Brownin liikkeen avulla muodostetun stokastisen differentiaaliyhtälön ratkaisun approksimaation keskineliövirheen olevan luokkaa $O(h^2)$ aikavaativuuden ollessa $O(h^{-2}(\log h)^2)$. Menetelmä laajennetaan Lévy-prosessin avulla muodostettuihin stokastisiin differentiaaliyhtälöihin, jolloin aikavaativuudella $O(n)$ ratkaisun approksimaation keskineliövirhe on $O(n^{-1/2})$ silloin, kun muodostavassa Lévy-prosessissa ei ole mukana diffuusiokomponenttia, ja $O(n^{-1/2}(\log n)^{3/2})$ muulloin.

Taustatietoina käydään läpi perustiedot todennäköisyysteoriasta ja stokastisista prosesseista. Jälkimmäisistä esitellään nimeltä Brownin liike, Poisson-prosessi sekä Lévy-prosessi. Prosessien esittelyn yhteydessä tutustutaan myös Lévy-Itô -hajotelmaan, joka kertoo kuinka Lévy-prosessi voidaan esittää Brownin liikkeen ja hyppyprosessin yhdistelmänä. Lisäksi käydään läpi stokastista integrointia Brownin liikkeen ja yleisemmin martingaalin suhteen, sekä Poisson-satunnaismitan suhteen. Lopuksi tutustutaan lyhyesti stokastisiin differentiaaliyhtälöihin ja niiden ratkaisuihin.

Abstract: Johanna Rantala, *A Multilevel Monte Carlo algorithm for SDEs with jumps*, Master's thesis in mathematics, 49 p. University of Jyväskylä, Department of Mathematics and Statistics, February 2019.

The multilevel Monte Carlo algorithm is an extension of the traditional Monte Carlo algorithm. It is a numerical method, which allows us to approximate the expected value of a random variable X. We use some appropriate discretization method to obtain approximations $\widehat{X}_1, \widehat{X}_2, \ldots, \widehat{X}_L$ of X such that each approximation is made with a finer grid. The more accuracy we want from our approximation, the more the computational cost grows. The multilevel method exploits evaluation at multiple levels of refining discretizations allowing us to achieve a better accuracy with lower cost.

In this thesis we use the Euler scheme to approximate the solution of the stochastic differential equation, and then we use the multilevel Monte Carlo algorithm to estimate the expected value of the solution. We prove that the mean squared error of the estimator is $O(h^2)$ with computational complexity $O(h^{-2}(\log h)^2)$ with stochastic differential equations driven by a Brownian motion. Lastly, we prove that with computational complexity $O(n)$, when the driving process is a Lévy process without Brownian component the error is $O(n^{-1/2})$ and with the Brownian component $O(n^{-1/2}(\log n)^{3/2})$.

As a background theory we introduce the basic concepts of probability and of stochastic processes, namely the Brownian motion, the Poisson processes and Lévy processes. We formulate the famous Lévy-Itô decomposition, which allows us to represent a Lévy process as the combination of a jump process and a Brownian motion. Additionally, we consider stochastic integration with respect to the Brownian motion, a martingale and the Poisson random measure. We use these to formulate the stochastic differential equations in two cases, driven by a Brownian motion or by a Lévy process.

Contents

1. Introduction

In this thesis we study an algorithm known as the multilevel Monte Carlo, which is an extension of the classical Monte Carlo method. Often computing something analytically is not possible or it is too expensive. In those cases we need numerical analysis. Numerical methods are often based on discretizations and evaluating needed quantities at certain points. When using them we need to find the balance between the accuracy of the result and the computational complexity, the latter of which is usually connected to time consumption of the algorithm. The multilevel Monte Carlo algorithm is used in many applications, for example in option pricing.

Our main goal is to analyze the numerical approximation of

$$S(g) := \mathbb{E}\big[g(Y)\big],$$

where $Y = (Y_t)_{t\in[0,1]}$ is the solution to a stochastic differential equation (SDE) and g is the so called payoff function with certain properties. Our main concerns are the error of the estimate and the time consumption, which we analyze in two different settings.

As a background for the algorithm we get familiar with some of the most well known stochastic processes, namely Poisson processes, Brownian motion and Lévy processes. Using these processes, we define a stochastic integral in various settings, and proceed to consider the stochastic differential equations.

In the first chapter we focus on the preliminary knowledge needed to understand the algorithm. The chapter is divided into two parts, which cover the basic theory of probabilities and stochastic processes. The next chapter covers stochastic integration and stochastic differential equations as needed in this thesis. The first and the second chapter will form the main theory behind our main goal, the multilevel Monte Carlo algorithm.

The third chapter is based on the work of Michael B. Giles [4]. We first formulate precisely the multilevel Monte Carlo algorithm to approximate the expected value of $g(Y_T)$, when Y_T is the solution of the SDE driven by a Brownian motion at time T. The main result of the chapter is that under certain conditions the mean squared error of the multilevel estimator is $O(h^2)$ with the computational complexity is $O(h^{-2}(\log h)^2)$.

In the last chapter we extend the algorithm to a more general case. We follow the work of Steffen Dereich and Felix Heidenreich [3] and consider the solutions to Lévy-driven SDEs. We consider Lévy processes with finite L'evy measure, and with the computation time n we obtain an error that is $O(n^{-1/2})$ in the case where the driving process has no Brownian component, and $O(n^{-1/2}(\log n)^{3/2})$ otherwise.

2. Preliminaries

In this section we introduce the basic concepts of probabilities and stochastic processes. We also present three of the most well known and used stochastic processes, the Poisson process, the Brownian motion and the Lévy process, and we discuss the connection between them.

2.1. **Probabilities.**

Definition 2.1. A collection $\mathcal{F}$ of subsets of Ω is a *σ-algebra* provided that

(i) $\emptyset \in \mathcal{F}$

(ii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
(iii) if $A_1, A_2, A_3, \ldots \in \mathcal{F}$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$

Example 2.2. It is easy to see that the smallest possible σ-algebra is $\{\emptyset, \Omega\}$.

Example 2.3. Let $\mathcal{G}$ be an arbitrary collection of subsets of Ω. Then $\sigma(\mathcal{G})$ denotes the smallest σ-algebra containing the collection $\mathcal{G}$.

If we set $\mathcal{G}$ to be system of all open subsets on $M \subseteq \mathbb{R}^d$ (or actually on any topological space), then $\sigma(\mathcal{G})$ is called *Borel σ-algebra* and it is denoted by $\mathcal{B}(M)$. The generating system can also be chosen to consist of all closed subsets from $M \subseteq \mathbb{R}$, or even of all open (closed, half open) cuboids.

Definition 2.4. Let $\mathcal{F}$ be a σ-algebra on Ω. A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is *a probability measure* provided that

(i) $\mathbb{P}(\Omega) = 1$
(ii) $\mathbb{P}(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$ for $A_1, A_2, \ldots \in \mathcal{F}$ with $A_k \cap A_l = \emptyset$ for $k \neq l$.

Definition 2.5. Let $\mathcal{F}$ be a σ-algebra on Ω and $\mathbb{P}$ a probability measure. The pair $(\Omega, \mathcal{F})$ is called *a measurable space*, and the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called *a probability space.*

Definition 2.6. Let $(\Omega, \mathcal{F})$ be a measurable space. A sequence $(\mathcal{F}_t)_{t\geq 0}$ of σ-algebras on Ω, such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $0 \leq s < t < \infty$, is called a *filtration.* The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t\geq 0}$ is called a *stochastic basis* or a *filtered probability space* and is denoted by $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

Definition 2.7. Let $(\Omega, \mathcal{F})$ be a measurable space. The function $f : \Omega \to \mathbb{R}$ is *a random variable*, if every pre-image of a Borel set is in $\mathcal{F}$, i.e. $f^{-1}(B) := \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F}$ for every $B \in \mathcal{B}(\mathbb{R})$. The function f is then said to be $\mathcal{F}$-measurable. In general, if $(\Omega, \mathcal{F})$ and (M, Σ) are measurable spaces, a map $f : \Omega \to M$ is called $(\mathcal{F}, \Sigma)$-measurable, provided that

$$f^{-1}(B) := \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F} \text{ for every } B \in \Sigma.$$

Definition 2.8. [1, p. 10] Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra and f be an integrable real valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists an integrable random variable g on $(\Omega, \mathcal{G}, \mathbb{P})$ such that

$$\int_B f d\mathbb{P} = \int_B g d\mathbb{P} \quad \text{for all } B \in \mathcal{G}.$$

The function g is called *conditional expectation* of f with respect to $\mathcal{G}$ and is denoted by

$$g = \mathbb{E}[f|\mathcal{G}].$$

When $f = (f_1, f_2, \ldots, f_d)$ takes values in $\mathbb{R}^d$ with $\mathbb{E}[|f|] < \infty$, we write

$$\mathbb{E}[f|\mathcal{G}] = \big(\mathbb{E}[f_1|\mathcal{G}], \mathbb{E}[f_2|\mathcal{G}], \ldots, \mathbb{E}[f_d|\mathcal{G}]\big).$$

The conditional expectation is unique up to null sets from $\mathcal{G}$.

In the following proposition we list some well known properties of the conditional expectation.

Proposition 2.9. *Let f, f_1 and f_2 be integrable $\mathbb{R}^d$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ be σ-algebras.*

(i) If f is $\mathcal{G}$-measurable, then $\mathbb{E}[f|\mathcal{G}] = f$ a.s.
(ii) Linearity: For a, $b \in \mathbb{R}^d$ we have $\mathbb{E}[af_1 + bf_2|\mathcal{G}] = a\mathbb{E}[f_1|\mathcal{G}] + b\mathbb{E}[f_2|\mathcal{G}]$ a.s.
(iii) Convexity: For a convex function g we have $g\big(\mathbb{E}[f|\mathcal{G}]\big) \leq \mathbb{E}\big[g(f)|\mathcal{G}\big]$ a.s.
(iv) Tower property: $\mathbb{E}[f|\mathcal{H}] = \mathbb{E}\big[\mathbb{E}[f|\mathcal{H}]\big|\mathcal{G}\big] = \mathbb{E}\big[\mathbb{E}[f|\mathcal{G}]\big|\mathcal{H}\big]$ a.s.

Proof. (i) Follows from the definition of conditional expectation.
(ii) Follows from

$$\begin{aligned}\int_B \mathbb{E}[af_1 + bf_2|\mathcal{G}]d\mathbb{P} &= \int_B af_1 + bf_2 d\mathbb{P} \\ &= a\int_B f_1 d\mathbb{P} + b\int_B f_2 d\mathbb{P} \\ &= a\int_B \mathbb{E}[f_1|\mathcal{G}]d\mathbb{P} + b\int_B \mathbb{E}[f_2|\mathcal{G}]d\mathbb{P} \quad \text{for all } B \in \mathcal{G}.\end{aligned}$$

(iii) Follows from Jensen's inequality.
(iv) Since $\mathbb{E}[f|\mathcal{H}]$ is $\mathcal{H}$-measurable and therefore $\mathcal{G}$-measurable, the first equality follows from (i). Because $\mathcal{H} \subseteq \mathcal{G}$, the second equality follows from the fact that for all $B \in \mathcal{H}$ we have

$$\int_B \mathbb{E}\big[\mathbb{E}[f|\mathcal{G}]\big|\mathcal{H}\big]d\mathbb{P} = \int_B \mathbb{E}[f|\mathcal{G}]d\mathbb{P} = \int_B f d\mathbb{P} = \int_B \mathbb{E}[f|\mathcal{H}]d\mathbb{P}.$$

□

Definition 2.10. A probability distribution F on $\mathbb{R}^d$ is said to be *infinitely divisible* if for any integer $n \geq 2$, there exist n i.i.d random variables $Z_1, \ldots, Z_n$ such that $Z_1 + \cdots + Z_n$ has distribution F.

2.2. **Stochastic processes.** Stochastic processes are an important tool for analyzing phenomena that are changing randomly over time. They can be defined as a collection $Z = (Z_t)_{t\geq 0}$ of random variables where the parameter t describes the time. The map $t \mapsto Z_t(\omega)$ for fixed ω is called a (sample) path.

Definition 2.11. For any stochastic process $Z = (Z_t)_{t\geq 0}$, we call the filtration $(\mathcal{F}_t^Z)_{t\geq 0}$ given by

$$\mathcal{F}_t^Z = \sigma(Z_s : s \leq t)$$

the *natural filtration* of Z. Here the notation $\sigma(\cdot)$ represents the smallest σ-algebra generated by a collection of random variables.

Definition 2.12. A stochastic process $Z = (Z_t)_{t\geq 0}$ is said to be $(\mathcal{F}_t)_{t\geq 0}$-*adapted* or *adapted*, if each Z_t is $\mathcal{F}_t$-measurable.

Remark 2.1. Note that the natural filtration of the process Z is actually the filtration generated by process Z. It is also the smallest filtration with respect to which the process Z is adapted.

Definition 2.13. The *predictable σ-algebra* is the σ-algebra $\mathcal{P}$ on $\Omega \times [0, \infty)$ that is generated by all left-continuous adapted processes considered as mappings from $\Omega \times [0, \infty)$ to $\mathbb{R}$. A process that is $\mathcal{P}$-measurable is called a *predictable process.*

Remark 2.2. [6, p. 16, Theorem 2.2] A predictable σ-algebra is also generated by a collection of random sets

$$A \times \{0\}, \text{ where } A \in \mathcal{F}_0, \text{ and } A \times (s, t], \text{ where } s < t, A \in \mathcal{F}_s.$$

Definition 2.14. A stochastic process Z is called *progressively measurable (with respect to* $(\mathcal{F}_t)_{t\geq 0}$*)* if the mapping $(s,\omega) \mapsto Z_s(\omega) : ([0,t]\times\Omega, \mathcal{B}([0,t])\otimes\mathcal{F}_t) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is measurable for every $t \geq 0$.

Proposition 2.15. *A predictable process is progressively measurable, and a progressively measurable process is also measurable and adapted. Additionally, an adapted process with all paths left- or right-continuous is progressively measurable.*

Proof. By Remark 2.2 the generating sets of the predictable σ-algebra are of form

$$A\times\{0\}, \text{ where } A\in\mathcal{F}_0, \text{ or } A\times(s,t], \text{ where } s<t, A\in\mathcal{F}_s.$$

It is clear, that each of these sets belongs in $\mathcal{B}\big([0,t]\big)\otimes\mathcal{F}_t$, which implies that predictable processes are progressively measurable. Measurability and adaptedness of a progressively measurable process follow from the definition of progressively measurable process. For the final statement we refer reader to [7, p. 5, Proposition 1.13]. □

Definition 2.16. Let X and Y be metric spaces. A function $f : X \to Y$ is called *cadlag* if it is right-continuous with left limits, i.e. if both left and right hand side limits exist and $f(x_0) = \lim_{x\searrow x_0} f(x)$ for each $x_0 \in X$. The process Z is cadlag if all its paths are cadlag.

Similarly, we define *caglad* functions to be left-continuous with right limits.

Remark 2.3. For any cadlag process Z we let Z_{t-} denote the left limit with $Z_{0-} := Z_0$ and for $t > 0$

$$Z_{t-} := \lim_{s\nearrow t} Z_s.$$

Additionally we let ΔZ_t denote the jump sizes

$$\Delta Z_t := Z_t - \lim_{s\nearrow t} Z_s = Z_t - Z_{t-}.$$

Proposition 2.17. [6, p. 17, Proposition 2.6] *If Z is a cadlag $(\mathcal{F}_t)_{t\geq 0}$-adapted process, then $(Z_{t-})_{t\geq 0}$ is a predictable process.*

Definition 2.18. The *jump times* T of a cadlag process Z are defined by $T_0 := 0$ and

$$T_k := \inf\{t > T_{k-1} : \Delta Z_t \neq 0\}.$$

Definition 2.19. The *jump measure* J_Z of process Z is defined by

$$J_Z(B) = \#\{(t,\Delta Z_t)\in B\}$$

with $B\in\mathcal{B}((0,\infty)\times\mathbb{R}^d)$. For every measurable set $A\subset\mathbb{R}^d$, $J_Z([t_1,t_2]\times A)$ counts the number of jump times of the process Z between t_1 and t_2 such that their jump sizes are in A.

Definition 2.20. Let $E\subset\mathbb{R}^d$. A *Radon measure* on $(E,\mathcal{B})$ is a measure μ such that $\mu(B) < \infty$ for every compact measurable set $B\in\mathcal{B}$.

Definition 2.21. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, $E\subset\mathbb{R}^d$ and μ a given positive Radon measure on $(E,\mathcal{E})$. A *Poisson random measure* on E with intensity measure μ is an integer valued random measure

$$N : \Omega\times\mathcal{E}\to\mathbb{N}, \quad (\omega,A)\mapsto N(\omega,A)$$

such that

(i) the measure $N(\omega, .)$ is an integer valued Radon measure on $(E, \mathcal{E})$ for almost all $\omega \in \Omega$.
(ii) the function $N(., A) = N(A)$ is a Poisson random variable with intensity $\mu(A)$ for each measurable set $A \subset E$, i.e.

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

for all $k \in \mathbb{N}$.
(iii) For disjoint measurable sets $A_1, \ldots, A_n \in \mathcal{E}$, the variables $N(A_1), \ldots, N(A_n)$ are independent.

Definition 2.22. Given the Poisson random measure N the *compensated Poisson random measure* $\tilde{N}$ is given by

$$\tilde{N}(A) = N(A) - \mu(A).$$

Definition 2.23. Let $\tau \geq 0$ be a positive random variable with

$$\{\tau \leq t\} \in \mathcal{F}_t$$

for all $t \geq 0$. Then τ is called a *stopping time* with respect to filtration $(\mathcal{F}_t)_{t \geq 0}$. Moreover,

$$\mathcal{F}_\tau := \big\{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F} \text{ for all } t \in \mathbb{R}\big\}.$$

Definition 2.24. Let $(Z_t)_{t \geq 0}$ be a stochastic process and τ a stopping time. Then the process $(Z_{\tau \wedge t})_{t \geq 0}$ is called a *stopped process*.

Definition 2.25. The process $Z = (Z_t)_{t \geq 0}$ is *square integrable*, if $\mathbb{E}[|Z_t|^2] < \infty$ for all $t \geq 0$.

Definition 2.26. Let $Z = (Z_t)_{t \geq 0}$ be an adapted and integrable stochastic process. Then Z is a
(i) *super-martingale* if $\mathbb{E}[Z_t | \mathcal{F}_s] \leq X_s$ a.s. for all $0 \leq s \leq t$,
(ii) *sub-martingale* if $\mathbb{E}[Z_t | \mathcal{F}_s] \geq X_s$ a.s. for all $0 \leq s \leq t$,
(iii) *martingale* if $\mathbb{E}[Z_t | \mathcal{F}_s] = X_s$ a.s. for all $0 \leq s \leq t$.

Remark 2.4. We will sometimes call a square integrable martingale, i.e. a martingale $(M_t)_{t \geq 0}$ with $M_t \in L^2$ for all $t \geq 0$, an L^2-martingale.

Proposition 2.27. *A martingale M has constant expectation, i.e. $\mathbb{E}[M_t] = \mathbb{E}[M_0]$ for all $t \geq 0$.*

Proof. This follows from the definition of the martingale and the tower property of the conditional expectation. □

Remark 2.5. If τ is a stopping time, then a martingale M stopped at time τ is also a martingale. It is called a stopped martingale.

Next we introduce some well known stochastic processes, which will be used later.

Definition 2.28. Let $W_1, W_2, \ldots$ be independent and exponentially distributed random variables with parameter $\lambda > 0$, $T_0 := 0$ and $T_n := \sum_{i=1}^n W_i$. The process $(N_t)_{t \geq 0}$ defined by

$$N_t = \sum_{n \geq 0} \mathbb{1}_{t \geq T_n}$$

is called *a Poisson process with intensity* λ.

Remark 2.6. [2, p. 55] The sequence $(T_i)_{i=0,1,2,\dots}$ denotes the jump times of the Poisson process $(N_t)_{t\geq 0}$. Thus, the process counts the amount of jumps that occur at the time interval $[0,t]$. We observe that it defines a measure N by setting

$$N(\omega, A) = \#\{i \geq 1, T_i(\omega) \in A\}$$

for any measurable set $A \subset [0,\infty)$. This measure is a Poisson random measure on $[0,\infty)$ with Lebesgue measure as intensity measure. We can represent the Poisson process as

$$N_t(\omega) = \int_0^t N(\omega, ds).$$

Proposition 2.29. [2, p. 48, Proposition 2.12] *Let the process $N = (N_t)_{t\geq 0}$ be a* Poisson process with intensity $\lambda > 0$. *Then*

(i) N_t *has independent increments*
(ii) $\mathbb{P}(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$ *for all* $n \in \mathbb{N}$
(iii) for any $t > s$, $N_t - N_s \overset{d}{=} N_{t-s}$
(iv) the paths $t \mapsto N_t(\omega)$ *are cadlag almost surely.*

Remark 2.7. The filtration generated by the Poisson process $(N_t)_{t\in\mathbb{R}_+}$ is denoted by

$$\mathcal{F}_t^N := \sigma(N_s : s \in [0,t])$$

for $t \geq 0$.

