Quadrature for Path-dependent Functionals of Lévy-driven Stochastic Differential Equations

Felix Heidenreich

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation.

1. Gutachter: Prof. Dr. Klaus Ritter
2. Gutachter: Prof. Dr. Michael Giles

Tag der Einreichung: 30. Mai 2012
Tag der Disputation: 26. Oktober 2012

D 386

Contents

1 Introduction

The main topic of this thesis is to define and analyze a multilevel Monte Carlo algorithm for path-dependent functionals of the solution of a stochastic differential equation (SDE)

$$\begin{aligned} \mathrm{d}Y_t &= a\,(Y_{t-})\ \mathrm{d}X_t, \qquad t \in [0,1], \\ Y_0 &= y_0, \end{aligned} \tag{1.1}$$

which is driven by a Lévy process X. To be more precise, we work with standard Lipschitz assumptions for the diffusion coefficient a, the driving process X is d_X-dimensional and square integrable and $Y = (Y_t)_{t\in[0,1]}$ is the d_Y-dimensional strong solution process of (1.1) with deterministic initial value $y_0 \in \mathbb{R}^{d_Y}$. We investigate the computation of expectations

$$S(f) = \mathrm{E}[f(Y)],$$

where $f : D[0,1] \to \mathbb{R}$ is a measurable function mapping from the path space of the solution process Y to the reals and is Lipschitz continuous with respect to the supremum norm. The entire assumptions of the underlying problem are defined in more detail in Chapter 3 in Assumptions A (3.2).

Problems of the above kind arise, for instance, in financial mathematics in the context of option pricing of European options. There, (the logarithm of) the underlying stock price is modeled by an SDE, and the function is given by the (discounted) payoff function of the considered option. We emphasize, that our setting allows to consider path-dependent options like Asian, lookback or barrier options.

We are interested in the relation of the error and the computational cost of randomized algorithms $\widehat{S}$, which approximately compute this value for the considered class of functions. For a given SDE we consider as error criterion the worst case of

the root mean square error over the given class of functionals F, i.e., we define the error $e(\widehat{S})$ of an algorithm $\widehat{S}$ by

$$e^2\left(\widehat{S}\right) = \sup_{f\in F} \mathrm{E}\left[\left|S(f) - \widehat{S}(f)\right|^2\right]. \tag{1.2}$$

The computational cost of an algorithm $\widehat{S}$, denoted $\mathrm{cost}(\widehat{S})$, should represent the runtime of the algorithm on a computer. We work in the real number model of computation, which means that we assume that arithmetic operations with real numbers and comparisons can be done in one time unit. We further suppose that evaluations of a are possible at any point $y \in \mathbb{R}^{d_Y}$ in constant time and evaluations of f are possible for piecewise constant functions in time units according to its number of breakpoints plus one. We also assume that sampling from the uniform distribution on $[0,1]$ and from the suitably restricted Lévy measure are possible in constant time. Here, the Lévy measure is restricted to regions bounded away from the origin by open balls of radius $h > 0$ and rescaled to a probability measure.

The most natural choice of an algorithm for the computation of $S(f)$ is the classical Monte Carlo simulation S^{MC}, whose output $S^{MC}(f)$ for a function f is the average over the function evaluations of an i.i.d. sequence $(Y_i)_{i=1,\dots,n}$ with distribution P_Y. The root mean square error of the latter is given by $n^{-1/2}$ times the standard deviation of $f(Y)$. In most computational problems, the distribution of Y is only implicitly given, e.g. in terms of the solution of an SDE, and has to be suitably approximated by a random element $\hat{Y}$ whose distribution can be simulated. The corresponding Monte Carlo algorithm $\widehat{S}^{MC}$ has the output

$$\widehat{S}^{MC}(f) = \frac{1}{n}\sum_{i=1}^{n} f\left(\hat{Y}_i\right),$$

for a given function f, where $(\hat{Y}_i)_{i=1,\dots,n}$ is an i.i.d. sequence with distribution $P_{\hat{Y}}$. The resulting root mean square error can then be decomposed into

$$e^2\left(\widehat{S}^{MC}(f)\right) = \Bigg|\underbrace{\mathrm{E}\left[f(Y)\right] - \mathrm{E}\left[f\left(\hat{Y}\right)\right]}_{=\mathrm{bias}(\widehat{S}^{MC}(f))}\Bigg|^2 + \frac{1}{n}\mathrm{var}\left(f\left(\hat{Y}\right)\right).$$

The first term is due to the usage of an approximation $\hat{Y}$ and is called the bias of the algorithm, while the second term is the statistical error of the Monte Carlo algorithm itself.

A standard idea to improve the trade-off between error and cost in this Monte Carlo algorithm setting is to find good approximations of Y, i.e., to find approximations $\hat{Y}$ with a small bias with respect to its computational cost, and to apply standard Monte Carlo. There are various weak approximation results available for the case that f depends only on the endpoint marginal of Y, see [36] for the standard Euler and [23] for the approximate Euler method. Both provide explicit error expansions in the first moments of the inverse of the step size. In this cases, the convergence order can be further improved by the Romberg extrapolation technique as described in [43] for the classical diffusion case. There are other higher-order weak approximation results available, e.g., in [44] via an Markov operator approach, in [46] for a jump-adapted scheme with Gaussian correction for the neglected small jumps and an additional approximation for the SDE between the jump times, or in the recent article [45] for a similar jump-adapted scheme with non-Gaussian correction for the neglected small jumps with matching first three or more moments.

In our setting of global errors, i.e., considering the supremum norm on the path instead of the endpoint, there are no higher order weak approximation schemes available such that we use a multilevel approach originating from an idea of Stefan Heinrich in [18] for parametric integration. A special case of such a multilevel scheme with only two levels appearing has been used in [26]. There, a control variate type variance reduction technique, called the statistical Romberg method, is used. The multilevel scheme for a Brownian SDE has then been defined and analyzed in [16] and a complexity analysis of the scheme can be found in [34]. The results demonstrate its superiority to all known Monte Carlo methods so far, by reaching convergence order $1/2$ in terms of the computational cost up to some logarithmic terms.

For the multilevel scheme, one needs a hierarchy of approximation schemes for the solution of the SDE, denoted $\hat{Y}^{(1)}, \hat{Y}^{(2)}, \ldots$, with accuracy increasing with the upper index. The idea is to split the expectation of a fixed accuracy level m into a

telescoping sum of all approximations with lower accuracy by

$$\mathrm{E}\left[f\left(\hat{Y}^{(m)}\right)\right] = \mathrm{E}\left[f\left(\hat{Y}^{(1)}\right)\right] + \sum_{k=2}^{m} \mathrm{E}\left[f\left(\hat{Y}^{(k)}\right) - f\left(\hat{Y}^{(k-1)}\right)\right],$$

and to compute the expectations on the right hand side by independent classical Monte Carlo methods. The resulting multilevel scheme features the bias of the highest accuracy, while the variance is given by the sum of variances of the right hand side variables. If we now properly couple the approximations of $\hat{Y}^{(k)}$ and $\hat{Y}^{(k-1)}$ for $k = 2, \ldots, m$, the variance of $(f(\hat{Y}^{(k)}) - f(\hat{Y}^{(k-1)}))$ decreases in k and the number of replications needed to guarantee a desired precision decreases as well. Along this way, we can substantially reduce the computational effort.

We will use an approximate Euler scheme, i.e. an Euler scheme with a compound Poisson approximation $\hat{X}^{(h,\epsilon)}$ of the driving process on a non-deterministic grid. The latter arises from X by neglecting jumps smaller $h > 0$ and taking a random time discretization including all times where X jumps with magnitude greater h and having step size at most $\varepsilon > 0$. For this scheme, strong error estimates are provided in Chapter 3, and we will take as weak error the one induced by the strong one.

The Euler scheme for a Lévy-driven SDE has already been analyzed in a variety of articles. In [27], the more general case of a semimartingale as driving process has been considered and the global strong approximation has been investigated to show uniform convergence on compacts in probability of the scheme. For Lévy processes the convergence rate obtained is approximately $1/2$ in terms of the step size of the Euler scheme. As the explicit simulation of increments of X is only possible in a few cases, there has quite recently been articles on the approximate Euler scheme, which uses a compound Poisson approximation of X. In [39], this scheme has been analyzed in terms of a limit theorem, showing convergence in law of the error process multiplied with a rate function of the right order, depending amongst other things on the Lévy measure. The weak approximation of the approximate Euler scheme has been analyzed in [23] while a strong approximation error recently has been presented in [15], where an additional Gaussian term compensating the neglected small jumps is considered, following the idea of [2]. The above references for the approximate Euler scheme are again only covering the case of endpoint marginals in the weak error case and the error appearing in the discretization points in the

strong approximation case.

Together with the strong error estimates from Chapter 3, the multilevel algorithm leads to upper bounds for the error of the underlying quadrature problem by considering for F the Lipschitz class of measurable functionals on the Skorohod space $D[0,1]$ of càdlàg functions, that are Lipschitz continuous with coefficient 1 with respect to supremum norm. We can summarize the main results of Chapter 4 in terms of the Blumenthal-Getoor index of the driving Lévy process, denoted by $\beta \in [0,2]$, which measures the frequency of occurrence of small jumps with sizes around the origin. For $\beta < 1$ and no Brownian component present, we almost reach convergence order $1/2$, which means, that there exists a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \leq n$ such that $e(\widehat{S}_n) \precsim n^{-1/2}$. Here, by $\precsim$, we denote a weak asymptotic upper bound, i.e. the inequality holds up to an unspecified positive constant. If X has a Brownian component, the order has an additional logarithmic term, in which case, we reach $e(\widehat{S}_n) \precsim n^{-1/2}\,(\log(n))^{3/2}$.

The higher β is, the more frequent the small jumps appear and the worse is our approach of neglecting them. For $\beta \geq 1$, we come arbitrary close to the convergence order $1/\beta - 1/2$, which unfortunately tends to zero for $\beta \to 2$. For this case, an improved algorithm is already defined in [12]. There, an additional Gaussian correction term, which recovers the covariance structure of the neglected small jumps, improves the weak error estimate such that a convergence order, which is arbitrary close to $(4-\beta)/(6\beta)$, can be obtained.

For the special subclass of Y being the Lévy process itself, we also provide a lower bound, which, up to a logarithmic term, recovers the order $1/2$, i.e. neglecting logarithmic terms, the multilevel algorithm is order optimal for $\beta < 1$.

The multilevel Monte Carlo algorithm was first introduced by Stefan Heinrich in [18] in a different setting, namely parametric integration of integral equations, and further presented in a special article on these algorithms in [19]. In 2006, Mike Giles introduced the multilevel scheme in [16] in the context of financial mathematics for the calculation of expectations with respect to (marginals of) diffusion processes given by the solution of an SDE with Brownian motion as driving process. Since then, there has been a wide range of applications and new fields of research, where the multilevel idea has been successfully installed. In the case of a Brownian SDE, we mention [4], where discontinuous functions are considered, [9], where the compu-

tation of Greeks via multilevel is outlined as well as [21] for the calculation of mean exit times. In [11], the authors show the asymptotical optimality of the multilevel scheme in a variety of computational problems where the desired distribution of Y is in a suitable scale of Gaussian distributions. Furthermore, the multilevel scheme also applies in the quadrature problem on the sequence space in [20], where it can also be combined with quasi Monte Carlo methods as already considered in [17], as well as in the context of SPDEs in [6].

In the case of a Lévy driven SDE, the results of [16] has been extended in [31], combining the strong approximation result of [15] and the weak approximation of [23]. The results hold for expectations with respect to marginals of the solution process and the assumptions on F and the diffusion coefficient are more restrictive than those presented here. The statement itself remains pretty similar to [16] up to the possibility of a non-linear growth of the computational cost with respect to the step-size of the deterministic grid of the Euler scheme. The particular case of a jump-diffusion as driving process is studied in [48], where, in contrast to our case, the jump-intensity also may depend on the value of the solution process.

This thesis is organized as follows: In Chapter 2, we shortly review the characteristics and some basic facts on Lévy processes needed in the following. We also introduce some examples of driving Lévy processes, which will be revisited after each chapter to apply the received results. In Chapter 3, we define the Euler approximation scheme with random time discretization used in our multilevel scheme. Strong error estimates for the latter with respect to the second moment of the supremum norm on $[0, 1]$ are presented. With their help, we define and analyze the multilevel scheme in Chapter 4. The asymptotic choice of parameters in the multilevel scheme as occurring in the proof, are separately stated. A different semi-heuristic method to implement the multilevel algorithm is outlined in Chapter 5 together with two numerical examples recovering our theoretical results.

1.1 Notation

Throughout this thesis, we will denote by $\langle ., . \rangle$ the standard scalar product on $\mathbb{R}^d$ and by $|.|$ the corresponding Euclidean norm for vectors and the Frobenius norm for matrices, i.e. for $A \in \mathbb{R}^{m \times n}$ we set $|A|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}|^2$. The operator norm

will be denoted by $|.|_{\text{op}}$ and for $h > 0$, the open ball around the origin with radius h is given by $B_h = \{x \in \mathbb{R}^d : |x| < h\}$. Furthermore, the space of $\mathbb{R}^d$-valued càdlàg functions on $[0,1]$ is denoted by $D[0,1]$, endowed with the Borel-σ-field of the Skorohod topology, which is equivalent to the trace of the product-σ-field on $\mathbb{R}^{[0,1]}$ in $D[0,1]$ or the σ-field induced by the projections on finite-dimensional marginals. We are interested in the global error of our approximations in the interval $[0,1]$, which is given by the sup norm, which we denote by $\|X\| = \sup_{t\in[0,1]} |X_t|$. We further denote by $\text{Lip}(1)$ the class of Borel measurable functions $f : D[0,1] \to \mathbb{R}$ which are Lipschitz continuous with coefficient 1 with respect to supremum norm. The indicator function of a set $A \in \mathbb{R}^d$ will be written by $\mathbb{1}_A : \mathbb{R}^d \to \{0,1\}$. We will further consider the following definitions of asymptotic relations. For positive functions f and g, we write $f \precsim g$ if $\limsup_{x\to a} f(x)/g(x) < \infty$, i.e. if there exists a constant $\kappa > 0$ such that $f(x) \leq \kappa \cdot g(x)$ for $x \to a$, where typically $a = 0$ or $a = \infty$. If both $f \precsim g$ and $g \precsim f$ holds, we write $f \asymp g$. The strong asymptotic equivalence $\lim f/g = 1$ will be denoted by $f \approx g$ and for a strong asymptotic upper bound we write $f \lesssim g$, which means $\lim f/g \leq 1$.

2 Lévy processes

In this chapter, we will give a very brief introduction to Lévy processes and present useful and neccessary results for the definition and the error analysis of our multilevel scheme. Most of the proofs and ideas can be found in monographs, in which cases the references are outlined in the text. Popular reading references for the introduction to Lévy processes are [1], [37], [41], [7] and [29] as well as for the use in finance the monographs [42] and [10].

2.1 Definition

In the sequel, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ always denotes a filtered probability space satisfying the usual hypthesis of right-continuity and completeness.

Definition 1 (Lévy process)**.** *On $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, an adapted stochastic process $X = (X_t)_{t\geq 0}$ is called a Lévy process if*

- *$X_0 = 0$ P-a.s.,*
- *X is P-a.s. càdlàg (right-continuous with left limits),*
- *X has independent increments, i.e. for all $n \in \mathbb{N}$ and all sequences $0 = t_0 \leq t_1 < t_2 < \ldots < t_n < \infty$, $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent,*
- *X has stationary increments, i.e. for all $0 \leq s < t < \infty$, $X_t - X_s$ is distributed like X_{t-s}.*

Remark 1. The augmented natural filtration $(\mathcal{F}_t^X)_{t\geq 0}$ of a Lévy process X, i.e. the natural filtration of X with $\mathcal{F}_0^X$ containing all sets of $\mathcal{F}$ with P-measure zero, always satisfies the usual hypothesis of right-continuity and completeness, see [37, p. 22] or [1, p. 88].

For an $\mathbb{R}^{d_X}$-valued Lévy process $X = (X_t)_{t\geq 0}$, the marginals X_t are always infinitely divisible for all $t \geq 0$. Due to the independent and stationary increments, the characteristic function can thus be given in terms of the characteristic exponent $\Psi : \mathbb{R}^{d_X} \to \mathbb{C}$ of X_1, which is $\Psi(u) = \log(\mathrm{E}[\exp(i\langle u, X_1\rangle)])$. It clearly holds

$$\mathrm{E}\left[\exp\left(i\langle u, X_t\rangle\right)\right] = \exp\left(t\Psi(u)\right) \tag{2.1}$$

for all $t \geq 0$ and all $u \in \mathbb{R}^{d_X}$. Furthermore, the Lévy-Khintchine formula, see [1, p. 29] provides an explicit form for the characteristic exponent Ψ of this infinitely divisible random variable. We denote by $b \in \mathbb{R}^{d_X}$ a vector, by $A \in \mathbb{R}^{d_X \times d_X}$ a positive semidefinit symmetric matrix and by ν a Lévy measure on $\mathbb{R}^{d_X}$, that is a Borel measure on $\mathbb{R}^{d_X}$ which satisfies $\nu(\{0\}) = 0$ and

$$\int_{\mathbb{R}^{d_X}} |x|^2 \wedge 1 \, \nu(\mathrm{d}x) < \infty.$$

Then, for the infinitely divisible random variable X_1, there exist b, A and ν as above such that for all $u \in \mathbb{R}^{d_X}$

$$\Psi(u) = i\langle b, u\rangle - \frac{1}{2}\langle u, Au\rangle + \int_{\mathbb{R}^{d_X}} \left[\exp(i\langle u, x\rangle) - 1 - i\langle u, x\rangle \mathbb{1}_{B_1}(x)\right] \nu(\mathrm{d}x). \tag{2.2}$$

Clearly, the distribution of X is completely determined by the three parameters b, A and ν. Here, b is called the drift, A is called the Gauss coefficient and ν is called the Lévy measure of X.

The inverse implication holds as well, i.e. for every infinitely divisible distribution μ on $\mathbb{R}^{d_X}$, there exists an $\mathbb{R}^{d_X}$-valued Lévy process X with distribution μ as marginal distribution at time $t = 1$, see [1, p. 65].

Observe that the truncation function $\mathbb{1}_{B_1}(x)$ assures the integrability of the integral in (2.2). In the literature, this truncation function is sometimes chosen in slightly different ways, e.g. by a continuous function which approximates the indicator function. This change has an obvious impact on the drift b.

In this work, we deal with square integrable Lévy processes, in which case the truncation function can be ommited as the Lévy measure then fulfills $\int |x|^2 \, \nu(\mathrm{d}x) < \infty$. In the sequel, we assume b to be the drift with no truncation function in the Lévy-Khintchine-representation, i.e. the characteristic function from (2.2) changes

to

$$\Psi(u) = i\langle b, u\rangle - \frac{1}{2}\langle u, Au\rangle + \int_{\mathbb{R}^{d_X}} [\exp(i\langle u, x\rangle) - 1]\ \nu(\mathrm{d}x).$$

2.2 The Lévy-Itô decomposition

To obtain a better feeling for the parameters of X, we provide here a sketch of the Lévy-Itô decomposition for Lévy processes. Roughly speaking, we can split any Lévy process into the sum of three independent parts, namely a Wiener process, an L_2 jump-martingale and a compound Poisson part with drift. In our case of square integrability, the compound Poisson part can as well be written as part of the L_2 jump-martingale part. A proof of this decomposition can be found, e.g., in the monographs of Protter [37, p. 31] or Applebaum [1, p. 126].

To be more precise, we at first take a closer look at the jumps of the Lévy process X. Therefore, we denote by

$$\Delta X_t = X_t - \lim_{s \nearrow t} X_s$$

the jump of X at time $t > 0$. For simplicity, we suppose the Lévy process to have càdlàg paths for all $\omega \in \Omega$. Now we count the jumps of X by defining the random measure

$$N(t, A)(\omega) = \#\{s \in [0, t] : \Delta X_s(\omega) \in A\} \tag{2.3}$$

for $\omega \in \Omega, A \in \mathcal{B}(\mathbb{R}^{d_X})$ and $t \geq 0$. Clearly, for fixed $\omega \in \Omega$ and $s \geq 0$, $N(s, .)(\omega)$ is a counting measure on $\mathcal{B}(\mathbb{R}^{d_X})$ counting the jumps of the realization $(X_t(\omega))_{t\geq 0}$ in the time interval $[0, s]$ with size in the set A. It turns out, that the average number of jumps with size in $A \in \mathcal{B}(\mathbb{R}^{d_X})$ in the unit time interval is given by the value of the Lévy measure

$$\nu(A) = \mathrm{E}[N(1, A)]$$

for the set A. As the paths of X are càdlàg, we deduce that there exist only finitely many jumps with size greater than some threshold greater zero in any finite time interval. This yields for a set A that is bounded below, i.e. $0 \notin \bar{A}$, that $N(t, A) < \infty$. Furthermore, for any set A with $\nu(A) < \infty$, the process $(N(t, A))_{t\geq 0}$ is a Poisson

process with intensity $\nu(A)$ and the integral $\int_A x\, N(t, \mathrm{d}x)$ then is a random finite sum, which gives rise to a compound Poisson process.

More specifically, we now consider complements of balls $B_h = \{x \in \mathbb{R}^{d_X} : |x| < h\}$ with sufficiently small $h > 0$ such that $\nu(B_h^c) > 0$. We put $\mu^{(h)}(\mathrm{d}x) = \nu|_{B_h^c}(\mathrm{d}x)/\nu(B_h^c)$ to be the Lévy measure restriced to the above complement and normalized. Then $\mu^{(h)}$ defines a Borel probability measure on $\mathbb{R}^{d_X}$ that determines the law of the jumps of the above mentioned compound Poisson process which consists of large jumps. More precisely, it holds

$$\int_{B_h^c} x\, N(t, \mathrm{d}x) \stackrel{d}{=} \sum_{i=1}^{N_t} \xi_i, \tag{2.4}$$

where $\stackrel{d}{=}$ denotes equality in distribution, $(N_t)_{t\geq 0}$ is a Poisson process with intensity $\nu(B_h^c)$ and $(\xi_i)_{i\in\mathbb{N}}$ is an i.i.d. sequence of random variables with distribution $\mu^{(h)}$ and independent of $(N_t)_{t\geq 0}$. To make sense of the jump process consisting of possibly countably infinitely many small jumps as well, we investigate the limit for $h \to 0$. Therefore, we want to switch to the complete space of L_2-martingales by subtracting the expectation which calculates to $\mathrm{E}\big[\int_{B_h^c} x\, N(t, \mathrm{d}x)\big] = F_0(h)t$, where we set

$$F_0(h) = \int_{B_h^c} x\, \nu(\mathrm{d}x).$$

We then consider the compensated process $L^{(h)} = (L_t^{(h)})_{t\geq 0}$, given by

$$L_t^{(h)} = \int_{B_h^c} x\, N(t, \mathrm{d}x) - tF_0(h), \tag{2.5}$$

which is an L_2-Lévy process with zero mean and thus an L^2-martingale. It turns out that for $h \to 0$, $L^{(h)}$ forms a Cauchy sequence, such that the jump part of X can be defined by its L_2-limit $L = (L_t)_{t\geq 0}$, which also holds P-*a.s.*, see, e.g., Applebaum [1, p. 121 ff.]. To further investigate the structure of X, we subtract the jump process L as well as the deterministic drift $b = \mathrm{E}[X_1]$ from the original process X. It turns out, that the remaining process

$$B_t = X_t - L_t - bt, \qquad t \geq 0,$$

is a Brownian motion with covariance matrix $A = \Sigma\Sigma^*$, which is independent of the process L. Altogether, with W denoting a d_X-dimensional Brownian motion independent of L, we can decompose a given Lévy process X into the sum

$$X_t = \Sigma W_t + L_t + bt \tag{2.6}$$

for $t \geq 0$. We add that the Lévy-Itô decomposition guarantees that every L_2-Lévy process has a representation (2.6).