Proposition 2.30. *The compensated Poisson process*

$$(N_t - \lambda t)_{t\geq 0}$$

is a martingale with respect to $(\mathcal{F}_t^N)_{t\geq 0}$.

Proof. We will prove this in a more general case in Proposition 2.35. □

Definition 2.31. Let $N = (N_t)_{t\geq 0}$ be a Poisson process with intensity λ, W_i i.i.d. random variables with distribution f and N be independent from $(W_i)_{i\geq 1}$. Then the process

$$C_t = \sum_{i=1}^{N_t} W_i$$

with the convention that $C_t = 0$ if $N_t = 0$, is called *compound Poisson process with intensity* $\lambda > 0$ *and jump size distribution* f *on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Remark 2.8. The jump size ΔC_t of the compound Poisson process at time t is given by the relation

$$\Delta C_t = W_{N_t} \Delta N_t$$

for $t \geq 0$ i.e. because $\Delta N_t \in \{0,1\}$, the process C can only have jumps when N has jumps, and the size of the jump is W_{N_t}. Clearly jump sizes are independent from jump times.

Proposition 2.32. [2, p. 71] *Let $C = (C_t)_{t\geq 0}$ be a compound Poisson process. Then*

(i) The sample paths of C are piecewise constant functions.
(ii) The jump times $(T_i)_{i=1,2,\dots}$ can be expressed as partial sums of independent exponential random variables with parameter λ.

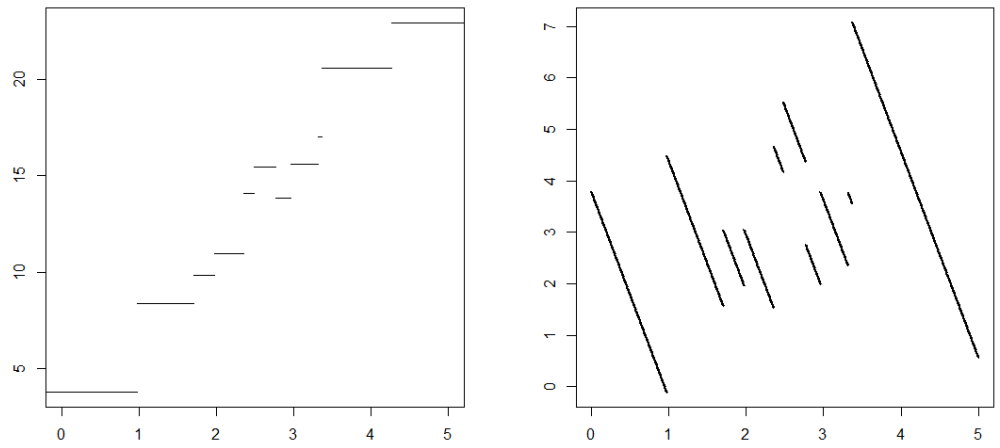

FIGURE 1. Typical sample paths of compound Poisson process and compensated Poisson process.

Remark 2.9. A typical sample path of a compound Poisson process alongside with a compensated compound Poisson process are shown in Figure 1.

Proposition 2.33. *The characteristic function of the increment $C_T - C_t$ for any $t \in [0,T]$ is*

$$\varphi_{C_T - C_t}(s) = \mathbb{E}\Big[\exp\big(is(C_T - C_t)\big)\Big] = \exp\Big(\lambda(T-t)\big(\varphi_{W_1}(s) - 1\big)\Big), \quad s \in \mathbb{R}^d.$$

Proof. By definition of the characteristic function and the compound Poisson process, and with conditioning we have

$$\begin{aligned}
\varphi_{C_T - C_t}(s) &= \mathbb{E}[\exp(is(C_T - C_t))] \\
&= \mathbb{E}[\exp(is(\sum_{k=1}^{N_T} W_k - \sum_{k=1}^{N_t} W_k))] \\
&= \mathbb{E}[\exp(is(\sum_{k=N_t+1}^{N_T} W_k))] \\
&= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathbb{E}\Big[\exp\Big(is(\sum_{k=n+1}^{m+n} W_k)\Big)\Big| N_t = n, N_T - N_t = m\Big] \\
&\qquad \cdot \mathbb{P}(N_t = n, N_T - N_t = m).
\end{aligned}$$

The Poisson process has independent and stationary increments which gives us

$$\begin{aligned}
\mathbb{P}(N_t = n, N_T - N_t = m) &= \mathbb{P}(N_t = n)\mathbb{P}(N_T - N_t = m) \\
&= e^{-\lambda t}\frac{(\lambda t)^n}{n!} e^{-\lambda(T-t)}\frac{\big(\lambda(T-t)\big)^m}{m!} \\
&= e^{-\lambda T}\frac{(\lambda t)^n}{n!}\frac{\big(\lambda(T-t)\big)^m}{m!},
\end{aligned}$$

and with the independence of the Poisson process N and the random variables $W_1, W_2, \dots$ we get that

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big(is(\sum_{k=n+1}^{m+n} W_k)\Big)\Big|N_t = n, N_T - N_t = m\Big] \\
&\quad= \mathbb{E}\Big[\exp\Big(is(\sum_{k=n+1}^{m+n} W_k)\Big)\Big] \\
&\quad= \prod_{k=n+1}^{m+n} \mathbb{E}\big[\exp(isW_k)\big] \\
&\quad= \mathbb{E}\big[\exp(isW_1)\big]^m.
\end{aligned}$$

Combining these with the Taylor expansion of e^x gives us

$$\begin{aligned}
&\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathbb{E}\Big[\exp\Big(is(\sum_{k=n+1}^{m+n} W_k)\Big)\Big|N_t = n, N_T - N_t = m\Big]\cdot\mathbb{P}(N_t = n, N_{T-t} = m) \\
&\quad= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathbb{E}\big[\exp(isW_1)\big]^m e^{-\lambda T}\frac{(\lambda t)^n}{n!}\frac{(\lambda(T-t))^m}{m!} \\
&\quad= e^{-\lambda T}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}\sum_{m=0}^{\infty}\mathbb{E}\big[\exp(isW_1)\big]^m\frac{(\lambda(T-t))^m}{m!} \\
&\quad= \exp(-\lambda T)\exp(\lambda t)\exp\Big(\mathbb{E}\big[e^{isW_1}\big]\lambda(T-t)\Big) \\
&\quad= \exp\Big(\lambda(T-t)\big(\mathbb{E}[e^{isW_1}]-1\big)\Big) \\
&\quad= \exp\Big(\lambda(T-t)\big(\varphi_{W_1}(s)-1\big)\Big).
\end{aligned}$$

□

Corollary 2.34. *By setting $t = 0$ we get the characteristic function of the process C at time T to be $\varphi_{C_T}(s) = \mathbb{E}\Big[\exp\big(is(C_T)\big)\Big] = \exp\big(\lambda T(\varphi_{W_1}(s)-1)\big)$.*

The Poisson process is actually a special case of the compound Poisson process: for a Poisson process the jump height distribution is $f = \delta_1$. Thus, the characteristic function of the increment of the Poisson process N is

$$\varphi_{N_T - N_t}(s) = \exp\big(\lambda(T-t)(\exp(is)-1)\big).$$

With the characteristic function it is easy to compute the expectation of the compound Poisson process $(C_t)_{t\geq 0}$ for fixed t as $\mathbb{E}[C_t] = \lambda t\mathbb{E}[W_1]$ and the variance as $\mathrm{Var}[C_t] = \lambda t\mathbb{E}[|W_1|^2] = \mathbb{E}[N_t]\mathbb{E}[|W_1|^2]$.

Proposition 2.35. *The compensated compound Poisson process*

$$\tilde{C}_t := C_t - \lambda t\mathbb{E}[W_1]$$

for $t \geq 0$ is a martingale with respect to the filtration $(\mathcal{F}_t^C)_{t\geq 0}$.

Proof. Let $t \geq s$. Note that the random variable $\tilde{C}_s$ is measurable with respect to $\mathcal{F}_s^C$ and that the increment $\tilde{C}_t - \tilde{C}_s$ is independent from $\mathcal{F}_s^C$. Thus we have that

$$\begin{aligned}\mathbb{E}[\tilde{C}_t|\mathcal{F}_s^C] = \mathbb{E}[\tilde{C}_t + \tilde{C}_s - \tilde{C}_s|\mathcal{F}_s^C] &= \mathbb{E}[\tilde{C}_t - \tilde{C}_s] + \tilde{C}_s \\ &= \mathbb{E}[C_t - C_s] - \lambda\mathbb{E}[W_1](t-s) + \tilde{C}_s \\ &= \tilde{C}_s\end{aligned}$$

□

Definition 2.36. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis and $B = (B_t)_{t\geq 0}$ a stochastic process with $B_t : \Omega \to \mathbb{R}^d$ for $d = 1, 2, \ldots$. A process B is called a *d-dimensional standard Brownian motion (with respect to filtration $(\mathcal{F}_t)_{t\geq 0}$)* provided that

(i) $B_0 \equiv 0$ almost surely,
(ii) B has almost surely continuous paths,
(iii) B is adapted,
(iv) for all $0 \leq s \leq t < \infty$ the increment $B_t - B_s$ is independent from $\mathcal{F}_s$ and
(v) the increment $B_t - B_s$ is $\mathcal{N}(0, (t-s)I_d)$ distributed

where I_d is the d-dimensional identity matrix.

Definition 2.37. [1, p. 48-49] Let $B = (B_t)_{t\geq 0}$ be a standard r-dimensional Brownian motion, $\mu \in \mathbb{R}^d$ and σ a $d \times r$-matrix. Then the process $X = (X_t)_{t\geq 0}$, where $X_t = \mu t + \sigma B_t$ is called *Brownian motion with a drift parameter μ and covariance matrix $\sigma\sigma*$.*

Proposition 2.38. [5, p. 179] *Let $M = (M_t)_{t\geq 0}$ be a right-continuous square integrable $\mathcal{F}$-martingale. Then there exists a right-continuous increasing process $[M] = ([M]_t)_{t\geq 0}$ such that for each $t \geq 0$ and each sequence of partitions $p_n = (0 = t_0^n \leq t_1^n \leq \cdots \leq t_0^{m_n} = T)$ of $[0, T]$ with $\Delta(p_n) \to 0$ as $n \to \infty$, we have*

$$S_n = \sum_{k=1}^{m_n} \left(M_{t_k^{(n)}} - M_{t_{k-1}^{(n)}} \right)^2 \to [M]_t \text{ in } L_1 \text{ as } n \to \infty.$$

Definition 2.39. The process $[M]$ of proposition (2.38) is called *quadratic variation* of the martingale M.

Example 2.40. The quadratic variation of the standard Brownian motion B is

$$[B]_t = t.$$

It can also be proven, that given an adapted and continuous process $Z = (Z_t)_{t\geq 0}$ such that $Z_0 = 0$ with $[Z]_t = t$, then the process Z is actually a Brownian motion. This is known as the Lévy's characterization of the Brownian motion.

Proposition 2.41. [5, p. 180] *If M is a square integrable, right-continuous $\mathcal{F}$-martingale, then*

$$M^2 - [M]$$

is a martingale.

2.3. **Lévy processes.** Lévy processes are the time-continuous counterpart of random walks from the discrete world. The compound Poisson process and the Brownian motion considered before are both core examples of Lévy processes, and they can also be seen as building blocks of more complicated ones. A typical sample path of a Lévy process is shown in Figure 2.

Definition 2.42. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis and let $X = (X_t)_{t\geq 0}$ be cadlag with values in $\mathbb{R}^d$. The process X is called *Lévy process* if and only if

(i) $X_0 = 0$ almost surely,
(ii) X is adapted,
(iii) for all $0 \leq s \leq t < \infty$ the increment $X_t - X_s$ is independent from $\mathcal{F}_s$ and
(iv) for all $0 \leq s < \infty$ and $0 < t < \infty$ the distribution of $X_{s+t} - X_s$ does not depend on s.
(v) for all $\varepsilon > 0$, $\lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \varepsilon) = 0$ (stochastic continuity).

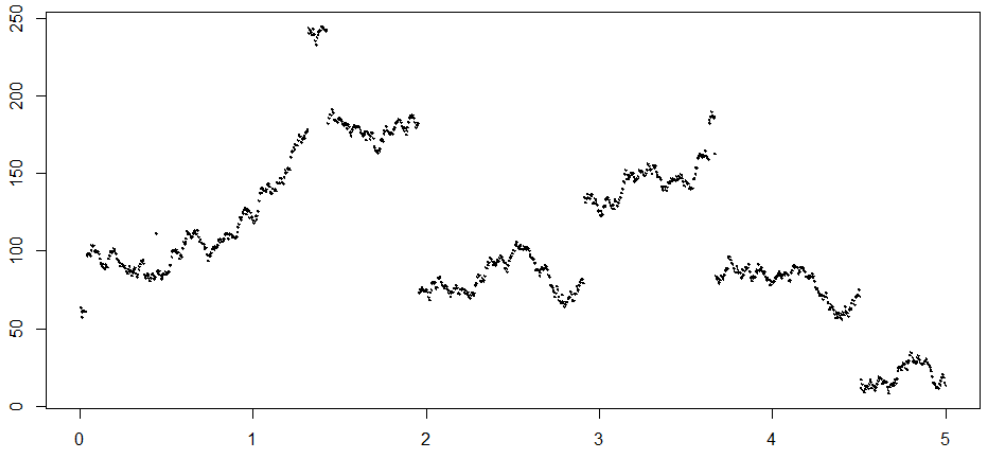

FIGURE 2. A typical sample path of a Lévy process.

Proposition 2.43. [2, p. 69, Proposition 3.1] *Let $X = (X_t)_{t\geq 0}$ be a Lévy process. Then for every t, X_t has an infinitely divisible distribution. Conversely, if F is an infinitely divisible distribution then there exists a Lévy process X such that the distribution of X_1 is given by F.*

Definition 2.44. Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$. The measure ν on $\mathbb{R}^d$ defined by

$$\nu(A) = \mathbb{E}[\#\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\}], \ A \in \mathcal{B}(\mathbb{R}^d)$$

is called the *Lévy measure* of X, i.e. it describes the expected number of times between 0 and 1 the process jumps such that the jump size belongs to the set A. Note that the Lévy measure is in general not a probability measure.

Example 2.45. The Lévy measure ν of a compound Poisson process $C = (C_t)_{t\geq 0}$ is given by $\nu(A) = \lambda f(A)$, where f is the jump size distribution of C. It is also easy to see that due to continuity of the paths the standard Brownian motion has Lévy measure $\nu(A) = 0$ for any $A \in \mathcal{B}(\mathbb{R}^d)$.

Proposition 2.46. [2, p. 75, Proposition 3.5] *Let $(C_t)_{t\geq 0}$ be a compound Poisson process with intensity λ and a jump size distribution f. Its jump measure J_C is a Poisson random measure on $\mathbb{R}^d \times [0,\infty)$ with intensity measure $\mu(dx \times dt) = \nu(dx)dt = \lambda f(dx)dt$.*

Proposition 2.47. [2, p. 71, Proposition 3.3] *A process X is a compound Poisson process if and only if it is a Lévy process and its sample paths are piecewise continuous functions.*

All information needed to determine the distribution of the Lévy process is uniquely described by the triplet $(\nu, \sigma\sigma*, b)$, where ν is the Lévy measure of the process, $\sigma\sigma*$ is a $d \times d$ matrix and b is a d-dimensional vector. It is called the characteristic triplet or the Lévy triplet of the Lévy process X.

THEOREM 2.48 (Lévy-Itô decomposition)**.** [1, p. 126, Theorem 2.4.16] *If X is a Lévy process, then there exists $b \in \mathbb{R}^d$, a Brownian motion σB with covariance matrix $\sigma\sigma*$ and an independent Poisson random measure N on $[0,\infty) \times (\mathbb{R}^d \setminus \{0\})$ with intensity measure $\nu(dx)dt$ such that*

$$X_t = bt + \sigma B_t + \int_{B_1} x\tilde{N}(t,dx) + \int_{B_1^c} xN(t,dx)$$

for each $t \geq 0$.

The Lévy-Itô decomposition shows us that a Lévy process X can be decomposed to a rather simple form. The decomposition consists of three components, first of which is a Brownian motion with a drift and the latter ones describe the jump parts of the process divided into small and large ones respectively. Small jumps can be tricky because there can be infinitely many of them possibly preventing their sum from converging. Hence, we use the compensated sum of small jumps, i.e. the process $(\int_{|x|<1} x\tilde{N}(t,dx))_{t\geq 0}$ which is an L^2-martingale. The process $(\int_{|x|\geq 1} xN(t,dx))_{t\geq 0}$ describing the large jumps is a coumpound Poisson process (see e.g. [1, p. 127], [2, p. 80]). If we make some further assumptions on the Lévy measure we can achieve an even simpler form.

Remark 2.10. If the measure ν is finite and $\int_{\mathbb{R}^d} |x|^2\nu(dx) < \infty$, then we can write the process X as

$$X_t = \left(b + \int_{B_1^c} x\nu(dx)\right)t + \sigma B_t + \int_{\mathbb{R}^d} x\tilde{N}(t,dx)$$

and it follows from Proposition 3.7 that the process $(L_t)_{t\geq 0}$ with $L_t = \int_{\mathbb{R}^d} x\tilde{N}(t,dx)$ is a square integrable martingale and

$$\mathbb{E}[L_t^2] = t\int_{\mathbb{R}^d} x^2\nu(dx).$$

Remark 2.11. Integration with respect to Poisson random measures is considered in section 3.3 more specifically.

THEOREM 2.49 (Lévy-Khinchin)**.** *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic triplet $(\nu, \sigma\sigma*, b)$. Then*

$$\mathbb{E}[e^{iz.X_t}] = e^{t\psi(z)}, \quad z \in \mathbb{R}^d$$

with

$$\psi(z) = -\frac{1}{2}z.(\sigma\sigma*)z + ib.z + \int_{\mathbb{R}^d}(e^{iz.x} - 1 - iz.x\mathbb{1}_{|x|\leq 1})\nu(dx).$$

Example 2.50. From the Lévy-Itô decomposition it is easy to see that the Lévy triplet of the standard Brownian motion B is $(0, I_d, 0)$ and hence its characteristic function is

$$\mathbb{E}\big[e^{iz.B_t}\big] = \exp\bigg(-\frac{1}{2}t\sum_{k=0}^{d} z_k^2\bigg).$$

3. Stochastic integrals and SDEs

In this section we will finish the background theory with the basics about stochastic integration and the SDEs. We assume that the reader is familiar with the fundamental measure theory. We begin by making some assumptions and continue with introducing the essentials of stochastic integrals with respect to Brownian motion. We will then extend the idea of stochastic integrals to cover integrals with respect to jump processes, more precisely to the compound Poisson processes, and sums of a compound Poisson process and a Brownian motion. Lastly, we present the stochastic differential equations to the extent of what is needed in this study. For more detailed theory we refer the reader to [7] for Brownian motion and [1] for more general cases.

We fix the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and from now on we will, unless said otherwise, assume that our stochastic basis satisfies the usual conditions, meaning that

(i) any set contained in a set of measure zero has also measure zero, or in other words, the space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete,
(ii) $\mathcal{F}_t = \bigcap_{s>t}\mathcal{F}_s$, or in other words, the filtration is right-continuous.

We note here that completing an incomplete space is straight-forward and so is enlarging a filtration to be a right-continuous one when the filtration is generated by a strong Markov process. It means that any stochastic basis can be transferred to satisfy the usual conditions without too many changes.

Before we begin with stochastic integration, we present two famous and useful inequalities, Gronwall's lemma and Doob's martingale inequalities.

Proposition 3.1 (Gronwall)**.** [1, p. 358, Proposition 6.1.4] *Let $f, g : [0, T] \to [0, \infty)$ with f locally bounded and g integrable. If there exists $C \geq 0$ such that, for all $t \in [0, T]$,*

$$f(t) \leq C + \int_0^t f(s)g(s)ds,$$

then we have

$$f(t) \leq C\exp\bigg(\int_0^t g(s)ds\bigg).$$

Lemma 3.2 (Doob)**.** [1, p. 87, Theorem 2.1.6; p. 86, Theorem 2.1.5] *Let $M = (M_t)_{t\geq 0}$ be a right-continuous martingale or a right-continuous positive sub-martingale. Let $M_t^* := \sup_{s\in[0,t]}|M_s|$. Then, for $\lambda, t \geq 0$ and $p \in (1, \infty)$,*

$$\lambda\mathbb{P}(M_t^* \geq \lambda) \leq \mathbb{E}[|M_t|]$$

and

$$\mathbb{E}[(M_t^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|M_t|^p].$$

3.1. **Stochastic integrals with respect to Brownian motion.** [7, p. 132-139] We begin by briefly reviewing a stochastic integral of simple processes with respect to Brownian motion. A simple process is a pathwise constant process with some additional properties. More precisely, we have a simple process K if for some partition $0 = t_0 < t_1 < \dots$ with $\lim_{n\to\infty} t_n = \infty$ and random variables $\xi_0, \xi_1, \dots$ with ξ_i being $\mathcal{F}_{t_i}$-measurable for all i with $\sup_{n\geq 0} |\xi_n(\omega)| < \infty$, for every ω, such that

$$K_t = \xi_0 \mathbb{1}_{t=0} + \sum_{i=1}^{\infty} \xi_{i-1} \mathbb{1}_{(t_{i-1}, t_i]}(t)$$

where $0 \leq t < \infty$. The stochastic integral of a simple process K with respect to the Brownian motion B is the process $(I_t(K))_{t\geq 0}$ with

$$I_t^B(K) = \sum_{i=1}^{\infty} \xi_{i-1}(B_{t_i \wedge t} - B_{t_{i-1} \wedge t}).$$

This integral can be generalizated from simple processes to progressively measurable processes with $\mathbb{E}\big[\int_0^T K_t^2 dt\big] < \infty$. It can be shown (see e.q. [7, p. 137, Proposition 2.8]) that for a progressively measurable process $(K_t)_{t\geq 0}$ there exists a sequence $((K_t)_{t\geq 0}^{(n)})_{n=0}^{\infty}$ of simple processes such that

$$\lim_{n\to\infty} \mathbb{E}\big[\int |K_t - K_t^{(n)}|^2 dt\big] = 0.$$

In other words, the space of simple processes is included in the space of progressively measurable processes, and the inclusion is dense. Now the stochastic integral of the process K with respect to the Brownian motion is the unique, square-integrable martingale $(J_t^B(K))_{t\geq 0}$ for which

$$\lim_{n\to\infty} \mathbb{E}\big[J_t^B(K) - I_t^B(K^{(n)})\big] = 0.$$

We write

$$\int_0^t K_s dB_s := J_t^B(K)$$

and additionally

$$\int_s^t K_s dB_s := J_t^B(K) - J_s^B(K)$$

for $\infty > t \geq s \geq 0$.

Since later we are interested in predictable processes in particular, we formulate the first isometry result for predictable processes. Recall that any predictable process is also progressively measurable and hence the definition of the integral with respect to Brownian motion we formulated holds for predictable processes.

Proposition 3.3 (isometry)**.** [2, p. 259, Proposition 8.6] *Let ϕ be a predictable process verifying*

$$\mathbb{E}\bigg[\int_0^t |\phi_s|^2 ds\bigg] < \infty.$$

Then $\int_0^t \phi_s dB_s$ *is is a square integrable martingale with* $\mathbb{E}\big[\int_0^t \phi_s dB_s\big] = 0$ *and*

$$\mathbb{E}\bigg[\bigg|\int_0^t \phi_s dB_s\bigg|^2\bigg] = \mathbb{E}\bigg[\int_0^t |\phi_s|^2 dt\bigg].$$

3.2. **Stochastic integrals with jumps.** [2, p. 249] The notion of stochastic integrals can be extended to various situations. In this study, we are interested in processes with jumps. For this setting, we need to define simple predictable processes.

Definition 3.4. A stochastic process $K = (K_t)_{t>=0}$ is called a *simple predictable process* if it can be represented as

$$K_t = K_0 \mathbb{1}_{\{t=0\}} + \sum_{i=0}^{n} K_i \mathbb{1}_{(T_i, T_{i+1}]}(t),$$

where $0 = T_0 \leq \ldots \leq T_{n+1} < \infty$ is a finite sequence of stopping times and K_i is $\mathcal{F}_{T_i}$-measurable with $|H_i| < \infty$ a.s.

In Proposition 2.35 we saw that the compensated compound Poisson process is actually a martingale. Similarly to the case with Brownian motion we now use the simple predictable processes as integrands and instead of Brownian motion we consider integration with respect to a martingale M, namely the process $\big(I_t^M(K)\big)_{t\geq 0}$ with

$$I_t^M(K) = K_0 M_0 + \sum_{i=0}^{n} K_i (M_{T_{i+1}\wedge t} - M_{T_i \wedge t}).$$

It can be shown that with M being a martingale and K a simple predictable process the integral process $\big(I_t^M(K)\big)_{t\geq 0}$ is also a martingale [2, p. 251, Proposition 8.1]. In many applications simple processes are not enough but we want to study also more complex integrands. However, it is essential for us to conserve the martingale preservation property. This leads us to caglad integrands.

Proposition 3.5. [2, p. 256, Proposition 8.4] *Let* M *be a martingale,* ϕ *a caglad process and* $p^n = (T_0^n = 0 < T_1^n < \cdots < T_{n+1}^n = T)$ *a sequence of random partitions of* $[0, T]$ *such that* $|p^n| = \sup_k |T_k^n - T_{k-1}^n| \to 0$ *almost surely when* $n \to \infty$*. Then*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \phi_{T_k^n}(M_{T_{k+1}^n \wedge t} - M_{T_k^n \wedge t}) = J_t^M(\phi) \quad \textit{in probability}$$

uniformly in t *on* $[0, T]$.

The martingale-preserving property can now be formulated as follows.

Proposition 3.6. [9, p. 56, Theorem 20] *Let the process* $M = (M_t)_{t\geq 0}$ *be a square integrable martingale and* $K = (K_t)_{t\geq 0}$ *an adapted, caglad and bounded process. Then the stochastic integral of* K *with respect to* M*, i.e. the process* $\big(I_t^M(K)\big)_{t\geq 0}$ *is also a square integrable martingale.*

Like in Brownian motion case, we write

$$\int_0^t K_s dM_s := J_t^M(K)$$

and additionally

$$\int_s^t K_s dM_s := J_t^M(K) - J_s^M(K)$$

for $\infty > t \geq s \geq 0$.

3.3. **Stochastic integrals with respect to Poisson random measures.** [2, p. 259-263][1, p. 214] In this case, we consider simple predictable functions instead of processes and then by taking limits we extend to more general ones in analogy to previous cases. Let N be a Poisson random measure on $[0,T] \times \mathbb{R}^d$ with intensity $\mu(dt, dx)$. We say that the function $K : \Omega \times [0,T] \times \mathbb{R}^d \to \mathbb{R}$ is simple and predictable if, for some $m, n \in \mathbb{N}$ there exist adapted random times $(T_i)_{i=0}^{n+1}$ such that $0 = T_0 \leq T_1 \leq \ldots \leq T_{n+1} = T$, bounded $\mathcal{F}_{T_i}$-measurable random variables $(K_{ij})_{j=1,\ldots,m}$ and disjoint subsets $(A_i)_{i=1}^n$ of $\mathbb{R}^d$ with $\mu\big([0,T] \times A_i\big) < \infty$ such that

$$K(t,y) = \sum_{i=1}^{n} \sum_{j=1}^{m} K_{ij} \mathbb{1}_{(T_i, T_{i+1}]}(t) \mathbb{1}_{A_j}(y). \tag{3.1}$$

The stochastic integral of K with respect to Poisson random measure N is defined as

$$\int_0^T \int_{\mathbb{R}^d} K(t,y) N(dt,dy) = \sum_{i=1}^{n} \sum_{j=1}^{m} K_{ij} \int_{(T_i, T_{i+1}] \times A_j} N(dt,dy)$$

and similarly the corresponding compensated integral is

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} K(t,y) \tilde{N}(dt,dy) &= \sum_{i=1}^{n} \sum_{j=1}^{m} K_{ij} \int_{(T_i, T_{i+1}] \times A_j} \tilde{N}(dt,dy) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} K_{ij} \bigg(\int_{(T_i, T_{i+1}] \times A_j} N(dt,dy) - \mu\big((T_i, T_{i+1}] \times A_j\big) \bigg). \end{aligned}$$

The space of simple predictable functions is dense in the space of predictable and square integrable functions (see [1, p. 218, Lemma 4.1.4]). Hence given a predictable random function K with

$$\mathbb{E}\bigg[\int_0^T \int_{\mathbb{R}^d} |K(t,y)|^2 \mu(dt,dy) \bigg] < \infty$$

there exists a sequence $\big(K^{(n)}\big)_{n=0}^{\infty}$ of simple predictable functions such that

$$\lim_{n \to \infty} \mathbb{E}\bigg[\int_0^T \int_{\mathbb{R}^d} |K^{(n)}(t,y) - K(t,y)|^2 \mu(dt,dy) \bigg] = 0.$$

This results in the following proposition about the compensated Poisson integral and the isometry formula.

Proposition 3.7. [2, p. 262, Proposition 8.8] *Let $K : \Omega \times [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be a predictable random function with*

$$\mathbb{E}\bigg[\int_0^T \int_{\mathbb{R}^d} |K(t,y)|^2 \mu(dt,dy) \bigg] < \infty.$$

Then the process $\left(\int_0^t \int_{\mathbb{R}^d} |K(t,y)|^2 \tilde{N}(dt,dy)\right)_{t\geq 0}$ *is a square integrable martingale and*

$$\mathbb{E}\left[\left|\int_0^T \int_{\mathbb{R}^d} K(t,y)\tilde{N}(dt,dy)\right|^2\right] = \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} |K(t,y)|^2 \mu(dt,dy)\right].$$

3.4. **Stochastic differential equations.** We begin with formulating the notion of stochastic differential equations in two cases, with respect to Brownian motion and with respect to a Lévy process.

Definition 3.8. Let $y_0 \in \mathbb{R}^d$, B be an r-dimensional Brownian motion and let functions $\mu : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^r$ be continuous. Then a continuous and adapted stochastic process $Y = (Y_t)_{t\geq 0}$ with values in $\mathbb{R}^d$ is a (strong) solution of the *stochastic differential equation* with respect to Brownian motion

$$dY_t = \sigma(t, Y_t)dB_t + \mu(t, Y_t)dt \quad \text{with} \quad Y_0 = y_0 \tag{3.2}$$

if

(i) $Y_0 = y_0$,
(ii) $Y_t = y_0 + \int_0^t \sigma(s, X_s)dB_s + \int_0^t \mu(s, X_s)ds$ for $t \geq 0$ a.s.

Remark 3.1. The function μ of the stochastic differential equation is called *drift coefficient* and the function σ *diffusion coefficient.*

In the following $|\cdot|$ will denote the Euclidean norm in $\mathbb{R}^d$, for $d \in \mathbb{N}$.

THEOREM 3.9 (Existence and uniqueness)**.** [7, p. 289, Theorem 2.9] *Let* y_0*,* B*,* σ *and* μ *be as stated in Definition 3.8 with* σ *and* μ *satisfying*

$$|\sigma(t,x) - \sigma(t,y)| + |\mu(t,x) - \mu(t,y)| \leq K|x-y|$$

and

$$|\sigma(t,x)|^2 + |\mu(t,x)|^2 \leq K^2(1+|x|^2)$$

for every $0 \leq t < \infty$*,* $x \in \mathbb{R}^d$*,* $y \in \mathbb{R}^d$ *with a positive constant* K*. Then there exists a continuous, adapted process* $Y = (Y_t)_{t\geq 0}$ *which is a (strong) solution of equation* (3.2)*. Moreover, this process is square integrable, i.e. for every* $T > 0$ *there exists a constant* C*, depending only on* K *and* T *such that*

$$\mathbb{E}[|Y_t|^2] \leq C(1+y_0^2)e^{Ct}$$

for $0 \leq t \leq T$*, and the solution is unique up to indistinguishability.*

When thinking about financial applications, the main problems concerning Brownian motion as driving process are the continuity and the scale invariance of the sample paths of the Brownian motion. Many real life events like share prices we want to simulate are not continuous at all. To be able to provide more accurate and realistic models for these phenomena we need to formulate also the concept of the stochastic differential equation with respect to Lévy processes.

Definition 3.10. [1, p. 377] Let $y_0 \in \mathbb{R}^r$ and $X = (X_t)_{t\geq 0}$ be a Lévy process with values in $\mathbb{R}^d$. Recall that by Theorem 2.48 (Lévy-Itô decomposition) X_t can be written as

$$X_t = bt + \sigma B_t + \int_{B_1} x\tilde{N}(t,dx) + \int_{B_1^c} xN(t,dx)$$

for $t \geq 0$, $b \in \mathbb{R}^d$ and a $d \times d$ matrix σ. Let also $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^r$ be a function. Then a cadlag and adapted $\mathbb{R}^r$-valued process $Y = (Y_t)_{t \geq 0}$ is a solution of the stochastic differential equation with respect to the Lévy process

$$dY_t = a(Y_{t-})dX_t \quad \text{with } Y_0 = y_0 \tag{3.3}$$

provided that

(i) $Y_0 = y_0$ and
(ii) $Y_t = y_0 + b\int_0^t a(Y_{s-})ds + \sigma\int_0^t a(Y_{s-})dB_s$
$+ \int_{B_1} a(Y_{s-})x\tilde{N}(t, dx)$
$+ \int_{B_1^c} a(Y_{s-})xN(t, dx)$.

THEOREM 3.11 (Existence and uniqueness)**.** [1, p. 377-378] *If there exists a positive constant K such that*

$$|a(y) - a(y')| \leq K|y - y'|$$

for all $y, y' \in \mathbb{R}^d$, then there exists a cadlag and adapted process $Y = (Y_t)_{t \geq 0}$ which is a unique solution to the equation (3.3)*.*

4. The multilevel Monte Carlo algorithm

In this section we consider an SDE with respect to a Brownian motion given in Definition 3.8. Our primary object of interest is the expected value of $g(Y_T)$, where $Y = (Y_t)_{t \in [0,T]}$ is a solution to given SDE and g is a Lipschitz payoff function with a Lipschitz constant C_g. In general we are not able to compute this analytically, not even when the existence and uniqueness of the solution are known. In these cases one can try to find a good enough approximate solution by using a discrete time net. With a finer time grid we achieve a smaller error but with a high computational cost and, on the other hand, by using a coarser grid the cost would be less but the error would grow. Using different time grids on different levels of evaluation is one solution to find the balance between these two options.

We begin by formulating a simple numerical method which is used to approximate the solution of the given SDE. This method is known as the Euler-Maruyama method or the Euler scheme.

Definition 4.1 (Euler-Maruyama)**.** [8, p. 305-307, 340-341] Let $Y = (Y_t)_{t \geq 0}$ be a solution to a Brownian motion driven stochastic differential equation (3.2) and let p denote a partition of m timepoints on $[0, T]$ with $0 = t_0 < \ldots < t_m = T$. We write $h = \max_n(t_n - t_{n-1})$ and call it the maximum timestep. We define the approximation of process Y as $\widehat{Y}_0^{(h)} = y_0$ and for $k = 0, \ldots, m-1$,

$$\widehat{Y}_{t_{k+1}}^{(h)} = \widehat{Y}_{t_k}^{(h)} + \mu(t_k, \widehat{Y}_{t_k}^{(h)})(t_{k+1} - t_k) + \sigma(t_k, \widehat{Y}_{t_k}^{(h)})(B_{t_{k+1}} - B_{t_k})$$

with $\widehat{Y}_t^{(h)} = \widehat{Y}_{t_k}^{(h)}$ for all $t \in [t_k, t_{k-1})$. For simplicity, we use an equidistant time discretization and set $t_k = kT/m$ for all $k > 0$, unless said otherwise. In the equidistant case every timestep is T/m and our approximation is $\widehat{Y}_0^{(h)} = y_0$ and for $k = 0, \ldots, n$

$$\widehat{Y}_{t_{k+1}}^{(h)} = \widehat{Y}_{t_k}^{(h)} + \mu(t_k, \widehat{Y}_{t_k}^{(h)})h + \sigma(t_k, \widehat{Y}_{t_k}^{(h)})(B_{t_{k+1}} - B_{t_k}).$$

Proposition 4.2. [8, p. 342, Theorem 10.2.2] *Let $Y = (Y_t)_{t\geq 0}$ be a solution to a Brownian motion driven stochastic differential equation (3.2), and let $\widehat{Y}^{(h)}$ be its Euler approximation with maximum timestep h. Assume that the conditions for the existence and uniqueness stated in Theorem 3.9 hold. Additionally, we assume that with some constant $K > 0$, which does not depend on h, we have that*

$$|\sigma(s,x) - \sigma(t,x)| + |\mu(s,x) - \mu(t,x)| \leq K\big(1 + |x|\big)|s-t|^{1/2} \tag{4.1}$$

for all $s,t \in [0,T]$ and $x, y \in \mathbb{R}^d$. Then for $\widehat{Y}^{(h)}$ the estimate

$$\mathbb{E}[|Y_T - \widehat{Y}_T^{(h)}|] \leq \tilde{K} h^{1/2}$$

holds, where the constant $\tilde{K} > 0$ does not depend on h.

Remark 4.1. From the proof of [8, p. 342, Theorem 10.2.2] we see that also the estimate

$$\mathbb{E}[|Y_T - \widehat{Y}_T^{(h)}|^2] \leq \tilde{K}^2 h$$

holds.

Definition 4.3. The *mean squared error MSE* of an estimator $\widehat{\theta}$ with parameter θ is

$$\mathrm{MSE}(\widehat{\theta}) = \mathbb{E}[(\widehat{\theta} - \theta)^2],$$

and it describes the systematic error of the estimator. We have

$$\mathrm{MSE}(\widehat{\theta}) = \mathbb{E}[(\widehat{\theta} - \theta)^2] = \mathrm{Var}(\widehat{\theta}) + \mathbb{E}[\widehat{\theta} - \theta]^2 = \mathrm{Var}(\widehat{\theta}) + \mathrm{bias}(\widehat{\theta}, \theta)^2,$$

and hence for finding the upper bound for MSE of the estimator $\widehat{S}$ it is enough to find an upper bound for its variance and bias.

The law of large numbers gives us the elementary and the most common way to estimate the expected value of a random variable, that is, by taking the average of N independent samples. If so, our estimate of $\mathbb{E}\big[g(Y_T)\big]$ would be

$$\widehat{Y} = \frac{1}{N} \sum_{i=1}^{N} g\big(\widehat{Y}_T^{(h)}(i)\big),$$

where $\widehat{Y}_T^{(h)}(i)$ is the ith independent copy of the Euler approximation of the process Y with the maximum timestep h at time T. The estimate $\widehat{Y}$ is the standard Monte Carlo estimate. The mean squared error of $\widehat{Y}$ is $O(1/N) + O(h^2)$ (see [4]).

In the multilevel approach we exploit the sequence of equidistant partitions $(p_l)_{l=0,\dots,L}$ on $[0,T]$ such that partition p_l has M^l timesteps, each of which is of length $M^{-l}T =: h_l$ for integer $M \geq 2$. This construction results in a geometric sequence of refining partitions where each sequence is M times finer than the previous one.

We consider the telescoping sum

$$\mathbb{E}\big[g(\widehat{Y}_T^{(h_L)})\big] = \mathbb{E}\big[g(\widehat{Y}_T^{(h_0)})\big] + \sum_{l=0}^{L} \mathbb{E}\big[g(\widehat{Y}_T^{(h_l)}) - g(\widehat{Y}_T^{(h_{l-1})})\big].$$

Each of these expectations on the right hand side of the equation are then estimated separately and on each level l we use N_l samples. More precisely, we denote the

estimator of $\mathbb{E}\big[g(\widehat{Y}_T^{(h_0)})\big]$ with $\widehat{S}_0$ and the estimator of $\mathbb{E}[g(\widehat{Y}_T^{(h_l)}) - g(\widehat{Y}_T^{(h_{l-1})})]$ with $\widehat{S}_l$ and use the average of N_l independent samples as estimator. That is,

$$\widehat{S}_0 = \frac{1}{N_0}\sum_{i=1}^{N_0} g\big(\widehat{Y}_T^{(h_0)}(i)\big)$$

and

$$\widehat{S}_l = \frac{1}{N_l}\sum_{i=1}^{N_l}\Big(g\big(\widehat{Y}_T^{(h_l)}(i)\big) - g\big(\widehat{Y}_T^{(h_{l-1})}(i)\big)\Big)$$

for $l = 1, 2, \ldots, L$. The estimator of $\mathbb{E}\big[g(Y_T)\big]$ is then $\widehat{S} := \sum_{l=0}^{L}\widehat{S}_l$. It is important that when we approximate $\widehat{S}_l$ we use the same Brownian motion for both $\widehat{Y}_T^{(h_l)}(i)$ and $\widehat{Y}_T^{(h_{l-1})}(i)$. That reduces variance and allows us to use fewer samples on finer levels. In practice, we would simulate the Brownian increments needed to approximate $\widehat{S}_l$, and then sum them up in groups of size M to obtain increments for $\widehat{S}_{l-1}$. In Figure 3 we see sample paths of a single Brownian motion approximated with $2^m + 1$ time points, $m = 1, 2, \ldots, 8$. Note that while the quantities $\widehat{Y}_T^{(h_l)}(i)$ and $\widehat{Y}_T^{(h_{l-1})}(i)$ inside $\widehat{S}_l$ depend on each others, the estimators $\widehat{S}_l$ are independent for all $l \geq 0$.