2.3 The Blumenthal-Getoor index

For a Lévy process X, the Blumenthal-Getoor index β measures the frequency in which the infinitely many small jumps appear. As it is a characteristic for the jump part, it depends on the Lévy measure ν of X and is defined by

$$\beta = \inf\left\{p > 0 : \int_{|x|<1} |x|^p \, \nu(\mathrm{d}x) < \infty\right\}. \tag{2.7}$$

For a Lévy measure ν we always have $\int_{|x|<1} |x|^2 \, \nu(\mathrm{d}x) < \infty$ such that β only takes values in $[0, 2]$. For a standard Brownian motion, the Lévy measure is zero and thus has Blumenthal-Getoor index 0. Processes with finitely many jumps in finite time intervals, i.e. compound Poisson processes, also have $\beta = 0$, while the inverse implication is not true in general, which we will see later in the examples.

There are other equivalent ways of defining β that fit into our setting regarding the function $\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x)$. The latter will be a crucial quantity in our main results, which links the strong approximation rate with the computational cost of the algorithm. More explicitly, with

$$\begin{aligned} g(h) &= \int_{|x|>h} 1 \, \nu(\mathrm{d}x) + \frac{1}{h^2} \int_{|x|\leq h} |x|^2 \, \nu(\mathrm{d}x) + \frac{1}{h} \left| b - \int_{h<|x|\leq 1} x \, \nu(\mathrm{d}x) \right| \\ &\geq \int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x), \end{aligned}$$

we can write the Blumenthal-Getoor index as the asymptotic size of the singularity

of g in the origin in terms of functions h^{-p} for $p > 0$, namely,

$$\beta = \inf \left\{ p > 0 : \limsup_{h \searrow 0} h^p g(h) = 0 \right\}.$$

For proofs and further reading on the B-G index, we refer to the paper [8] of Blumenthal and Getoor from 1961, where the index was first mentioned.

2.4 Itô isometry

For our strong error estimates in Chapter 3, we state here a consequence of the Itô isometry for Lévy processes, which itself can be found, e.g., in [37, p. 161 ff.]. To this end, for a square integrable $\mathbb{R}^{d_X}$-valued Lévy martingale $L = (L_t)_{t \geq 0}$ with Lévy measure ν and covariance matrix of its Brownian component $A = \Sigma\Sigma^*$, we define the self-adjoint operator $Q : \mathbb{R}^{d_X} \to \mathbb{R}^{d_X}$ by

$$Qx = Ax + \int \langle x, y \rangle y \, \nu(\mathrm{d}y),$$

for $x \in \mathbb{R}^{d_X}$. Then, for a previsible $\mathbb{R}^{d_Y \times d_X}$ valued process $(H_s)_{s \geq 0}$ and a stopping time τ with

$$\mathrm{E}\left[\int_0^\tau \left| H_s Q^{\frac{1}{2}} \right|_{\mathrm{op}}^2 \, \mathrm{d}s \right] < \infty,$$

the stopped process $(\int_0^{\tau \wedge t} H_s \mathrm{d}L_s)_{t \geq 0}$ is a uniformly square integrable $\mathbb{R}^{d_Y}$-valued martingale with

$$\mathrm{E}\left[\left| \int_0^\tau H_s \, \mathrm{d}L_s \right|^2 \right] = \mathrm{E}\left[\int_0^\tau \left| H_s Q^{\frac{1}{2}} \right|_{\mathrm{op}}^2 \, \mathrm{d}s \right].$$

Observe, that for the operator norm of a self-adjoint operator T it holds

$$|T|_{\mathrm{op}} = \sup_{|x| \leq 1} |\langle Tx, x \rangle|,$$

see [47, p.239], such that we have $|Q^{\frac{1}{2}}|_{\mathrm{op}}^2 = |Q|_{\mathrm{op}}$ and we can estimate the above

expectation by

$$\begin{aligned}\mathrm{E}\left[\int_0^\tau \left|H_s Q^{\frac{1}{2}}\right|_{\mathrm{op}}^2 \,\mathrm{d}s\right] &\leq \mathrm{E}\left[\int_0^\tau |H_s|_{\mathrm{op}}^2 \cdot \left|Q^{\frac{1}{2}}\right|_{\mathrm{op}}^2 \,\mathrm{d}s\right] \\ &\leq \mathrm{E}\left[\int_0^\tau |H_s|^2 \cdot |Q|_{\mathrm{op}} \,\mathrm{d}s\right].\end{aligned}$$

The operator norm of Q is bounded from above by

$$\begin{aligned}|Q|_{\mathrm{op}} &= \sup_{|x|\leq 1} \langle Qx, x\rangle \\ &= \sup_{|x|\leq 1} \left\langle Ax + \int \langle x, y\rangle y \,\nu(\mathrm{d}y), x \right\rangle \\ &\leq \sup_{|x|\leq 1} \langle \Sigma\Sigma^* x, x\rangle + \sup_{|x|\leq 1} \int \langle x, y\rangle^2 \nu(\mathrm{d}y) \\ &\leq |\Sigma|^2 + \int |y|^2 \,\nu(\mathrm{d}y),\end{aligned}$$

where we have used linearity of the integral, the Cauchy Schwarz inequality and the fact that the Frobenius norm dominates the operator norm induced by the Euclidean norm.

Altogether we can bound the L_2-norm of the above stochastic integral by

$$\mathrm{E}\left[\left|\int_0^\tau H_s \,\mathrm{d}L_s\right|^2\right] \leq \mathrm{E}\left[\int_0^\tau |H_s|^2 \cdot \left(|\Sigma|^2 + \int |y|^2 \,\nu(\mathrm{d}y)\right) \,\mathrm{d}s\right]. \tag{2.8}$$

2.5 SDEs driven by Lévy processes

Recall from (1.1) that in this thesis we consider SDEs of the form

$$\begin{aligned}\mathrm{d}Y_t &= a\left(Y_{t-}\right) \,\mathrm{d}X_t, \qquad t \in [0,1], \\ Y_0 &= y_0,\end{aligned}$$

on a filtered probability space with deterministic initial value $y_0 \in \mathbb{R}^{d_Y}$, a diffusion coefficient $a : \mathbb{R}^{d_Y} \to \mathbb{R}^{d_Y \times d_X}$ and a Lévy process X, adapted to the filtration $\mathcal{F}$. Theory for this kind of SDEs can be found, e.g., in [1, p. 377 ff.] or in [37, p. 255 ff.], where the more general case of a semimartingale is considered as driving process. We

define a process $Y = (Y_t)_{t\in[0,1]}$ to be a strong solution to the SDE, if Y is adapted to $\mathcal{F}$ and satisfies the integral equation

$$Y_t = y_0 + \int_0^t a(Y_{s-})\,\mathrm{d}X_s,$$

P-a.s. for all $t \in [0,1]$. This of course implies that the stochastic integral on the right hand side exists. For this specific case of deterministic finite initial value y_0, a sufficient condition for a strong solution process to exist is the following Lipschitz condition on a. For all $y, y' \in \mathbb{R}^{d_Y}$, it holds

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

for a constant $K > 0$, where you should recall that $|.|$ denotes the Euclidean norm for vectors and the Frobenius norm for matrices, respectively. Under this condition, a strong solution Y, exists. Furthermore, this solution is càdlàg and unique P-a.s. in the pathwise sense.

2.6 Examples

Next we will show some interesting examples of Lévy processes different from a simple Poisson process or a Brownian motion. Here, we introduce these processes and after each of the following chapters, the results will be applied to these processes.

Stable Lévy processes

We begin with the definition of a strictly stable distribution. A random variable Z is called strictly stable if there exists a sequence $(c_n)_{n\in\mathbb{N}}$ with $c_n > 0$ such that for each $n \in \mathbb{N}$ it holds

$$Z_1 + Z_2 + \ldots + Z_n \overset{d}{=} c_n Z, \tag{2.9}$$

where $Z_1, \ldots, Z_n$ are i.i.d. copies of Z. It turns out that $c_n = n^{1/\alpha}$ for $\alpha \in (0,2]$. We thus call the distribution of Z strictly α-stable. These distributions have been of interest in modelling, e.g., for telephone noise and in finance, because of their heavy

tails and their scaling property. Stable laws are also of interest in the context of limit theorems as the distribution of the limits. Clearly, the strictly stable distributions are infinitely divisible and thus provide possible choices for the marginal distribution of a Lévy process at time $t = 1$. We call a Lévy process X with X_1 being strictly α-stable distributed an α-stable Lévy process. The latter is a subclass of stable processes, which are not neccessarily of Lévy type and which are thoroughly studied in [40].

Resulting from the scaling property (2.9) with $c_n = n^{1/\alpha}$, strictly stable processes possess self-similarity, i.e. for any $a > 0$

$$\left(\frac{X_{at}}{a^{1/\alpha}}\right)_{t\geq 0} \stackrel{d}{=} (X_t)_{t\geq 0}.$$

Recall that a standard Brownian motion W provides self-similarity with $\alpha = 2$ which yields that a Brownian motion is the special case of a strictly stable Lévy process with stability index $\alpha = 2$. In the following, we restrict ourselves to $\alpha \in (0, 2)$, because we are mainly interested in processes besides Brownian motions.

The Blumenthal-Getoor index of an α-stable process with $\alpha \in (0, 2)$ is $\beta = \alpha$, while, due to the frequency of large jumps, only moments of order less than α exist. The Lévy measure ν of a one-dimensional α-stable process can be represented via a Lebesgue density, denoted in the following by f_ν, by

$$\nu(\mathrm{d}x) = \underbrace{\left(\mathbb{1}_{(0,\infty]}(x)\frac{A_+}{|x|^{1+\alpha}} + \mathbb{1}_{[-\infty,0)}(x)\frac{A_-}{|x|^{1+\alpha}}\right)}_{:=f_\nu(x)} \mathrm{d}x,$$

for $x \in \mathbb{R}\backslash\{0\}$, where $A_+, A_- \geq 0$ with $A_+ + A_- > 0$. We will apply our numerical scheme to this one-dimensional stable processes, where we slightly change the Lévy measure such that its second moment exists, but the stable like behaviour of X remains valid for the small jumps. One such possibility is to temper the large jumps with an exponential decay, i.e. by multiplying the Lebesgue density with $\exp(-\lambda|x|)$ with some $\lambda > 0$. For the index of stability varying in $\{0.5, 0.8, 1.2\}$, we show some approximations of the trajectories of tempered stable processes with $A_+ = A_- = \lambda = 2$ in Figures 2.1, 2.2 and 2.3.

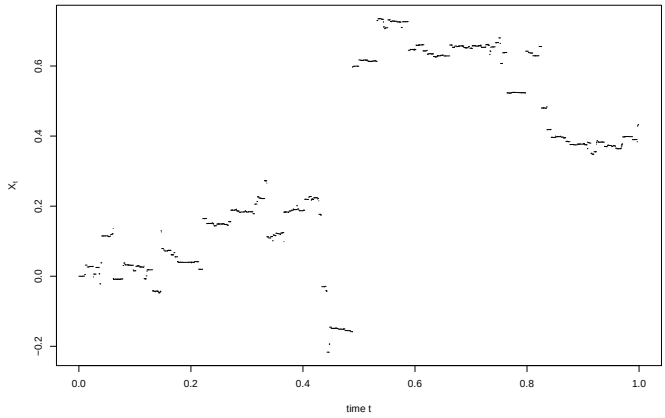

Figure 2.1: Tempered stable process with $\alpha = 0.5$.

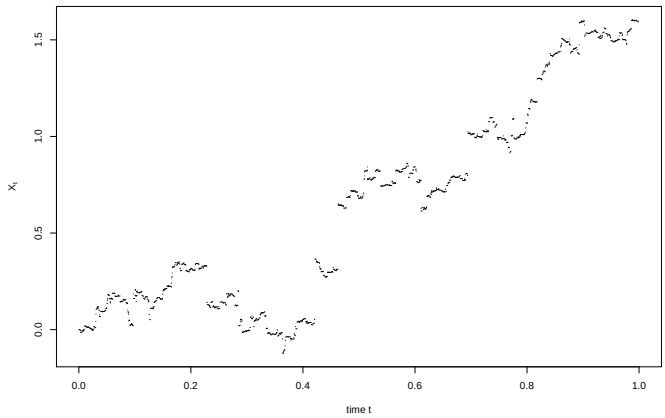

Figure 2.2: Tempered stable process with $\alpha = 0.8$.

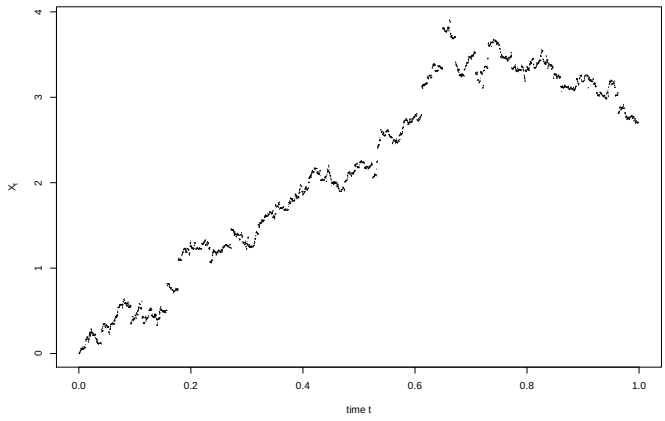

Figure 2.3: Tempered stable process with $\alpha = 1.2$.

Inverse Gaussian process (IG)

We first define the distribution, which literally should be the inverse of a Gaussian distribution, denoted by $\mathrm{IG}(a, b)$, with parameters $a, b > 0$. Therefore, for W denoting a standard Brownian motion, we define by $T^{(a,b)}$ the time, when the process $(W_t + bt)$ reaches the level a for the first time, i.e.

$$T^{(a,b)} = \inf\{t > 0 : W_t + bt \geq a\}.$$

Then the distribution of $T^{(a,b)}$ is called inverse Gaussian with parameters a and b, abbreviated by $\mathrm{IG}(a, b)$. Thus, it does not describe the inverse of a Gaussian distribution literally, but it describes the distribution, when, for a Wiener process, instead of considering the level at a certain time, we consider the time at which a certain level is reached. It turns out, that $T^{(a,b)}$ is infinitely divisible with characteristic function

$$\mathrm{E}\left[\exp(iuT^{(a,b)})\right] = \exp\left(-a\left(\sqrt{-2iu + b^2} - b\right)\right).$$

It thus leads to a Lévy process X with X_1 having an inverse Gaussian distribution with parameter a and b. It furthermore has the following scaling property. For $c > 0$ and an $\mathrm{IG}(a, b)$ random variable X the random variable cX is $\mathrm{IG}(\sqrt{c}a, b/\sqrt{c})$ distributed. The Lévy measure of an inverse Gaussian process can be given in terms

of a Lebesgue density by

$$\nu_{IG}(\mathrm{d}x) = \frac{a}{\sqrt{2\pi}x^{3/2}} \cdot \exp\left(-\frac{1}{2}b^2 x\right) \cdot \mathbb{1}_{x>0} \cdot \mathrm{d}x$$

for $x \in \mathbb{R}$. The inverse Gaussian process is a Lévy subordinator, i.e. a pure jump Lévy process having only positive jumps and non-negative drift, which is equivalent for being entirely positive. An approximation of a trajectory of an IG process with $a = \sqrt{2\pi}$ and $b = 2$ is given in Figure 2.4. Such processes can be used to create new processes by substituting the time of a given Lévy process by the subordinator, see e.g. [10]. This leads to a new Lévy process. Especially in finance, this method is used, e.g., to model the logarithm of the stock price by a Brownian motion observed in "business time", which is then given by the subordinator. A Brownian motion subordinated by an inverse Gaussian process is then called a normal inverse Gaussian process. Another special case of subordination is the following example.

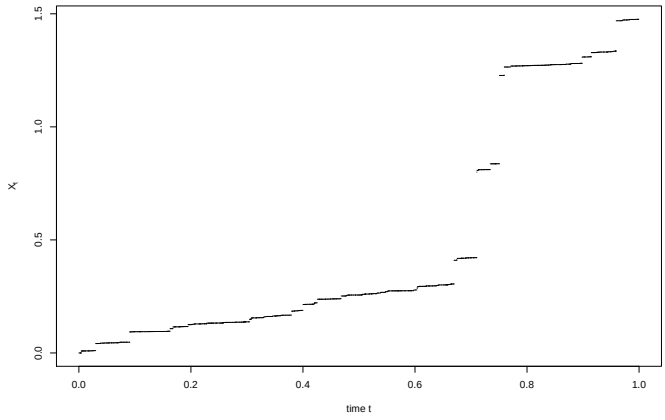

Figure 2.4: Inverse Gaussian process with $a = \sqrt{2\pi}$ and $b = 2$.

Variance Gamma process (VG)

The interest in this process arised within the work [30], where it was proposed as a new source of uncertainty for the logarithm of the stock price besides Brownian motion in an own market model. It came along with advantages like heavy tailedness and a good empirical fit while preserving nice technical properties like, e.g., the

possibility to a multivariate extension. A Variance Gamma process $(X_t)_{t\geq 0}$ is a Lévy process resulting from subordinating a Brownian motion with a Gamma subordinator. More explicitly, a Brownian motion $(W_t)_{t\geq 0}$ is timeshifted by an independent Gamma Lévy process $(T_t)_{t\geq 0}$. The latter is the Lévy process T with T_1 being Gamma distributed, which is of course an infinitely divisible distribution. We will use the following parametrization: For $a, b > 0$ the Lebesgue density of the distribution of T is given by

$$f_{T_t}(x) = \frac{b^{at}}{\Gamma(at)} x^{at-1} \exp(-bx),$$

for $x > 0$, where Γ denotes the Gamma function. The name arises from the fact that the variance of the process X is conditionally determined by the Gamma process T. The VG process is then given by $(X_t)_{t\geq 0} = (W_{T_t})_{t\geq 0}$. The Lévy measure ν of a VG process X can be given in terms of the Lebesgue density

$$f_\nu(x) = \frac{a}{|x|} \exp(-\sqrt{2b}|x|)$$

for $x \in \mathbb{R}\backslash\{0\}$.

3 Strong approximation

Crucial for the error analysis of the multilevel Monte Carlo scheme will be a weak error result as well as the variance of the estimator, for which we need a strong error result of the underlying approximation schemes. Here, we only consider the weak error, that is implicitly given by the strong one using the Lipschitz continuity of the functionals. In the following we will at first define the underlying approximation scheme, which essentially consists of an Euler scheme on a random time discretization where the Lévy process will be approximated by a compound Poisson process. After that, we will state strong convergence results for these approximation schemes.

3.1 The approximation scheme

We first recall to the reader the SDE (1.1), from whose solution process Y the computational problem arises.

$$\begin{aligned} \mathrm{d}Y_t &= a(Y_{t-})\mathrm{d}X_t, \qquad t \in [0,1], \\ Y_0 &= y_0, \end{aligned}$$

with a Lipschitz continuous function $a : \mathbb{R}^{d_Y} \to \mathbb{R}^{d_Y \times d_X}$ and a deterministic initial value $y_0 \in \mathbb{R}^{d_Y}$. The driving Lévy process X is supposed to be square integrable, i.e. the Lévy measure ν satisfies

$$\int |x|^2 \, \nu(\mathrm{d}x) < \infty,$$

and thus X decomposes to

$$X_t = L_t + \Sigma W_t + bt,$$

for $t \geq 0$, where $b \in \mathbb{R}^{d_X}$, ΣW is a d_X-dimensional Brownian motion with covariance matrix $\Sigma\Sigma^*$, which is independent of the L_2-jump martingale L. Remember, that the drift b is chosen such that L is a martingale, i.e. with no truncation function in the integral of the Lévy-Khintchine-formula (2.2).

As we are not able to simulate the process L on a given time discretization in many cases, we will approximate it by a compound Poisson process $L^{(h)}$, arising by neglecting the jumps smaller than some threshold $h > 0$. Thereby, the underlying idea to avoid huge global errors is to choose a random time discretization which includes at least all times, where L jumps with absolute size greater h, such that all big jumps are simulated at the right position. Along this way, the discretization has to incorporate at least the following stopping times $(T_j^{(h)})_{j\geq 0}$ given by $T_0^{(h)} = 0$ and

$$T_j^{(h)} = \inf\{t > T_{j-1}^{(h)} : |\Delta L_t| \geq h\}, \quad j \geq 1.$$

Up to now, the discretization is completely random and so we have no deterministic control on the mesh size of our grid. Because of the Brownian component, which can cause errors of size according to the size of the gap in between two discretization times, we refine the given time discretization such that two consecutive discretization points are at most $\varepsilon > 0$ apart. This new time grid $(T_j^{(h,\varepsilon)})_{j\geq 0}$ is defined by $T_0^{(h,\varepsilon)} = 0$ and

$$T_j^{(h,\varepsilon)} = \inf\{T_k^{(h)} > T_{j-1}^{(h,\varepsilon)} : k \in \mathbb{N}\} \wedge (T_{j-1}^{(h,\varepsilon)} + \varepsilon) \tag{3.1}$$

for $j \geq 1$.

Summarizing, X is approximated at the discretization times $T_j = T_j^{(h,\varepsilon)}$ by $\hat{X}_0^{(h,\epsilon)} = 0$ and

$$\hat{X}_{T_j}^{(h,\epsilon)} = \hat{X}_{T_{j-1}}^{(h,\epsilon)} + \Sigma(W_{T_j} - W_{T_{j-1}}) + \Delta L_{T_j}^{(h)} + (b - F_0(h))(T_j - T_{j-1})$$

for $j \geq 1$, where we recall that $F_0(h) = \int_{B_h^c} x\,\nu(\mathrm{d}x)$. Observe that with this choice

$$\hat{X}_{T_j}^{(h,\epsilon)} = \Sigma W_{T_j} + L_{T_j}^{(h)} + bT_j,$$

i.e., by $\hat{X}^{(h,\epsilon)}$, we essentially consider the process defined by

$$X_t^{(h)} = \Sigma W_t + L_t^{(h)} + bt$$

for $t \geq 0$ on the random grid $(T_j)_{j\in\mathbb{N}}$. For the SDE (1.1) the resulting approximate Euler scheme is defined by $\hat{Y}_0^{(h,\epsilon)} = y_0$ and

$$\hat{Y}_{T_j}^{(h,\epsilon)} = \hat{Y}_{T_{j-1}}^{(h,\epsilon)} + a(\hat{Y}_{T_{j-1}}^{(h,\epsilon)})(\hat{X}_{T_j}^{(h,\epsilon)} - \hat{X}_{T_{j-1}}^{(h,\epsilon)}) \tag{3.2}$$

for $j \geq 1$. Between two discretization points, we define the process $\hat{Y}^{(h,\epsilon)}$ to be piecewise constant by setting $\hat{Y}_t^{(h,\epsilon)} = \hat{Y}_{T_j}^{(h,\epsilon)}$ for $t \in [T_j, T_{j+1})$. It is of course also possible to choose interpolation schemes of higher order between the grid points, e.g., the piecewise linear interpolation. We focus on the piecewise constant approximations as these higher order interpolation schemes can only improve our strong approximation results in terms of constants, while the evaluation of functionals of the solution path becomes more difficult.

3.2 Strong error estimates

All restrictions on the SDE needed for the strong error estimates are summarized in the following

Assumption (A)**.** For a fixed $K < \infty$, the function $a : \mathbb{R}^{d_Y} \to \mathbb{R}^{d_Y \times d_X}$ satisfies

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

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

$$|a(y_0)| \leq K,\ 0 < \int |x|^2 \, \nu(\mathrm{d}x) \leq K^2,\ |\Sigma| \leq K \ \text{ and } \ |b| \leq K.$$

To prove error bounds for approximations of Y under Assumption (A), we will need the following lemma showing the boundedness of the solution process by something only depending on K. The proof of this lemma is a standard way to prove estimates of moments of stochastic differential equations. Essentially, it is a combination of Doob's inequality and Gronwall's lemma.