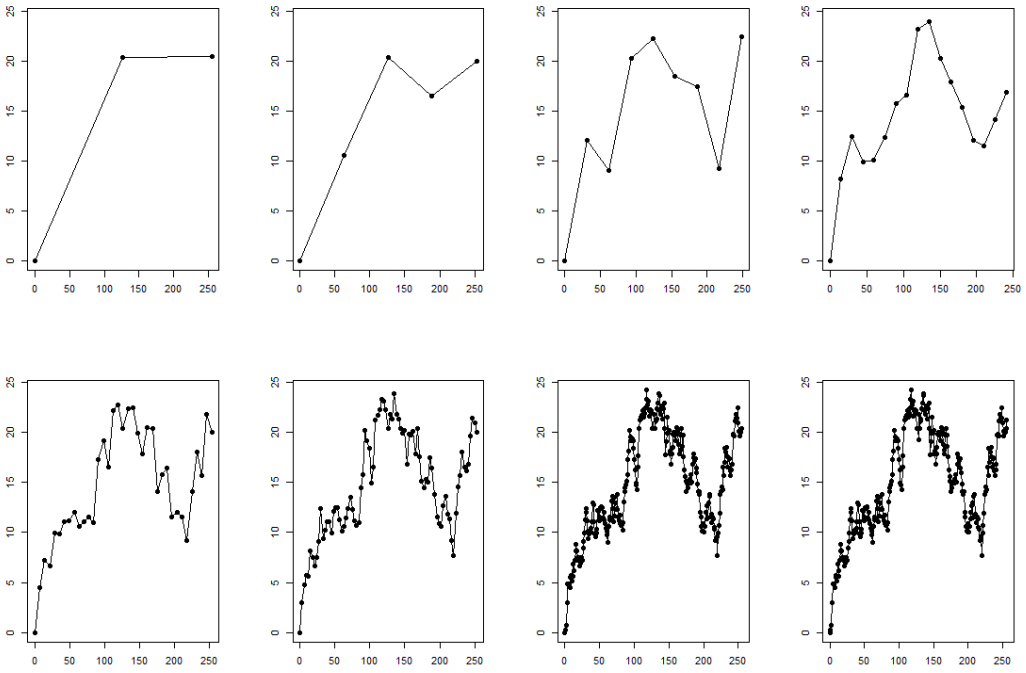

FIGURE 3. Approximated sample paths of the same Brownian motion with each time grid twice as fine as the previous one.

With simple computations, we see that

$$\mathbb{E}\big[\widehat{S}\big] = \mathbb{E}\Big[\sum_{l=0}^{L}\widehat{S}_l\Big] = \mathbb{E}\big[g(\widehat{Y}_T^{(h_0)})\big] + \sum_{l=1}^{L}\mathbb{E}\big[g(\widehat{Y}_T^{(h_l)}) - g(\widehat{Y}_T^{(h_{l-1})})\big] = \mathbb{E}\big[\widehat{S}_L\big]$$

and

$$\mathrm{Var}\left[\widehat{S}\right] = \sum_{l=0}^{L} \mathrm{Var}\left[\widehat{S}_l\right].$$

Recall that the function g is Lipschitz with a Lipschitz constant C_g. Hence by Remark 4.1 we have that

$$\begin{aligned}
\mathrm{Var}\left[\widehat{S}_l\right] &= \frac{1}{N_l}\mathrm{Var}\left[g(\widehat{Y}_T^{(h_l)}) - g(\widehat{Y}_T^{(h_{l-1})})\right] \\
&\leq \frac{1}{N_l}\mathbb{E}\Big[\big|g(\widehat{Y}_T^{(h_l)}) - g(\widehat{Y}_T^{(h_{l-1})})\big|^2\Big] \\
&\leq \frac{C_g^2}{N_l}\mathbb{E}\Big[\big|\widehat{Y}_T^{(h_l)} - \widehat{Y}_T^{(h_{l-1})}\big|^2\Big] \\
&\leq 2\frac{C_g^2}{N_l}\mathbb{E}\Big[\big|\widehat{Y}_T^{(h_l)} - Y_T\big|^2\Big] + \mathbb{E}\Big[\big|Y_T - \widehat{Y}_T^{(h_{l-1})}\big|^2\Big] \\
&\leq 2\frac{C_g^2}{N_l}\tilde{K}^2(h_l + h_{l-1}) \\
&\leq \frac{\kappa_1}{N_l}h_l
\end{aligned}$$

for some constant $\kappa_1 > 0$ and for each $l = 1, 2, \ldots, L$. Similarly, since y_0 is a constant, for $\widehat{S}_0$ we have

$$\mathrm{Var}\left[\widehat{S}_0\right] = \frac{1}{N_0}\mathrm{Var}\left[g(\widehat{Y}_T^{h_0}) - g(y_0)\right] \leq \frac{C_g^2\tilde{K}^2}{N_0}h_0,$$

and hence for some $\kappa > 0$ we have that

$$\mathrm{Var}[\widehat{S}_l] \leq \kappa N_l^{-1}h_l \tag{4.2}$$

for all $l = 0, 1, \ldots, L$.

About the computational cost: We assume that all the arithmetic operations can be done in one time unit. Additionally we assume that drawing samples from distributions and evaluating the value of the given function can be done in constant time. It follows that a evaluating a piecewise constant function f is $O(\kappa n)$, where $\kappa > 0$ is a constant and n is the number of breakpoints. Therefore the computational cost of the algorithm is proportional to $\sum_{l=0}^{L} N_l h_l^{-1}$. The optimal choices for N_l are achieved by minimizing the variance for fixed computational cost. For more details, we refer reader to [4].

4.1. **Complexity.**

THEOREM 4.4. [4, Giles, Theorem 3.1] *Let $P = g(Y_T)$ denote a functional of the solution Y of a Brownian motion driven SDE (3.2) at time T for a given Brownian path B_t, and let $\widehat{P}_l$ denote the corresponding approximation $g(\widehat{Y}_T^{(h_l)})$ using a numerical equidistant time discretization with timestep $h_l = M^{-l}T$. If there exists independent estimators $\widehat{S}_l$ based on N_l Monte Carlo samples, and positive constants $\alpha \geq \frac{1}{2}$, β, c_1, c_2 and c_3 such that*

(i) $\mathbb{E}[\widehat{P}_l - P] \leq c_1 h_l^{\alpha}$,
(ii) $\mathbb{E}[\widehat{S}_0] = \mathbb{E}[\widehat{P}_0]$ *and* $\mathbb{E}[\widehat{S}_l] = \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$ *for all* $l > 0$,
(iii) $\mathrm{Var}[\widehat{S}_l] \leq c_2 N_l^{-1} h_l^{\beta}$,

(iv) $C_l \leq c_3 N_l h_l^{-1}$,

where C_l *denotes the computational complexity of* $\widehat{S}_l$*, then there exists a positive constant* c_4 *such that for any* $\varepsilon < e^{-1}$*, there are values* L *and* N_l *for which the multilevel estimator* $\widehat{S}$ *of* $\mathbb{E}[P]$*, defined as*

$$\widehat{S} = \sum_{l=0}^{L} \widehat{S}_l,$$

has an MSE with bound

$$\mathrm{MSE}(\widehat{S}) = \mathbb{E}\big[(\widehat{S} - \mathbb{E}[P])^2\big] < \varepsilon^2,$$

with a computational complexity C *with bound*

$$C \leq \begin{cases} c_4 \varepsilon^{-2}, & \beta > 1 \\ c_4 \varepsilon^{-2} (\log \varepsilon)^2, & \beta = 1 \\ c_4 \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

Proof. First, we prove that the bias is bounded with $\frac{1}{2}\varepsilon^2$. Then, we continue by proving that the variance has the same upper bound, which when combined with the bias upper bound leads us to the MSE bounded by ε^2. We consider the variance for different values of β simultaneously with computational complexity.

Recall that the ceiling function of x, denoted by $\lceil x \rceil$ gives us the smallest integer which is equal or larger than x, that is $x \leq \lceil x \rceil < x + 1$. We begin our proof by choosing L to be

$$L = \left\lceil \frac{\log(\sqrt{2} c_1 T^\alpha \varepsilon^{-1})}{\alpha \log M} \right\rceil.$$

From the definition of the ceiling function with some elementary computations we see that

$$\begin{aligned} & \frac{\log(\sqrt{2} c_1 T^\alpha \varepsilon^{-1})}{\alpha \log M} \leq L < \frac{\log(\sqrt{2} c_1 T^\alpha \varepsilon^{-1})}{\alpha \log M} + 1 \\ \iff & L - 1 < \frac{\log(\sqrt{2} c_1 T^\alpha \varepsilon^{-1})}{\alpha \log M} \leq L \\ \iff & \log(M^{(L-1)\alpha}) < \log(\sqrt{2} c_1 T^\alpha \varepsilon^{-1}) \leq \log(M^{\alpha L}) \\ \iff & \frac{1}{\sqrt{2}} M^{-\alpha} \varepsilon < c_1 (\frac{T}{M^L})^\alpha \leq \frac{1}{\sqrt{2}} \varepsilon \\ \iff & \frac{1}{\sqrt{2}} M^{-\alpha} \varepsilon < c_1 h_L^\alpha \leq \frac{1}{\sqrt{2}} \varepsilon. \end{aligned} \tag{4.3}$$

Assumptions (ii) and (i) combined with (4.3) imply that

$$\begin{aligned}
\operatorname{bias}(\widehat{S}, P)^2 &= \big(\mathbb{E}[\widehat{S}] - \mathbb{E}[P]\big)^2 \\
&= \left(\sum_{l=0}^{L} \mathbb{E}[\widehat{S}_l] - \mathbb{E}[P]\right)^2 \\
&= \left(\mathbb{E}[\widehat{P}_0] + \sum_{l=1}^{L} \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}] - \mathbb{E}[P]\right)^2 \\
&= \left(\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P]\right)^2 \\
&\leq (c_1 h_L^{\alpha})^2 \leq \frac{1}{2}\varepsilon^2,
\end{aligned}$$

which is the estimate we wanted for bias.

Next we find an upper limit for the sum of the sequence $(h_l^{-1})_{l=0,\dots,L}$. Using a standard result for the geometric series we have that

$$\begin{aligned}
\sum_{l=0}^{L} h_l^{-1} &= \frac{1}{h_L}\sum_{l=0}^{L} \frac{M^l}{M^L} = \frac{1}{h_L}\sum_{l=0}^{L} \frac{1}{M^{L-l}} = \frac{1}{h_L}\sum_{l=0}^{L} \frac{1}{M^l} \\
&= \frac{1}{h_L}\left(\frac{M^{L+1}-1}{M^{L+1}-M^L}\right) \\
&< \frac{1}{h_L}\left(\frac{M}{M-1}\right),
\end{aligned}$$

and furthermore with (4.3) we have

$$\frac{1}{h_L} < M\left(\frac{c_1\sqrt{2}}{\varepsilon}\right)^{1/\alpha}.$$

Note that when $\alpha \geq \frac{1}{2}$ we have $\varepsilon^{-1/\alpha} \leq \varepsilon^{-2}$ and that we assumed $\varepsilon < e^{-1}$. Now when combined with previous inequalities we have

$$\sum_{l=0}^{L} h_l^{-1} < \frac{M^2}{M-1}\left(\frac{c_1\sqrt{2}}{\varepsilon}\right)^{1/\alpha} \leq \frac{M^2}{M-1}(c_1\sqrt{2})^{1/\alpha}\varepsilon^{-2}. \tag{4.4}$$

It remains to find the variance of the estimator $\widehat{Y}$ and the computational complexity C. We proceed by considering β with different values. For each β we find the upper bound for the variance and then check the computational complexity.

Case $\beta = 1$:

We choose $N_l = \lceil 2\varepsilon^{-2}(L+1)c_2 h_l \rceil$. Then by assumption (iii) the properties of the ceiling function the variance is bounded with

$$\operatorname{Var}[\widehat{S}] = \sum_{l=0}^{L} \operatorname{Var}[\widehat{S}_l] \leq \sum_{l=0}^{L} \frac{c_2 h_l}{N_l} \leq \sum_{l=0}^{L} \frac{c_2 h_l \varepsilon^2}{2(L+1)c_2 h_l} \leq \frac{1}{2}\varepsilon^2,$$

which is the bound we wanted for the variance. Defining each N_l as ceiling function gives us

$$N_l \leq 2\varepsilon^{-2}(L+1)c_2 h_l + 1$$

and hence by assumption (iv) the computational complexity C of $\widehat{Y}$ is bounded by

$$\begin{aligned} C &\leq \sum_{l=0}^{L} C_l \leq c_3 \sum_{l=0}^{L} \frac{N_l}{h_l} \\ &\leq c_3 \sum_{l=0}^{L} \frac{2\varepsilon^{-2}(L+1)c_2 h_l + 1}{h_l} \\ &\leq c_3 2\varepsilon^{-2}(L+1)^2 c_2 + c_3 \sum_{l=0}^{L} h_l^{-1}. \end{aligned} \tag{4.5}$$

Note that for L we have from the properties of the ceiling function that

$$L \leq \frac{\log(\sqrt{2}c_1 T^\alpha \varepsilon^{-1})}{\alpha \log M} + 1 \leq \frac{\log(\sqrt{2}c_1 T^\alpha)}{\alpha \log M} + \frac{\log(\varepsilon^{-1})}{\alpha \log M} + 1,$$

and that from assumption $\varepsilon < e^{-1}$ we have $1 < \log(\varepsilon^{-1})$. It follows that

$$L + 1 \leq \log(\varepsilon^{-1}) \left(\frac{\log(\sqrt{2}c_1 T^\alpha)}{\log(\varepsilon^{-1})\alpha \log M} + \frac{1}{\alpha \log M} + \frac{2}{\log(\varepsilon^{-1})} \right) \leq c_5 \log(\varepsilon^{-1}), \tag{4.6}$$

where

$$c_5 = \frac{1}{\alpha \log M} + \max\left(0, \frac{\log(\sqrt{2}c_1 T^\alpha)}{\alpha \log M}\right) + 2.$$

Hence, from (4.5) together with (4.6) and (4.4) it follows that

$$\begin{aligned} C &\leq c_3 2\varepsilon^{-2}(L+1)^2 c_2 + c_3 \sum_{l=0}^{L} h_l^{-1} \\ &\leq c_3 2\varepsilon^{-2} c_5^2 (\log \varepsilon^{-1})^2 c_2 + c_3 \frac{M^2}{M-1} (\sqrt{2}c_1)^{1/\alpha} \varepsilon^{-2} \\ &\leq \varepsilon^{-2} (\log \varepsilon^{-1})^2 \left(2c_3 c_5^2 c_2 + c_3 \frac{M^2}{M-1} (\sqrt{2}c_1)^{1/\alpha} \right) \\ &= c_4 \varepsilon^{-2} (\log \varepsilon^{-1})^2, \end{aligned}$$

where $c_4 = 2c_3 c_5^2 c_2 + c_3 \frac{M^2}{M-1} (\sqrt{2}c_1)^{1/\alpha}$. Note that $(\log \varepsilon^{-1})^2 = (-\log \varepsilon)^2 = (\log \varepsilon)^2$.

Case $\beta > 1$:

Set

$$N_l = \left\lceil 2\varepsilon^{-2} c_2 T^{\frac{\beta-1}{2}} (1 - M^{-\frac{\beta-1}{2}})^{-1} h_l^{\frac{\beta+1}{2}} \right\rceil.$$

Again, for geometric series for some $s \neq 0$ we have

$$\sum_{l=0}^{L} h_l^s = T^s \sum_{l=0}^{L} (M^s)^l \leq T^s \frac{M^s}{M^s - 1} = T^s (1 - M^{-s})^{-1}, \tag{4.7}$$

with which, by choosing $s = (\beta - 1)/2$, we have for the variance by assumption (iii) and from the choice of N_l that

$$\begin{aligned}\mathrm{Var}[\widehat{Y}] = \sum_{l=0}^{L} \mathrm{Var}[\widehat{Y}_l] &\leq c_2 \sum_{l=0}^{L} \frac{h_l^\beta}{N_l} \\ &\leq \frac{1}{2}\varepsilon^2 T^{-\frac{\beta-1}{2}}(1 - M^{-\frac{\beta-1}{2}}) \sum_{l=0}^{L} h_l^{\frac{\beta-1}{2}} \\ &\leq \frac{1}{2}\varepsilon^2.\end{aligned}$$

From the properties of the ceiling function we have

$$N_l < 2\varepsilon^{-2} c_2 T^{\frac{\beta-1}{2}}(1 - M^{-\frac{\beta-1}{2}})^{-1} h_l^{\frac{\beta+1}{2}} + 1,$$

which together with the assumption (iv) gives us the following upper bound for computational complexity

$$C \leq \sum_{l=0}^{L} c_3 N_l h_l^{-1} \leq c_3 2\varepsilon^{-2} c_2 T^{\frac{\beta-1}{2}}(1 - M^{-\frac{\beta-1}{2}})^{-1} \sum_{l=0}^{L} h_l^{\frac{\beta-1}{2}} + c_3 \sum_{l=0}^{L} h_l^{-1}. \tag{4.8}$$

By choosing $s = (\beta - 1)/2$ in (4.7) we have

$$\sum_{l=0}^{L} h_l^{\frac{\beta-1}{2}} \leq T^{\frac{\beta-1}{2}}(1 - M^{-\frac{\beta-1}{2}})^{-1}$$

and with (4.4) we have

$$c_3 \sum_{l=0}^{L} h_l^{-1} \leq c_3 \frac{M^2}{M-1}(c_1\sqrt{2})^{1/\alpha}\varepsilon^{-2}.$$

We can now estimate (4.8) further by

$$\begin{aligned}C &\leq c_3 2\varepsilon^{-2} c_2 T^{\frac{\beta-1}{2}}(1 - M^{-\frac{\beta-1}{2}})^{-1} \sum_{l=0}^{L} h_l^{\frac{\beta-1}{2}} + c_3 \sum_{l=0}^{L} h_l^{-1} \\ &\leq 2c_2 c_3 \varepsilon^{-2} T^{\beta-1}(1 - M^{-\frac{\beta-1}{2}})^{-2} + c_3 \frac{M^2}{M-1}(\sqrt{2}c_1)^{1/\alpha}\varepsilon^{-2} \\ &\leq c_4 \varepsilon^{-2},\end{aligned}$$

where $c_4 = 2c_2c_3T^{\beta-1}(1 - M^{-\frac{\beta-1}{2}})^{-2} + c_3\frac{M^2}{M-1}(\sqrt{2}c_1)^{1/\alpha}$.

Case $\beta < 1$:

We set $N_l = \left\lceil 2\varepsilon^{-2} c_2 h_L^{-\frac{1-\beta}{2}}(1 - M^{-\frac{1-\beta}{2}})^{-1} h_l^{\frac{\beta+1}{2}} \right\rceil$. As in previous cases, we have a geometric series and therefore

$$\begin{aligned}\sum_{l=0}^{L} h_l^{-\frac{1-\beta}{2}} &= h_L^{-\frac{1-\beta}{2}} \sum_{l=0}^{L} (M^{L-l})^{-\frac{1-\beta}{2}} \\ &= h_L^{-\frac{1-\beta}{2}} \sum_{l=0}^{L} (M^{-\frac{1-\beta}{2}})^{l} \\ &< h_L^{-\frac{1-\beta}{2}}(1 - M^{-\frac{1-\beta}{2}})^{-1}.\end{aligned} \tag{4.9}$$

By assumption (iii), properties of the ceiling function and (4.9) we have

$$\sum_{l=0}^{L} \mathrm{Var}[\widehat{Y}_l] < \sum_{l=0}^{L} c_2 N_l^{-1} h_l^{\beta} < \frac{1}{2}\varepsilon^2 h_L^{\frac{1-\beta}{2}} (1 - M^{-\frac{1-\beta}{2}}) \sum_{l=0}^{L} h_l^{-\frac{1-\beta}{2}} < \frac{1}{2}\varepsilon^2.$$

We have

$$N_l < 2\varepsilon^{-2} c_2 h_L^{-\frac{1-\beta}{2}} (1 - M^{-\frac{1-\beta}{2}})^{-1} h_l^{\frac{\beta+1}{2}} + 1,$$

from which it follows by assumption (iv) that

$$\begin{aligned} C &\leq \sum_{l=0}^{L} c_3 N_l h_l^{-1} \\ &< 2\varepsilon^{-2} c_2 c_3 h_L^{-\frac{1-\beta}{2}} (1 - M^{-\frac{1-\beta}{2}})^{-1} \sum_{l=0}^{L} h_l^{-\frac{1-\beta}{2}} + c_3 \sum_{l=0}^{L} h_l^{-1}. \end{aligned} \tag{4.10}$$

By inequality (4.9) we have that

$$h_L^{-\frac{1-\beta}{2}} (1 - M^{-\frac{1-\beta}{2}})^{-1} \sum_{l=0}^{L} h_l^{-\frac{1-\beta}{2}} < h_L^{-(1-\beta)} (1 - M^{-\frac{1-\beta}{2}})^{-2} \tag{4.11}$$

and inequality (4.3) gives that

$$h_L^{-(1-\beta)} < (\sqrt{2}c_1)^{\frac{1-\beta}{\alpha}} M^{1-\beta} \varepsilon^{-\frac{1-\beta}{\alpha}}. \tag{4.12}$$

Combining the inequalities (4.11) and (4.4) with the cost estimate (4.10) we see that

$$\begin{aligned} C &< 2\varepsilon^{-2} c_2 c_3 h_L^{-\frac{1-\beta}{2}} (1 - M^{-\frac{1-\beta}{2}})^{-1} \sum_{l=0}^{L} h_l^{-\frac{1-\beta}{2}} + c_3 \sum_{l=0}^{L} h_l^{-1} \\ &< 2\varepsilon^{-2} c_2 c_3 h_L^{-(1-\beta)} (1 - M^{-\frac{1-\beta}{2}})^{-2} + c_3 \frac{M^2}{M-1} (\sqrt{2}c_1)^{1/\alpha} \varepsilon^{-2} \end{aligned}$$

and the equation (4.12) together with the fact that $\varepsilon^{-2} < \varepsilon^{-2-(1-\beta)/\alpha}$ for $\varepsilon < e^{-1}$ gives us that

$$\begin{aligned} &2\varepsilon^{-2} c_2 c_3 h_L^{-(1-\beta)} (1 - M^{-\frac{1-\beta}{2}})^{-2} + c_3 \frac{M^2}{M-1} (\sqrt{2}c_1)^{1/\alpha} \varepsilon^{-2} \\ &< \varepsilon^{-2-\frac{1-\beta}{\alpha}} 2c_2 c_3 (\sqrt{2}c_1)^{\frac{1-\beta}{\alpha}} M^{1-\beta} (1 - M^{-\frac{1-\beta}{2}})^{-2} + c_3 \frac{M^2}{M-1} (\sqrt{2}c_1)^{1/\alpha} \varepsilon^{-2} \\ &< c_4 \varepsilon^{-2-\frac{1-\beta}{\alpha}}, \end{aligned}$$

where

$$c_4 = 2c_2 c_3 (\sqrt{2}c_1)^{\frac{1-\beta}{\alpha}} M^{1-\beta} (1 - M^{-\frac{1-\beta}{2}})^{-2} + c_3 \frac{M^2}{M-1} (\sqrt{2}c_1)^{1/\alpha}.$$

□

Remark 4.2. The multilevel algorithm presented in Section 4 satisfies the assumptions of Theorem 4.4. By Proposition 4.2 the condition (i) follows from

$$\mathbb{E}\Big[g\big(\widehat{Y}_T^{(h_l)}\big) - g\big(Y_T\big)\Big] \leq \kappa h_l^{1/2},$$

and condition (ii) is clear from the way the algorithm was constructed. Variance condition (iii) was discussed in (4.2) and finally, computational cost on each level is connected to the number of evaluations N_l and timestep h_l, i.e. $C_l \leq \kappa N_l h_l^{-1}$ for some $\kappa > 0$.