Lemma 1. *Under Assumption (A), there exists a constant $\kappa > 0$ depending only on K, such that*

$$\mathrm{E}\left[\sup_{t\in[0,1]} |Y_t - y_0|^2\right] < \kappa.$$

Proof. The idea is to use Gronwall's inequality like it is stated in the Appendix 8. Therefore define

$$z(t) = \mathrm{E}\left[\sup_{s\in[0,t]} |Y_s - y_0|^2\right], \tag{3.3}$$

for $t \in [0,1]$. Then $z(t) < \infty$ for $t \in [0,1]$ by a result of [1, p. 373] together with our assumption of second moments of X. Remembering the definition of Y, we decompose the integral equation into the martingale part and the drift part. For $t \in [0,1]$,

$$\begin{aligned} Y_t &= y_0 + \int_0^t a(Y_{s-})\,\mathrm{d}X_s \\ &= y_0 + \int_0^t a(Y_{s-})b\,\mathrm{d}s + \int_0^t a(Y_{s-})\,\mathrm{d}(\Sigma W_s + L_s). \end{aligned}$$

Then with the inequality $(a+b)^2 \le 2(a^2+b^2)$ for $a, b \ge 0$ we derive

$$\begin{aligned} z(t) &= \mathrm{E}\left[\sup_{s\in[0,t]} \left|\int_0^s a(Y_{u-})b\,\mathrm{d}u + \int_0^s a(Y_{u-})\,\mathrm{d}(\Sigma W_u + L_u)\right|^2\right] \\ &\le 2\,\mathrm{E}\left[\sup_{s\in[0,t]} \left|\int_0^s a(Y_{u-})b\,\mathrm{d}u\right|^2\right] + 2\,\mathrm{E}\left[\sup_{s\in[0,t]} \left|\int_0^s a(Y_{u-})\,\mathrm{d}(\Sigma W_u + L_u)\right|^2\right]. \end{aligned}$$

The next step is to bound the above two expectations by integrals of $z(s)$ w.r.t. the Lebesgue measure. Therefore observe for the first expression, that due to Jensen's

inequality and Fubini's theorem, we have

$$\begin{aligned}
\mathrm{E}\left[\sup_{s\in[0,t]}\left|\int_0^s a(Y_{u-})b\,\mathrm{d}u\right|^2\right] &\leq \mathrm{E}\left[\sup_{s\in[0,t]}\int_0^s |a(Y_{u-})b|^2\,\mathrm{d}u\right] \\
&\leq \mathrm{E}\left[\int_0^t |a(Y_{u-})|^2\cdot|b|^2\,\mathrm{d}u\right] \\
&\leq K^2\int_0^t \mathrm{E}\left[|a(Y_{u-})|^2\right]\,\mathrm{d}u,
\end{aligned}$$

where we have used the assumptions to bound the absolute value of the drift b. Furthermore, we derive again due to the assumptions, that

$$\begin{aligned}
\mathrm{E}\left[|a(Y_{u-})|^2\right] &\leq \mathrm{E}\left[|a(Y_{u-})-a(y_0)|^2+|a(y_0)|^2\right] \\
&\leq K^2\,\mathrm{E}\left[|Y_{u-}-y_0|^2\right]+K^2 \\
&\leq K^2\,(z(u)+1).
\end{aligned}$$

Altogether, the first expression can be bounded by

$$\begin{aligned}
\mathrm{E}\left[\sup_{s\in[0,t]}\left|\int_0^s a(Y_{u-})b\,\mathrm{d}u\right|^2\right] &\leq K^4\int_0^t (z(u)+1)\,\mathrm{d}u \\
&\leq K^4+K^4\int_0^t z(u)\,\mathrm{d}u.
\end{aligned}$$

For the second expression, we use Doob's supremum inequality for L_2-martingales to only consider the second moment in the endpoint, see, e.g., [24, p. 225]. Then the upper bound (2.8) coming from the Itô isometry, Assumptions (A) and Fubini's

theorem yield as upper bound

$$\begin{aligned}
\mathrm{E}\left[\left|\int_0^t a(Y_{u-})\,\mathrm{d}(\Sigma W_u + L_u)\right|^2\right] &\leq \mathrm{E}\left[\int_0^t |a(Y_{u-})|^2\Big(|\Sigma|^2 + \int |y|^2\,\nu(\mathrm{d}y)\Big)\,\mathrm{d}u\right] \\
&\leq 2K^2\,\mathrm{E}\left[\int_0^t |a(Y_{u-})|^2\,\mathrm{d}u\right] \\
&= 2K^2\int_0^t \mathrm{E}\left[|a(Y_{u-})|^2\right]\,\mathrm{d}u \\
&\leq 2K^4\int_0^t (z(u)+1)\,\mathrm{d}u \\
&\leq 2K^4 + 2K^4\int_0^t z(u)\,\mathrm{d}u.
\end{aligned}$$

Inserting both estimates into the upper bound of $z(t)$, we get

$$\begin{aligned}
z(t) &\leq 2K^4 + 2K^4\int_0^t z(u)\,\mathrm{d}u + 16K^4 + 16K^4\int_0^t z(u)\,\mathrm{d}u \\
&= 18K^4 + 18K^4\int_0^t z(u)\,\mathrm{d}u.
\end{aligned}$$

Gronwall's lemma then implies

$$\mathrm{E}\left[\sup_{s\in[0,1]} |Y_s - y_0|^2\right] = z(1) \leq 18K^4\cdot\exp(18K^4),$$

which finishes the proof. □

Before we consider our piecewise constant approximation process $\hat{Y}^{(h,\epsilon)}$, we need the error analysis of another approximating process $\bar{Y} = \bar{Y}^{(h,\varepsilon)}$, which will be the solution to the integral equation

$$\bar{Y}_t = y_0 + \int_0^t a(\bar{Y}_{\iota(s)-})\mathrm{d}X_s^{(h)},$$

with $\iota(t) = \sup([0,t]\cap\mathbb{T})$, and $\mathbb{T} = (T_j)_{j\in\mathbb{Z}_+}$. The strong error estimates for our

approximation schemes will always be given in terms of the function

$$F(h) = \int_{B_h} |x|^2 \, \nu(\mathrm{d}x)$$

for $h > 0$. Roughly speaking $F(h)$ measures the L_2-error which occurs by neglecting jumps of size smaller than h. Recall that we use this compound Poisson approximation of the jump part of the driving process X. Clearly, F is monotone increasing in h and bounded by K^2 by our Assumption (A). Furthermore, we always have $\lim_{h \downarrow 0} F(h) = 0$ with asymptotic behavior determined by the B-G index β. More precisely, we have for any $\beta' > \beta$ that $F(h) \precsim h^{2-\beta'}$ for $h \downarrow 0$.

Theorem 1. *If Assumption (A) from Chapter 3.2 holds, there exists a constant $\kappa > 0$ depending only on K such that for all $\varepsilon \in (0, 1]$ and $h > 0$ with $\nu(B_h^c) \leq 1/\varepsilon$, we have*

$$\mathrm{E}\left[\sup_{t \in [0,1]} |Y_t - \bar{Y}_t|^2\right] \leq \kappa(\varepsilon + F(h)) \tag{3.4}$$

in the general case and

$$\mathrm{E}\left[\sup_{t \in [0,1]} |Y_t - \bar{Y}_t|^2\right] \leq \kappa(F(h) + |b - F_0(h)|^2 \varepsilon^2) \tag{3.5}$$

in the case without a Brownian component, i.e. $\Sigma = 0$.

Proof. We will use a similar idea as in Lemma 1. Therefore, we put $Z_t = Y_t - \bar{Y}_t$ and $\bar{Z}_t = Y_t - \bar{Y}_{\iota(t)}$ for $t \geq 0$. Then, for a fixed stopping time τ we define the error in $[0, t]$ up to the stopping time by

$$z_\tau(t) = \mathrm{E}\left[\sup_{s \in [0, t \wedge \tau]} |Z_s|^2\right]. \tag{3.6}$$

The stopping time τ provides $z_\tau(t) < \infty$ for $t \in [0, 1]$ by stopping for too large values of $|Z_t|$. The goal now is to derive an estimate for $z_\tau(t)$ in the way it is needed to apply Gronwall's inequality, i.e. for constants $\alpha_1, \alpha_2 > 0$, not depending on the

choice of τ, and all $t \in [0,1]$, we want

$$z_\tau(t) \leq \alpha_2 + \alpha_1 \int_0^t z_\tau(s)\mathrm{d}s.$$

Then, using a localizing sequence, i.e. an increasing sequence of stopping times $(\tau_n)_{n\geq 1}$ with $\tau_n \to \infty$ P-$a.s.$ and each τ_n satisfying $z_{\tau_n}(1) < \infty$, we can deduce with a montone convergence argument that

$$\mathrm{E}\left[\sup_{s\in[0,1]} |Y_s - \bar{Y}_s|^2\right] = \mathrm{E}\left[\sup_{s\in[0,1]} |Z_s|^2\right] = \lim_{n\to\infty} z_{\tau_n}(1) \leq \alpha_2 \exp(\alpha_1). \tag{3.7}$$

So we now start to derive an estimate for (3.6). We define by $\bar{L}_t = L_t - L_t^{(h)}$ for $t \geq 0$ the remaining term of the driving process not occurring in the compound Poisson approximation. Essentially, $\bar{L}$ then is the jump martingale of compensated jumps smaller h. Then we can decompose Z in a local martingale and a finite variation part by

$$Z_t = M_t + \bar{M}_t + V_t,$$

for $t \geq 0$. Hereby, the local martingales M and $\bar{M}$ are given by

$$M_t = \int_0^t (a(Y_{s-}) - a(\bar{Y}_{\iota(s)-}))\,\mathrm{d}(\Sigma W_s + L_s^{(h)})$$

and

$$\bar{M}_t = \int_0^t a(Y_{s-})\,\mathrm{d}\bar{L}_s,$$

respectively. The remaining finite variation process V is then given by

$$V_t = \int_0^t (a(Y_{s-}) - a(\bar{Y}_{\iota(s)-}))b\,\mathrm{d}s.$$

Now using the inequality $(a+b+c)^2 \leq 4(a^2+b^2+c^2)$ for $a, b, c \geq 0$, we clearly have

$$z_\tau(t) \leq 4\,\mathrm{E}\left[\sup_{s\in[0,t\wedge\tau]} |M_s|^2\right] + 4\,\mathrm{E}\left[\sup_{s\in[0,t\wedge\tau]} |\bar{M}_s|^2\right] + 4\,\mathrm{E}\left[\sup_{s\in[0,t\wedge\tau]} |V_s|^2\right],$$

and it remains to bound the three expectations on the right hand side.

For the first one, Doob's supremum inequality, (2.8), Assumption (A) and Fubini's theorem imply

$$\begin{aligned}
\mathrm{E}\left[\sup_{s\in[0,t\wedge\tau]} |M_s|^2\right] &\leq 4\,\mathrm{E}\left[|M_{t\wedge\tau}|^2\right] \\
&\leq 4\,\mathrm{E}\left[\int_0^{t\wedge\tau} \left|a(Y_{s-}) - a(\bar{Y}_{\iota(s)-})\right|^2 (|\Sigma|^2 + \int_{B_h^c} |y|^2\,\nu(\mathrm{d}y))\,\mathrm{d}s\right] \\
&\leq 8K^2\,\mathrm{E}\left[\int_0^{t\wedge\tau} \left|a(Y_{s-}) - a(\bar{Y}_{\iota(s)-})\right|^2 \mathrm{d}s\right] \\
&\leq 8K^4\,\mathrm{E}\left[\int_0^{t\wedge\tau} \left|Y_{s-} - \bar{Y}_{\iota(s)-}\right|^2 \mathrm{d}s\right] \\
&= 8K^4 \int_0^t \mathrm{E}\left[\mathbb{1}_{\{s\leq\tau\}} |\bar{Z}^\tau_{s-}|^2\right]\,\mathrm{d}s,
\end{aligned}$$

where $\bar{Z}^\tau$ denotes the process $\bar{Z}$ stopped at τ, i.e., $\bar{Z}^\tau_t = \bar{Z}_{\tau\wedge t}$. For the second expectation, we derive with similar arguments, that

$$\begin{aligned}
\mathrm{E}\left[\sup_{s\in[0,t\wedge\tau]} |\bar{M}_s|^2\right] &\leq 4\,\mathrm{E}\left[|\bar{M}_{t\wedge\tau}|^2\right] \\
&\leq 4\,\mathrm{E}\left[\int_0^{t\wedge\tau} |a(Y_{s-})|^2 \int_{B_h} |y|^2\,\nu(\mathrm{d}y)\,\mathrm{d}s\right] \\
&\leq 8\,F(h)K^2\,\mathrm{E}\left[\int_0^{t\wedge\tau} (|Y_{s-} - y_0|^2 + 1)\,\mathrm{d}s\right] \\
&\leq 8\,F(h)K^2 \int_0^t (\mathrm{E}\left[|Y_{s-} - y_0|^2\right] + 1)\,\mathrm{d}s.
\end{aligned}$$

By Lemma 1, $\mathrm{E}\,[|Y_{s-} - y_0|^2]$ is bounded by a constant depending only on K such that we can essentially bound the second expectation by a constant multiple of $F(h)$.

The third expectation can be estimated by

$$\begin{aligned}
\mathrm{E}\left[\sup_{s\in[0,t\wedge\tau]}|V_s|^2\right] &= \mathrm{E}\left[\sup_{s\in[0,t\wedge\tau]}\left|\int_0^s (a(Y_{u-})-a(\bar{Y}_{\iota(u)-}))b\,\mathrm{d}u\right|^2\right] \\
&\le \mathrm{E}\left[\sup_{s\in[0,t\wedge\tau]}\int_0^s \left|(a(Y_{u-})-a(\bar{Y}_{\iota(u)-}))b\right|^2\,\mathrm{d}u\right] \\
&\le \mathrm{E}\left[\int_0^{t\wedge\tau} \left|(a(Y_{u-})-a(\bar{Y}_{\iota(u)-}))\right|^2\,|b|^2\,\mathrm{d}u\right] \\
&\le K^4\int_0^t \mathrm{E}\left[\mathbb{1}_{\{u\le\tau\}}\left|\bar{Z}^\tau_{u-}\right|^2\right]\mathrm{d}u.
\end{aligned}$$

Putting the three estimates together, we conclude that there exists a constant $\kappa_1 > 0$ depending only on K such that for $t \in [0,1]$,

$$z_\tau(t) \le \kappa_1\left(F(h) + \int_0^t \mathrm{E}\left[\mathbb{1}_{\{s\le\tau\}}\left|\bar{Z}^\tau_{s-}\right|^2\right]\mathrm{d}s\right). \tag{3.8}$$

For Gronwall's lemma we need an upper bound in terms of an integral of $s \mapsto z_\tau(s)$. Therefore observe that $\bar{Z}_s = Z_s + \bar{Y}_s - \bar{Y}_{\iota(s)}$ to derive

$$\begin{aligned}
\mathrm{E}\left[\mathbb{1}_{\{s\le\tau\}}\left|\bar{Z}^\tau_{s-}\right|^2\right] &= \mathrm{E}\left[\mathbb{1}_{\{s\le\tau\}}\left|Z^\tau_{s-}+\bar{Y}_{s-}-\bar{Y}_{\iota(s)-}\right|^2\right] \\
&\le 2\,\mathrm{E}\left[\mathbb{1}_{\{s\le\tau\}}\left|Z^\tau_{s-}\right|^2\right] + 2\,\mathrm{E}\left[\mathbb{1}_{\{s\le\tau\}}\left|\bar{Y}_{s-}-\bar{Y}_{\iota(s)-}\right|^2\right].
\end{aligned}$$

We have $\bar{Y}_{s-} - \bar{Y}_{\iota(s)-} = a(\bar{Y}_{\iota(s)-})(X^{(h)}_{s-} - X^{(h)}_{\iota(s)})$ and thus

$$\begin{aligned}
\mathrm{E}\left[\mathbb{1}_{\{s\le\tau\}}\left|\bar{Y}_{s-}-\bar{Y}_{\iota(s)-}\right|^2\right] &\le \mathrm{E}\left[\mathbb{1}_{\{s\le\tau\}}\left|a(\bar{Y}_{\iota(s)-})\left(X^{(h)}_{s-}-X^{(h)}_{\iota(s)}\right)\right|^2\right] \\
&\le \mathrm{E}\left[\mathbb{1}_{\{s\le\tau\}}\left|a(\bar{Y}_{\iota(s)-})\right|^2\left|X^{(h)}_{s-}-X^{(h)}_{\iota(s)}\right|^2\right] \\
&\le 2K^2\,\mathrm{E}\left[\left(\left|\bar{Y}^\tau_{\iota(s)-}-y_0\right|^2+1\right)\left|X^{(h)}_{s-}-X^{(h)}_{\iota(s)}\right|^2\right] \\
&= 2K^2\,\mathrm{E}\left[\left|\bar{Y}^\tau_{\iota(s)-}-y_0\right|^2+1\right]\,\mathrm{E}\left[\left|X^{(h)}_{s-}-X^{(h)}_{\iota(s)}\right|^2\right]
\end{aligned}$$

by the strong Markov property of X, see, e.g. [1, p. 97]. The first term can be further

bounded, using $|\bar{Y}_{\iota(s)-} - y_0| \leq |Y_{\iota(s)-} - y_0| + |Z_{\iota(s)-}|$, by

$$\mathrm{E}\left[\left|\bar{Y}^\tau_{\iota(s)-} - y_0\right|^2\right] \leq 2\left(\mathrm{E}\left[\left|Y^\tau_{\iota(s)-} - y_0\right|^2\right] + \mathrm{E}\left[\left|Z^\tau_{\iota(s)-}\right|^2\right]\right),$$

where $\mathrm{E}\left[\left|Y^\tau_{\iota(s)} - y_0\right|^2\right]$ is bounded by a constant depending on K by Lemma 1. For the second term, observe, that in $]\iota(s), s[$ no jumps of $X^{(h)}$ occur, i.e.,

$$X^{(h)}_{s-} - X^{(h)}_{\iota(s)} = \Sigma\left(W_s - W_{\iota(s)}\right) + (b - F_0(h))\,(s - \iota(s)),$$

such that the expectation calculates to

$$\begin{aligned}
\mathrm{E}\left[\left|X^{(h)}_{s-} - X^{(h)}_{\iota(s)}\right|^2\right] &= \mathrm{E}\left[\left|\Sigma\left(W_s - W_{\iota(s)}\right) + (b - F_0(h))(s - \iota(s))\right|^2\right] \\
&\leq 2\,\mathrm{E}\left[\left|\Sigma\left(W_s - W_{\iota(s)}\right)\right|^2\right] + 2\,\mathrm{E}\left[|b - F_0(h)|^2\,|s - \iota(s)|^2\right] \\
&\leq 2\left(|\Sigma|^2\,\varepsilon + |b - F_0(h)|^2\,\varepsilon^2\right),
\end{aligned}$$

where we used that W is independent of the stopping times in $\mathbb{T}$. Further notice, that the last term is uniformly bounded in h as by the Cauchy-Schwarz inequality

$$|F_0(h)|^2 = \left|\int_{B_h^c} x\,\nu(\mathrm{d}x)\right|^2 \leq \nu(B_h^c)\int |x|^2\,\nu(\mathrm{d}x) \leq K^2/\varepsilon.$$

Clearly, we can bound $\mathrm{E}\left[\mathbb{1}_{\{s\leq\tau\}}|Z^\tau_{s-}|^2\right]$ and $\mathrm{E}\left[\left|Z^\tau_{\iota(s)-}\right|^2\right]$ by $z_\tau(s)$, respectively. Then putting the estimates together we get for a constant $\kappa_2 > 0$ depending only on K that

$$\mathrm{E}\left[\mathbb{1}_{\{s\leq\tau\}}\left|\bar{Z}^\tau_{s-}\right|^2\right] \leq \kappa_2\left(z_\tau(s) + |\Sigma|^2\,\varepsilon + |b - F_0(h)|^2\,\varepsilon^2\right).$$

Inserting this into (3.8), we get

$$z_\tau(t) \leq \kappa_3\left(F(h) + |\Sigma|^2\,\varepsilon + |b - F_0(h)|^2\,\varepsilon^2 + \int_0^t z_\tau(s)\,\mathrm{d}s\right)$$

for a constant κ_3 depending only on K. By (3.7) we thus derived the upper bound

$$\mathrm{E}\left[\sup_{t\in[0,1]}|Y_t-\bar{Y}_t|^2\right]\leq\kappa_4\left(F(h)+|\Sigma|^2\,\varepsilon+|b-F_0(h)|^2\,\varepsilon^2\right),$$

i.e., if $\Sigma=0$, the second term cancels and we are done. For the general case, observe that

$$|\Sigma|^2\,\varepsilon+|b-F_0(h)|^2\,\varepsilon^2\leq K^2\varepsilon+(|b|^2+|F_0(h)|^2)\varepsilon^2\leq K^2(\varepsilon^2+2\varepsilon)\leq\kappa_5\varepsilon, \tag{3.9}$$

for a constant κ_5 depending only on K. The result then follows immediately. □

Now we come back to the actual approximation process $\hat{Y}^{(h,\epsilon)}$. Essentially, the difference to $\bar{Y}$ is, that $\hat{Y}^{(h,\epsilon)}$ is taken piecewise constant between two discretization points $T_j^{(h,\varepsilon)}$ and $T_{j+1}^{(h,\varepsilon)}$, i.e. the approximation scheme $\bar{Y}$ is some kind of continuous Euler scheme. Under the same assumptions on the process X, the approximation parameter h and ε, and the coefficients a and b in terms of a universal constant K given in the Assumptions (A), we get a global approximation result similar to the result of Theorem 1. The changes are due to the error induced by the Brownian motion between two discretization points. More precisely we get the following

Theorem 2. *If Assumption (A) from Chapter 3.2 holds, there exists a constant $\kappa>0$ depending only on K such that for all $\varepsilon\in(0,1]$ and $h>0$ with $\nu(B_h^c)\leq 1/\varepsilon$, we have*

$$\mathrm{E}\left[\sup_{t\in[0,1]}|Y_t-\hat{Y}_t^{(h,\epsilon)}|^2\right]\leq\kappa\left(\varepsilon\log\left(\frac{\exp(1)}{\varepsilon}\right)+F(h)\right) \tag{3.10}$$

in the general case and

$$\mathrm{E}\left[\sup_{t\in[0,1]}|Y_t-\hat{Y}_t^{(h,\epsilon)}|^2\right]\leq\kappa(F(h)+|b-F_0(h)|^2\varepsilon^2) \tag{3.11}$$

in the case without a Brownian component, i.e., $\Sigma=0$.

Remark 2. The Euler scheme for a Lévy-driven SDE has already been analyzed in [27]. The authors consider semimartingales as driving processes and show uniform

convergence on compacts in probability with the help of a global strong approximation. For the case of X being a Lévy process the convergence rate obtained is approximately $1/2$ in terms of the step size of the Euler scheme. Observe that this Euler scheme employs increments of the driving semimartingale.

As the explicit simulation of increments of X is only possible in a few cases, there has been a rising interest in the approximate Euler scheme, which uses a compound Poisson approximation of X. In [39], this scheme has been analyzed in terms of a limit theorem, showing convergence in law of the error process multiplied with a rate function of the right order, depending amongst other things on the Lévy measure.

A strong approximation error then very recently has been presented in [15]. The author studies the L_2-error appearing in the discretization points under an approximate Euler scheme in the one-dimensional case, where an additional Gaussian term compensating the neglected small jumps is considered, following the idea of [2]. The resulting error bounds are approximately of the same size as those from Theorem 2.

Proof. For the proof, we will write $\hat{Y}$ for $\hat{Y}^{(h,\epsilon)}$. By

$$\mathrm{E}\left[\sup_{t\in[0,1]}|Y_t-\hat{Y}_t|^2\right]\leq 2\left(\mathrm{E}\left[\sup_{t\in[0,1]}|Y_t-\bar{Y}_t|^2\right]+\mathrm{E}\left[\sup_{t\in[0,1]}|\bar{Y}_t-\hat{Y}_t|^2\right]\right),$$

and Theorem 1 it suffices to find a upper bound for the second term. Therefore observe that for each discretization point $T_j\in\mathbb{T}$ we have $\bar{Y}_{T_j}=\hat{Y}_{T_j}$. Furthermore, as $\hat{Y}$ is chosen piecewise constant between two points in $\mathbb{T}$, we conclude, that

$$\bar{Y}_t-\hat{Y}_t=\bar{Y}_t-\bar{Y}_{\iota(t)}=\underbrace{a(\bar{Y}_{\iota(t)})(b-F_0(h))(t-\iota(t))}_{=:A_t}+\underbrace{a(\bar{Y}_{\iota(t)})\Sigma(W_t-W_{\iota(t)})}_{=:B_t}. \quad (3.12)$$

With (3.12), the problem splits into

$$\begin{aligned}\mathrm{E}\left[\sup_{t\in[0,1]}\left|\bar{Y}_t-\hat{Y}_t\right|^2\right]&=\mathrm{E}\left[\sup_{t\in[0,1]}|A_t+B_t|^2\right]\\&\leq 2\left(\mathrm{E}\left[\sup_{t\in[0,1]}|A_t|^2\right]+\mathrm{E}\left[\sup_{t\in[0,1]}|B_t|^2\right]\right)\end{aligned}$$

The first expression can be bounded by

$$\begin{aligned}
\mathrm{E}\left[\sup_{t\in[0,1]}|A_t|^2\right] &\leq \mathrm{E}\left[\sup_{t\in[0,1]}\left|a(\bar{Y}_{\iota(t)})\right|^2|b-F_0(h)|^2|t-\iota(t)|^2\right] \\
&\leq 2K^2\mathrm{E}\left[\sup_{t\in[0,1]}\left|\bar{Y}_{\iota(t)}-y_0\right|^2+1\right]|b-F_0(h)|^2\varepsilon^2 \\
&\leq \kappa_1|b-F_0(h)|^2\varepsilon^2,
\end{aligned}$$

with a constant κ_1 only depending on K, as we have the boundedness of

$$\begin{aligned}
\mathrm{E}\left[\sup_{t\in[0,1]}\left|\bar{Y}_{\iota(t)}-y_0\right|^2\right] &\leq \mathrm{E}\left[\sup_{t\in[0,1]}\left|\bar{Y}_t-y_0\right|^2\right] \\
&\leq 2\left(\mathrm{E}\left[\sup_{t\in[0,1]}\left|\bar{Y}_t-Y_t\right|^2\right]+\mathrm{E}\left[\sup_{t\in[0,1]}|Y_t-y_0|^2\right]\right)
\end{aligned}$$

by a constant depending on K by Theorem 1 and Lemma 1, respectively.

For the second term, we derive

$$\begin{aligned}
\mathrm{E}\left[\sup_{t\in[0,1]}|B_t|^2\right] &\leq \mathrm{E}\left[\sup_{t\in[0,1]}\left|a(\bar{Y}_{\iota(t)})\right|^2|\Sigma|^2\left|W_t-W_{\iota(t)}\right|^2\right] \\
&\leq 2K^2|\Sigma|^2\,\mathrm{E}\left[\sup_{t\in[0,1]}\left(\left|\bar{Y}_{\iota(t)}-y_0\right|^2+1\right)\left|W_t-W_{\iota(t)}\right|^2\right] \\
&= 2K^2|\Sigma|^2\,\mathrm{E}\left[\max_{j\in\mathbb{N}}\left(\left(\mathbb{1}_{\{T_j<1\}}\left|\bar{Y}_{T_j}-y_0\right|^2+1\right)V_j\right)\right],
\end{aligned}$$

where we denote for every $j\in\mathbb{N}$ the relative supremum of W between T_j and T_{j+1} by

$$V_j=\sup_{t\in[T_j,T_{j+1}\wedge 1)}\left|W_t-W_{T_j}\right|^2,$$

and used that for $t\in[T_j,T_{j+1})$ we have $\iota(t)=T_j$. Observe that $V_j=0$ for every j with $T_j\geq 1$. For the expectation of the maximum of the latter products we can now apply Lemma 3 from the Appendix for a fixed but arbitrary number r of stopping times. Here we choose as filtration $\mathcal{G}_j=\mathcal{F}_{T_j}$, and the random variables

$U_j = \mathbb{1}_{\{T_j<1\}} \left|\bar{Y}_{T_j} - y_0\right|^2 + 1$, which is $\mathcal{G}_j$ measurable and V_j as above, which is $\mathcal{G}_{j+1}$ measurable and independent of $\mathcal{G}_j$. Letting the number r tend to infinity, we can use a monotone convergence argument to obtain the upper bound

$$\mathrm{E}\left[\sup_{t\in[0,1]} |B_t|^2\right] \leq \kappa_2 \left|\Sigma\right|^2 \mathrm{E}\left[\max_{j\in\mathbb{N}} \mathbb{1}_{\{T_j<1\}} \left|\bar{Y}_{T_j} - y_0\right|^2 + 1\right] \mathrm{E}\left[\max_{j\in\mathbb{N}} V_j\right].$$

Here, κ_2 is a constant depending on K. The first expectation has already been discussed and is known to be bounded by a constant depending only on K. For the second expectation, we need to bound the second moment of the maximal fluctuation of a standard Brownian motion for a step size tending to zero. Therefore, for a function $f : [0,1] \to \mathbb{R}^d$ we denote by w_f its modulus of continuity, i.e., for $\delta \in [0,1]$, we define

$$w_f(\delta) = \sup_{\substack{t,s\in[0,1]\\ |t-s|\leq\delta}} |f(t) - f(s)|.$$

For the Brownian motion, Lévy's modulus of continuity states that

$$\lim_{\delta\to 0} \frac{w_W(\delta)}{\sqrt{2\delta\log(1/\delta)}} = 1 \qquad P-a.s.,$$

and can be found, e.g. in [38] or [25]. Here we need to bound the second moment of the latter, which is one of the topics of [14]. They show for p-th moments of the modulus of continuity of a standard Brownian motion that

$$\mathrm{E}\left[w_W(\delta)^p\right] \leq c(p)(\delta\log(2/\delta))^{p/2},$$

for all $\delta \in (0,1]$, where $c(p) > 0$ is a constant depending on p. We denote by φ the monotone increasing function

$$\varphi : [0,1] \to [0,\infty), \delta \mapsto \sqrt{\delta\log\left(\frac{\exp(1)}{\delta}\right)}.$$

Then with [14], we can deduce, that there exist a constant $\kappa_3 > 0$ such that

$$\mathrm{E}\left[w_W(\delta)^2\right] \leq \kappa_3 \varphi(\delta)^2,$$

for all $\delta \in (0,1]$. We derive

$$\begin{aligned}\mathrm{E}\left[\max_{T_j \in \mathbb{T}} \sup_{t \in [T_j, T_{j+1})} \left|W_t - W_{T_j}\right|^2\right] &\leq \mathrm{E}\left[\sup_{\substack{t,s \in [0,1] \\ |t-s| \leq \varepsilon}} |W_t - W_s|^2\right] \\ &= \mathrm{E}\left[w_W(\varepsilon)^2\right] \leq \kappa_3 \varphi(\varepsilon)^2.\end{aligned}$$

Inserting this into our inequalities we deduce that there exists a constant κ_4 depending on K such that

$$\mathrm{E}\left[\sup_{t \in [0,1]} |B_t|^2\right] \leq \kappa_4 |\Sigma|^2 \varphi(\varepsilon)^2.$$

By putting together the estimates, we conclude that there exists a constant κ depending only on K such that

$$\begin{aligned}\mathrm{E}\left[\sup_{t \in [0,1]} \left|\bar{Y}_t - \hat{Y}_t\right|^2\right] &\leq \kappa \left(|b - F_0(h)|^2 \varepsilon^2 + |\Sigma|^2 \varphi(\varepsilon)^2\right) \\ &= \kappa \left(|b - F_0(h)|^2 \varepsilon^2 + |\Sigma|^2 \varepsilon \log\left(\frac{\exp(1)}{\varepsilon}\right)\right).\end{aligned}$$

Together with the result from Theorem 1, we are done in the case $\Sigma = 0$. For the general case, observe that we already derived a bound for $|b - F_0(h)|^2 \varepsilon^2$ by the Cauchy-Schwarz inequality and our Assumptions (A) which is linear in ε in (3.9). Observe further that $\varepsilon \leq \varepsilon \log(\exp(1)/\varepsilon)$ for ε sufficiently small to finish the proof. □

3.3 Examples

We continue here with the examples of driving Lévy processes presented in Chapter 2. We apply the adaptive approximation scheme from above and show how to

simulate the various resulting compound Poisson processes on a computer.

Stable Lévy processes

Remember from Section 2.6, that the Lévy measure ν of an α-stable process has the Lebesgue density

$$f_\nu(x) = \mathbb{1}_{(0,\infty]}(x)\frac{A_+}{|x|^{1+\alpha}} + \mathbb{1}_{[-\infty,0)}(x)\frac{A_-}{|x|^{1+\alpha}},$$

for $x \in \mathbb{R}\backslash\{0\}$, where $A_+, A_- \geq 0$ with $A_+ + A_- > 0$ and $\alpha \in (0,2)$. To apply our strong approximation results, the process has to be square integrable. One way to ensure square integrability is to temper the big jumps of our process with an exponential decay, another one is to neglect the jumps above a given size. The process then only possesses stable like behavior for the small jumps with sizes around the origin.

Truncated α-stable processes

We first consider the truncation method to ensure the square integrability of an α-stable process. Therefore we truncate the jumps bigger than some given size $u > 0$ in absolute value. The Lévy measure is then given by

$$\nu(\mathrm{d}x) = \left(\mathbb{1}_{(0,u]}(x)\frac{A_+}{|x|^{1+\alpha}} + \mathbb{1}_{[-u,0)}(x)\frac{A_-}{|x|^{1+\alpha}}\right)\cdot \mathrm{d}x,$$

for $x \in \mathbb{R}\backslash\{0\}$. For a fixed $h > 0$ the simulation of the restricted and normalized Lévy measure $\nu|_{B_h^c}/\nu(B_h^c)$ is done by inversion as its cumulative distribution function can be explicitly calculated.

We first consider the case of a **non-symmetric** Lévy measure ν. The jump in-

tensity for jumps greater or equal to h then calculates to

$$\begin{aligned}\nu(B_h^c) &= \int_{-u}^{-h} \frac{A_-}{|x|^{1+\alpha}}\,\mathrm{d}x + \int_h^u \frac{A_+}{|x|^{1+\alpha}}\,\mathrm{d}x \\ &= A_- \left[-\frac{1}{\alpha}x^{-\alpha}\right]_h^u + A_+ \left[-\frac{1}{\alpha}x^{-\alpha}\right]_h^u \\ &= \frac{A_+ + A_-}{\alpha}(h^{-\alpha} - u^{-\alpha}).\end{aligned}$$

Here, we assume of course, that $h < u$. Otherwise $\nu(B_h^c) = 0$.

For simulations via inversion, we have to calculate the distribution function of the desired distribution and the inverse function thereof. Therefore, we put $\bar{A} = A_-/(A_+ + A_-)$. Then the distribution function is given by

$$\frac{1}{\nu(B_h^c)}\nu|_{B_h^c}(]-\infty, t]) = \begin{cases} A_-((-t)^{-\alpha} - u^{-\alpha})/(\nu(B_h^c)\,\alpha) & ,-u \le t \le -h, \\ \bar{A} & ,-h \le t \le h, \\ \bar{A} + A_+(h^{-\alpha} - t^{-\alpha})/(\nu(B_h^c)\,\alpha) & ,h \le t \le u. \end{cases}$$

As this distribution function is not bijective, we cannot directly formulate the inverse function. Instead, we distinguish two cases in our simulation. For a uniform distributed $U \sim \mathcal{U}([0,1])$, we get $Z \sim \nu|_{B_h^c}/\nu(B_h^c)$ by defining

$$Z = \begin{cases} -\left(\left(\frac{\nu(B_h^c)\,\alpha U + A_- u^{-\alpha}}{A_-}\right)^{-\frac{1}{\alpha}}\right) & ,U \le \bar{A}, \\ \left(-\left(\frac{\nu(B_h^c)\,\alpha(U-\bar{A}) - A_+ h^{-\alpha}}{A_+}\right)\right)^{-\frac{1}{\alpha}} & ,U > \bar{A}. \end{cases}$$

For the case of a **symmetric** Lévy measure ν, i.e. $A_+ = A_- = c > 0$, the intensity of jumps bigger than h for $0 < h < u$ is given by

$$\nu(B_h^c) = \frac{2c}{\alpha}(h^{-\alpha} - u^{-\alpha}).$$

For the simulation, due to symmetry, it suffices to sample from $\nu|_{[h,\infty)}/\nu([h,\infty))$ which is done by inversion and to multiply this with an independent Bernoulli random sign, i.e., a random variable V with $P(V = 1) = 1/2 = P(V = -1)$. For $U \sim \mathcal{U}([0,1])$ being uniformly distributed on $[0,1]$ and V distributed as above and

independent of U, we obtain

$$V \cdot (h^{-\alpha} - U \cdot (h^{-\alpha} - u^{-\alpha}))^{-1/\alpha} \sim \nu|_{B_h^c}/\nu(B_h^c).$$

Tempered symmetric α-stable processes

For simplicity, we only consider the symmetric case, i.e. $A_+ = A_- = c > 0$. For some $\lambda > 0$ the Lévy measure is then given by the Lebesgue density

$$f_\nu(x) = \frac{c}{|x|^{1+\alpha}} \exp(-\lambda|x|),$$

for $x \in \mathbb{R}\backslash\{0\}$. The simulation of the approximating compound Poisson process is then done by rejection. Therefore observe that ν is symmetric and the Lebesgue density of $\nu|_{[h,\infty)}$ is bounded by a constant multiple of the density of an exponential distribution with

$$f_\nu|_{B_h^c}(x) \leq \frac{c}{h^{1+\alpha}} \exp(-\lambda|x|) =: u(x).$$

Due to the symmetry, it suffices to simulate $\nu|_{[h,\infty)}/\nu([h,\infty))$ and an independent Bernoulli random sign as above. For $V \sim \text{Exp}(\lambda)$ and $U \sim \mathcal{U}([0,1])$ independent of V, consider $Z = u(V) \cdot U$. If $V < h$ or $f_\nu(V) < Z$ we reject, otherwise, if $f_\nu(V) \geq Z$ we take V as output.

Using integration by parts, the frequency of jumps calculates to

$$\begin{aligned}
\nu(B_h^c) &= 2c \int_h^\infty \frac{1}{x^{1+\alpha}} \exp(-\lambda x)\,\mathrm{d}x \\
&= 2c \left(\frac{\exp(-\lambda h)}{\alpha h^\alpha} - \frac{\lambda}{\alpha} \int_h^\infty \frac{1}{x^\alpha} \exp(-\lambda x)\,\mathrm{d}x \right) \\
&= 2c \left(\frac{\exp(-\lambda h)}{\alpha h^\alpha} \left(1 + \frac{\lambda}{1-\alpha} h \right) - \frac{\lambda^2}{1-\alpha} \int_h^\infty \frac{1}{x^{\alpha-1}} \exp(-\lambda x)\,\mathrm{d}x \right).
\end{aligned}$$

Thus, repeating integration by parts, we inductively derive

$$\nu(B_h^c) = 2c \frac{\exp(-\lambda h)}{\alpha h^\alpha} \left(\sum_{n=0}^\infty \frac{(\lambda h)^n}{\prod_{i=1}^n (i-\alpha)} \right).$$

The calculations of $\nu(B_h^c)$ for the finite number of values for $h > 0$ needed in the

algorithm are done numerically by deterministic quadrature rules.

The Variance Gamma process

Remember from Section 2.6 that the Lévy measure ν of a VG process X can be represented via a Lebesgue density by

$$\nu(\mathrm{d}x) = \frac{a}{|x|} \exp(-\sqrt{2b}|x|) \cdot \mathrm{d}x$$

for $x \in \mathbb{R}\backslash\{0\}$. For $h > 0$ we now want to simulate the approximating compound Poisson process that arises by neglecting the jumps with size smaller than h. Therefore we have to calculate $\nu(B_h^c)$ and simulate the probability measure $\nu|_{B_h^c}/\nu(B_h^c)$. The latter is again done by the rejection method. As before in the tempered stable case, the measure ν is symmetric and the Lebesgue density of $\nu|_{(h,\infty)}$ is bounded by a constant multiple of the density of an exponential distribution

$$f_\nu|_{B_h^c}(x) \leq \frac{a}{h} \exp(-\sqrt{2b}|x|) =: u(x).$$

We again use symmetry such that only realizations of $\nu|_{[h,\infty)}/\nu([h,\infty))$ and an independent Bernoulli random sign are needed. For $V \sim \mathrm{Exp}(\sqrt{2b})$ and $U \sim \mathcal{U}([0,1])$ independent of V, we consider $Z = u(V) \cdot U$. If $V < h$ or $f_\nu(V) < Z$ we reject, otherwise, if $f_\nu(V) \geq Z$ we take V as output.

For the simulation of the jump times we have to calculate the exponential integral

$$\nu(B_h^c) = 2 \int_h^\infty \frac{a}{x} \exp(-\sqrt{2b}x)\, \mathrm{d}x.$$

This value is as before approximated numerically for a given size $h > 0$.

Remark 3. For the acceptance rejection simulations in the tempered stable and the VG case, observe that the average cost per jump simulation is given by the ratio of the total measure on $[h,\infty)$ for the bounding density to the total measure on $[h,\infty)$ for the original density. Taking into account the considerations about the computational cost in the next chapter, this ratio has to be uniformly bounded in

h for $h \to 0$. To this end, the bounding densities have to be changed to

$$u(x) = \begin{cases} \frac{c}{x^{1+\alpha}} \exp(-\lambda h), & h \leq x < \lambda^{-1}, \\ c\lambda^{1+\alpha} \exp(-\lambda x), & \lambda^{-1} < x, \end{cases}$$

for $\alpha \in (0, 2)$ in the tempered stable case and with $\alpha = 0$, $c = a$ and $\lambda = \sqrt{2b}$ in the VG case. With these choices, the computational cost for one jump simulation is uniformly bounded by a constant, where the simulations with respect to the bounding densities are done by suitable inversion methods.

4 The multilevel algorithm

We will at first recall the considered quadrature problem. We want to compute expectations of a class F of functions $f : D[0,1] \to \mathbb{R}$ with respect to the distribution of the strong solution Y to the SDE (1.1). We denote these expectations by

$$S(f) = \mathrm{E}\left[f(Y)\right],$$

and consider for F the class of measurable functions $f : D[0,1] \to \mathbb{R}$ that are Lipschitz continuous with coefficient 1 with respect to the supremum norm, denoted by $\mathrm{Lip}(1)$. To compute $S(f)$, we consider randomized algorithms $\widehat{S}$ with output $\widehat{S}(f)$ for a function f from the Lipschitz class $\mathrm{Lip}(1)$.

We measure the error $e(\widehat{S})$ of the algorithm $\widehat{S}$ by a worst case of the root mean square error over all $f \in \mathrm{Lip}(1)$, i.e., by

$$e^2\left(\widehat{S}\right) = \sup_{f \in \mathrm{Lip}(1)} \mathrm{E}\left[\left|S(f) - \widehat{S}(f)\right|^2\right]. \tag{4.1}$$

Our goal is to relate the error with the computational cost of the algorithm, denoted by $\mathrm{cost}(\widehat{S})$. Here, we as well take the worst case over all functions $f \in \mathrm{Lip}(1)$ and average the random computational cost for $f \in \mathrm{Lip}(1)$, denoted by $\mathrm{cost}(\widehat{S}, f)$, i.e.,

$$\mathrm{cost}\left(\widehat{S}\right) = \sup_{f \in \mathrm{Lip}(1)} \mathrm{E}\left[\mathrm{cost}\left(\widehat{S}, f\right)\right].$$

The computational cost represents, up to constants, the expected runtime of the algorithm on a computer. Therefore recall, that we suppose that arithmetic operations with real numbers and comparisons can be done at cost one. We further suppose that evaluations of a are possible at any point $y \in \mathbb{R}^{d_Y}$ at cost one and evaluations

of f are possible for piecewise constant functions at cost given by their number of breakpoints. We also assume that sampling from the uniform distribution on $[0,1]$ and from suitably restricted Lévy measures are possible at cost one.

For a Monte Carlo algorithm $\widehat{S}$, we denote the bias of $\widehat{S}(f)$ by

$$\operatorname{bias}\big(\widehat{S}(f)\big) = \mathrm{E}\left[S(f) - \widehat{S}(f)\right].$$

The mean square error of $\widehat{S}$ can always be decomposed into the squared bias and the variance, namely

$$\mathrm{E}\left[\left|S(f) - \widehat{S}(f)\right|^2\right] = \Big(\operatorname{bias}\big(\widehat{S}(f)\big)\Big)^2 + \operatorname{var}\Big(\widehat{S}(f)\Big). \tag{4.2}$$

In a typical application of an MC algorithm, the bias is the error induced by approximating a desired distribution Y by an approximation $\hat{Y}$ and so making systematically at least the error of the bias. Additionally, we have a statistical error by using a randomized algorithm for the desired expectation. In the following, we will recall the classical Monte Carlo approach, which will be helpful for the numerical experiments of this work. We then motivate the introduction of multilevel in this setting.

4.1 Classical Monte Carlo and multilevel Monte Carlo

For a better understanding of these concepts, we will present the classical Monte Carlo approach for the case, that we have to approximate Y. Therefore, we consider a sequence of approximations $(\hat{Y}^{(k)})_{k\geq 1}$ that can be simulated and that converge to Y in a suitable way specified later. We then choose one approximation $\hat{Y} = \hat{Y}^{(k)}$ and take n independent copies $\hat{Y}_1, \dots, \hat{Y}_n$ thereof. The latter means that $\hat{Y}_1, \dots, \hat{Y}_n$ form an *i.i.d.* sequence with each $\hat{Y}_i$ having the distribution of $\hat{Y}$. The Monte Carlo output is defined by the arithmetical mean

$$\widehat{S}^{MC}(f) = \frac{1}{n}\sum_{i=1}^{n} f\big(\hat{Y}_i\big)$$

based on this n independent copies of $\hat{Y}$. In this case the bias is given by

$$\operatorname{bias}\big(\widehat{S}^{MC}(f)\big) = \mathrm{E}\left[f(Y) - f(\hat{Y})\right]$$

and the variance of this arithmetical mean of independent copies is given by

$$\operatorname{var}\left(\widehat{S}^{MC}(f)\right) = \frac{1}{n}\operatorname{var}\left(f(\hat{Y})\right).$$

The variance of the random variable $f(\hat{Y})$ is a constant value depending on the approximation scheme $\hat{Y}$. Due to the definition of $\widehat{S}^{MC}(f)$, there are only two possibilities to calibrate the algorithm, namely by the number of replications n to change the statistical error, and by the choice of the approximation $\hat{Y}$ from our sequence of approximations to change the bias. Thus, it is a standard approach to create approximation schemes with a small bias for all $f \in F$ and then choose n asymptotically optimal in the following way.

For a sequence of approximation schemes $(\hat{Y}^{(k)})_{k\geq 1}$, we say that $(\hat{Y}^{(k)})_{k\geq 1}$ has weak convergence order $\gamma > 0$ if there exists a constant $\kappa > 0$ such that

$$\sup_{f\in F}\left|\mathrm{E}\left[f(Y) - f(\hat{Y}^{(k)})\right]\right| \leq \kappa \cdot \left(\sup_{f\in F}\mathrm{E}\left[\operatorname{cost}\left(f(\hat{Y}^{(k)}), f\right)\right]\right)^{-\gamma}, \tag{4.3}$$

for all $k \geq 1$. For simplicity, we suppose as well $\operatorname{var}\left(f(\hat{Y}^{(k)})\right) \leq \kappa^2$ for all $f \in F$ and $k \geq 1$. If we now choose an approximation $\hat{Y}^{(k)}$ with computational cost m from the above sequence, the worst case mean squared error of the classical Monte Carlo algorithm becomes

$$e^2\left(\widehat{S}^{MC}\right) \leq \kappa^2\left(m^{-2\gamma} + \frac{1}{n}\right).$$

The overall computational cost is then given by $\operatorname{cost}(\widehat{S}^{MC}) = m \cdot n$. For a given cost bound N, we can asymptotically choose the replication number n and the cost m for one approximation to be

$$m = N^{\frac{1}{1+2\gamma}} \qquad \text{and} \qquad n = N^{\frac{2\gamma}{1+2\gamma}},$$

such that we obtain a worst case mean square error of order $1/(2+1/\gamma)$. Being more explicit, there exists a sequence of classical Monte Carlo algorithms $(\widehat{S}_N^{MC})_{N\in\mathbb{N}}$ using approximation schemes with weak convergence order γ with $\mathrm{cost}(\widehat{S}_N^{MC}) \leq N$ and

$$e\left(\widehat{S}_N^{MC}\right) \precsim N^{-\frac{1}{2+1/\gamma}}. \tag{4.4}$$

Remark 4. Even with an approximation scheme of higher order, we cannot reach the order $1/2$ with the classical Monte Carlo algorithm in the biased case. Moreover, the approximation schemes with a sufficiently large γ, which are needed here, come along with more assumptions on the functionals f and the diffusion coefficient a. This essentially means that we have to simplify the problem by considering smaller classes F and fewer possibilities for the SDE to reach the above orders of convergence.