5. Application to Lévy-driven SDE

In this section we consider a d-dimensional L^2-Lévy process $X = (X_t)_{t\geq 0}$ characterized by triplet $(\nu, \sigma\sigma^*, b)$, and the solution $Y = (Y_t)_{t\geq 0}$ to the stochastic integral equation

$$Y_t = y_0 + \int_0^t a(Y_{s-})dX_s, \tag{5.1}$$

where $y_0 \in \mathbb{R}^d$ and the function a has the properties formulated in Assumption 1 below. Equation (5.1) corresponds to the SDE in Definition 3.10. By Theorem 3.11 the properties in Assumption 1 imply the existence and uniqueness of the solution. We begin by formulating the multilevel Monte Carlo algorithm for the evaluation of

$$S(g) = \mathbb{E}\big[g(Y)\big],$$

where $g : D[0,1] \to \mathbb{R}$. Here $D[0,1]$ is the space of cadlag functions, more precisely

$$D[0,1] = \{f : [0,1] \to \mathbb{R}^r, f \text{ cadlag }\},$$

which is usually called Skorokhod space. On this space we consider the supremum norm denoted by

$$\|f\| = \sup_{t\in[0,1]} |f(t)|$$

for $f \in D[0,1]$ and we use the notation

$$\mathrm{Lip}(1) = \{H : D[0,1] \to \mathbb{R}, |H(f) - H(g)| \leq \|f - g\|\}$$

for the set of Borel-measurable functions that are Lipschitz continuous with Lipschitz constant one. This section is based on the article [3].

Assumption 1. For fixed $0 < K < \infty$, the function $a : \mathbb{R}^r \to \mathbb{R}^r \times \mathbb{R}^d$ satisfies

$$|a(y) - a(y')| \leq K|y - y'|$$

for all $y, y' \in \mathbb{R}^{d_Y}$. Furthermore, we have

$$|a(y_0)| \leq K,$$

$$0 < \int |x|^2 \nu(\mathrm{dx}) \leq K^2,$$

$$\sigma \in \mathbb{R}^{d\times d} \text{ with } |\sigma| \leq K,$$

where $|\cdot|$ denotes here a sub-multiplicate matrix norm and

$$b \in \mathbb{R}^d \text{ with } |b| \leq K$$

We recall that the Lévy-Itô decomposition formulated in Remark 2.10 grants us the representation

$$X_t = \sigma B_t + L_t + bt,$$

where $\sigma B_t + bt$ is a Brownian motion with a drift b and covariance $\sigma\sigma*$, and $L = (L_t)_{t\geq 0}$ is a Lévy process and a martingale with

$$L_t = \int_{\mathbb{R}^d} x\tilde{N}(t, dx) = \int_{\mathbb{R}^d} xN(t, dx) - t\int_{\mathbb{R}^d} x\nu(dx),$$

where $N(x,t)$ is a Poisson random measure with intensity $\nu(x)t$. For $h > 0$ we define

$$L_t^{(h)} := \int_{B_h^{\mathsf{c}}} xN(t,dx) - t\int_{B_h^{\mathsf{c}}} x\nu(dx), \tag{5.2}$$

and set $F_0(h) := \int_{B_h^{\mathsf{c}}} x\nu(dx)$. Then the process $C^{(h)} = (C_t^{(h)})_{t\geq 0}$ with $C_t^{(h)} = \int_{B_h^{\mathsf{c}}} N(t,dx)$ is a compound Poisson process with intensity $\lambda^{(h)} = \nu(B_h^{\mathsf{c}})$ and jump size distribution $\mu^{(h)} = \nu/\lambda^{(h)}$. The process $L^{(h)} = (L_t^{(h)})_{t\geq 0}$ is a martingale and the process B is independent of L. By the isometry formula we have that

$$\begin{aligned}
\lim_{h\to 0}\mathbb{E}\Big[\Big|\int_{\mathbb{R}^d} x\tilde{N}(t,dx) - L_t^{(h)}\Big|^2\Big] &= \lim_{h\to 0}\mathbb{E}\Big[\Big|\int_{\mathbb{R}^d} x\tilde{N}(t,dx) - \int_{B_h^{\mathsf{c}}} x\tilde{N}(t,dx)\Big|^2\Big] \\
&= \lim_{h\to 0}\mathbb{E}\Big[\Big|\int_{B_h} x\tilde{N}(t,dx)\Big|^2\Big] \\
&= \lim_{h\to 0}\mathbb{E}\Big[t\int_{B_h}|x|^2\nu(dx)\Big] = 0
\end{aligned}$$

and hence $L_t^{(h)} \to L_t$ in $L^2(\Omega,\mathcal{F},\mathbb{P})$ as $h \to 0$.

In Section 4 we used the Euler scheme with equidistant time discretization, but with Lévy-SDE we proceed more carefully. When evaluating the process X_t we distinguish between "small" and "big" jumps, i.e. jumps smaller than h and at least h. For the Euler scheme we want to find a time discretization which allows us to describe the behaviour of the solution well enough. In other words, we want to approximate the process using a time grid which does not let the process change too much between time steps. The times of the jumps that are at least h form the basis of our discretization, which is then refined such that each step is at most ε. We will refer to this refined discretization with $(T_j^{(h,\varepsilon)})_{j=0}^{\infty}$. Furthermore, we will denote the approximation of the process X by $\widehat{X}^{(h,\varepsilon)}$ and the approximation of the process Y by $\widehat{Y}^{(h,\varepsilon)}$ with parameters h and ε described more specifically below.

5.1. **Approximation of the process X.** We begin by creating a time discretization based on the jump times of the process X. First we deal with jumps that can be seen as the big jumps of the process X, i.e. jumps that are of size at least h. From the definition of $L^{(h)}$ we see that the jumps $\Delta L_t^{(h)}$ have values in $\{0\} \cup B_h^{\mathsf{c}}$. We denote the jump times of $L^{(h)}$ with $T_0^{(h)} = 0$ and

$$T_k^{(h)} = \inf\{t > T_{k-1}^{(h)} : \Delta L_t^{(h)} \neq 0\}$$

for $k \geq 1$. Recall from the properties of a compound Poisson process that the time differences $T_k^{(h)} - T_{k-1}^{(h)}$ form an i.i.d. sequence of random variables that are exponentially distributed with parameter $\lambda^{(h)}$ and that this sequence is independent of the sequence of jump heights denoted by $\Delta L_{T_k^{(h)}}^{(h)}$. From (5.2) we see that on every interval $[T_k^{(h)}, T_{k-1}^{(h)})$ the process $L^{(h)}$ is linear with the slope $-F_0(h)$.

It is not enough to consider only the big jumps of the process, but we need to handle more minor changes as well. Therefore we refine our time discretization such that our step sizes are ε at most. Hence our new discretization, which is the one we

will use, is defined as $T_0^{h,\varepsilon} = 0$ and

$$T_j^{(h,\varepsilon)} = \inf\{T_k^{(h)} > T_{j-1}^{(h,\varepsilon)} : k \in \mathbb{N}\} \wedge (T_{j-1}^{(h,\varepsilon)} + \varepsilon). \tag{5.3}$$

Similarly to the original Giles algorithm with the case of Brownian motion as driving process we will here consider the simulation of the joint distribution of $\widehat{X}^{(h,\varepsilon)}$ and $\widehat{X}^{(h',\varepsilon')}$ for different values of $h' > h > 0$ and $\varepsilon > \varepsilon' > 0$. Due to independence of processes $\Delta L^{(h')}$ (jumps bigger than h') and $\Delta(L^{(h)} - L^{(h')})$ (jumps between h and h') it holds that

$$\Delta L_t^{(h')} = \Delta L_t^{(h)} \cdot \mathbb{1}_{\{|\Delta L_t^{(h)}| > h'\}},$$

it is enough to simulate both jump times and heights of the process $L^{(h)}$ and then extract the ones we need to form the process $L^{(h')}$ to get the joint distribution of $(L^{(h)}, L^{(h')})$.

Finally, at each timepoint $T_j = T_j^{(h,\varepsilon)}$ we approximate the process X by $\widehat{X}_0^{(h,\varepsilon)} = 0$ and

$$\widehat{X}_{T_j}^{(h,\varepsilon)} = \widehat{X}_{T_{j-1}}^{(h,\varepsilon)} + \sigma(B_{T_j} - B_{T_{j-1}}) + \Delta L_{T_j}^{(h)} + (b - F_0(h))(T_j - T_{j-1})$$

for $j \geq 1$. Note that

$$\widehat{X}_{T_j}^{(h,\varepsilon)} = \sigma B_{T_j} + L_{T_j}^{(h)} + bT_j,$$

which can be proven with induction. Case $j = 1$ is trivial, and for arbitrary $j \geq 1$

$$\begin{aligned}
\widehat{X}_{T_j}^{(h,\varepsilon)} &= \widehat{X}_{T_{j-1}}^{(h,\varepsilon)} + \sigma(B_{T_j} - B_{T_{j-1}}) + \Delta L_{T_j}^{(h)} + (b - F_0(h))(T_j - T_{j-1}) \\
&= L_{T_{j-1}}^{(h)} + \sigma B_{T_j} + \Delta L_{T_j}^{(h)} - F_0(h)(T_j - T_{j-1}) + bT_j \\
&= \sigma B_{T_j} + L_{T_j} + bT_j
\end{aligned}$$

since $L_{T_j} = L_{T_{j-1}} + \Delta L_{T_j}^{(h)} - F_0(h)(T_j - T_{j-1})$.

In conclusion, to simulate the coupled process $(\widehat{X}^{(h,\varepsilon)}, \widehat{X}^{(h',\varepsilon')})$ we need to draw samples only from $\mu^{(h)}$ and an exponential distribution to form $T_k^{(h)}$ and $L_{T_k^{(h)}}^{(h)}$ for $(L^{(h)}, L^{(h')})$ respectively, as well as from normal distribution to simulate the Brownian motion component at times $(T_j^{(h,\varepsilon)})_{j\in\mathbb{N}}$ and $(T_j^{(h',\varepsilon')})_{j\in\mathbb{N}}$.

5.2. **Approximation of the process Y.** To approximate the solution Y of the SDE (5.1) we will use the Euler scheme presented in Definition 4.1 with the difference that the driving process is X, approximated with $\widehat{X}^{(h,\varepsilon)}$. We use the random time discretizaton $(T_j)_{j\in\mathbb{N}} = (T_j^{(h,\varepsilon)})_{j\in\mathbb{N}}$ created in previous subsection and the approximation of the process Y at times T_j is defined by $\widehat{Y}_{T_0}^{(h,\varepsilon)} = \widehat{Y}_0^{(h,\varepsilon)} = y_0$ and

$$\widehat{Y}_{T_j}^{(h,\varepsilon)} = \widehat{Y}_{T_{j-1}}^{(h,\varepsilon)} + a\big(\widehat{Y}_{T_{j-1}}^{(h,\varepsilon)}\big)\big(\widehat{X}_{T_j}^{(h,\varepsilon)} - \widehat{X}_{T_{j-1}}^{(h,\varepsilon)}\big)$$

for $j \geq 1$. Additionally, we define $\widehat{Y}_t^{(h,\varepsilon)} = \widehat{Y}_{T_j}^{(h,\varepsilon)}$ for all $t \in [T_j, T_{j+1})$ so that the approximation $\widehat{Y} = \widehat{Y}^{(h,\varepsilon)}$ is piecewise constant with respect to our time discretization $(T_j)_{j\in\mathbb{N}}$. It is clear that to get the approximation for Y we only need to approximate the driving process X. For the multilevel algorithm we will use coupled Euler schemes $(\widehat{Y}^{(h,\varepsilon)}, \widehat{Y}^{(h',\varepsilon')})$ with $h' > h > 0$ and $\varepsilon' > \varepsilon > 0$. It is again important that we use the same simulation of the Brownian motion path for both $\widehat{Y}^{(h,\varepsilon)}$ and $\widehat{Y}^{(h',\varepsilon')}$ in the coupled scheme.

5.3. **The algorithm.** To create the telescoping sum like in Section 4, we fix two positive and decreasing sequences $(\varepsilon_k)_{k\in\mathbb{N}}$ and $(h_k)_{k\in\mathbb{N}}$. Additionally, for shorter notation we denote $\widehat{Y}_t^{(k)} = \widehat{Y}_t^{(h_k,\varepsilon_k)}$ for all $t \geq 0$. This time, we denote the number of levels with $m \in \mathbb{Z}_+$ and the payoff function with $g : D[0,1] \to \mathbb{R}$ where g is measurable and $g(\widehat{Y}^{(k)})$ is integrable for all $k = 1, \dots, m$. Now we can write

$$\mathbb{E}\big[g(\widehat{Y}^{(m)})\big] = \mathbb{E}\big[g(\widehat{Y}^{(1)})\big] + \sum_{k=2}^{m} \mathbb{E}\big[g(\widehat{Y}^{(k)}) - g(\widehat{Y}^{(k-1)})\big].$$

Like previously, we approximate each expectation separately with classical Monte Carlo with the help of approximations from sections 5.1 and 5.2, and for $k = 1, \dots, m$ denote with n_k the number of samples on each level. The corresponding multilevel Monte Carlo algorithm is given by

$$\widehat{S}(g) = \frac{1}{n_1}\sum_{i=1}^{n_1} g\big(\widehat{Y}^{(1)}(i)\big) + \sum_{k=2}^{m} \frac{1}{n_k}\sum_{i=1}^{n_k} \Big[g\big(\widehat{Y}^{(k)}(i)\big) - g\big(\widehat{Y}^{(k-1)}(i)\big)\Big] \tag{5.4}$$

where $\widehat{Y}^{(1)}(i)$ is the ith independent copy of $\widehat{Y}^{(1)}$ and for each k the $\big(\widehat{Y}^{(k)}(i), \widehat{Y}^{(k-1)}(i)\big)$ is ith independent copy of the coupled Euler scheme $(\widehat{Y}^{(k)}, \widehat{Y}^{(k-1)})$. By construction the estimators $\frac{1}{n_1}\sum_{i=1}^{n_1} g\big(\widehat{Y}^{(1)}(i)\big)$ and $\frac{1}{n_k}\sum_{i=1}^{n_k} \Big[g\big(\widehat{Y}^{(k)}(i)\big) - g\big(\widehat{Y}^{(k-1)}(i)\big)\Big]$ for $k = 2, \dots, m$ are independent.

5.3.1. *The error of the algorithm.* We are interested in the mean squared error of the estimator $\widehat{S}(g)$. Recall from the Definition 4.3 that

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{Var}(\widehat{\theta}) + \mathrm{bias}(\widehat{\theta}, \theta)^2,$$

which means that variance and bias can be estimated separately. It is easy to see that

$$\begin{aligned}
\mathbb{E}\big[\widehat{S}(g)\big] &= \mathbb{E}\bigg[\frac{1}{n_1}\sum_{i=1}^{n_1} g\big(\widehat{Y}^{(1)}(i)\big) + \sum_{k=2}^{m} \frac{1}{n_k}\sum_{i=1}^{n_k} \Big(g\big(\widehat{Y}^{(k)}(i)\big) - g\big(\widehat{Y}^{(k-1)}(i)\big)\Big)\bigg] \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbb{E}\Big[g\big(\widehat{Y}^{(1)}(i)\big)\Big] + \sum_{k=2}^{m} \frac{1}{n_k}\sum_{i=1}^{n_k} \mathbb{E}\Big[g\big(\widehat{Y}^{(k)}(i)\big) - g\big(\widehat{Y}^{(k-1)}(i)\big)\Big] \\
&= \mathbb{E}\big[g(\widehat{Y}^{(1)})\big] + \sum_{k=2}^{m} \mathbb{E}\big[g(\widehat{Y}^{(k)}) - g(\widehat{Y}^{(k-1)})\big] \\
&= \mathbb{E}\big[g(\widehat{Y}^{(m)})\big]
\end{aligned}$$

and hence for $g \in \mathrm{Lip}(1)$ we compute

$$\begin{aligned}
\mathrm{bias}\big(\widehat{S}(g), S(g)\big)^2 &= \mathbb{E}\big[\widehat{S}(g) - S(g)\big]^2 \\
&= \mathbb{E}\big[g(\widehat{Y}^{(m)}) - g(Y)\big]^2 \\
&\leq \mathbb{E}\big[\,\big\|Y - Y^{(m)}\big\|\,\big]^2.
\end{aligned}$$

For variance with $g \in \mathrm{Lip}(1)$ we compute

$$\begin{aligned}\mathrm{Var}\left[\widehat{S}(g)\right] &= \mathrm{Var}\Big[\frac{1}{n_1}\sum_{i=1}^{n_1} g\big(\widehat{Y}^{(1)}(i)\big)\Big] + \mathrm{Var}\Big[\sum_{k=2}^{m}\frac{1}{n_k}\sum_{i=1}^{n_k} g\big(\widehat{Y}^{(k)}(i)\big) - g\big(\widehat{Y}^{(k-1)}(i)\big)\Big] \\ &= \frac{1}{n_1}\mathrm{Var}\Big[g\big(\widehat{Y}^{(1)}\big)\Big] + \sum_{k=2}^{m}\frac{1}{n_k}\mathrm{Var}\Big[g\big(\widehat{Y}^{(k)}\big) - g\big(\widehat{Y}^{(k-1)}\big)\Big] \\ &\le \frac{1}{n_1}\mathbb{E}\Big[\big|g\big(\widehat{Y}^{(1)}\big) - g(y_0)\big|^2\Big] + \sum_{k=2}^{m}\frac{1}{n_k}\mathbb{E}\Big[\Big|g\big(\widehat{Y}^{(k)}\big) - g\big(\widehat{Y}^{(k-1)}\big)\Big|^2\Big] \\ &\le \frac{1}{n_1}\mathbb{E}\Big[\big\|\widehat{Y}^{(1)} - y_0\big\|^2\Big] + \sum_{k=2}^{m}\frac{1}{n_k}\mathbb{E}\Big[\big\|\widehat{Y}^{(k)} - \widehat{Y}^{(k-1)}\big\|^2\Big],\end{aligned}$$

and see that the mean squared error of $\widehat{S}(g)$ can be estimated by

$$\begin{aligned}\mathrm{MSE}\left(\widehat{S}(g)\right) \le \mathbb{E}\big[\,\big\|Y - Y^{(m)}\big\| \\ + \frac{1}{n_1}\mathbb{E}\big[\big\|\widehat{Y}^{(1)} - y_0\big\|^2\big]\big]^2 + \sum_{k=2}^{m}\frac{1}{n_k}\mathbb{E}\big[\big\|\widehat{Y}^{(k)} - \widehat{Y}^{(k-1)}\big\|^2\big].\end{aligned}$$

Note that the upper bound we achieved does not depend on the function g and hence for the worst case error it holds that

$$\begin{aligned}e^2(\widehat{S}) &= \sup_{g\in\mathrm{Lip}(1)} \mathrm{MSE}\left(\widehat{S}(g)\right) \\ &\le \mathbb{E}\big[\,\big\|Y - Y^{(m)}\big\| + \frac{1}{n_1}\mathbb{E}\big[\big\|\widehat{Y}^{(1)} - y_0\big\|^2\big] \\ &\qquad + \sum_{k=2}^{m}\frac{1}{n_k}\mathbb{E}\big[\big\|\widehat{Y}^{(k)} - \widehat{Y}^{(k-1)}\big\|^2\big]. \qquad (5.5)\end{aligned}$$

5.3.2. *The cost of the algorithm.* We make the same assumptions of computation times we made in Subsection 4. Hence the computational complexity of the algorithm $\widehat{S}$ is

$$\mathrm{cost}(\widehat{S}) = \sum_{k=1}^{m} n_k \mathbb{E}\big[\Upsilon(\widehat{Y}^{(k)})\big],$$

where the function $\Upsilon(\cdot)$ gives us the breakpoints of an piecewise constant function.