The other rather classical way to improve the error of a biased Monte Carlo algorithm in relation to its computational effort, is to reduce the variance by changing the estimator, i.e. by not taking the classical Monte Carlo estimator. The multilevel algorithm is such a variance reduction technique. In the case of X being a standard Brownian motion, it has been shown by [16] that up to some logarithmic terms, the convergence order $1/2$ in terms of the computational cost can be achieved. A further advantage of this method is that it assumes f and a to be only Lipschitz whereby for the approximation $\hat{Y}$, it suffices to choose a simple Euler scheme.

Multilevel Monte Carlo

For the multilevel algorithm, we use a whole hierarchy of approximation schemes $\hat{Y}^{(1)}, \ldots, \hat{Y}^{(m)}$ with accuracy and computational effort increasing with the upper index. For threshold parameters $(h_k, \varepsilon_k)_{k=1,\ldots,m}$ we will use the non-linear approximation schemes defined in Chapter 3 by setting $\hat{Y}^{(k)} = \hat{Y}^{(h_k,\varepsilon_k)}$. It is clear from the above, that threshold parameters with $h_1 \geq \ldots \geq h_m \geq 0$ and $\varepsilon_1 \geq \ldots \geq \varepsilon_m > 0$ are a reasonable choice. The expectation of the approximation with the highest accuracy can be written by the following telescoping sum

$$\mathrm{E}\left[f\left(\hat{Y}^{(m)}\right)\right] = \mathrm{E}\left[f\left(\hat{Y}^{(1)}\right)\right] + \sum_{k=2}^{m} \mathrm{E}\left[f\left(\hat{Y}^{(k)}\right) - f\left(\hat{Y}^{(k-1)}\right)\right]. \tag{4.5}$$

Each expectation on the right hand side can now be estimated independently by the standard Monte Carlo approach, i.e., by independently averaging independent copies $(D_i^{(k)})_{i=1,\dots,n_k}$ of

$$D^{(k)} = \begin{cases} f\left(\hat{Y}^{(1)}\right) & \text{for } k = 1, \\ f\left(\hat{Y}^{(k)}\right) - f\left(\hat{Y}^{(k-1)}\right) & \text{for } k \geq 2, \end{cases}$$

Here, it is important, that for $k \geq 2$, the approximation schemes $(\hat{Y}^{(k)}, \hat{Y}^{(k-1)})$ in $D^{(k)}$ are coupled via X in a way that the variances of the D_k decrease with k. The coupling of $\hat{Y}^{(k)}$ and $\hat{Y}^{(k-1)}$, such that they suitably depend on the same realization of the driving Lévy process X, is shown in detail in Section 4.2. Altogether, the multilevel Monte Carlo estimator is given by

$$\widehat{S}^{ML}(f) = \sum_{k=1}^{m} \frac{1}{n_k} \sum_{i=1}^{n_k} D_i^{(k)}. \tag{4.6}$$

Due to the telescoping sum (4.5), this estimator has the bias of the approximation scheme with the highest accuracy

$$\operatorname{bias}\big(\widehat{S}^{ML}(f)\big) = \mathrm{E}\left[f(Y) - f(\hat{Y}^{(m)})\right],$$

and the variance decomposes due to Bienaymé's formula to

$$\operatorname{var}\left(\widehat{S}^{ML}(f)\right) = \sum_{k=1}^{m} \frac{1}{n_k} \operatorname{var}\left(D^{(k)}\right).$$

With this decomposition, we can use the decreasing variances to reduce the error in relation to the computational cost. Therefore observe that, while we have to spend more independent copies of $D^{(k)}$ for the lower levels, which are computationally cheap, we need less independent copies of $D^{(k)}$ for the higher levels near the highest accuracy m to obtain the same variance contribution to the error. Balancing this out, we can optimize the upper bound of the convergence order, i.e. the relation of error and cost.

Essentially this algorithm is very familiar to a known variance reduction technique, namely it is some kind of a repeated control variate method. Usually, in the control

variate method, we split the desired random variable Z into $Z - Z'$ and Z' such that $\mathrm{E}[Z] = \mathrm{E}[Z - Z'] + \mathrm{E}[Z']$, where $\mathrm{E}[Z']$ can be calculated directly or at least with little computational effort, and the remaining part $Z - Z'$ has a considerably smaller variance than Z.

To use our results of Chapter 3, we will work with the following upper bounds for the bias and the variance of the multilevel algorithm to properly choose the approximations $\hat{Y}^{(k)}$ as well as the parameters m and $n_1, \ldots, n_m$. Observe that for every functional f from the Lipschitz class F, we have

$$\begin{aligned}\mathrm{bias}\big(\widehat{S}^{ML}(f)\big)^2 &= \left|\mathrm{E}\left[f(Y) - f(\hat{Y}^{(m)})\right]\right|^2 \\ &\leq \mathrm{E}\left[\left|f(Y) - f(\hat{Y}^{(m)})\right|^2\right] \\ &\leq \mathrm{E}\left[\left\|Y - \hat{Y}^{(m)}\right\|^2\right],\end{aligned}$$

where we used Jensen's inequality and the Lipschitz continuity of f. For the variance, we also deduce with the Lipschitz continuity of f, that

$$\begin{aligned}\mathrm{var}\left(\widehat{S}^{ML}(f)\right) &= \sum_{k=1}^{m} \frac{1}{n_k}\mathrm{var}(D^{(k)}) \\ &\leq \frac{1}{n_1}\mathrm{var}\left(f(\hat{Y}^{(1)}) - f(y_0)\right) + \sum_{k=2}^{m} \frac{1}{n_k}\mathrm{E}\left[\left|f(\hat{Y}^{(k)}) - f(\hat{Y}^{(k-1)})\right|^2\right] \\ &\leq \frac{1}{n_1}\mathrm{E}\left[\left\|\hat{Y}^{(1)} - y_0\right\|^2\right] + \sum_{k=2}^{m} \frac{1}{n_k}\mathrm{E}\left[\left\|\hat{Y}^{(k)} - \hat{Y}^{(k-1)}\right\|^2\right]\end{aligned}$$

Both upper bounds hold for all $f \in F$ such that the worst case error defined in (1.2) is bounded by

$$\begin{aligned}e^2\big(\widehat{S}^{ML}\big) \leq \mathrm{E}\left[\left\|Y - \hat{Y}^{(m)}\right\|^2\right] &+ \frac{1}{n_1}\mathrm{E}\left[\left\|\hat{Y}^{(1)} - y_0\right\|^2\right] \\ &+ \sum_{k=2}^{m} \frac{1}{n_k}\mathrm{E}\left[\left\|\hat{Y}^{(k)} - \hat{Y}^{(k-1)}\right\|^2\right].\end{aligned} \tag{4.7}$$

Remember the considerations about the cost of an algorithm $\widehat{S}$ in the beginning of this chapter. If we apply the Euler schemes $\hat{Y}^{(k)}$, an important number for the computational cost is the number of breakpoints of the piecewise constant approximation process $\hat{Y}^{(k)}$ of Y, denoted in the following by $\Upsilon(\hat{Y}^{(k)})$. By summing the cost for arithmetic operations, simulations and function evaluations, we deduce that $\hat{Y}^{(k)}$ has computational cost proportional to $\Upsilon(\hat{Y}^{(k)})$. We now conclude that the computational cost of the multilevel algorithm is given by

$$\operatorname{cost}\left(\widehat{S}^{ML}\right) = \kappa \sum_{k=1}^{m} n_k \operatorname{E}\left[\Upsilon\left(\hat{Y}^{(k)}\right)\right], \tag{4.8}$$

for a constant $\kappa > 0$, as the cost for simulations, function evaluations and arithmetical operations for $D^{(k)}$ are given in terms of constant multiples of $\Upsilon(\hat{Y}^{(k)})$. The latter is the finer approximation scheme on this level and induces the coarser one. Before we present our main results, we first discuss in the next chapter, how to simulate coupled Euler schemes $(\hat{Y}^{(h,\epsilon)}, \hat{Y}^{(h',\varepsilon')})$ for $h' > h > 0$ and $\varepsilon' > \varepsilon > 0$ in a way that given $\hat{Y}^{(h,\epsilon)}$ we can deduce $\hat{Y}^{(h',\varepsilon')}$ based on the same realization of X with no additional simulation effort. This coupling is the most important ingredient of the algorithm as well for the variance of the $D^{(k)}$ to decrease as for the cost reduction in the simulation of the algorithm.

4.2 The coupled approximation

We obtain the coupled approximations $(\hat{Y}^{(h,\epsilon)}, \hat{Y}^{(h',\varepsilon')})$ by applying the Euler scheme (3.2) to the coupled driving processes $\hat{X}^{(h,\epsilon)}$ and $\hat{X}^{(h',\varepsilon')}$ with their random discretization times $T_j^{(h,\varepsilon)}$ and $T_j^{(h',\varepsilon')}$, respectively. Thus we need a coupled simulation of the approximation schemes $\hat{X}^{(h,\epsilon)}$ and $\hat{X}^{(h',\varepsilon')}$, where the threshold parameters satisfy $h' > h > 0$ and $\varepsilon' > \varepsilon > 0$, which means that $\hat{X}^{(h,\epsilon)}$ is the finer approximation scheme. The jump processes $\Delta L^{(h')}$ and $\Delta(L^{(h)} - L^{(h')})$ are independent with values in $\{0\} \cup B_{h'}^c$ and $\{0\} \cup B_h^c \backslash B_{h'}^c$, respectively, see [1, p. 116], i.e., the jumps of $L^{(h)}$ are given by the independent sum $\Delta L^{(h)} = \Delta L^{(h')} + \Delta(L^{(h)} - L^{(h')})$, and therefore the jumps of the process $L^{(h')}$ can be obtained from those of $L^{(h)}$ by taking

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

We conclude that the simulation of the joint distribution of $(L^{(h)}, L^{(h')})$ only requires samples from the jump times and jump heights $T_k^{(h)}$ and $\Delta L^{(h)}_{T_k^{(h)}}$, respectively, which amounts to sampling from $\nu|_{B_h^c}/\nu(B_h^c)$ and from an exponential distribution. To simulate the Brownian components of the coupled processes $(\hat{X}^{(h,\epsilon)}, \hat{X}^{(h',\varepsilon')})$, we refine the sequence of jump times $T_k^{(h)}$ to get $(T_j^{(h,\varepsilon)})_{j\in\mathbb{N}_0}$ and $(T_j^{(h',\varepsilon')})_{j\in\mathbb{N}_0}$, respectively. Since W and L are independent, the process W is easily simulated at all times $(T_j^{(h,\varepsilon)})_{j\in\mathbb{N}_0}$ and $(T_j^{(h',\varepsilon')})_{j\in\mathbb{N}_0}$ that are in $[0,1]$ by sampling from a normal distribution.

4.3 Main results

Our main findings are summarized in the following Theorems 3, 4 and 5. The choices of the parameters m and $(n_k, \varepsilon_k, h_k)_{k=1,\dots,m}$ for the algorithm $\widehat{S}^{ML}$ to satisfy Theorems 3, 4 and 5 are stated separately in Chapter 4.5, but can as well be found in the course of the proofs. For the main results, we will use a decreasing and invertible function $g : (0,\infty) \to (0,\infty)$, which satisfies

$$\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x) = \frac{F(h)}{h^2} + \nu(B_h^c) \leq g(h), \tag{4.9}$$

for all $h > 0$. Essentially, this function provides a properly scaled relation of error and cost for the jump threshold parameter h. Then, we can optimize the error by choosing h in relation to the maximal step size ε via the inverse of g such that Theorem 2 is applicable, i.e. such that $\nu(B_h^c) \leq 1/\varepsilon$ holds.

Theorem 3. *Let $g : (0,\infty) \to (0,\infty)$ denote the decreasing and invertible function satisfying (4.9). If the driving process X has no Brownian component, i.e., $\Sigma = 0$, and if there exists $\gamma > 0$ such that*

$$g(h) \precsim \frac{1}{h\,(\log(1/h))^{1+\gamma}} \tag{4.10}$$

as $h \to 0$, then there exists a family $(\widehat{S}_n)_{n\in\mathbb{N}}$ of multilevel algorithms satisfying $\mathrm{cost}(\widehat{S}_n) \leq n$ and

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

Proof. We choose the step size approximately dyadically decreasing by taking

$$\varepsilon_k = 2^{-k} \qquad \text{and} \qquad h_k = g^{-1}(2^k) = g^{-1}(\varepsilon_k^{-1}),$$

for $k = 0, \dots, m$. Here, the parameters h_0 and ε_0 does not appear in the algorithm itself but are chosen only for the proof. With this choice and the definition of g in (4.9) we have

$$\nu(B_{h_k}^c) \leq \frac{F(h_k)}{h_k^2} + \nu(B_{h_k}^c) \leq g(h_k) = g\big(g^{-1}(\varepsilon_k^{-1})\big) = \frac{1}{\varepsilon_k},$$

and $\varepsilon_k \leq 1$ for all $k = 1, \dots, m$ as required for the results of Theorem 2. For $\Sigma = 0$, we deduce for the terms appearing in (4.7) for the worst case mean square error of $\widehat{S}^{ML}$ that

$$\begin{aligned}\frac{1}{n_1}\mathrm{E}\left[\left\|\hat{Y}^{(1)} - y_0\right\|^2\right] &\leq \frac{1}{n_1} 2\left(\mathrm{E}\left[\left\|\hat{Y}^{(1)} - Y\right\|^2\right] + \mathrm{E}\left[\|Y - y_0\|^2\right]\right) \\ &\leq \frac{1}{n_1}\kappa_1\left(F(h_1) + |b - F_0(h_1)|^2\varepsilon_1^2\right),\end{aligned}$$

for a constant κ_1 depending on K, where we have used Theorem 2 and Lemma 1. The latter term can be uniformly bounded as done before in the proof of Theorem 1 such that we have an estimate in terms of a constant multiple of $1/n_1$. For the terms with $k \geq 2$, observe that we always have

$$F(h_k) = \int_{B_{h_k}} |x|^2 \, \nu(\mathrm{d}x) \leq \int_{B_{h_{k-1}}} |x|^2 \, \nu(\mathrm{d}x) = F(h_{k-1})$$

as it holds $h_{k-1} \geq h_k > 0$. Furthermore the step size parameter are also ordered descending in k such that $\varepsilon_{k-1}^2 \geq \varepsilon_k^2 > 0$. The term $|b - F_0(h)|^2$ can be bounded uniformly by using $|b|^2 \leq K^2$, which is one of the Assumptions (A), and by an

estimate for $|F_0(h)|^2$. For $v \in (0,1)$ we have

$$\begin{aligned}
|F_0(h)| = \left| \int_{B_h^c} x\, \nu(\mathrm{d}x) \right| &\leq \int_{B_h^c} |x|\, \nu(\mathrm{d}x) \\
&\leq \int |x| \frac{v}{v}\, \nu(\mathrm{d}x) \leq \frac{1}{v} \int |x|(v \vee |x|)\, \nu(\mathrm{d}x) \\
&\leq \frac{1}{v} \int_{B_v^c} |x|^2\, \nu(\mathrm{d}x) + \int_{B_v} |x|\, \nu(\mathrm{d}x) \\
&\leq \frac{1}{v} \int |x|^2\, \nu(\mathrm{d}x) + \int_0^v \nu(B_u^c)\, \mathrm{d}u
\end{aligned}$$

for any $h > 0$. The first term is finite for a fixed $v \in (0,1)$ as the integral is finite by Assumption (A). For the second term, we now use the assumptions on the function g such that for a constant κ_2, we have

$$\int_0^v \nu(B_u^c)\, \mathrm{d}u \leq \int_0^v g(u)\, \mathrm{d}u \leq \kappa_2 \int_0^v \frac{1}{u(\log(1/u))^{1+\gamma}}\, \mathrm{d}u,$$

where the latter integral is finite. Thus, we can bound each term with $k = 2, \ldots, m$ by

$$\begin{aligned}
\frac{1}{n_k} \mathrm{E}\left[\left\| \hat{Y}^{(k)} - \hat{Y}^{(k-1)} \right\|^2 \right] &\leq \frac{1}{n_k} 2 \left(\mathrm{E}\left[\left\| Y - \hat{Y}^{(k)} \right\|^2 \right] + \mathrm{E}\left[\left\| Y - \hat{Y}^{(k-1)} \right\|^2 \right] \right) \\
&\leq \frac{1}{n_k} \kappa_3 \left(F(h_{k-1}) + \varepsilon_{k-1}^2 \right),
\end{aligned}$$

where we have used Theorem 2, the observations above and κ_3 is a constant depending only on K. A similar bound holds for the first term in (4.7) for a constant κ_4 by

$$\mathrm{E}\left[\left\| Y - \hat{Y}^{(m)} \right\|^2 \right] \leq \kappa_4 (F(h_m) + \varepsilon_m^2).$$

We again used Theorem 2 and the assumption (4.10) for the boundedness of $|b - F_0(h)|^2$. Altogether we have an estimate for (4.7) for a constant κ_5 depending on K

in terms of

$$e^2(\widehat{S}^{ML}) \leq \kappa_5 \sum_{k=1}^{m+1} \frac{1}{n_k}(F(h_{k-1}) + \varepsilon_{k-1}^2),$$

where we set $n_{m+1} = 1$ for the proof, which is not appearing as parameter in the algorithm itself. To set the two terms in the same size, observe, that (4.10) implies for the inverse of g, that

$$\frac{1}{y^{3/2}} \precsim g^{-1}(y) \precsim \frac{1}{y(\log(y))^{1+\gamma}} \tag{4.11}$$

for $y \to \infty$. Then we have on the one hand, that

$$F(h_k) \leq g(h_k)h_k^2 = 2^k(g^{-1}(2^k))^2,$$

and on the other hand with (4.11) we have

$$\begin{aligned}\varepsilon_k^2 = \varepsilon_k^{-1}(2^{-k})^3 &= \varepsilon_k^{-1}\left(\frac{1}{(2^k)^{3/2}}\right)^2 \\ &\precsim \varepsilon_k^{-1}(g^{-1}(2^k))^2 = 2^k(g^{-1}(2^k))^2.\end{aligned}$$

We end up with the error bound

$$e^2(\widehat{S}^{ML}) \leq \kappa_6 \sum_{k=1}^{m+1} \frac{1}{n_k} 2^{k-1}(g^{-1}(2^{k-1}))^2$$

for a constant κ_6.

To get the relation of error and cost, we have to set the parameter m and $(n_k)_{k=1,\dots,m}$ appropriately. Therefore we choose for a fixed $C \geq 1/g^{-1}(1)$ the replication parameter by

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

and to ensure $n_k \in \mathbb{N}$ we set $m = m(C) = \inf\{k \in \mathbb{N} : g^{-1}(2^k)C < 1\} - 1$. With this

choice and (4.11), we derive with $1/n_k \leq 2/(g^{-1}(2^{k-1})C)$ that

$$\begin{aligned}\frac{1}{n_k}2^{k-1}(g^{-1}(2^{k-1}))^2 &\leq 2^{k-1}g^{-1}(2^{k-1})\frac{2}{C}\\ &\precsim \frac{2^{k-1}}{2^{k-1}(\log(2^{k-1}))^{1+\gamma}}\frac{1}{C}\\ &\precsim \frac{1}{(k-1)^{1+\gamma}}\frac{1}{C}\end{aligned}$$

where the last term is summable in k and thus the resulting sum is uniformly bounded. This implies for the error that there exists a constant κ_7 depending on K and the function g, such that

$$e^2(\widehat{S}^{ML}) \leq \kappa_7 \frac{1}{C}$$

Next we consider the computational cost of $\widehat{S}^{ML}$. Therefore observe that the expected number of breakpoints of $\hat{Y}^{(k)}$ is bounded by

$$\mathrm{E}\left[\Upsilon(\hat{Y}^{(k)})\right] \leq \frac{1}{\varepsilon_k} + \nu(B^c_{h_k}) \leq 2^{k+1}.$$

Then we can derive the cost bound

$$\begin{aligned}\mathrm{cost}\left(\widehat{S}^{ML}\right) &= \sum_{k=1}^{m} n_k \mathrm{E}\left[\Upsilon(\hat{Y}^{(k)})\right]\\ &\leq \sum_{k=1}^{m} g^{-1}(2^{k-1})C\, 2^{k+1}\\ &= 4C\sum_{k=1}^{m} 2^{k-1}g^{-1}(2^{k-1}).\end{aligned}$$

The latter sum is bounded as observed before by (4.11) such that for a constant κ_8 depending on g and K we have

$$\mathrm{cost}\left(\widehat{S}^{ML}\right) \leq \kappa_8 C.$$

Altogether we obtain a sequence of multilevel algorithms with cost bound C and an asymptotic worst case root mean square error of size $1/\sqrt{C}$ as proposed in the

Theorem. □

The result of Theorem 3 covers all cases, where the driving Lévy process has Blumenthal-Getoor index smaller than 1 and does not comprise a Gaussian component, i.e., it holds $\Sigma = 0$. This essentially is equivalent for the Lévy process to have paths of finite variation, see, e.g., [10]. Next, we investigate the cases with Gaussian component, i.e. $\Sigma \neq 0$, while we keep the Blumenthal-Getoor index small.

Theorem 4. *Let $g : (0,\infty) \to (0,\infty)$ again denote the decreasing and invertible function satisfying (4.9). If there exists $\gamma \geq 1/2$ such that*

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

as $h \to 0$, then there exists a family $(\widehat{S}_n)_{n\in\mathbb{N}}$ of multilevel algorithms satisfying $\mathrm{cost}(\widehat{S}_n) \leq n$ and

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

Proof. The idea of the proof is the same as for the proof of Theorem 3. This time we can modify the assumptions wlog to

$$\frac{\sqrt{\log(1/h)}}{h} \precsim g(h) \precsim \frac{(\log(1/h))^{\gamma}}{h} \tag{4.12}$$

for $h \to 0$ and some $\gamma \geq 1/2$. We choose the threshold parameter as before by

$$\varepsilon_k = 2^{-k} \qquad \text{and} \qquad h_k = g^{-1}(2^k) = g^{-1}(\varepsilon_k^{-1}),$$

for $k = 0, \ldots, m$ such that the results of Theorem 2 hold and we can use the general ones, i.e. the case with arbitrary Σ. We again search estimates for the terms in (4.7). As before we have the term with the coarsest approximation to be bounded by a constant multiple of $1/n_1$ by noticing that

$$\frac{1}{n_1}\mathrm{E}\left[\left\|\hat{Y}^{(1)} - y_0\right\|^2\right] \leq \frac{1}{n_1}2\left(\mathrm{E}\left[\left\|\hat{Y}^{(1)} - Y\right\|^2\right] + \mathrm{E}\left[\|Y - y_0\|^2\right]\right) \leq \frac{1}{n_1}\kappa_1$$

for a constant κ_1 which depends on K because of Lemma 1 and because the term

$(F(h_1)+\varepsilon_1 \log(\exp(1)/\varepsilon_1))$ obtained from Theorem 2 is as well bounded by a constant depending on K. The terms with $k \geq 2$ are approximated by

$$\begin{aligned}\frac{1}{n_k}\mathrm{E}\left[\left\|\hat{Y}^{(k)} - \hat{Y}^{(k-1)}\right\|^2\right] &\leq \frac{1}{n_k}2\left(\mathrm{E}\left[\left\|Y - \hat{Y}^{(k)}\right\|^2\right] + \mathrm{E}\left[\left\|Y - \hat{Y}^{(k-1)}\right\|^2\right]\right) \\ &\leq \frac{1}{n_k}\kappa_2\left(F(h_{k-1}) + \varepsilon_{k-1}\log\left(\frac{\exp(1)}{\varepsilon_{k-1}}\right)\right),\end{aligned}$$

for a constant κ_2 depending on K. Here we used that

$$\varepsilon_k \log\left(\frac{\exp(1)}{\varepsilon_k}\right) \leq \varepsilon_{k-1}\log\left(\frac{\exp(1)}{\varepsilon_{k-1}}\right)$$

by the monotonicity of $\varepsilon \mapsto \varepsilon\log(\exp(1)/\varepsilon)$ together with $\varepsilon_k \leq \varepsilon_{k-1}$, and, as we observed before, that $F(h_k) \leq F(h_{k-1})$. Putting things together and setting again $n_{m+1} = 1$ we obtain as upper bound for the worst case error

$$e^2(\widehat{S}^{ML}) \leq \kappa_3 \sum_{k=1}^{m+1} \frac{1}{n_k}\left(F(h_{k-1}) + \varepsilon_{k-1}\log\left(\frac{\exp(1)}{\varepsilon_{k-1}}\right)\right),$$

where κ_3 is again a constant depending on K. By the choice of ε_k and h_k we have as before $F(h_k) \leq 2^k(g^{-1}(2^k))^2$ and now

$$\varepsilon_k \log\left(\frac{\exp(1)}{\varepsilon_k}\right) \leq 2^{-k}\log(2^k\exp(1)).$$

The assumption (4.12) implies for the inverse of g that

$$\frac{\sqrt{\log(y)}}{y} \precsim g^{-1}(y) \precsim \frac{(\log(y))^\gamma}{y} \tag{4.13}$$

for $y \to \infty$. With this we derive

$$\sqrt{\log(2^k\exp(1))} \precsim 2^k g^{-1}(2^k)$$

such that the worst case error can be bounded like in the proof of Theorem 3 by

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

for a constant κ_4. Thus we choose the parameter in a similar way as in the proof before by

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

and $n_k = n_k(C) = \lfloor g^{-1}(2^{k-1})C \rfloor$ for $k = 1, \ldots, m$. This time we restrict the parameter $C \geq \exp(1) \vee 1/g^{-1}(1)$. With this choice and (4.13) we obtain

$$\begin{aligned} e^2(\widehat{S}^{ML}) &\leq \frac{\kappa_5}{C} \sum_{k=1}^{m+1} 2^{k-1} g^{-1}(2^{k-1}) \\ &\precsim \frac{1}{C} \sum_{k=1}^{m+1} 2^{k-1} \frac{(\log(2^{k-1}))^\gamma}{2^{k-1}} \\ &\precsim \frac{1}{C} \sum_{k=0}^{m} k^\gamma \leq \frac{1}{C} m^{\gamma+1} \end{aligned}$$

for a constant κ_5. Remembering the choice of m as the largest natural number such that $Cg^{-1}(2^m) \geq 1$. The latter is equivalent to $2^m \leq g(1/C)$ which implies

$$m \leq \frac{\log\left(g\left(\frac{1}{C}\right)\right)}{\log(2)} \precsim \log(C)$$

for $C \to \infty$ by the assumption (4.12) where we can asymptotically neglect the double log term. With this observation we conclude that

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

For the cost we obtain with the cost bound of the last proof and the just derived

asymptotic upper bound that for a constant κ_6, we have

$$\begin{aligned}\operatorname{cost}\left(\widehat{S}^{ML}\right) &\leq 4C\sum_{k=1}^{m} 2^{k-1}g^{-1}(2^{k-1}) \\ &\leq \kappa_6 C\left(\log(C)\right)^{\gamma+1}.\end{aligned}$$

To compute the decay of the error with respect to the computational cost, we define another parameter n, which shall be asymptotically linear with respect to the cost. Therefore we define n such that

$$C = \frac{n}{\kappa_6(\log(n))^{\gamma+1}}$$

for sufficiently large $n \geq \exp(1)$. Then we conclude to have

$$\begin{aligned}\operatorname{cost}\left(\widehat{S}^{ML}\right) &\leq \kappa_6 C\left(\log(C)\right)^{\gamma+1} \\ &= \kappa_6 \frac{n}{\kappa_6(\log(n))^{\gamma+1}}\left(\log\left(\frac{n}{\kappa_6(\log(n))^{\gamma+1}}\right)\right)^{\gamma+1} \\ &\leq \frac{n}{(\log(n))^{\gamma+1}}\left(\log(n)\right)^{\gamma+1} = n\end{aligned}$$

for sufficiently large n. Inserting n into our error bound, we derive

$$\begin{aligned}e^2\left(\widehat{S}^{ML}\right) &\precsim \frac{1}{C}(\log(C))^{\gamma+1} \\ &= \frac{\kappa_6\left(\log(n)\right)^{\gamma+1}}{n}\left(\log\left(\frac{n}{\kappa_6(\log(n))^{\gamma+1}}\right)\right)^{\gamma+1} \\ &\precsim \frac{\left(\log(n)\right)^{2(\gamma+1)}}{n}.\end{aligned}$$

Now taking the square root, we obtain the proposed error of $e(\widehat{S}^{ML}) \precsim \frac{(\log(n))^{\gamma+1}}{\sqrt{n}}$ with computational cost $\operatorname{cost}(\widehat{S}^{ML}) \leq n$. □

The two previous theorems apply for $\beta < 1$. Thereby, Theorem 4 shows, that if X has a Gaussian component, the error is bounded by $n^{-1/2}(\log(n))^{3/2}$, which matches the result of [16] in the Brownian diffusion case. It is also possible to consider $\gamma \leq 1/2$ with the same computations, if there is no Gaussian component present.

For the remaining cases with $\Sigma \neq 0$, i.e. for $\beta \geq 1$, the following Theorem applies.

Theorem 5. *Let $g : (0,\infty) \to (0,\infty)$ again denote the decreasing and invertible function satisfying (4.9). If, for all sufficiently small $h > 0$, there exists $\gamma > 1$ with*

$$g\left(\frac{\gamma}{2}h\right) \geq 2g(h), \tag{4.14}$$

then there exists a family $(\widehat{S}_n)_{n\in\mathbb{N}}$ of multilevel algorithms with $\mathrm{cost}(\widehat{S}_n) \leq n$ and

$$e(\widehat{S}_n) \precsim \sqrt{n}\, g^{-1}(n).$$

Proof. By defining ε_k and h_k as before by

$$\varepsilon_k = 2^{-k} \qquad \text{and} \qquad h_k = g^{-1}(2^k) = g^{-1}(\varepsilon_k^{-1}),$$

for $k = 1, \ldots, m$, our setting so far is identical to the one of Theorem 4 such that the cost bound

$$\mathrm{cost}(\widehat{S}^{ML}) \leq \sum_{k=1}^{m} n_k 2^{k+1}$$

remains valid. The same holds true for the error bound

$$e^2(\widehat{S}^{ML}) \leq \kappa_1 \sum_{k=1}^{m+1} \frac{1}{n_k} \left(F(h_{k-1}) + \varepsilon_{k-1} \log\left(\frac{\exp(1)}{\varepsilon_{k-1}}\right)\right)$$

with $n_{m+1} = 1$ and a constant κ_1, where we have $F(h_k) \leq 2^k (g^{-1}(2^k))^2$ and

$$\varepsilon_k \log\left(\frac{\exp(1)}{\varepsilon_k}\right) \leq 2^{-k} \log\left(2^k \exp(1)\right).$$

We can find an equivalent formulation for assumption (4.14) in terms of g^{-1} by inserting $g^{-1}(u)$ instead of h which tends to zero as well for $u \to \infty$. Then we have for $u > 0$ sufficiently large that

$$\frac{\gamma}{2} g^{-1}(u) \leq g^{-1}(2u). \tag{4.15}$$

We will use this to bound the right hand-side of the error estimate uniformly. Therefore observe at first, that with (4.15), we have for $l, k \in \mathbb{N}$ with $l \geq k$ that

$$g^{-1}(2^k) \leq \kappa_2 \left(\frac{2}{\gamma}\right)^{l-k} g^{-1}(2^l) \tag{4.16}$$

for a constant κ_2 depending on g. Furthermore it is clear that the asymptotic behavior of g^{-1} is given in terms of

$$\left(\frac{\gamma}{2}\right)^k \precsim g^{-1}(2^k)$$

for $k \to \infty$. For $\gamma > 1$ we can thus bound the second term of the error estimate by

$$\begin{aligned} 2^{-k} \log\left(2^k \exp(1)\right) &\leq \left(\frac{\gamma}{2}\right)^k \log\left(2^k \exp(1)\right) \\ &\precsim g^{-1}(2^k)\, k \precsim g^{-1}(2^k)\gamma^k \\ &\precsim 2^k \left(g^{-1}(2^k)\right)^2. \end{aligned}$$

Putting this into our error estimate there exists a constant κ_3 such that

$$e^2(\widehat{S}^{ML}) \leq \kappa_3 \sum_{k=1}^{m+1} \frac{1}{n_k} 2^{k-1} \left(g^{-1}(2^{k-1})\right)^2.$$

We choose the n_k in a similar way as in the proof before by

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

for $k = 1, \ldots, m$, where this time

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

for a parameter C with $C \geq 2/g^{-1}(1)$. With this choice and (4.16) we obtain

$$\begin{aligned} e^2(\widehat{S}^{ML}) &\leq \kappa_3 \frac{2}{C} \sum_{k=1}^{m+1} 2^{k-1} g^{-1}(2^{k-1}) \\ &\leq \kappa_3 \kappa_2 \frac{2}{C} \sum_{k=1}^{m+1} 2^{k-1} \left(\frac{2}{\gamma}\right)^{m+2-k} g^{-1}(2^{m+1}) \\ &\leq \kappa_3 \kappa_2 \frac{2}{C} 2^{m+1} g^{-1}(2^{m+1}) \sum_{k=1}^{m+1} \gamma^{-(m+2-k)} \\ &\leq \kappa_4 \frac{1}{C} 2^{m+1} g^{-1}(2^{m+1}) \end{aligned}$$

for a constant κ_4. Remembering the definition of m as the largest integer for which $Cg^{-1}(2^m) \geq 2$ holds, we get for sufficiently large m with (4.15) that

$$Cg^{-1}(2^{m+1}) \geq \frac{\gamma}{2} Cg^{-1}(2^m) \geq \gamma > 1,$$

which yields

$$e^2(\widehat{S}^{ML}) \leq \kappa_4 \frac{1}{C} 2^{m+1} g^{-1}(2^{m+1}) \leq \kappa_4 2^{m+1} \left(g^{-1}(2^{m+1})\right)^2.$$

For the cost bound similar reasoning implies

$$\begin{aligned} \operatorname{cost}(\widehat{S}^{ML}) &\leq 4 \sum_{k=1}^{m} Cg^{-1}(2^{k-1}) 2^{k-1} \\ &\leq \kappa_5 Cg^{-1}(2^{m+1}) 2^{m+1} \leq 2\kappa_5 2^{m+1}, \end{aligned}$$

where we additionally use the fact that for $k > m$, we have $Cg^{-1}(2^k) < 2$ by the definition of m.

Like in the proof of Theorem 4, what remains to do is to define a parameter n, which is asymptotically linear in the cost of $\widehat{S}^{ML}$ such that we can calculate the asymptotic error with respect to n. Therefore, for $n \geq 2\kappa_5$ we choose $C > 0$ such that

$$m = \left\lfloor \log_2 \left(\frac{n}{2\kappa_5}\right) \right\rfloor - 1.$$

Then it holds $\mathrm{cost}(\widehat{S}^{ML}) \leq n$ and the error calculates to

$$\begin{aligned} e^2(\widehat{S}^{ML}) &\leq \kappa_4 2^{m+1} \left(g^{-1}(2^{m+1})\right)^2 \\ &\leq \kappa_4 \frac{n}{2\kappa_5} \left(g^{-1}\left(\frac{n}{2\kappa_5}\right)\right)^2 \\ &\leq \frac{\gamma\kappa_4}{4} \frac{n}{\kappa_5} \left(g^{-1}\left(\frac{n}{\kappa_5}\right)\right)^2 \precsim n(g^{-1}(n))^2 \end{aligned}$$

for $n \to \infty$, where we once again used (4.15). Taking the square root finishes the proof. □

Next we present a corollary, which helps to apply the theorems stated so far by only knowing the Blumenthal-Getoor index β of the driving Lévy process. Therefore, recall the definition of β for a Lévy process X with non-zero Lévy measure ν as already defined in (2.7) by

$$\beta = \inf\left\{p > 0 : \int_{B_1} |x|^p \, \nu(\mathrm{d}x) < \infty\right\}.$$

Clearly $\beta \in [0, 2]$. Now, instead of calculating the decreasing, invertible function g dominating $\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x)$ for each Lévy process, we present Corollary 1, which sets a relation between β of the driving Lévy process X and the order of convergence of our computational problem with respect to the computational cost. It is clearly not surprising that the higher the Blumenthal-Getoor index is, the smaller is the order of convergence. This relies on the fact that β expresses the frequency of the occurrence of small jumps. The higher this index is, the harder is the approximation problem itself. The result can be expressed as follows.

Corollary 1. *Let the driving Lévy process X have Blumenthal-Getoor index β. Then there exists a family $(\widehat{S}_n)_{n\in\mathbb{N}}$ of multilevel algorithms with $\mathrm{cost}(\widehat{S}_n) \leq n$ such that*

$$\sup\{\gamma \geq 0 : e(\widehat{S}_n) \precsim n^{-\gamma}\} \geq \left(\frac{1}{\beta} - \frac{1}{2}\right) \wedge \frac{1}{2}.$$

Before we present the proof of the corollary, we comment on the result as it is very compactly stated.

Remark 5. The statement presents a lower bound for the supremum of the order of convergence, which means that we do not necessarily reach this order of convergence, but at least come arbitrary close to the term on the right hand-side. Thus Corollary 1 can be reformulated in the following way: For $\beta \leq 1$, there exists a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \leq n$ such that for any $\gamma < 1/2$ we have $e(\widehat{S}_n) \precsim n^{-\gamma}$. For $\beta > 1$, the order of convergence decreases, but we still reach for any $\gamma < 1/\beta - 1/2$ that there exists a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \leq n$ such that $e(\widehat{S}_n) \precsim n^{-\gamma}$.

Proof. For $\beta = 2$, the assertion is trivial and thus we consider $\beta < 2$. For fixed such β, we analyze the function

$$\tilde{g}(h) = \int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x)$$

for small $h > 0$ to apply one of the theorems above. Observe that for $h \in (0, 1]$, we can decompose $\tilde{g}$ into

$$\tilde{g}(h) = \int_{B_h} \frac{|x|^2}{h^2} \, \nu(\mathrm{d}x) + \int_{B_1 \setminus B_h} 1 \, \nu(\mathrm{d}x) + \int_{B_1^c} 1 \, \nu(\mathrm{d}x).$$

The third integral is bounded, as

$$\int_{B_1} 1 \, \nu(\mathrm{d}x) = \nu(B_1) < \infty, \tag{4.17}$$

independent of the value of h, such that it is asymptotically negligible. For the other two integrals, we take $\beta' \in (\beta, 2)$. Then by the definition of β, we have

$$\int_{B_1} |x|^{\beta'} \, \nu(\mathrm{d}x) < \infty.$$

Using this, we can bound the other two integrals by

$$\int_{B_h} \frac{|x|^2}{h^2} \, \nu(\mathrm{d}x) \leq \int_{B_h} \frac{|x|^{\beta'}}{h^{\beta'}} \, \nu(\mathrm{d}x) \leq h^{-\beta'} \int_{B_1} |x|^{\beta'} \, \nu(\mathrm{d}x) \leq C h^{-\beta'} \tag{4.18}$$

and

$$\int_{B_1\setminus B_h} 1\,\nu(\mathrm{d}x) \le \int_{B_1\setminus B_h} \frac{|x|^{\beta'}}{h^{\beta'}}\,\nu(\mathrm{d}x) \le h^{-\beta'}\int_{B_1} |x|^{\beta'}\,\nu(\mathrm{d}x) \le Ch^{-\beta'} \tag{4.19}$$

for the same constant $C > 0$, respectively. With β' as above and (4.17),(4.18) and (4.19), we can always find a decreasing and invertible function $g : (0,\infty) \to (0,\infty)$ that dominates $\tilde{g}$ and such that

$$g(h) = Ch^{-\beta''} \tag{4.20}$$

for all sufficiently small $h > 0$ and any $\beta'' > \beta'$. In this case, we can use Theorem 5, where the assumption on g is to satisfy

$$g\left(\frac{\gamma}{2}h\right) \ge 2g(h)$$

for some $\gamma > 1$ and all sufficiently small $h > 0$. For g as in (4.20), these assumption are equivalent to

$$1 < \gamma \le 2^{1-\frac{1}{\beta''}}.$$

Here, $\beta'' > 1$ implies the existence of $\gamma = 2^{1-1/\beta''} > 1$ and thus we take $\beta'' \in ((\beta' \vee 1, 2]$. Theorem 5 now implies that there exists a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \le n$ and

$$e(\widehat{S}_n) \precsim \sqrt{n}g^{-1}(n).$$

For g given by (4.20) we derive $g^{-1}(y) = \left(\frac{y}{C}\right)^{-1/\beta''}$ for sufficiently large $y > 0$, which yields

$$e(\widehat{S}_n) \precsim \sqrt{n}\left(\frac{n}{C}\right)^{-\frac{1}{\beta''}} = n^{\frac{1}{2}}\,C^{\frac{1}{\beta''}}\,n^{-\frac{1}{\beta''}} \precsim n^{-\left(\frac{1}{\beta''}-\frac{1}{2}\right)}.$$

For $\beta > 1$, we can now choose $\beta'' > \beta$ arbitrary close to β such that the infimum over those β'' is β. It follows that the supremum over $1/\beta'' - 1/2$ is given by $1/\beta - 1/2$.

For $\beta < 1$, we can choose $\beta' = 1$ and can then again choose the $\beta'' > 1$ arbitrary

close to one which yields order of convergence $1/\beta'' - 1/2 < 1/2$ such that with the same considerations as above we deduce that the supremum of convergence orders is $1/2$.

For $\beta = 1$, one can again choose the value $\beta'' > \beta$ arbitrary close to β and still find a β' with $\beta'' > \beta' > \beta$ such that the conclusions above hold and result in a supremum of convergence orders of value $1/2$, which finishes the proof. □

4.4 Lower bound for Lévy processes

Now the natural question arises, if it is possible to construct algorithms that perform better then the multilevel algorithm in the above sense. Therefore, we are interested in a lower bound for the error in relation to the computational cost, i.e. an error, which occurs for any randomized algorithm for the computational problem outlined at the beginning of this chapter. Therefore, to combine already existing results of [11] and [3], we will relax the computational problem by taking a one dimensional Lévy process itself as desired distribution, i.e. $Y = X$, and we will weaken the definition of the cost by only counting the number of function evaluations. To be more precise, we shortly present the setting of [11]. Here, the worst case error of any randomized algorithm $\widehat{S}$, that terminates in finite time, is defined as in our setting by

$$e\left(\widehat{S}\right) = \sup_{f \in F} \left(\mathrm{E}\left[\left| S(f) - \widehat{S}(f) \right|^2 \right] \right)^{\frac{1}{2}}.$$

We are interested in the so-called full sampling case, where the algorithm is allowed to evaluate any functional $f \in F$ at any point $x \in \mathcal{X}$ at constant computational cost 1. Here, $\mathcal{X}$ is the underlying Banach space in which the desired distribution Y has its values in, i.e. in our setting $\mathcal{X} = D([0,1])$ and the evaluation of any kind of approximation $\hat{Y}^{(k)}$ has unit cost. Thus, in the case of full sampling, we count as cost the number of function evaluations of f in $\widehat{S}$, denoted by $\mathrm{cost}_{\mathrm{full}}\left(\widehat{S}, f\right)$. The computational cost of the algorithm $\widehat{S}$ is given as before by the worst case of the average cost over the function class F,

$$\mathrm{cost}_{\mathrm{full}}\left(\widehat{S}\right) = \sup_{f \in F} \mathrm{E}\left[\mathrm{cost}_{\mathrm{full}}\left(\widehat{S}, f\right) \right].$$

The resulting n-th minimal error is defined by

$$e_n = \inf\left\{e\left(\widehat{S}\right) : \mathrm{cost}_{\mathrm{full}}\left(\widehat{S}\right) \leq n\right\}.$$

Clearly, as we have $\mathrm{cost}_{\mathrm{full}}(\widehat{S}) \leq \mathrm{cost}(\widehat{S})$ for our computational cost function given in (4.8), it holds

$$\inf\left\{e(\widehat{S}) : \mathrm{cost}\left(\widehat{S}\right) \leq n\right\} \geq e_n,$$

and thus, the lower bound in the setting of [11] provides a lower bound for our setting. The latter is given in terms of m-th quantization numbers of order $r = 1$. Before we can define these, we have to define the L_1-Wasserstein distance $\mathcal{W}^{(1)}$ between two probability measures μ and $\hat{\mu}$. As a consequence of the Kantorovich-Rubinstein theorem and the fact, that we consider $F = \mathrm{Lip}(1)$ on the separable space $\mathcal{X} = D[0,1]$, the L_1-Wasserstein distance can be defined by

$$\mathcal{W}^{(1)}(\mu, \hat{\mu}) = \sup_{f \in F}\left|\int_{\mathcal{X}} f(x)\,\mu(\mathrm{d}x) - \int_{\mathcal{X}} f(x)\,\hat{\mu}(\mathrm{d}x)\right|.$$

Then the m-th quantization number of order $r = 1$ for a probability measure μ on $\mathcal{X}$ is defined by

$$q_m^{(1)} = \inf\left\{\mathcal{W}^{(1)}(\mu, \hat{\mu}) : |\mathrm{supp}(\hat{\mu})| \leq m\right\},$$

Given the above definitions, we can now state Corollary 2 from [11] which will imply the lower bound of interest.

Theorem 6. *Let $f : [0,\infty) \to (0,\infty)$ be convex and differentiable. If*

$$\limsup_{n\to\infty} \frac{q_n^{(1)}}{f(n)} \geq 1 \qquad \textit{and} \qquad \lim_{n\to\infty} q_n^{(1)} = 0,$$

then we have

$$\limsup_{N\to\infty} e_N / \left(N^{\frac{1}{2}} \cdot |f'|(4N+3)\right) \geq \frac{1}{8}.$$

Now observe that with the results from Theorem 1.5 in [3], for non-vanishing

Gaussian component, i.e. $\Sigma \neq 0$ we have

$$q_n^{(1)} \gtrsim (\log(n))^{-\frac{1}{2}}.$$

With the same result and the assumption that the integral from (4.9) prevails the following asymptotic lower bound, $\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x) \gtrsim h^{-\alpha}$ for $\alpha \in (0,2)$, we have

$$q_n^{(1)} \gtrsim (\log(n))^{-\frac{1}{\alpha}}.$$

For the first case taking $\alpha = 2$, we can now define the convex and differentiable function f needed in the result above by

$$f(n) = C\,(\log(n))^{-\frac{1}{\alpha}}$$

with a constant $C > 0$ such that $\limsup_{n\to\infty} q_n^{(1)}/f(n) \geq 1$ is fulfilled. Then the asymptotic of the derivative is $|f'|(4N+3) \asymp (\log(N))^{-(1+\frac{1}{\alpha})} \cdot N^{-1}$ and the result above yields the following

Theorem 7. *For Y being a real-valued Lévy process satisfying $\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x) \gtrsim h^{-\alpha}$ for $\alpha \in (0,2)$ or having non-vanishing Gaussian component in which case we set $\alpha = 2$, the N-th minimal error e_N fulfills*

$$\limsup_{N\to\infty} N^{\frac{1}{2}} \, (\log(N))^{1+\frac{1}{\alpha}} \cdot e_N > 0. \tag{4.21}$$

Thus, in the case of Blumenthal-Getoor index $\beta < 1$ the multilevel algorithm achieves optimal order up to logarithmic terms. For the case $\beta \geq 1$ the multilevel algorithm can be further improved, see [12], but no asymptotic optimal algorithms are known so far.