5.4. **Results and proofs.** In this section our main goal is to analyze the error in the worst case and the cost of the multilevel algorithm presented in Section 5.3. We will have the error estimates in terms of the function

$$F(h) := \int_{B_h} |x|^2 \nu(dx)$$

for $h > 0$. We will first present and prove some auxiliary results.

Lemma 5.1. [3, Lemma 1] *Under Assumption 1, we have*

$$\mathbb{E}\big[\sup_{t\in[0,1]} |Y_t - y_0|^2\big] \le \kappa,$$

where κ is a finite constant depending on K only.

Proof. Let $t \in [0,1]$. By the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ we have

$$\begin{aligned}
\mathbb{E}\Big[\big|Y_t - y_0\big|^2\Big] &= \mathbb{E}\Big[\Big|\int_0^t a(Y_{s-})dX_s\Big|^2\Big] \\
&= \mathbb{E}\Big[\Big|\int_0^t a(Y_{s-})\sigma dB_s + b\int_0^1 a(Y_{s-})ds + \int_0^t a(Y_{s-})dL_s\Big|^2\Big] \\
&\leq \kappa_1 \mathbb{E}\Big[\Big|\int_0^t a(Y_{s-})\sigma dB_s\Big|^2\Big] \\
&\qquad + \mathbb{E}\Big[\Big|\int_0^t a(Y_{s-})bds\Big|^2\Big] + \mathbb{E}\Big[\Big|\int_0^t a(Y_{s-})dL_s\Big|^2\Big], \qquad (5.6)
\end{aligned}$$

where κ_1 is an appropriate positive constant. We analyze each of the expectations on the right hand side of the last inequality separately. By Proposition 3.3 and Assumption 1 we have

$$\mathbb{E}\Big[\Big|\int_0^t a(Y_{s-})\sigma dB_s\Big|^2\Big] \leq \mathbb{E}\Big[|\sigma|^2\int_0^t |a(Y_{s-})|^2 ds\Big] \leq K^2\mathbb{E}\Big[\int_0^t |a(Y_{s-})|^2 ds\Big], \qquad (5.7)$$

since $|a(Y_{s-})\sigma| \leq |a(Y_{s-})||\sigma|$ holds for a sub-multiplicate matrix norm. By the monotonicity of the integral and Assumption 1 we obtain

$$\mathbb{E}\Big[\Big|\int_0^t a(Y_{s-})bds\Big|^2\Big] \leq K^2\mathbb{E}\Big[\int_0^t |a(Y_{s-})|^2 ds\Big].$$

For the last term by Proposition 3.7 and Assumption 1 we have

$$\begin{aligned}
\mathbb{E}\Big[\Big|\int_0^t a(Y_{s-})dL_s\Big|^2\Big] &= \mathbb{E}\Big[\Big|\int_0^t\int_{\mathbb{R}^d} a(Y_{s-})x d\tilde{N}(ds,dx)\Big|^2\Big] \\
&= \mathbb{E}\Big[\int_0^t\int_{\mathbb{R}^d} |a(Y_{s-})x|^2 \nu(dx)ds\Big] \\
&= \mathbb{E}\Big[\int_0^t |a(Y_{s-})|^2 \int_{\mathbb{R}^d} |x|^2\nu(dx)ds\Big] \\
&\leq K^2\mathbb{E}\Big[\int_0^t |a(Y_{s-})|^2 ds\Big].
\end{aligned}$$

All the terms have the same upper bound, which by Assumption 1 and Fubini's Theorem we evaluate with

$$\begin{aligned}
K^2\mathbb{E}\Big[\int_0^t |a(Y_{s-})|^2 ds\Big] &\leq 2K^2\Big(\mathbb{E}\Big[\int_0^t |a(Y_{s-}) - a(y_0)|^2 ds\Big] + \mathbb{E}\Big[\int_0^t |a(y_0)|^2 ds\Big]\Big) \\
&\leq 2K^4\Big(\mathbb{E}\Big[\int_0^t |Y_{s-} - y_0|^2 ds\Big] + 1\Big) \\
&\leq 2K^4\Big(\int_0^t \mathbb{E}\big[|Y_{s-} - y_0|^2\big] ds + 1\Big). \qquad (5.8)
\end{aligned}$$

By combining (5.8) together with (5.6) we have

$$\mathbb{E}\big[|Y_t - y_0|^2\big] \leq \kappa_2\Big(\int_0^t \mathbb{E}\big[|Y_{s-} - y_0|^2\big] ds + 1\Big)$$

for a constant κ_2 depending only on K. Therefore, by Lemma 3.2 (Doob's inequality) and Proposition 3.1 (Gronwall's lemma) we obtain

$$\mathbb{E}\big[\sup_{t\in[0,1]}|Y_t - y_0|^2\big] \le 4\mathbb{E}\big[|Y_1 - y_0|^2\big] \le \kappa,$$

where κ is an appropriate constant depending only on K. □

At this point we introduce an auxiliary process $\bar{Y} = (\bar{Y}_t)_{t\ge 0}$, which will be used in the following proofs and results. Recall that $L_t^{(h)} \to L_t$ in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ as $h \to 0$ and that for fixed h the process $(L_t^{(h)})_{t\ge 0}$ is a compensated compound Poisson process. We denote $L^{(h)}$ by L' and let L'' denote the process $(L_t - L'_t)_{t\ge 0}$, in which case $L = L' + L''$. Furthermore, we denote the previous evaluation point before or at time t by $l(t)$, or, more precisely, we define

$$l(t) = \sup[0,t] \cap \mathbb{T},$$

where $\mathbb{T}$ is the set of jump times $(T_k)_{k=0,1,\dots}$ defined by $T_k = T_k^{(h,\varepsilon)}$ as in Section 5.1. We denote $\bar{X} = (\sigma B_t + L'_t + bt)_{t\ge 0}$ and let $\bar{Y}$ be the solution to the corresponding integral equation

$$\bar{Y}_t = y_0 + \int_0^t a(\bar{Y}_{l(s-)})d\bar{X}_s.$$

Proposition 5.2. [3, Proposition 1] *Under Assumption 1, there exists a constant κ depending only on K such that for any $\varepsilon \in (0,1]$ and $h > 0$ with $\nu(B_h^{\complement}) \le 1/\varepsilon$ we have*

(i) in the case without Brownian component, i.e. $\sigma = 0$

$$\mathbb{E}\big[\sup_{t\in[0,1]}|Y_t - \bar{Y}_t|^2\big] \le \kappa[F(h) + |b - F_0(h)|^2\varepsilon^2]$$

(ii) in the general case

$$\mathbb{E}\big[\sup_{t\in[0,1]}|Y_t - \bar{Y}_t|^2\big] \le \kappa(\varepsilon + F(h)).$$

Proof. For $t \geq 0$ we use the notation $Z_t := Y_t - \bar{Y}_t$ and $Z'_t := Y_t - \bar{Y}_{l(t)}$. Note first that

$$\begin{aligned}
Z_t = Y_t - \bar{Y}_t &= \int_0^t a(Y_{s-})dX_s - \int_0^t a(Y_{l(s-)})d\bar{X}_s \\
&= \int_0^t \big(a(Y_{s-}) - a(\bar{Y}_{l(s-)})\big)\sigma dB_s \\
&\quad + \int_0^t a(Y_{s-})dL_s - \int_0^t a(\bar{Y}_{l(s-)})dL'_s \\
&\quad + b\int_0^t a(Y_{s-}) - a(\bar{Y}_{l(s-)})ds \\
&= \int_0^t a(Y_{s-}) - a(\bar{Y}_{l(s-)})\sigma dB_s \\
&\quad + \int_0^t a(Y_{s-}) - a(\bar{Y}_{l(s-)})dL'_s + \int_0^t a(Y_{s-})dL''_s \\
&\quad + b\int_0^t a(Y_{s-}) - a(\bar{Y}_{l(s-)})ds \\
&= M_t + b\int_0^t a(Y_{s-}) - a(\bar{Y}_{l(s-)})ds, \qquad (5.9)
\end{aligned}$$

where

$$\begin{aligned}
M_t = \int_0^t \big(a(Y_{s-}) - a(\bar{Y}_{l(s-)})\big)\sigma dB_s \\
+ \int_0^t a(Y_{s-}) - a(\bar{Y}_{l(s-)})dL'_s + \int_0^t a(Y_{s-})dL''_s. \qquad (5.10)
\end{aligned}$$

First we analyze the integral terms of M_t separately. By Proposition 3.3, Assumption1 and Fubini's Theorem we have

$$\begin{aligned}
\mathbb{E}\Big[\Big|\int_0^t \big(a(Y_{s-}) - a(\bar{Y}_{l(s-)})\big)\sigma dB_s\Big|^2\Big] &\leq |\sigma|^2\mathbb{E}\Big[\int_0^t |a(Y_{s-}) - a(\bar{Y}_{l(s-)})|^2 ds\Big] \\
&\leq K^4\mathbb{E}\Big[\int_0^t |Y_{s-} - \bar{Y}_{l(s-)}|^2 ds\Big] \\
&\leq K^4\int_0^t \mathbb{E}[|Z'_{s-}|^2]ds \qquad (5.11)
\end{aligned}$$

and similarly by Proposition 3.7, Assumption 1 and Fubini we have

$$\begin{aligned}
\mathbb{E}\Big[\Big|\int_0^t a(Y_{s-}) - a(\bar{Y}_{l(s-)})dL'_s\Big|^2\Big] &= \mathbb{E}\Big[\Big|\int_0^t \int_{B_h^c} \big(a(Y_{s-}) - a(\bar{Y}_{l(s-)})\big)x\tilde{N}(ds,dx)\Big|^2\Big] \\
&= \mathbb{E}\Big[\int_0^t \int_{B_h^c} |a(Y_{s-}) - a(\bar{Y}_{l(s-)})|^2|x|^2\nu(dx)ds\Big] \\
&= \mathbb{E}\Big[\int_0^t |a(Y_{s-}) - a(\bar{Y}_{l(s-)})|^2 ds \int_{B_h^c} |x|^2\nu(dx)\Big] \\
&\leq K^4\int_0^t \mathbb{E}[|Z'_{s-}|^2]ds. \qquad (5.12)
\end{aligned}$$

The last one follows similarly

$$\begin{aligned}\mathbb{E}\Big[\int_0^t a(Y_{s-})dL''_s\Big]^2 &= \mathbb{E}\Big[\int_0^t |a(Y_{l(s-)})|^2 ds \int_{B_h} |x|^2\nu(dx)\Big] \\ &= F(h)\mathbb{E}\Big[\int_0^t \big(|a(Y_{l(s-)}) - a(y_0)| + |a(y_0)|\big)^2 ds\Big] \\ &= F(h)K^2\mathbb{E}\Big[\int_0^t |Y_{l(s-)} - y_0|^2 ds + 1\Big]. \end{aligned} \tag{5.13}$$

Note that the process $(M_t)_{t\geq 0}$ is an L_2-martingale. Using Proposition (3.2) together with (5.11), (5.12) and (5.13) we have

$$\begin{aligned}\mathbb{E}\big[\sup_{s\in[0,t]} |M_s|^2\big] &\leq 4\mathbb{E}\big[|M_t|^2\big] \\ &= 8K^4\int_0^t \mathbb{E}[|Z'_{s-}|^2]ds + 4F(h)K^2\mathbb{E}\Big[\int_0^t |Y_{l(s-)} - y_0|^2 ds + 1\Big]. \end{aligned} \tag{5.14}$$

By using Hölder's inequality, Assumption 1 and Fubini's Theorem the remaining part of Z_t in (5.9) can be bounded by

$$\begin{aligned}&\mathbb{E}\bigg[\Big|b\int_0^t a(Y_{s-}) - a(\bar{Y}_{l(s-)})ds\Big|^2\bigg] \\ &\leq |b^2 t|\mathbb{E}\bigg[\int_0^t \Big|a(Y_{s-}) - a(\bar{Y}_{l(s-)})\Big|^2 ds\bigg] \\ &\leq K^4\mathbb{E}\bigg[\int_0^t \Big|Z'_{s-}\Big|^2 ds\bigg] \\ &\leq K^4\int_0^t \mathbb{E}\Big[\Big|Z'_{s-}\Big|^2\Big]ds\end{aligned}$$

for $t \in [0,1]$. Now, by using (5.9), the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ and that $Z'_t = Z_t + \bar{Y}_t - \bar{Y}_{l(t)}$, we can estimate the process Z by

$$\begin{aligned}\mathbb{E}\Big[\sup_{u\in[0,t]} |Z_u|^2\Big] &\leq 2\mathbb{E}\Big[\sup_{u\in[0,t]} |M_u|^2\Big] + 2\mathbb{E}\Big[\sup_{u\in[0,t]} \Big|b\int_0^u \big(a(Y_{s-}) - a(\bar{Y}_{l(s-)})\big)ds\Big|^2\Big] \\ &\leq 18K^4\int_0^t \mathbb{E}\Big[|Z'_{s-}|^2\Big]ds + 8K^2F(h)\int_0^t \mathbb{E}\Big[\big(|Y_{s-} - y_0| + 1\big)^2\Big]ds \\ &\leq 18K^4\int_0^t \mathbb{E}\Big[|Z_{s-} + \bar{Y}_t - \bar{Y}_{l(s-)}|^2\Big]ds \\ &\quad + 8K^2F(h)\int_0^t \mathbb{E}\Big[\big(|Y_{s-} - y_0| + 1\big)^2\Big]ds \\ &\leq 36K^4\int_0^t \mathbb{E}\Big[|Z_{s-}|^2 + |\bar{Y}_t - \bar{Y}_{l(s-)}|^2\Big]ds \\ &\quad + 8K^2F(h)\int_0^t \mathbb{E}\Big[\big(|Y_{s-} - y_0| + 1\big)^2\Big]ds.\end{aligned}$$

Since by Lemma 5.1 the expression $\mathbb{E}\big[\sup_{s\in[0,t]}(|Y_s - y_0| + 1)^2\big]$ is bounded by a constant we have that

$$z(t) := \mathbb{E}\Big[\sup_{s\in[0,t]}|Z_s|^2\Big] \le \kappa_1\Big(\int_0^t \mathbb{E}\Big[|Z_{s-}|^2 + |\bar{Y}_t - \bar{Y}_{l(s-)}|^2\Big]ds + F(h)\Big)$$
$$\le \kappa_1\Big(\int_0^t z(s) + \mathbb{E}\Big[|\bar{Y}_t - \bar{Y}_{l(s-)}|^2\Big]ds + F(h)\Big), \tag{5.15}$$

where κ_1 is a constant that depends only on K.

It remains to find an estimate for $\mathbb{E}\big[|\bar{Y}_t - \bar{Y}_{l(t)}|^2\big]$. First recall that $l(t)$ is the last evaluation point before or at time t and thus the process L' has no jumps on interval $\big(l(t), t\big]$ and it is linear with slope $-F_0(h)$. We see that

$$\bar{Y}_t - \bar{Y}_{l(t)} = \sigma\int_{l(t)}^t a(\bar{Y}_{l(s-)})dB_s + \int_{l(t)}^t a(\bar{Y}_{l(s-)})dL'_s + b\int_{l(t)}^t a(\bar{Y}_{l(s-)})ds$$
$$= a(\bar{Y}_{l(t)})\Big(\sigma(B_t - B_{l(t)}) - F_0(h)\big(t - l(t)\big) + b\big(t - l(t)\big)\Big)$$
$$= a(\bar{Y}_{l(t)})\Big(\sigma(B_t - B_{l(t)}) + \big(b - F_0(h)\big)\big(t - l(t)\big)\Big).$$

Note that from Assumption 1 we obtain

$$\mathbb{E}\Big[\big|a(\bar{Y}_{l(t)})\big|^2\Big] \le \mathbb{E}\Big[\big(|a(\bar{Y}_{l(t)}) - a(y_0)| + |a(y_0)|\big)^2\Big]$$
$$\le K^2\mathbb{E}\Big[\big(|\bar{Y}_{l(t)} - y_0| + 1\big)^2\Big]$$
$$\le K^2\mathbb{E}\Big[\big(|Y_{l(t)} - y_0| + |Z_{l(t)}| + 1\big)^2\Big],$$

where the last inequality follows from

$$|\bar{Y}_{l(t)} - y_0| \le |\bar{Y}_{l(t)} - Y_{l(t)}| + |Y_{l(t)} - y_0| = |Y_{l(t)} - y_0| + |Z_{l(t)}|.$$

From the strong Markov property of Brownian motion([7, Theorem 2.6.16]) it follows that $\big(B_{l(t)+s} - B_{l(t)}\big)_{s\ge 0}$ is a Brownian motion independent from $\mathcal{F}_{l(t)}$. Hence, by using the fact that our step size is ε at most and additionally twice the inequality $(a+b)^2 \le 2a^2 + 2b^2$, it follows that

$$\mathbb{E}\big[|\bar{Y}_t - \bar{Y}_{l(t)}|^2\big] \le \mathbb{E}\Big[\Big|a(\bar{Y}_{l(t)})\Big|^2\Big|\Big(\sigma(B_t - B_{l(t)}) - \big(b - F_0(h)\big)\big(t - l(t)\big)\Big)\Big|^2\Big]$$
$$\le 2K^2\mathbb{E}\Big[\big(|Y_{l(t)} - y_0| + |Z_{l(t)}| + 1\big)^2\Big]\Big(|\sigma|^2\varepsilon + \big|b - F_0(h)\big|^2\varepsilon^2\Big)$$
$$\le 4K^2\Big(\mathbb{E}\Big[\big(|Y_{l(t)} - y_0| + 1\big)^2\Big] + \mathbb{E}\Big[|Z_{l(t)}|^2\Big]\Big)\Big(|\sigma|^2\varepsilon + \big|b - F_0(h)\big|^2\varepsilon^2\Big).$$

The Cauchy-Schwarz inequality together with our assumption $\nu(B_h^{\mathsf{c}}) \le 1/\varepsilon$ yields that

$$|F_0(h)|^2 = \Big|\int_{B_h^{\mathsf{c}}} x\nu(dx)\Big|^2 \le \nu(B_h^{\mathsf{c}})\int_{B_h^{\mathsf{c}}}|x|^2\nu(dx) \le \frac{K^2}{\varepsilon}, \tag{5.16}$$

and we also know by Lemma 5.1 that $\mathbb{E}\big[\big(|Y_{l(t)} - y_0| + 1\big)^2\big]$ is uniformly bounded. Consequently, for $t \in [0,1]$ we have

$$\mathbb{E}\big[|\bar{Y}_t - \bar{Y}_{l(t)}|^2\big] \le \kappa_2\big[|\sigma|^2\varepsilon + |b - F_0(h)|^2\varepsilon^2 + z(t)\big],$$

where κ_2 is an appropriate constant that depends on K only. This estimate combined with equation (5.15) gives us

$$z(t) \leq \kappa_3 \Big[\int_0^t z(s)ds + |\sigma|^2\varepsilon + |b - F_0(h)|^2\varepsilon^2 + F(h) \Big]$$

where again κ_3 is a constant depending only on K. We see that our estimate for z is of the form

$$z(t) \leq \alpha_1 \int_0^t z(s)ds + \alpha_2.$$

If $\sigma = 0$, the statement of the proposition follows from Gronwall's inequality (Proposition 3.1). The general case is an immediate consequence of the estimates $|\sigma|^2\varepsilon \leq K^2\varepsilon$ from Assumption 1 and

$$\begin{aligned} |b - F_0(h)|^2\varepsilon^2 &\leq 2\varepsilon^2(|b|^2 + |F_0(h)|^2) \\ &\leq 2\varepsilon^2K^2 + 2\varepsilon K^2 \\ &\leq 4K^2\varepsilon \end{aligned}$$

where we used (5.16). □

Lemma 5.3. [3, Lemma 2] *Let $r \in \mathbb{N}$ and $(\mathcal{G}_j)_{j=0,1,\dots,r}$ denote a filtration. Moreover, for $j = 0, \dots, r-1$ let U_j and V_j denote non-negative random variables such that U_j is $\mathcal{G}_j$-measurable, and V_j is $\mathcal{G}_{j+1}$-measurable and independent of $\mathcal{G}_j$. Then one has*

$$\mathbb{E}\left[\max_{j=0,\dots,r-1} U_jV_j \right] \leq \mathbb{E}\left[\max_{j=0,\dots,r-1} U_j \right] \mathbb{E}\left[\max_{j=0,\dots,r-1} V_j \right].$$