4.5 Asymptotic choice of parameters

For the convenience of the reader, we now summarize the parameters which achieve the asymptotic orders of convergence provided in Theorem 3, 4 and 5 and which are hidden in the proofs. The choices will always correspond to algorithms with computational cost at most n.

Remember, that the algorithm $\widehat{S}^{ML}$ is completely determined by the parameters m and $(n_k, \varepsilon_k, h_k)_{k=1,\dots,m}$. Further recall that all theorems depend on an invertible and decreasing function $g : (0,\infty) \to (0,\infty)$ satisfying

$$\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x) \leq g(h),$$

for all $h > 0$. We always define our threshold parameters for any $k \in \mathbb{N}$ by

$$\varepsilon_k = 2^{-k} \quad \text{and} \quad h_k = g^{-1}(2^k).$$

For the remaining, we introduce an auxiliary parameter C, which differs in the three theorems, but is always determined in terms of the cost size n. With respect to this parameter C, the replication parameters n_k are chosen to be

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

for $k = 1, \dots, m$. The choice of the highest level m also differs in the three theorems.

For Theorem 3, we set $C = n$ and the highest level m is set to

$$m = \inf\{k \in \mathbb{N} : \; C \, h_k < 1\} - 1.$$

Here we additionally assume without loss of generality, that $h^{-2/3} \precsim g(h)$.

In Theorem 4, the relation of n and C is given by

$$C = \frac{n}{(\log(n))^{\gamma+1}},$$

while the finest approximation level again is $m = \inf\{k \in \mathbb{N} : \; C \, h_k < 1\} - 1$. Here, we have an additional assumption on the asymptotic behavior of g in terms of the lower bound $h^{-1}\sqrt{\log(1/h)} \precsim g(h)$. This can also be made without loss of generality.

In the last Theorem 5, the assumption on g is given in terms of a scaling decay and not by a pure asymptotic, so that the parameter C as well depends on g by

$$C = 1/g^{-1}(n).$$

The corresponding highest level is given by

$$m = \inf\{k \in \mathbb{N} : \ C\, h_k < 2\} - 1.$$

These parameters optimize, up to constant multiples, the error estimate given by equation (4.7) together with the strong approximation results of Theorem 2 and the computational cost for multilevel algorithms given in (4.8) which are a consequence of our algorithmic setting given in Chapter 1 and at the very beginning of this chapter.

4.6 Examples

We now apply our results of Theorems 3, 4 and 5 to our examples of driving Lévy processes and to *jump-diffusions*. Here, a jump-diffusion is the independent sum of a Brownian motion and a compound Poisson process. Numerical results for an example with truncated stable processes and an example in the Barndorff-Nielsen Shephard model are presented in an own chapter after the theoretical results. There, we will also introduce a way to choose the parameters of the algorithm during the course of computation by suitable bias and variance estimations, which approximately ensure to reach a desired precision.

Jump-diffusions

In finance, special cases of jump-diffusions are used, e.g., in the Merton or the Kou model, to model the log price process, see [28] and [32] for details. The compound Poisson process has finite Lévy measure and thus the integral $\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x)$ is bounded by a constant. Hence, the trade-off between cost and error is the same as in the pure Gaussian case, i.e. there exists a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \leq n$ and $e(\widehat{S}_n) \precsim n^{-1/2}(\log(n))^{3/2}$.

Stable Lévy processes

First recall that the Lévy measure of an α-stable process is given by

$$\nu(\mathrm{d}x) = \left(\mathbb{1}_{(0,\infty]}(x) \frac{A_+}{|x|^{1+\alpha}} + \mathbb{1}_{[-\infty,0)}(x) \frac{A_-}{|x|^{1+\alpha}} \right) \cdot \mathrm{d}x,$$

for $x \in \mathbb{R}\backslash\{0\}$ and $\alpha \in (0,2)$, with $A_+, A_- \geq 0$ such that $A_+ + A_- > 0$. Due to the assumption of second moments, we consider processes where the jumps have α-stable behavior around 0 while the big jumps are either truncated at some given size $u > 0$ or tempered with an exponential decay. Then the influence of the big jumps to the function g as defined in (4.9) is only a constant, independent of $h > 0$. For small $h \to 0$, we thus derive the asymptotic bound

$$\begin{aligned} \int \frac{|x|^2}{h^2} \wedge 1 \; \nu(\mathrm{d}x) &= h^{-2} \int_{B_h} |x|^2 \, \nu(\mathrm{d}x) + \int_{B_1 \setminus B_h} 1 \, \nu(\mathrm{d}x) + \nu(B_1^c) \\ &\precsim h^{-2} \int_{B_h} |x|^{1-\alpha} \, \mathrm{d}x + h^{-\alpha} \int_{B_1 \setminus B_h} |x|^{-1} \, \mathrm{d}x \\ &\precsim h^{-\alpha} + h^{-\alpha} \log\left(\frac{1}{h}\right) \precsim h^{-\alpha'}, \end{aligned}$$

for any $\alpha' > \alpha$, where the constants neglected by the asymptotic treatment only depend on the parameters of the Lévy density. With these asymptotic bounds for g, we can now apply our theoretical results.

For $\alpha < 1$ Theorem 3 can be used because (4.10) is fulfilled for $\gamma > 0$ and $h \to 0$. The asymptotic error for a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \leq n$ thus behaves like $e(\widehat{S}_n) \precsim n^{-\frac{1}{2}}$.

In the case with $\alpha > 1$ Theorem 5 can be used with $\gamma = 2^{\frac{\alpha'-1}{\alpha'}} > 1$ which ensures the existence of a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \leq n$ which fulfill $e(\widehat{S}_n) \precsim n^{-(\frac{1}{\alpha'}-\frac{1}{2})}$ for $\alpha' > \alpha$ arbitrary close.

The case $\alpha = 1$ can be treated via Theorem 4 with $\gamma = 1$ which yields $e(\widehat{S}_n) \precsim n^{-\frac{1}{2}}(\log(n))^2$ for a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \leq n$.

Variance Gamma processes

We first recall that the Lévy measure of a VG process is given by

$$\nu(\mathrm{d}x) = \frac{a}{|x|} \exp\left(-\sqrt{2b}|x|\right) \cdot \mathrm{d}x,$$

for $x \in \mathbb{R}\backslash\{0\}$.

To apply our results and obtain the order of convergence for the case of a VG process as driving Lévy process, we need an asymptotic upper bound for $\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x)$. Therefore, observe that we can decompose the latter integral for $0 < h \leq 1$ like in the proof of Corollary 1 into

$$\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x) = \int_{B_h} \frac{|x|^2}{h^2} \, \nu(\mathrm{d}x) + \int_{B_1 \backslash B_h} 1 \, \nu(\mathrm{d}x) + \int_{B_1^c} 1 \, \nu(\mathrm{d}x).$$

Inserting the density of the Lévy measure, the first term can be estimated by

$$\begin{aligned}
\int_{B_h} \frac{|x|^2}{h^2} \frac{a}{|x|} \exp\left(-\sqrt{2b}|x|\right) \mathrm{d}x &\leq h^{-\frac{1}{2}} 2 \int_0^h \frac{a}{\sqrt{x}} \exp\left(-\sqrt{2b}x\right) \mathrm{d}x \\
&= h^{-\frac{1}{2}} 4 \int_0^{\sqrt{h}} a \exp\left(-\sqrt{2b}x^2\right) \mathrm{d}x \\
&\leq h^{-\frac{1}{2}} 4a \int_0^1 \exp\left(-\sqrt{2b}x^2\right) \mathrm{d}x \leq C \, h^{-\frac{1}{2}}
\end{aligned}$$

for a constant $C > 0$. For the second term we derive for the same constant C that

$$\begin{aligned}
\int_{B_1 \backslash B_h} \frac{a}{|x|} \exp\left(-\sqrt{2b}|x|\right) \mathrm{d}x &\leq \int_{B_1 \backslash B_h} \left(\frac{|x|}{h}\right)^{\frac{1}{2}} \frac{a}{|x|} \exp\left(-\sqrt{2b}|x|\right) \mathrm{d}x \\
&= h^{-\frac{1}{2}} 4 \int_{\sqrt{h}}^1 a \exp\left(-\sqrt{2b}x^2\right) \mathrm{d}x \\
&\leq h^{-\frac{1}{2}} 4a \int_0^1 \exp\left(-\sqrt{2b}x^2\right) \mathrm{d}x \leq C \, h^{-\frac{1}{2}}.
\end{aligned}$$

The last term can be approximated analogously by

$$\begin{aligned}\int_{B_1^c} \frac{a}{|x|} \exp\left(-\sqrt{2b}|x|\right) \mathrm{d}x &\leq \int_{B_1^c} \frac{a}{|x|^{\frac{1}{2}}} \exp\left(-\sqrt{2b}|x|\right) \mathrm{d}x \\ &\leq 2\int_1^\infty \frac{a}{x^{\frac{1}{2}}} \exp\left(-\sqrt{2b}x\right) \mathrm{d}x \\ &\leq 4a\int_0^\infty \exp\left(-\sqrt{2b}x^2\right) \mathrm{d}x = 4a\frac{\sqrt{\pi}\,(2b)^{\frac{1}{4}}}{2},\end{aligned}$$

where the latter integral is determined as an integral with respect to a Gaussian distribution with variance $(2\sqrt{2b})^{-1}$. Putting together the estimates from above, we can bound the integral by

$$\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x) \precsim h^{-\frac{1}{2}},$$

for $h \to 0$. Thus, having no Brownian component occurring, we can use Theorem 3 and deduce as order of convergence 1/2, i.e., there exists a sequence of multilevel algorithms $(\widehat{S}_n)_{n\in\mathbb{N}}$ with $\mathrm{cost}(\widehat{S}_n) \leq n$ such that $e(\widehat{S}_n) \precsim n^{-\frac{1}{2}}$.

5 Numerical simulations

In Chapter 3, we have already explained ways to simulate the processes under investigation. What remains to do in the implementation of a multilevel algorithm on a computer is to determine the parameters m, $(h_k, \varepsilon_k)_{k=1,\dots,m}$ and $(n_k)_{k=1,\dots,m}$. By the theoretical results of Chapter 4, this can be done asymptotically for a given upper bound of the computational cost by the choices of Section 4.5. However, in most applications, we are interested in first place in bounding the error of the multilevel algorithm, and in second place, but still of great importance, to have an idea of the computational cost. In the following, we present a way to implement the multilevel algorithm such that the latter should be achieved. Hereby, we fix the choice of threshold parameters $(h_k, \varepsilon_k)_{k\in\mathbb{N}}$ for the levels while the choices of m and the replication parameters n_k are done during the course of computation via bias and variance estimations. Along this way, we are also abled to approximate the remaining time of the algorithm to terminate.

5.1 Bias and variance estimation for MLMC

In the following, $\delta > 0$ denotes the precision we want to achieve with our algorithm, e.g. $\delta = 10^{-p}$ for $p \in \mathbb{N}$. Remember, that we have to determine the highest level m as well as the various threshold parameter $(h_k, \varepsilon_k)_{k=1,\dots,m}$ and the replication numbers $(n_k)_{k=1,\dots,m}$ to run the multilevel algorithm. At first we fix our level parameters $(h_k, \varepsilon_k)_{k\in\mathbb{N}}$ such that

$$\varepsilon_k = 2^{-k} \quad \text{and} \quad \nu(B_{h_k}^c) = 2^k, \tag{5.1}$$

if possible, and in an approximate way if the Lévy measure itself has to be approximated numerically. With this choice the average number of jumps we have to

simulate on level k is 2^k as it has already been proposed in the classical setting in [16] for $M = 2$. For our first example of truncated symmetric α-stable processes with parameters $c, u > 0$ and $\alpha \in (0, 2)$, the Lévy measure of jump sizes greater h can be directly calculated to

$$\nu\left(B_h^c\right) = \frac{2c}{\alpha}\left(h^{-\alpha} - u^{-\alpha}\right),$$

and so it is possible to choose

$$h_k = \left(\frac{2^k \alpha}{2c} + u^{-\alpha}\right)^{-1/\alpha}$$

according to (5.1). Observe, that in the case of variance gamma, inverse Gaussian or tempered stable processes the threshold parameters are chosen approximately fulfilling (5.1). This can be done independently of the algorithm itself. The integrals involved are handled by suitable deterministic quadrature rules.

Remembering the decomposition of our L_2-error into the bias and the variance of $\widehat{S}^{ML}(f)$ via

$$\mathrm{E}\left[\left|S(f) - \widehat{S}^{ML}(f)\right|^2\right] = \underbrace{\left|\mathrm{E}\left[f(Y) - f\left(\hat{Y}^{(m)}\right)\right]\right|^2}_{=\mathrm{bias}(\widehat{S}^{ML}(f))} + \mathrm{var}\left(\widehat{S}^{ML}(f)\right),$$

the next step to implement the algorithm is the determination of the highest level m and the replication numbers $n_1, \ldots, n_m$. Given the desired precision δ for the root mean square error, the algorithm should fulfill

$$\left|\mathrm{bias}\left(\widehat{S}^{ML}(f)\right)\right|^2 \leq \frac{\delta^2}{2} \tag{5.2}$$

and

$$\mathrm{var}\left(\widehat{S}^{ML}(f)\right) \leq \frac{\delta^2}{2}. \tag{5.3}$$

In applications we usually do not have a reference solution, so we have to estimate the remaining bias after level m. To do so, we use a heuristic argument explained in the following. We will frequently use the following notation for the expectations

of the differences on consecutive levels:

$$\text{bias}_k := \begin{cases} \mathrm{E}\left[f\left(\hat{Y}^{(k)}\right) - f\left(\hat{Y}^{(k-1)}\right)\right] & , k \geq 2, \\ \mathrm{E}\left[f\left(\hat{Y}^{(1)}\right)\right] & , k = 1. \end{cases}$$

Due to Jensen's inequality and our strong approximation result we have convergence of the approximations in L_1, i.e. $\mathrm{E}[f(\hat{Y}^{(k)})] \to \mathrm{E}[f(Y)]$ for $k \to \infty$ such that we can decompose the latter into

$$S(f) = \mathrm{E}[f(Y)] = \sum_{k=1}^{\infty} \text{bias}_k.$$

Consequently, the bias of our multilevel estimator has the representation

$$\begin{aligned} \text{bias}\left(\widehat{S}^{ML}(f)\right) &= \text{bias}\left(f\left(\hat{Y}^{(m)}\right)\right) = \mathrm{E}\left[f(Y) - f\left(\hat{Y}^{(m)}\right)\right] \\ &= \sum_{k=m+1}^{\infty} \mathrm{E}\left[f\left(\hat{Y}^{(k)}\right) - f\left(\hat{Y}^{(k-1)}\right)\right] = \sum_{k=m+1}^{\infty} \text{bias}_k. \end{aligned}$$

Monte Carlo simulations for the latter expectations suggest the hypothesis that there exist constants $\kappa_{\text{bias}} \in \mathbb{R}$ and $\varrho_{\text{bias}} \in]0,1[$ such that eventually

$$|\text{bias}_k| \leq \kappa_{\text{bias}} \cdot \varrho_{\text{bias}}^k.$$

In fact, the strong approximation result of Theorem 2 assures this hypothesis asymptotically. This decay implies

$$\text{bias}\left(\widehat{S}^{ML}(f)\right) \leq \sum_{k=m+1}^{\infty} |\text{bias}_k| \leq \sum_{k=m+1}^{\infty} \kappa_{\text{bias}} \cdot \varrho_{\text{bias}}^k,$$

which yields the following upper bound for the bias:

$$\text{bias}\left(\widehat{S}^{ML}(f)\right) \leq \kappa_{\text{bias}} \cdot \varrho_{\text{bias}}^{m+1} \cdot \sum_{k=0}^{\infty} \varrho_{\text{bias}}^k = \frac{\kappa_{\text{bias}}}{1 - \varrho_{\text{bias}}} \varrho_{\text{bias}}^{m+1}.$$

With the above estimate, we can choose the highest level m fulfilling (5.2) for

given $\delta > 0$ approximately by

$$m = \left\lceil \frac{\log((1-\varrho_{\text{bias}})\delta) - \log\left(\sqrt{2}\cdot\kappa_{\text{bias}}\right)}{\log(\varrho_{\text{bias}})} - 1 \right\rceil. \tag{5.4}$$

To estimate the parameters κ_{bias} and ϱ_{bias} of the exponential decay of the bias we apply a classical log-linear regression, i.e. we apply a best linear L_2-fit to the dependence of the logarithm of the absolute values of the empirical means of $(\text{bias}_k)_{k\geq 2}$ against the levels on the first few levels with $k \geq 2$. We then choose κ_{bias} and ϱ_{bias} to be the exponential of the intercept and the slope of the regression, respectively.

To run the multilevel algorithm we still have to define the number of replications on each of the levels, namely the parameters $(n_k)_{k=1,\dots,m}$. Our goal here is to minimize the cost of the algorithm under the constraint (5.3) for the given $\delta > 0$. Similar to the bias considerations, we at first fix some notation for reasons of simplicity. We denote by

$$\text{var}_k := \begin{cases} \text{var}\left(f\left(\hat{Y}^{(k)}\right) - f\left(\hat{Y}^{(k-1)}\right)\right) & , k \geq 2, \\ \text{var}\left(f\left(\hat{Y}^{(1)}\right)\right) & , k = 1 \end{cases}$$

the variances of the differences of coupled consecutive approximations. Then constraint (5.3) is given by

$$\text{var}\left(\widehat{S}^{ML}(f)\right) = \sum_{k=1}^{m} \frac{\text{var}_k}{n_k} \leq \frac{\delta^2}{2}.$$

The computational cost of the algorithm is given as before by

$$\text{cost}\left(\widehat{S}^{ML}\right) = \sum_{k=1}^{m} n_k \text{E}\left[\Upsilon(\hat{Y}^{(k)})\right]$$

with $\Upsilon(\hat{Y}^{(k)})$ as before, denoting the number of breakpoints of $\hat{Y}^{(k)}$.

The cost can be estimated while simulating the trajectories by counting the number of breakpoints and taking the mean, or, in the case without Brownian component in the Lévy process, it can be directly calculated by $\text{E}[\Upsilon(\hat{Y}^{(k)})] = \nu(B^c_{h_k}) + 2$ which is approximately $2^k + 2$ by the choice of h_k. Here the additional 2 originates from

always taking 0 and the endpoint $T = 1$ as discretization points as well, although there exists no jump at these points almost surely. By the choices of (ε_k, h_k), we also always have at least the cost bound $\mathrm{E}[\Upsilon(\hat{Y}^{(k)})] \leq 2^{k+1}$.

The values for var_k are estimated on the first few levels by the empirical variances of the means of bias_k. With the hypothesis that

$$\mathrm{var}_k \leq \kappa_{\mathrm{var}} \cdot \varrho_{\mathrm{var}}^k, \tag{5.5}$$

we once again solve a log-linear regression of the empirical values for var_k against the levels for $k \geq 2$. With the resulting values for κ_{var} and ϱ_{var} we can calculate estimators for var_k for the bigger levels up to level m using (5.5).

Given values for $(\mathrm{var}_k)_{k=1,\dots,m}$ and $(\mathrm{E}[\Upsilon(\hat{Y}^{(k)})])_{k=1,\dots,m}$, the replication parameters $(n_k)_{k=1,\dots,m}$ should solve the minimization problem

$$\begin{aligned} \min_{n \in \mathbb{N}^m} \; & f(n) = \sum_{k=1}^{m} n_k \mathrm{E}\left[\Upsilon(\hat{Y}^{(k)})\right] \\ \text{s.t.} \; & g(n) = \sum_{k=1}^{m} \frac{\mathrm{var}_k}{n_k} \leq \frac{\delta^2}{2}. \end{aligned}$$

For simplicity, we solve this problem for $n \in \mathbb{R}^m$ and equality in the constraint. Then the method of Lagrange multipliers can be used with Lagrange function $\Lambda(\lambda, n) = f(n) + \lambda(g(n) - \frac{\delta^2}{2})$. Setting $\nabla\Lambda = 0$ we obtain the following necessary conditions for the extremum

$$n_k = \sqrt{\frac{\lambda \mathrm{var}_k}{\mathrm{E}[\Upsilon(\hat{Y}^{(k)})]}}, \quad k = 1, \dots, m,$$

with

$$\sqrt{\lambda} = \frac{2}{\delta^2} \sum_{k=1}^{m} \sqrt{\mathrm{var}_k \cdot \mathrm{E}\left[\Upsilon(\hat{Y}^{(k)})\right]}.$$

This yields the choice of

$$n_k = \left\lceil \frac{2}{\delta^2} \sqrt{\frac{\mathrm{var}_k}{\mathrm{E}[\Upsilon(\hat{Y}^{(k)})]}} \sum_{k=1}^{m} \sqrt{\mathrm{var}_k \cdot \mathrm{E}\left[\Upsilon(\hat{Y}^{(k)})\right]} \right\rceil$$

to ensure that the constraint in the optimization is fulfilled, and that on the finest level we still have $n_m \geq 1$.

Before we now present the results of numerical simulations of the multilevel algorithm, we remember how to choose asymptotic confidence intervals in the classical Monte Carlo setting. The latter appears for instance in the estimation of the root mean square error of our algorithm as well as in the variance and bias estimation procedure. In this cases, confidence intervals are given in the following way.

Asymptotic confidence intervals for classical MC

We consider the problem of computing the expectation of a r.v. Z. In our setting, reasonable choices of Z are given by $Z = (f(\hat{Y}^{(k)}) - f(\hat{Y}^{(k-1)}))^p$ for $p = 1, 2$ or $Z = (\widehat{S}(f) - S(f))^2$, where the latter is, e.g., used to estimate the mean squared error of an algorithm $\widehat{S}$ for a given functional f. We can then always consider the classical Monte Carlo estimator

$$\hat{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i,$$

where $(Z_i)_{i \in \mathbb{N}}$ is an *i.i.d.* sequence of copies of Z. With the central limit theorem, we can conclude that for large n, the arithmetical mean $\hat{Z}$ is approximately Gaussian. The unknown variance is estimated by the empirical variance

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \hat{Z})^2.$$

Then an asymptotic confidence interval with confidence level 0.95 is given by

$$\left[\hat{Z} - \frac{\hat{\sigma}}{\sqrt{n}} \cdot 1.96, \hat{Z} + \frac{\hat{\sigma}}{\sqrt{n}} \cdot 1.96 \right],$$

where 1.96 is the 0.975-quantile of the standard normal distribution $\mathcal{N}(0,1)$. Proofs and references for classical confidence intervals can be found in most textbooks on statistics. For this setting we refer to the recent monograph on Monte Carlo algorithms [33].

Treatment of the examples

Before we present the specific examples, we will describe our general procedure.

We will always start with the exact description of the problem, i.e., we define the parameters of the SDE and the functional under consideration. After that, we start the bias and variance estimation procedure as described in Section 5.1. To this end, we have to choose the threshold parameters for the Euler scheme. This is always done such that (approximately)

$$\varepsilon_k = 2^{-k} \quad \text{and} \quad \nu(B_{h_k}^c) = 1/\varepsilon_k = 2^k.$$

For small levels k we estimate bias_k and var_k by the empirical mean of 1000 simulations and proceed our log-linear regression. If the confidence intervals chosen according to the section before are not convincing for the regression lines, we perform more simulations until the confidence regions match the corresponding regression lines.

With the estimates for the decay of $(\text{bias}_k)_{k\geq 2}$ and $(\text{var}_k)_{k\geq 2}$, we can determine the parameters m and $(n_k)_{k=1,\dots,m}$ depending on δ. Then, for several values of δ, the corresponding choices of replication numbers are determined and plotted.