Proof. Without loss of generality we assume that (U_j) is monotonically increasing. Otherwise, we can prove the result for $(\tilde{U}_j)$ given by $\tilde{U}_j = \max_{k \leq j} U_k$ and note that

$$\begin{aligned} \mathbb{E}\left[\max_{j=0,\dots,r-1} U_jV_j \right] &\leq \mathbb{E}\left[\max_{j=0,\dots,r-1} \tilde{U}_jV_j \right] \\ &\leq \mathbb{E}\left[\max_{j=0,\dots,r-1} \tilde{U}_j \right] \mathbb{E}\left[\max_{j=0,\dots,r-1} V_j \right]. \\ &= \mathbb{E}\left[\max_{j=0,\dots,r-1} U_j \right] \mathbb{E}\left[\max_{j=0,\dots,r-1} V_j \right]. \end{aligned}$$

We proceed by induction. For $r = 1$ we have

$$\mathbb{E}[U_0V_0] = \mathbb{E}\big[\mathbb{E}[U_0V_0|\mathcal{G}_0]\big] = \mathbb{E}\big[U_0\mathbb{E}[V_0|\mathcal{G}_0]\big] = \mathbb{E}[U_0]\mathbb{E}[V_0]$$

because of the independence of V_0 and $\mathcal{G}_0$. Now let $r \geq 1$ be arbitrary. Note that by induction hypothesis and monotonicity of U we can deduce that

$$\mathbb{E}\big[\max_{j=1,\dots r} U_jV_j \big] \leq \mathbb{E}\big[\max_{j=1,\dots,r} U_j \big] \mathbb{E}\big[\max_{j=1,\dots,r} V_j \big] = \mathbb{E}[U_r]\mathbb{E}\big[\max_{j=1,\dots,r} V_j \big]$$

and the monotonicity of U and the properties of conditional expectation imply that

$$\begin{aligned}\mathbb{E}\Big[(U_0V_0 - \max_{j=1,\dots r} U_jV_j)^+\Big] &= \mathbb{E}\Big[\mathbb{E}\big[(U_0V_0 - \max_{j=1,\dots r} U_jV_j)^+|\mathcal{G}_0\big]\Big]\\ &\le \mathbb{E}\Big[U_0\mathbb{E}\big[(V_0 - \max_{j=1,\dots r} V_j)^+|\mathcal{G}_0\big]\Big]\\ &= \mathbb{E}\Big[U_0\mathbb{E}\big[(V_0 - \max_{j=1,\dots r} V_j)^+\big]\Big]\\ &= \mathbb{E}[U_0]\mathbb{E}\big[(V_0 - \max_{j=1,\dots r} V_j)^+\big].\end{aligned}$$

Now it is easy to see that

$$\begin{aligned}\mathbb{E}\big[\max_{j=0,\dots,r} U_jV_j\big] &= \mathbb{E}\big[\max_{j=1,\dots r} U_jV_j\big] + \mathbb{E}\big[(U_0V_0 - \max_{j=1,\dots r} U_jV_j)^+\big]\\ &\le \mathbb{E}[U_r]\mathbb{E}\big[\max_{j=1,\dots,r} V_j\big] + \mathbb{E}[U_0]\mathbb{E}\big[(V_0 - \max_{j=1,\dots r} V_j)^+\big]\\ &\le \mathbb{E}[U_r]\Big(\mathbb{E}\big[\max_{j=1,\dots,r} V_j\big] + \mathbb{E}\big[(V_0 - \max_{j=1,\dots r} V_j)^+\big]\Big)\\ &= \mathbb{E}\big[\max_{j=0,\dots,r} U_j\big]\mathbb{E}\big[\max_{j=0,\dots,r} V_j\big].\end{aligned}$$

□

In the following results we denote $\widehat{Y}_t^{(h,\varepsilon)}$ with the short form $\widehat{Y}_t$.

Lemma 5.4. *The processes $(\bar{Y}_t)_{t\ge0}$ and $(\widehat{Y}_t)_{t\ge0}$ coincide almost surely for all times $(T_k)_{k=0,1,\dots}$ defined by $T_k = T_k^{(h,\varepsilon)}$ as in (5.3).*

Proof. Again, recall that L'_t is linear with slope $-F_0(h)$ on each interval $[T_{k-1}, T_k)$ for each $k \in \mathbb{N}$ and thus

$$L'_{T_k} = L'_{T_{k-1}} + \Delta L'_{T_k} - F_0(h)(T_k - T_{k-1}).$$

Additionally, we have that $\Delta L_{T_k} = \Delta L'_{T_k}$ almost surely.

It is clear that $\bar{Y}_{T_0} = \bar{Y}_0 = y_0 = \widehat{Y}_{T_0}$ and hence by induction we have

$$\begin{aligned}\bar{Y}_{T_k} &= y_0 + \int_0^{T_k} a(\bar{Y}_{l(s-)})d\bar{X}_s = \bar{Y}_{T_{k-1}} + \int_{T_{k-1}}^{T_k} a(\bar{Y}_{l(s-)})d\bar{X}_s\\ &= \bar{Y}_{T_{k-1}} + a(\bar{Y}_{T_{k-1}})\Big(\sigma(B_{T_k} - B_{T_{k-1}}) + \Delta L'_{T_k} + \big(b - F_0(h)\big)(T_k - T_{k-1})\Big)\\ &= \widehat{Y}_{T_{k-1}} + a(\widehat{Y}_{T_{k-1}})\Big(\sigma(B_{T_k} - B_{T_{k-1}}) + \Delta L_{T_k} + \big(b - F_0(h)\big)(T_k - T_{k-1})\Big)\\ &= \widehat{Y}_{T_{k-1}} + a(\widehat{Y}_{T_{k-1}})\Big(\widehat{X}_{T_k} - \widehat{X}_{T_{k-1}}\Big) = \widehat{Y}_{T_k}\end{aligned}$$

for $k \in \mathbb{N}$. □

THEOREM 5.5. [3, Theorem 2] *Under Assumption 1, there exists a constant κ depending only on K such that for any $\varepsilon \in (0,1]$ and $h > 0$ with $\nu(B_h) \le \frac{1}{\varepsilon}$, one has*

(i) in the case without Brownian component, i.e. $\sigma = 0$

$$\mathbb{E}[\sup_{t\in[0,1]} |Y_t - \widehat{Y}_t^{(h,\varepsilon)}|^2] \le \kappa(F(h) + |b - F_0(h)|^2\varepsilon^2)$$

(ii) in the general case

$$\mathbb{E}[\sup_{t\in[0,1]}|Y_t-\widehat{Y}_t^{(h,\varepsilon)}|^2]\le\kappa(\varepsilon\log(e/\varepsilon)+F(h)).$$

Proof. We will begin the proof by finding an appropriate upper bound for $\mathbb{E}\big[\sup_{t\in[0,1]}|\bar{Y}_t-\widehat{Y}_t|^2\big]$, which can then be combined with Proposition 5.2 to obtain the statement. Recall that in Section 5.2 we defined $\widehat{Y}_t=\widehat{Y}_{l(t)}$ and then by Lemma 5.4 we see that

$$\bar{Y}_t-\widehat{Y}_t=\bar{Y}_t-\bar{Y}_{l(t)}=a(\bar{Y}_{l(t)})\Big(\sigma(B_t-B_{l(t)})+\big(b-F_0(h)\big)\big(t-l(t)\big)\Big).$$

We analyze this in two parts by defining processes $(A_t)_{t\ge0}$ and $(D_t)_{t\ge0}$ with

$$A_t:=a(\bar{Y}_{l(t)})\sigma(B_t-B_{l(t)})$$

and

$$D_t:=a(\bar{Y}_{l(t)})(b-F_0(h))\big(t-l(t)\big).$$

We defined our time grid such that the mesh size is ε at most and hence

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,1]}|D_t|^2\Big]&\le\mathbb{E}\Big[\sup_{t\in[0,1]}|a(\bar{Y}_{l(t)})|^2\Big]|b-F_0(h)|^2\varepsilon^2\\&\le K^2\mathbb{E}\Big[\sup_{t\in[0,1]}(|\bar{Y}_{l(t)}-y_0|+1)^2\Big]|b-F_0(h)|^2\varepsilon^2\\&\le K^2\mathbb{E}\big[(\big\|\bar{Y}-y_0\big\|+1)^2\big]|b-F_0(h)|^2\varepsilon^2.\end{aligned}\tag{5.17}$$

For the process A we see that

$$\mathbb{E}\Big[\sup_{t\in[0,1]}|A_t|^2\Big]\le K^2|\sigma|^2\mathbb{E}\Big[\sup_{t\in[0,1]}(|\bar{Y}_{l(t)}-y_0|+1)^2|B_t-B_{l(t)}|^2\Big]$$

and for $r\in\mathbb{N}$ we get

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,1\wedge T_r]}|A_t|^2\Big]&\le K^2|\sigma|^2\mathbb{E}\Big[\sup_{t\in[0,1\wedge T_r]}(|\bar{Y}_{l(t)}-y_0|+1)^2|B_t-B_{l(t)}|^2\Big]\\&=K^2|\sigma|^2\mathbb{E}\Big[\max_{j=0,\dots,r-1}\Big(\mathbb{1}_{\{T_j<1\}}(|\bar{Y}_{T_j}-y_0|+1)^2\Big)\Big(\sup_{t\in[T_j,T_{j+1}\wedge1)}|B_t-B_{T_j}|^2\Big)\Big].\end{aligned}$$

We apply Lemma 5.3 with $U_j=\mathbb{1}_{\{T_j<1\}}(|\bar{Y}_{T_j}-y_0|+1)^2$, $V_j=\sup_{t\in[T_j,T_{j+1})}|B_t-B_{T_j}|^2$ and $\mathcal{G}_j=\mathcal{F}_{T_j}$ with $j=0,1,\dots$ and obtain

$$\begin{aligned}&K^2|\sigma|^2\mathbb{E}\Big[\max_{j=0,\dots,r-1}\Big(\mathbb{1}_{\{T_j<1\}}(|\bar{Y}_{T_j}-y_0|+1)^2\Big)\Big(\sup_{t\in[T_j,T_{j+1}\wedge1)}|B_t-B_{T_j}|^2\Big)\Big]\\&\quad\le K^2|\sigma|^2\mathbb{E}\Big[\max_{j=0,\dots,r-1}\mathbb{1}_{\{T_j<1\}}(|\bar{Y}_{T_j}-y_0|+1)^2\Big]\mathbb{E}\Big[\max_{j=0,\dots,r-1}\sup_{t\in[T_j,T_{j+1}\wedge1)}|B_t-B_{T_j}|^2\Big]\\&\quad=K^2|\sigma|^2\mathbb{E}\Big[\sup_{t\in[0,1\wedge T_r]}(|\bar{Y}_{l(t)}-y_0|+1)^2\Big]\mathbb{E}\Big[\max_{j=0,\dots,r-1}\sup_{t\in[T_j,T_{j+1}\wedge1)}|B_t-B_{T_j}|^2\Big].\end{aligned}$$

To estimate the expectation with Brownian motions we use Lévy's modulus of continuity theorem ([7, p. 114]). A function $g(\cdot)$ is called a modulus of continuity for

the function $f : [0,T] \to \mathbb{R}$ if $0 \le s < t \le T$ and $|t-s| \le \delta$ imply $|f(t)-f(s)| \le g(\delta)$ for all sufficiently small positive δ. From Lévy's Theorem it follows that with

$$g(\delta) := \sqrt{2\delta \log(1/\delta)}, \quad \delta > 0,$$

$cg(\delta)$ is a modulus of continuity for almost every Brownian path on $[0,1]$ if $c > 1$. Consider $c = \sqrt{e/2}$, and note that then $cg(\delta/e) = \sqrt{e/2}\sqrt{2\delta/e \log(1/\delta)} = \sqrt{\delta \log(e/\delta)}$. Therefore the term

$$\|B\|_\varphi := \sup_{0 \le s < t \le 1} \frac{|B_t - B_s|}{\varphi(t-s)}$$

is almost surely finite for $\varphi : [0,1] \to [0,\infty), \delta \mapsto \sqrt{\delta \log(e/\delta)}$ and, by Fernique's theorem, $\mathbb{E}\,\|B\|_\varphi^2$ is finite. Since on interval $[0,1]$ the function φ is increasing and the mesh size of our timegrid is ε at most, we have that

$$\begin{aligned}\mathbb{E}\Big[\max_{j=0,\dots,r-1} \sup_{t \in [T_j, T_{j+1}\wedge 1)} |B_t - B_{T_j}|^2\Big] &\le \mathbb{E}\Big[\,\|B\|_\varphi^2 \sup_{t \in [0,1\wedge T_r]} \varphi(t-l(t))^2\Big] \\ &\le \mathbb{E}\big[\,\|B\|_\varphi^2\,\big]\varphi(\varepsilon)^2.\end{aligned}$$

Now we see that

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,1]} |A_t|^2\Big] &= \lim_{r\to\infty} \mathbb{E}\Big[\sup_{t\in[0,1\wedge T_r]} |A_t|^2\Big] \\ &\le \lim_{r\to\infty} K^2|\sigma|^2\mathbb{E}\Big[\sup_{t\in[0,1\wedge T_r]} (|\bar{Y}_{l(t)} - y_0| + 1)^2\Big]\mathbb{E}\big[\,\|B\|_\varphi^2\,\big]\varphi(\varepsilon)^2 \\ &\le K^2|\sigma|^2\mathbb{E}\Big[\sup_{t\in[0,1]} (|\bar{Y}_{l(t)} - y_0| + 1)^2\Big]\mathbb{E}\big[\,\|B\|_\varphi^2\,\big]\varphi(\varepsilon)^2\end{aligned}$$

and by combining this estimation of the process A with the estimation of the process D in (5.17) we get

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,1]} |\bar{Y}_t - \widehat{Y}_t|^2\Big] &\le \mathbb{E}\Big[\sup_{t\in[0,1]} |A_t + B_t|^2\Big] \le 2\mathbb{E}\Big[\sup_{t\in[0,1]} |A_t|^2\Big] + 2\mathbb{E}\Big[\sup_{t\in[0,1]} |B_t|^2\Big] \\ &\le 2K^2\mathbb{E}\big[(\big\|\bar{Y} - y_0\big\| + 1)^2\big]\Big(|\sigma|^2\mathbb{E}\big[\,\|B\|_\varphi^2\,\big]\varphi(\varepsilon)^2 + |b - F_0(h)|^2\varepsilon^2\Big).\end{aligned}$$

By Proposition 5.2 and Lemma 5.1 $\mathbb{E}\big[(\big\|\bar{Y} - y_0\big\| + 1)^2\big]$ is bounded from above by some constant depending on K only and hence there exists a constant κ_1 depending only on K such that

$$\mathbb{E}\Big[\sup_{t\in[0,1]} |\bar{Y}_t - \widehat{Y}_t|^2\Big] \le \kappa_1\big(|\sigma|^2\varphi(\varepsilon)^2 + |b - F_0(h)|^2\varepsilon^2\big).$$

It remains to analyze

$$\mathbb{E}\Big[\sup_{t\in[0,1]} |Y_t - \widehat{Y}_t|^2\Big] \le 2\mathbb{E}\Big[\sup_{t\in[0,1]} |Y_t - \bar{Y}_t|^2\Big] + 2\mathbb{E}\Big[\sup_{t\in[0,1]} |\bar{Y}_t - \widehat{Y}_t|^2\Big].$$

With Proposition 5.2 and some constant κ_2 we get that

$$\mathbb{E}\Big[\sup_{t\in[0,1]} |Y_t - \widehat{Y}_t|^2\Big] \le \kappa_2\big(F(h) + |b - F_0(h)|^2\varepsilon^2\big)$$

for the case with $\sigma = 0$ and for general case with constants κ_3, κ_4 and κ_5

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,1]}|Y_t-\widehat{Y}_t|^2\Big] &\le \kappa_3\big(\varepsilon + F(h) + \varphi(\varepsilon)^2 + |b-F_0(h)|^2\varepsilon^2\big)\\ &\le \kappa_4\big(\varepsilon + F(h) + \varepsilon\log(e/\varepsilon) + \varepsilon\big)\\ &\le \kappa_5\big(F(h) + \varepsilon\log(e/\varepsilon)\big),\end{aligned}$$

where we used (5.16), i.e. the fact that $|F_0(h)|^2 \le K^2/\varepsilon$. □

Lemma 5.6. *For any $h > 0$ it holds*

$$\int_{B_h}|x|\nu(dx) \le \int_0^h \nu(B_u^{\mathsf{c}})du$$

Proof. We show the claim with Fubini's Theorem:

$$\begin{aligned}\int_{B_h}|x|\nu(\mathrm{dx}) &= \int_{\mathbb{R}^d}|x|\mathbb{1}_{\{x\in B_h\}}\nu(\mathrm{dx})\\ &\le \int_{\mathbb{R}^d}|x|\wedge h\nu(\mathrm{dx})\\ &= \int_{\mathbb{R}^d}\bigg(\int_0^{h\wedge|x|}1\mathrm{d}u\bigg)\nu(\mathrm{dx})\\ &= \int_{\mathbb{R}^d}\bigg(\int_0^{h}\mathbb{1}_{\{|x|>u\}}\mathrm{d}u\bigg)\nu(\mathrm{dx})\\ &= \int_{\mathbb{R}^d}\bigg(\int_0^{h}\mathbb{1}_{\{x\in B_u^{\mathsf{c}}\}}\mathrm{d}u\bigg)\nu(\mathrm{dx})\\ &= \int_0^h\bigg(\int_{\mathbb{R}^d}\mathbb{1}_{\{x\in B_u^{\mathsf{c}}\}}\nu(\mathrm{dx})\bigg)\mathrm{d}u\\ &= \int_0^h\nu(B_u^{\mathsf{c}})\mathrm{d}u.\end{aligned}$$

□

5.4.1. *Complexity.* In the following Theorem, for some positive functions f and g, we write $f \precsim g$ meaning $\limsup_{x\to a} f(x)/g(x) < \infty$ where $a \in \mathbb{R}\cup\{\infty\}$. Note that if $a \in \mathbb{R}$, then $f \precsim g$ if and only if there exists $\delta > 0$ and $\kappa > 0$ such that $f(x) \le \kappa g(x)$ for all $|a - x| < \delta$.

We consider a decreasing and invertible function $g : (0,\infty) \to (0,\infty)$ such that

$$\int_{\mathbb{R}^d}\frac{|x|^2}{h^2}\wedge 1\nu(\mathrm{dx}) \le g(h) \quad \text{for all} \quad h > 0.$$

THEOREM 5.7. [3, Theorem 1].

(i) If the driving process X has no Brownian component, i.e., $\sigma = 0$, and if there exists $\gamma > 0$ such that

$$g(h) \precsim \frac{1}{h(\log 1/h)^{1+\gamma}} \tag{5.18}$$

as $h \searrow 0$, then there exists a sequence of multilevel algorithms $\widehat{S}_n$ as introduced in Section 5.3 such that cost$(\widehat{S}_n) \leq n$ and

$$e(\widehat{S}_n) \precsim \frac{1}{\sqrt{n}}.$$

(ii) If there exists $\gamma \geq \frac{1}{2}$ such that

$$g(h) \precsim \frac{(\log 1/h)^\gamma}{h}$$

as $h \searrow 0$, then there exists a sequence of multilevel algorithms $\widehat{S}_n$ as introduced in Section 5.3 such that cost$(\widehat{S}_n) \leq n$ and

$$e(\widehat{S}_n) \precsim \frac{1}{\sqrt{n}}(\log n)^{\gamma+1}.$$

Proof. Proof of part (i)

Without loss of generality we can assume that

$$\frac{1}{h^{2/3}} \precsim g(h) \precsim \frac{1}{h(\log 1/h)^{1+\gamma}}, \tag{5.19}$$

since otherwise we can modify the function g by adding the term $\frac{1}{h^{2/3}}$, in which case the new function g would have all the wanted properties.

Our algorithm is uniquely described by parameters m and $(n_k, h_k, \varepsilon_k)_{k=1,2,\ldots,m}$. We begin by defining $h_k := g^{-1}(2^k)$ and $\varepsilon_k := 2^{-k} \leq 1$ for $k = 0, 1, \ldots$. Note that for all $x \in B_h^{\mathsf{c}}$ it holds that $|x|^2 \geq h^2$ and hence

$$\begin{aligned} \nu(B_h^{\mathsf{c}}) &= \int_{B_h^{\mathsf{c}}} 1\nu(\mathrm{dx}) \\ &\leq \int_{B_h^{\mathsf{c}}} \frac{|x|^2}{h^2} \wedge 1\nu(\mathrm{dx}) + \int_{B_h} \frac{|x|^2}{h^2} \wedge 1\nu(\mathrm{dx}) \leq g(h) \end{aligned} \tag{5.20}$$

and in particular

$$\nu(B_{h_k}^{\mathsf{c}}) \leq g(h_k) = \frac{1}{\varepsilon_k}. \tag{5.21}$$

Recall that $F(h) = \int_{B_h} |x|^2 \nu(dx)$. Furthermore, note that the inequality

$$F(h_k) \leq F(h_{k-1}) \tag{5.22}$$

follows from the facts that the function F is increasing and $h_{k-1} > h_k$. With Hölder's inequality we have

$$\begin{aligned}
|F_0(h_k) - F_0(h_{k-1})|^2 \varepsilon_{k-1}^2 &= \left| \int_{B_{h_k}^{\complement} \setminus B_{h_{k-1}}^{\complement}} x \nu(\mathrm{dx}) \right|^2 \varepsilon_{k-1}^2 \\
&\leq \left(\int_{B_{h_k}^{\complement} \setminus B_{h_{k-1}}^{\complement}} 1 \nu(\mathrm{dx}) \right) \left(\int_{B_{h_k}^{\complement} \setminus B_{h_{k-1}}^{\complement}} |x|^2 \nu(\mathrm{dx}) \right) \varepsilon_{k-1}^2 \\
&\leq \left(\int_{B_{h_k}^{\complement}} 1 \nu(\mathrm{dx}) \right) \left(\int_{B_{h_{k-1}}} |x|^2 \nu(\mathrm{dx}) \right) \varepsilon_{k-1}^2 \\
&= \nu(B_{h_k}^{\complement}) F(h_{k-1}) \varepsilon_{k-1}^2 \\
&\leq \frac{1}{\varepsilon_k} F(h_{k-1}) \varepsilon_{k-1}^2 \\
&= 2^{-k+2} F(h_{k-1})
\end{aligned}$$

and thus

$$\begin{aligned}
|b - F_0(h_{k-1})|^2 \varepsilon_{k-1}^2 &= |b - F_0(h_k) + F_0(h_k) - F_0(h_{k-1})|^2 \varepsilon_{k-1}^2 \\
&\leq 2|b - F_0(h_k)|^2 \varepsilon_{k-1}^2 + 2|F_0(h_k) - F_0(h_{k-1})|^2 \varepsilon_{k-1}^2 \\
&\leq 2|b - F_0(h_k)|^2 \varepsilon_{k-1}^2 + 2^{-k+3} F(h_{k-1}). \qquad (5.23)
\end{aligned}$$

Now by Theorem 5.5 and equation (5.23), we have for $k > 1$ that

$$\begin{aligned}
\mathbb{E}\big[\big\|\widehat{Y}^{(k)} - \widehat{Y}^{(k-1)}\big\|^2\big] &= \mathbb{E}\Big[\big(\big\|Y - \widehat{Y}^{(k)}\big\| + \big\|Y - \widehat{Y}^{(k-1)}\big\|\big)^2\Big] \\
&\leq 2\mathbb{E}\big[\big\|Y - \widehat{Y}^{(k)}\big\|^2\big] + 2\mathbb{E}\big[\big\|Y - \widehat{Y}^{(k-1)}\big\|^2\big] \\
&\leq \kappa_1\big(F(h_k) + |b - F_0(h_k)|^2 \varepsilon_k^2 + F(h_{k-1}) + |b - F_0(h_{k-1})|^2 \varepsilon_{k-1}^2\big) \\
&\leq \kappa_2\big(2F(h_{k-1}) + |b - F_0(h_k)|^2 \varepsilon_k^2 + |b - F_0(h_{k-1})|^2 \varepsilon_{k-1}^2\big) \\
&\leq \kappa_3\big(F(h_{k-1}) + |b - F_0(h_k)|^2 \varepsilon_{k-1}^2\big) \qquad (5.24)
\end{aligned}$$

for some constants $\kappa_1, \kappa_2, \kappa_3 > 0$.

Now by (5.5), Lemma 5.1, Theorem 5.5 and (5.24) we can estimate the error of the algorithm for some constant $\kappa_4 > 0$ by

$$\begin{aligned}
e^2(\widehat{S}) &\leq \mathbb{E}[\big\|Y - Y^{(m)}\big\|^2] + \sum_{k=2}^{m} \frac{1}{n_k} \mathbb{E}[\big\|Y^{(k)} - Y^{(k-1)}\big\|^2] + \frac{1}{n_1} \mathbb{E}[\big\|Y^{(1)} - y_0\big\|^2] \\
&\leq \kappa_4 \sum_{k=1}^{m+1} \frac{1}{n_k} [F(h_{k-1}) + |b - F_0(h_k)|^2 \varepsilon_{k-1}^2],
\end{aligned}$$

where we set $n_{m+1} = 1$.

By Lemma 5.6 and Assumption (5.18) with an appropriate constant $\kappa_5 > 0$ and sufficiently small $r \in (0,1)$, we have

$$\begin{aligned}
|F_0(h_k)| &= \Big| \int_{B_{h_k}^{\mathsf{c}}} x\nu(\mathrm{dx}) \Big| \le \int_{\mathbb{R}^d} |x|\nu(\mathrm{dx}) \le \frac{1}{r} \int_{\mathbb{R}^d} |x|(r \vee |x|)\nu(\mathrm{dx}) \\
&= \frac{1}{r} \int_{B_r^{\mathsf{c}}} |x|^2 \nu(\mathrm{dx}) + \int_{B_r} |x|\nu(\mathrm{dx}) \\
&\le \frac{1}{r} \int_{B_r^{\mathsf{c}}} |x|^2 \nu(\mathrm{dx}) + \int_0^r \nu(B_u^{\mathsf{c}})\mathrm{d}u \\
&\le \frac{1}{r} \int |x|^2 \nu(\mathrm{dx}) + \int_0^r g(u)\mathrm{d}u \\
&\le \frac{1}{r} \int |x|^2 \nu(\mathrm{dx}) + \kappa_5 \int_0^r \frac{1}{u(\log(1/u))^{1+\gamma}} \mathrm{d}u,
\end{aligned}$$

where both integrals are finite. Additionally, we have

$$\begin{aligned}
|F(h_k)| = \int_{B_{h_k}} |x|^2 \nu(\mathrm{dx}) &= h_k^2 \int_{B_{h_k}} \frac{|x|^2}{h_k^2} \nu(\mathrm{dx}) \\
&\le h_k^2 \int \frac{|x|^2}{h_k^2} \wedge 1 \nu(\mathrm{dx}) \\
&\le h_k^2 g(h_k) = g^{-1}(2^k) 2^k. \qquad (5.25)
\end{aligned}$$

Note that for (5.19), with some $M > 0$, we see that for $y = \frac{1}{Mh^{2/3}}$

$$\begin{aligned}
\frac{1}{h^{2/3}} \precsim g(h) &\iff \frac{1}{h^{2/3}} \le M g(h) \\
&\iff \frac{1}{y^{3/2}} \le M^{3/2} g^{-1}(y) \\
&\iff \frac{1}{y^{3/2}} \precsim g^{-1}(y) \quad \text{as} \quad y \to \infty. \qquad (5.26)
\end{aligned}$$

Additionally, from (5.19) for some $M > 0$ we have

$$g(h) \le \frac{M}{h(\log 1/h)^{1+\gamma}} \quad \text{as} \quad h \to 0$$

and therefore we have, setting $h = \frac{1}{y}$,

$$g\Big(\frac{1}{y}\Big) \le \frac{My}{(\log y)^{1+\gamma}} \quad \text{as} \quad y \to \infty.$$

We apply the decreasing function g^{-1} to the inequality.

$$\frac{1}{y} \ge g^{-1}\Big(\frac{My}{(\log y)^{1+\gamma}}\Big).$$

We choose $y := z(\log z)^{1+\gamma}/M$, and note that then $y \to \infty \iff z \to \infty$. By the properties of the logarithm and the fact that $\log((\log z)^{1+\gamma}) - (\log M) \ge 0$ for large z we have

$$\frac{M}{z(\log z)^{1+\gamma}} \ge g^{-1}\Big(\frac{z(\log z)^{1+\gamma}}{\big(\log z + \log((\log z)^{1+\gamma}) - \log M\big)^{1+\gamma}}\Big) \ge g^{-1}(z)$$

as $z \to \infty$. Therefore we have

$$g^{-1}(y) \precsim \frac{1}{y \log(y)^{1+\gamma}} \quad \text{as} \quad y \to \infty. \tag{5.27}$$

Hence, with equation (5.26), we have $2^k g^{-1}(2^k)^2 \succsim 2^{-2k} = \varepsilon_k^2$ as k tends to infinity. By equation (5.25) and in view of the fact that $|F_0(h_k)|$ is bounded, there exists a constant κ_5 such that

$$e^2(\widehat{S}) \leq \kappa_4 \sum_{k=1}^{m+1} \frac{1}{n_k} [F(h_{k-1}) + |b - F_0(h_k)|^2 \varepsilon_{k-1}^2] \leq \kappa_5 \sum_{k=1}^{m+1} \frac{1}{n_k} 2^{k-1} g^{k-1}(2^{-1})^2.$$

Next we will fix the parameters m and $n_1, \ldots, n_m$. For some given parameter $Z \geq \frac{1}{g^{-1}(1)}$, we choose

$$m = m(Z) = \inf\{k \in \mathbb{N} : Z g^{-1}(2^k) < 1\} - 1$$

and set

$$n_k = n_k(Z) = \lfloor Z g^{-1}(2^{k-1}) \rfloor$$

for $k = 1, \ldots, m$ and $n_{m+1} = 1$. Now from definition of m we know that $Z g^{-1}(2^{k-1}) \geq 1$ for all $k = 1, \ldots, m+1$, which implies $1/n_k = 1/\lfloor Z g^{-1}(2^{k-1}) \rfloor \leq 2/Z g^{-1}(2^{k-1})$ and therefore

$$e^2(\widehat{S}) \leq \kappa_5 \sum_{k=1}^{m+1} \frac{1}{n_k} 2^{k-1} g^{-1}(2^{k-1})^2 \leq \frac{2\kappa_5}{Z} \sum_{k=1}^{m+1} 2^{k-1} g^{-1}(2^{k-1}). \tag{5.28}$$

By (5.27) we have that $2^k g^{-1}(2^k) \precsim k^{-(1+\gamma)}$ and hence for some constant $M > 0$ we have

$$\sum_{k=1}^{m+1} 2^{k-1} g^{-1}(2^{k-1}) \leq g^{-1}(1) + M \sum_{k=2}^{m+1} (k-1)^{-(1+\gamma)} \tag{5.29}$$

so we see that the latter sum in (5.28) is uniformly bounded for all m. Hence, there exists a constant κ_6 depending only on g and K such that

$$e^2(\widehat{S}) \leq \kappa_6 \frac{1}{Z}. \tag{5.30}$$

Finally, we will analyze the cost of the algorithm. To be able to do that, we first need to evaluate the expected number of breakpoints of $\widehat{Y}^{(k)}$. Recall that there are two sources of breakpoints, the large and the small jumps. The large jumps were defined as jumps greater than h_k, and hence by Definition 2.44 of the Lévy measure we see that the expected number of large jumps is $\nu(B_{h_k}^{\mathfrak{c}})$. After setting the large jumps we then added small ones such that the jump size remained less than ε_k, and hence with equation (5.21) we see that

$$\mathbb{E}\big[\Upsilon(\widehat{Y}^{(k)})\big] \leq \frac{1}{\varepsilon_k} + \nu(B_{h_k}^{\mathfrak{c}}) \leq 2^{k+1}.$$

Thus, by using again (5.29) we have

$$\begin{aligned}\mathrm{cost}(\widehat{S}) &\leq \sum_{k=1}^{m} n_k \mathbb{E}\big[\Upsilon(Y^{(k)})\big] \leq \sum_{k=1}^{m} 2^{k+1} n_k \\ &\leq 4Z \sum_{k=1}^{m} 2^{k-1} g^{-1}(2^{k-1}) \\ &\leq \kappa_7 Z \end{aligned} \tag{5.31}$$

for some appropriate constant $\kappa_7 > 0$. Then we choose $Z = n$ in (5.30) and (5.31) and the result follows.

Proof of part (ii)

As in part (i), we can assume without loss of generality that g satisfies

$$\frac{\sqrt{\log 1/h}}{h} \precsim g(h) \precsim \frac{(\log 1/h)^{\gamma}}{h} \tag{5.32}$$

and again, we choose $h_k := g^{-1}(2^k)$ and $\varepsilon_k := 2^{-k}$ for $k = 0, 1, \ldots$. Note that we still have that $\nu(B_{h_k}^{\mathsf{c}}) \leq g(h_k) = 1/\varepsilon_k$, $\varepsilon_k \leq 1$ and that $F(h_k) \leq g(h_k) h_k^2 = 2^k g^{-1}(2^k)^2$ as explained in (5.21) and (5.25). By a simple computation we see that $\varepsilon_k \log(e/\varepsilon_k) \leq \varepsilon_{k-1} \log(e/\varepsilon_{k-1})$ for $k \geq 1$ and hence, by Theorem 5.5 and equation (5.22) we have that

$$\begin{aligned}\mathbb{E}\big[\big\|\widehat{Y}^{(k)} - \widehat{Y}^{(k-1)}\big\|^2\big] &\leq 2\mathbb{E}\big[\big\|Y - \widehat{Y}^{(k)}\big\|^2\big] + 2\mathbb{E}[\big\|Y - \widehat{Y}^{(k-1)}\big\|^2] \\ &\leq \kappa_1(\varepsilon_k \log(e/\varepsilon_k) + F(h_k) + \varepsilon_{k-1}\log(e/\varepsilon_{k-1}) + F(h_{k-1})) \\ &\leq \kappa_2(F(h_{k-1}) + \varepsilon_{k-1}\log(e/\varepsilon_{k-1}))\end{aligned}$$

for some constants $\kappa_1, \kappa_2 > 0$. Now, like in part (i), we can estimate the error of the algorithm by

$$\begin{aligned}e^2(\widehat{S}) &\leq \mathbb{E}[\Big\|Y - \widehat{Y}^{(m)}\Big\|^2] + \sum_{k=2}^{m} \frac{1}{n_k}\mathbb{E}[\big\|Y^{(k)} - Y^{(k-1)}\big\|^2] + \frac{1}{n_1}\mathbb{E}[\big\|Y^{(1)} - y_0\big\|^2] \\ &\leq \kappa_3 \sum_{k=1}^{m+1} \frac{1}{n_k}\left[F(h_{k-1}) + \varepsilon_{k-1}\log\left(\frac{e}{\varepsilon_{k-1}}\right)\right] \\ &\leq \kappa_3 \sum_{k=1}^{m+1} \frac{1}{n_k}\Big[2^{k-1} g^{-1}(2^{k-1})^2 + 2^{-(k-1)}\log(e2^{k-1})\Big]\end{aligned}$$

for some $\kappa_3 > 0$. With similar computations as in (5.27), which we do not repeat, we see that from (5.32) it follows that

$$\frac{\sqrt{\log y}}{y} \precsim g^{-1}(y) \precsim \frac{(\log y)^{\gamma}}{y} \text{ as } y \to \infty \tag{5.33}$$

By a simple computation we have that

$$2^{-k}\log(e2^k) = 2^k e\Big(\frac{\sqrt{\log(e2^k)}}{e2^k}\Big)^2 \precsim 2^k g^{-1}(e2^k)^2 \leq 2^k g^{-1}(2^k)^2$$

where the last step follows from the fact that the inverse of function g is decreasing. Thus, we can find a constant κ_4 such that

$$e^2(\widehat{S}) \leq \kappa_4 \sum_{k=1}^{m+1} \frac{1}{n_k} 2^{k-1} g^{-1}(2^{k-1})^2.$$

Again, to finish the evaluation of the error, for $Z \geq e \vee \big(1/g^{-1}(1)\big)$ we choose $m = m(Z) = \inf\{k \in \mathbb{N} : Zg^{-1}(2^k) < 1\} - 1$ and set $n_k = n_k(Z) = \lfloor Zg^{-1}(2^{k-1}) \rfloor$ for $k = 1, \ldots, m$. Like in previous part, note that $1/n_k \leq 2/Zg^{-1}(2^{k-1})$. Now with (5.33) we see that

$$\begin{aligned} e^2(\widehat{S}) &\leq 2\kappa_4 \frac{1}{Z} \sum_{k=1}^{m+1} 2^{k-1} g^{-1}(2^{k-1}) \\ &\leq \kappa_5 \frac{1}{Z} \sum_{k=1}^{m+1} \big(\log(2^{k-1})\big)^\gamma \\ &\leq \kappa_6 \frac{1}{Z} \sum_{k=1}^{m+1} (k-1)^\gamma \\ &\leq \kappa_6 \frac{1}{Z} m^{\gamma+1} \end{aligned}$$

for $\kappa_5, \kappa_6 > 0$. Furthermore, from the choice of m it follows that $Zg^{-1}(2^m) \geq 1$, and it is easy to see that $\log\big((\log Z)^\gamma\big) \leq \log Z$ for large Z. Thus, by (5.32) with some constant $M > 0$ we obtain

$$\begin{aligned} m &\leq \log_2\left(g\Big(\frac{1}{Z}\Big)\right) = \frac{\log\big(g(1/Z)\big)}{\log(2)} \\ &\leq \frac{M}{\log(2)} \log\Big(Z\big(\log Z\big)^\gamma\Big) \\ &\leq \frac{M}{\log(2)} \Big(\log Z + \log\big((\log Z)^\gamma\big)\Big) \\ &\leq \frac{2M}{\log(2)} \log(Z) \\ &\precsim \log(Z) \end{aligned} \tag{5.34}$$

as $Z \to \infty$. Hence, we can estimate the error with

$$e^2(\widehat{S}) \leq \kappa_7 \frac{(\log Z)^{\gamma+1}}{Z}$$

for some positive constant κ_7. Similarly, with the definition of n_k and the equations (5.33) and (5.34) we have

$$\begin{aligned}\operatorname{cost}(\widehat{S}) &\leq \sum_{k=1}^{m} 2^{k+1} n_k \leq \sum_{k=1}^{m} 2^{k+1} Z g^{-1}(2^{k-1}) \\ &\leq \kappa_8 Z \sum_{k=1}^{m} (k-1)^{\gamma} \leq \kappa_8 Z m^{\gamma+1} \\ &\leq \kappa_9 Z (\log Z)^{\gamma+1}\end{aligned}$$

with $\kappa_8, \kappa_9 > 0$. Finally, we choose

$$Z = Z(n) = \frac{1}{2\kappa_9} \frac{n}{(\log n)^{\gamma+1}}$$

for $n \geq e$ sufficiently large such that $Z \geq e \vee \big(1/g^{-1}(1)\big)$. Then

$$\lim_{n\to\infty} \frac{\log(Z)}{\log(n)} = \lim_{n\to\infty} \frac{\log(1/2\kappa_9)}{\log(n)} + \frac{\log(n)}{\log(n)} - \frac{\log((\log n)^{\gamma+1})}{\log(n)} = 1$$

which implies

$$\begin{aligned}\lim_{n\to\infty} \frac{\operatorname{cost}(\widehat{S})}{n/2} &\leq \lim_{n\to\infty} \frac{\kappa_9 Z (\log Z)^{\gamma+1}}{n/2} \\ &= \lim_{n\to\infty} \frac{n/2}{n/2} \frac{(\log Z)^{\gamma+1}}{(\log n)^{\gamma+1}} = 1\end{aligned}$$

and hence, we have for sufficiently large n that $\operatorname{cost}(\widehat{S}) \leq n$. Additionally, for the error we have

$$\begin{aligned}\limsup_{n\to\infty} \frac{\kappa_7 (\log Z)^{\gamma+1}}{Z} \frac{n}{(\log n)^{2(\gamma+1)}} &= \limsup_{n\to\infty} \frac{2\kappa_7\kappa_9 (\log Z)^{\gamma+1}}{n} \frac{n}{(\log n)^{\gamma+1}} \\ &= 2\kappa_7\kappa_9 \limsup_{n\to\infty} \frac{(\log Z)^{\gamma+1}}{(\log n)^{\gamma+1}} < \infty,\end{aligned}$$

which implies

$$e^2(\widehat{S}) \precsim \frac{(\log n)^{2(\gamma+1)}}{n}.$$

□

6. Afterword

This text covers the basic idea behind the multilevel Monte Carlo algorithm applied to a functional of a solution to an SDE with jumps. The algorithm is based on the method where one approximates the solution to the SDE with refining time grids. In general, finer grid leads to more time consuming algorithm. The multilevel approach exploits the grids with refining grid size to achieve better accuracy with lower cost. On each level it uses a discretization method. Here only the Euler scheme was introduced, but actually any discretization with certain properties would do. One example is the Milstein scheme, which has been widely studied. Both schemes are simple examples of Taylor approximations.

The multilevel algorithm provides considerably lower error rates compared to traditional Monte Carlo. Yet, it is not significantly more complicated to put into

practice. The article [4] in which the Section 4 is based on is one of the very first papers of this subject. The multilevel algorithm is a current topic in research and there are new applications rising up at steady pace.

References

[1] David Applebaum. *Lévy processes and stochastic calculus*. Cambridge University Press, 2009.

[2] Rama Cont and Peter Tankov. *Financial modelling with jump processes*. Chapman and Hall, 2004.

[3] Steffen Dereich and Felix Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Processes and their Applications*, 121:1565–1587, 2011.

[4] Michael B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[5] Mircea Grigoriu. *Stochastic calculus: applications in science and engineering*. Birkhäuser Boston, 2002.

[6] Jean Jacod and Albert Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, 2002.

[7] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2000.

[8] Peter E. Kloeden and Eckhard Platen. *Numerical Solutions of Stochastic Differential Equations*. Springer Verlag, Berlin, 1999.

[9] Phillip E. Protter. *Stochastic Integration and Differential equations: A New Approach*. Springer Verlag, Berlin, 1995.