The goal of our simulation experiment is to recover the relation between the root mean square error $(\mathrm{E}[S(f) - \widehat{S}^{ML}(f)]^2)^{1/2}$ and the computational cost of the multilevel algorithm $\mathrm{E}[\text{cost}(\widehat{S}^{ML}(f), f)]$. Therefore, we average the squared error and the cost of 1000 independent multilevel Monte Carlo simulations for the different precisions and corresponding replication numbers. Here, the unknown value $S(f)$ is replaced by the output of a master computation, given in terms of a multilevel output with an accuracy that is approximately one digit higher than the smallest δ. The results are presented in a log-log plot together with the linear regression lines whose slopes determine the empirical orders of convergence.

We then compare the multilevel algorithm with the classical Monte Carlo one. In

the latter case, we do not have a heuristic control of the bias and so, if possible, we use the strong approximation error of Theorem 2 as an upper bound. All unknown constants appearing in the error estimate as well as the unknown variance are assumed to be 1. With this considerations, we determine the parameters h, ε and the replication number n in order to reach precision $\delta > 0$ for the root mean square error. The empirical orders of convergence of the classical Monte Carlo algorithm are also obtained from 1000 simulations where the error is again determined with respect to the output of the master computation.

Observe that for small precisions $\delta > 0$ all simulations from the bias and variance estimation procedure can be included in our Euler scheme not yielding any additional cost. This can be easily seen in the graphics showing the replication numbers in Figure 5.2 and 5.5.

5.2 Lookback option and truncated stable processes

Our first numerical example will be a lookback option with strike $K = 1$, that is

$$f(Y) = \left(\sup_{t \in [0,1]} Y_t - 1 \right)_+ .$$

The process Y thereby satisfies the following stochastic differential equation

$$\begin{aligned} \mathrm{d}Y_t &= Y_{t-} \, \mathrm{d}X_t, \qquad t \in [0,1], \\ Y_0 &= 1, \end{aligned}$$

with X being a symmetric, truncated α-stable Lévy process as defined in Section 3.3 with index of stability α varying in $\{0.5, 0.8, 1.2\}$, while $u = 1$ and $c = 0.1$.

We choose the threshold parameters for the Euler scheme by

$$\varepsilon_k = 2^{-k} \quad \text{and} \quad h_k = \left(\frac{2^k \alpha}{2c} + u^{-\alpha} \right)^{-1/\alpha},$$

which is the unique value with $\nu(B^c_{h_k}) = 2^k$. Then, to reach a desired precision δ

measured in the root mean square error, we apply our bias and variance estimation procedure as defined in Section 5.1, i.e. we estimate bias_k and var_k for small levels k (here $k \leq 4$ or $k \leq 5$) by the empirical mean of 1000 simulations and proceed our log-linear regression, which can be seen in Figure 5.1. The corresponding decay parameter are $\kappa_{\text{bias}} = 0.06652732$, $\varrho_{\text{bias}} = 0.3097419$, $\kappa_{\text{var}} = 0.0161213$ and $\varrho_{\text{var}} = 0.1561709$ for $\alpha = 0.5$, $\kappa_{\text{bias}} = 0.08665627$, $\varrho_{\text{bias}} = 0.472498$, $\kappa_{\text{var}} = 0.02141904$ and $\varrho_{\text{var}} = 0.3830893$ for $\alpha = 0.8$ and $\kappa_{\text{bias}} = 0.09401562$, $\varrho_{\text{bias}} = 0.6638959$, $\kappa_{\text{var}} = 0.03301414$ and $\varrho_{\text{var}} = 0.6731316$ for $\alpha = 1.2$. With this estimates for the decay of the bias and the variance, we can determine the parameters m and $(n_k)_{k=1,\dots,m}$. For several values of δ, the corresponding choice of replication numbers is shown in Figure 5.2.

Next, we average the squared error and the cost of 1000 independent multilevel Monte Carlo simulations for the precisions and corresponding replication numbers given in Figure 5.2. Here, the unknown values for $S(f)$ are replaced by outputs of master computations. For $\alpha = 0.5$, a multilevel algorithm with accuracy $\delta = 0.00002$ provides as master computation output the value 0.1530069. In the case $\alpha = 0.8$, we compute 0.2120798 and for $\alpha = 1.2$ we obtain 0.3388049 for accuracies $\delta = 0.0001$ and 0.001, respectively. The results are presented in a log-log plot in Figure 5.3 together with the linear regression lines whose slopes determine the empirical orders of convergence. Thereby, the multilevel experiments are represented by the solid lines and corresponding points. In these experiments, the empirical orders of convergence are close to the asymptotic results from Corollary 1. For $\alpha = 0.5$ and $\alpha = 0.8$, the empirical orders are 0.47 and 0.46, where in both cases the asymptotic result is $1/2 - \varepsilon$. For stability index $\alpha = 1.2$, the orders are 0.38 empirically and $1/3 - \varepsilon$ according to Corollary 1.

Let us now compare the multilevel Monte Carlo algorithm with the classical Monte Carlo scheme. In the latter case, we do not have a heuristic control of the bias and so we use the strong approximation error of Theorem 2 as an upper bound. All unknown constants appearing in the error estimate as well as the unknown variance are assumed to be 1. With this considerations, we determine the parameters h, ε and the replication number n in order to reach precision $\delta > 0$ for the root mean

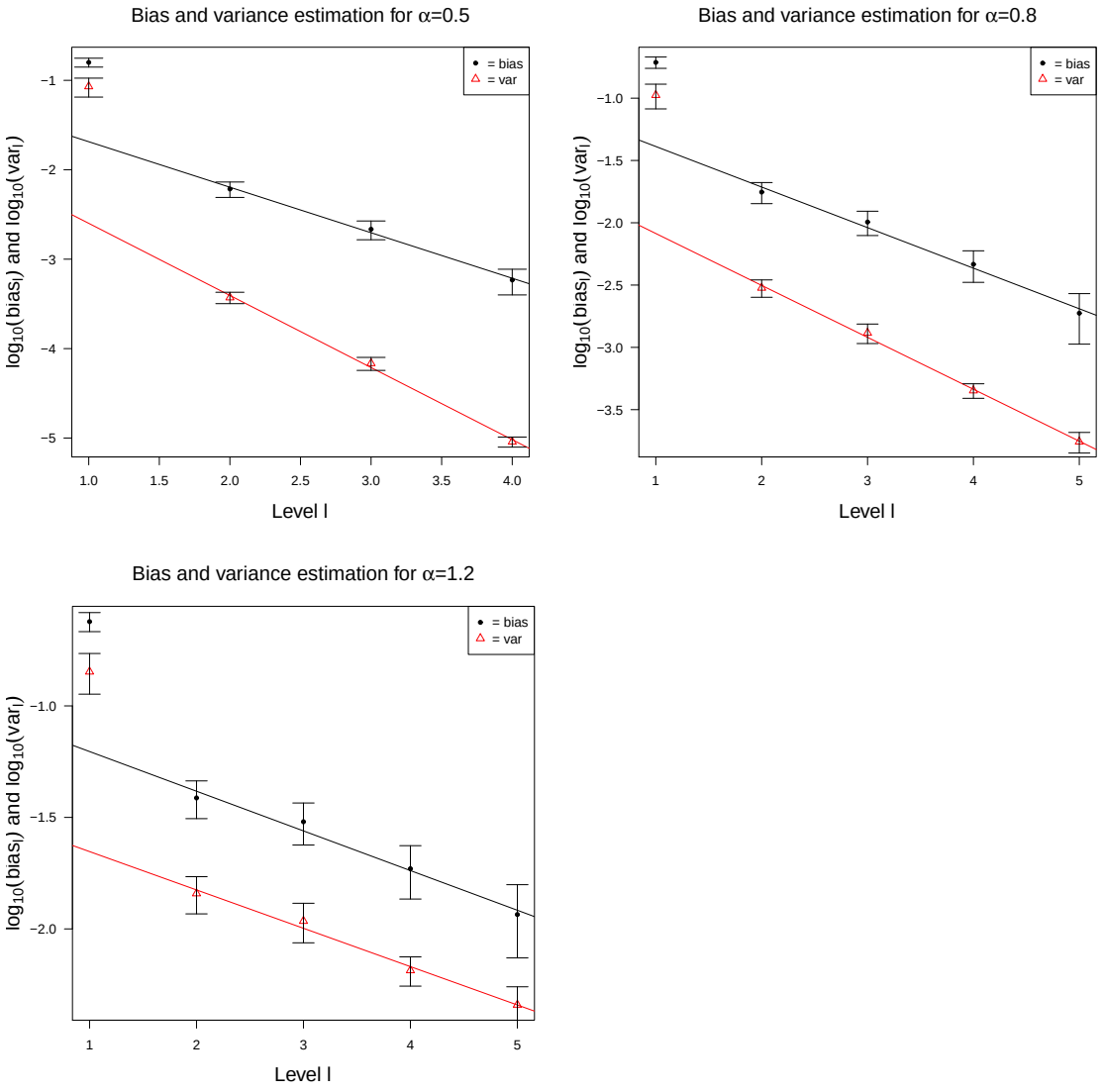

Figure 5.1: Estimates of bias_k and var_k with corresponding regression lines.

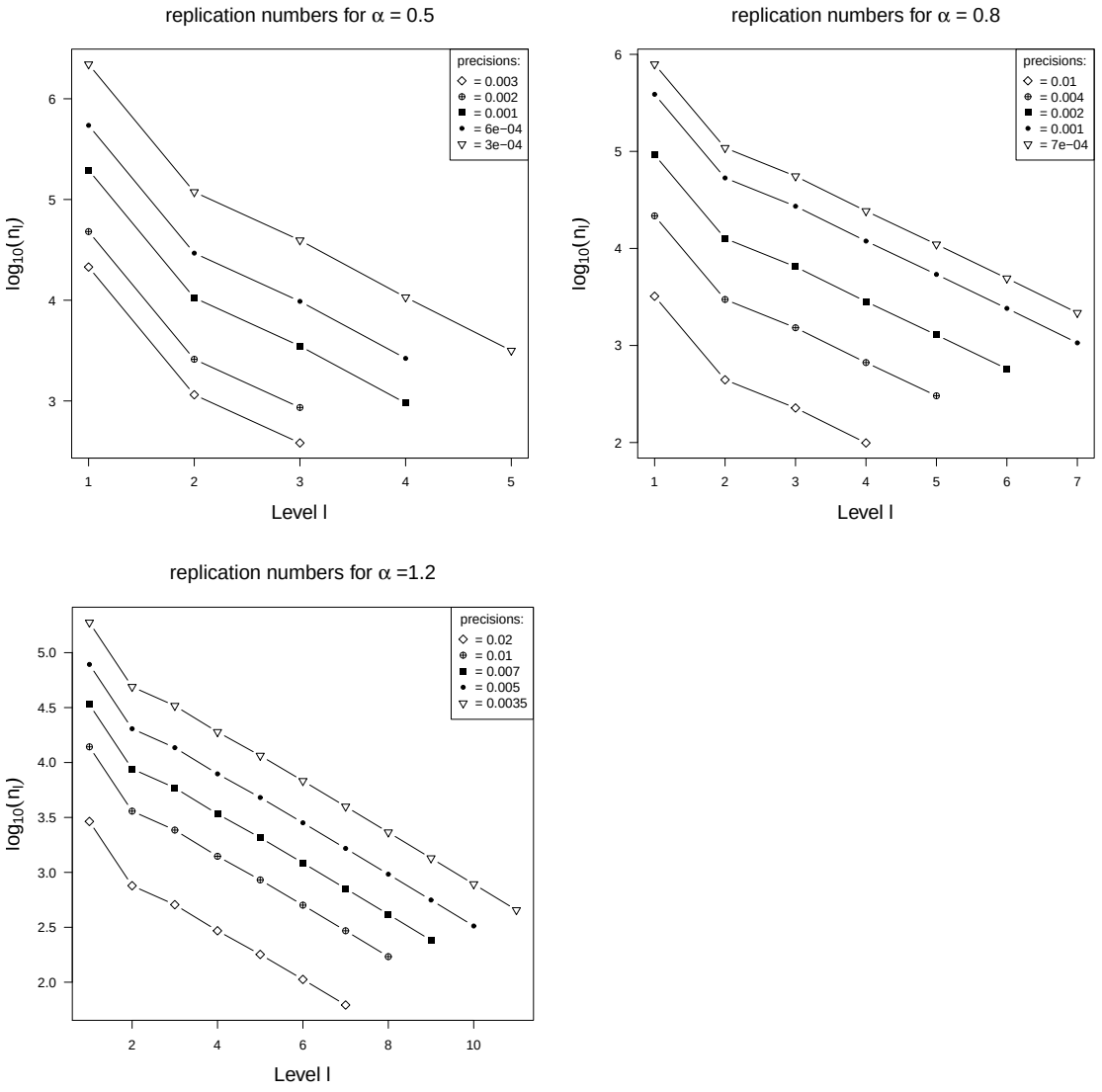

Figure 5.2: Replication numbers for different precisions δ.

square error to

$$h = \left(\frac{(2-\alpha)}{4c}\delta^2\right)^{\frac{1}{2-\alpha}}, \quad \varepsilon = \frac{1}{\nu(B_h^c)}, \quad \text{and } n = \left\lceil \frac{2}{\delta^2} \right\rceil.$$

The empirical orders of convergence obtained from 1000 simulations of the classical Monte Carlo algorithm are given by $0.43, 0.31$ and 0.23 for $\alpha = 0.5, 0.8$ and 1.2, and are given in terms of the dashed regression lines and corresponding points in Figure 5.3. Here you can also see, that the convergence orders as well as the absolute errors are always worse than those of the corresponding multilevel algorithms. During the bias and variance estimation step, we obtain the following empirical orders of convergence of the bias, see (4.3), of our approximation schemes. For $\alpha = 0.5$, the order is approximately 1.7, while for $\alpha = 0.8$ and 1.2 we obtain orders 1.1 and 0.6. These weak orders of convergence partially explain the root mean square convergence orders received by our experiments via result (4.4). Then, a well-balanced Monte Carlo experiment could already receive orders of $0.39, 0.34$ and 0.27 for the worst case setting, respectively. The higher the Blumenthal-Getoor index is, i.e. the harder the problem is itself, the more we benefit from the multilevel idea, since higher levels are needed to reach a desired precision, whereby the latter of course induce higher computational cost.

5.3 Geometric Asian option in the Barndorff-Nielsen Shephard (BNS) model

This choice of financial model and option is considered in [35] and a semi-explicit pricing formula for it has been developed in [22]. The option of interest is a continuously monitored average price geometric Asian option, i.e., we are interested in

$$f(X) = \left(S_0 \exp\left(\frac{1}{T}\int_0^T X_t \,\mathrm{d}t\right) - K\right)_+,$$

for a fixed strike $K > 0$ and a fixed endpoint T.

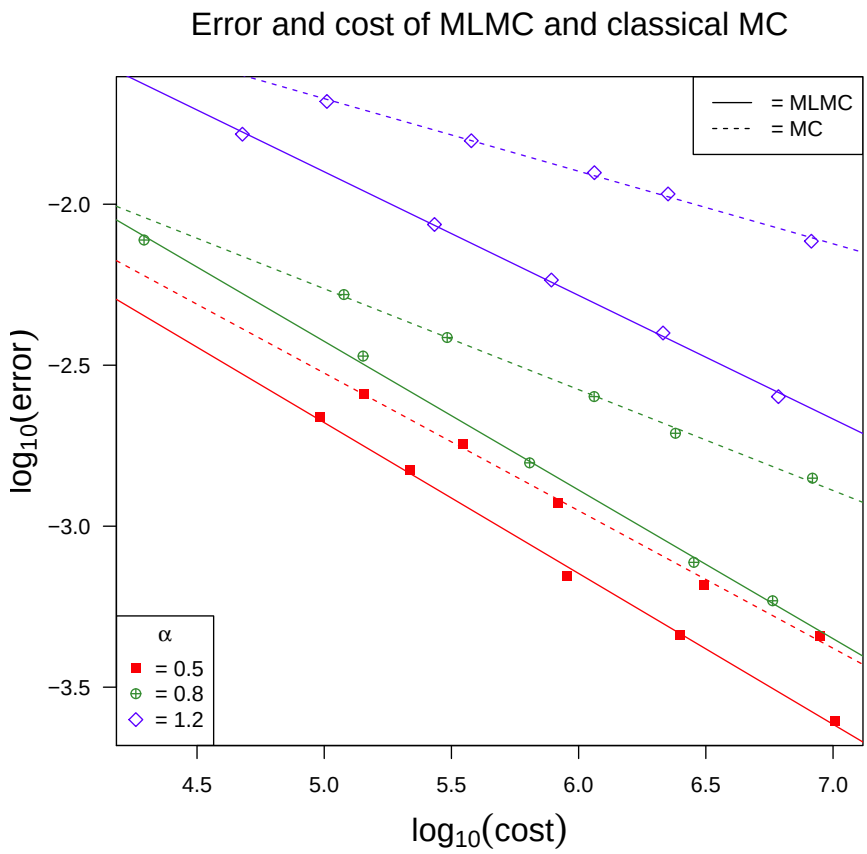

Figure 5.3: Cost and error of multilevel (MLMC) and classical (MC) Monte Carlo.

Definition of the BNS model

Suppose to be given a subordinator $(Z_t)_{t\geq 0}$ called the background driving Lévy process (BDLP) and an independent standard Brownian motion $(W_t)_{t\geq 0}$.

The financial market is modeled by a stock $(S_t)_{t\geq 0}$ and a bond $(B_t)_{t\geq 0}$ satisfying

$$S_t = S_0 \exp(X_t) \qquad \text{and} \qquad B_t = \exp(rt),$$

where $r \geq 0$ and $S_0 > 0$. The logarithmic stock price X is modeled by the SDE

$$\begin{aligned} \mathrm{d}X_t &= (\mu + \beta V_{t-})\mathrm{d}t + \sqrt{V_{t-}}\mathrm{d}W_t + \varrho \mathrm{d}Z_{\lambda t}, \\ X_0 &= 0, \end{aligned} \tag{5.6}$$

where $\mu, \beta \in \mathbb{R}$ and $\varrho \leq 0$. Hereby, the volatility term is given by a stochastic process V which satisfies the Langevin equation

$$\begin{aligned} \mathrm{d}V_t &= -\lambda V_{t-}\mathrm{d}t + \mathrm{d}Z_{\lambda t}, \\ V_0 &= v_0, \end{aligned} \tag{5.7}$$

with deterministic reals $v_0, \lambda > 0$.

Properties

- The parameters of the model are $r, S_0, \mu, \beta, \varrho, \lambda, v_0$ and the BDLP Z, as there is a direct correspondence between Z and the distribution of V.

- The volatility process, i.e. the solution V of the Langevin equation, can be explicitly calculated and is the Ornstein-Uhlenbeck (OU) process

  $$V_t = \exp(-\lambda t)v_0 + \int_0^t \exp(-\lambda(t-s))\mathrm{d}Z_{\lambda s}.$$

  As Z is a subordinator, the process V is strictly positive, more precisely, for any $t \geq 0$, it holds $V_t \geq \exp(-\lambda t)v_0$. With this fact, we can change the square root in (5.6) to a globally Lipschitz function which leads to the same solution, but fits perfectly into our assumptions on the SDE.

- If we change to an equivalent martingale measure the parameters in the dynamics in (5.6) change to

$$\mathrm{d}X_t = \left(r - \lambda\kappa(\varrho) - \frac{1}{2}V_{t-}\right)\mathrm{d}t + \sqrt{V_{t-}}\,\mathrm{d}W_t + \varrho\,\mathrm{d}Z_{\lambda t}, \tag{5.8}$$

  where $\kappa(u) = \log(\mathrm{E}[\exp(uZ_1)])$ is the cumulant generating function of Z.

Remark 6. During the simulation, the approximating Euler Scheme for the volatility can become negative, which yields numerical problems for the square root of the latter. As explained above, the process V up to time T is always greater or equal to the deterministic part $C = \exp(-\lambda T)v_0$. Thus, we can substitute the square root in (5.6) by a function that is continuous and equal to the square root for values greater or equal to C. A candidate for this is e.g. the function ϕ, which is defined constant for values smaller equal to C, namely

$$\phi(x) = \begin{cases} \sqrt{x}, & x \geq C, \\ \sqrt{C}, & \text{otherwise.} \end{cases}$$

Choice of the BDLP

There are two notions of OU processes defined by Barndorff-Nielsen and Shephard in [5]. The first one is defined in terms of the infinitely divisible distribution D characterizing the BDLP, which means $Z_1 \overset{d}{=} D$. In this case we call V an OU-D process.

We will use the following second notion. Therefore, it is used that the process V is strictly stationary, i.e. there exists a distribution D, called the stationary law, such that, if the starting point V_0 has distribution D, then the law of all marginals V_t for $t \geq 0$ is also D. The process V with stationary distribution D is then called a D-OU process. It turns out, that the only candidates for D are self-decomposable distributions. The latter form a subclass of infinitely divisible distributions that furthermore fulfill

$$\phi(u) = \phi(cu) \cdot \phi_c(u)$$

for all $u \in \mathbb{R}$ and all $c \in (0, 1)$, where ϕ is the characteristic function of D and $\{\phi_c : c \in (0, 1)\}$ is some family of characteristic functions.

The Inverse Gaussian (IG)-OU process

For the definition of an inverse Gaussian Lévy process we refer to Section 2.6 and just remember that the Lévy measure of an IG(a, b)-process is given by

$$\nu_{IG}(\mathrm{d}x) = \frac{a}{\sqrt{2\pi}x^{3/2}} \cdot \exp\left(-\frac{1}{2}b^2x\right) \cdot \mathbb{1}_{x>0} \cdot \mathrm{d}x.$$

As the IG distribution is self-decomposable, it can be chosen as the stationary distribution of an IG-OU process. The corresponding BDLP Z than has Lévy measure

$$\nu_Z(\mathrm{d}x) = \frac{a}{2\sqrt{2\pi}} \cdot \left(\frac{1}{x^{\frac{3}{2}}} + \frac{b^2}{x^{\frac{1}{2}}}\right) \cdot \exp\left(-\frac{1}{2}b^2x\right) \cdot \mathbb{1}_{x>0} \cdot \mathrm{d}x,$$

see [42], and is thus the sum of two independent Lévy processes, namely

$$Z_t = Z_t^{(1)} + Z_t^{(2)}, \quad t \geq 0,$$

where $(Z_t^{(1)})_{t\geq 0}$ is an $IG(a/2, b)$ process and $(Z_t^{(2)})_{t\geq 0}$ has Lévy measure

$$\nu_{Z^{(2)}}(\mathrm{d}x) = \frac{ab^2}{2\sqrt{2\pi}} \cdot x^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}b^2x\right) \cdot \mathbb{1}_{x>0} \cdot \mathrm{d}x.$$

The latter is a compound Poisson process with distribution given by

$$Z_t^{(2)} \stackrel{d}{=} \frac{1}{b^2} \cdot \sum_{n=1}^{N_t} V_n^2,$$

where $(N_t)_{t\geq 0}$ is a Poisson process with parameter $ab/2$, independent of the $i.i.d.$ sequence $(V_n)_{n\geq 1}$ of standard normal random variables, see Lemma 2 in the Appendix. The cumulant generating function $\kappa(\varrho)$ of Z, needed in the drift coefficient of the SDE (5.8) is given by $\kappa(\varrho) = a\varrho/\sqrt{b^2 - 2\varrho}$.

Numerical results

The parameters of the SDE (5.8) are chosen according to the fitted parameters of [35] by $\varrho = -4.7039$, $\lambda = 2.4958$, $b = 11.98$, $a = 0.0872$, $v_0 = 0.064262^2$ and $r = 0.0319$. The remaining parameters are not explicitly stated such that we cannot

use their solutions as references and we choose $T = 1$ and $K = S_0 = 80$. The threshold parameters are chosen approximately such that $\nu_Z(B^k_{h_k}) = 2^k$ with the help of a quadrature rule from the GNU Scientific library, namely gsl_integration_qag. This function applies a 41 point Gauss-Kronrod rule, which is a Gauss quadrature rule, adaptively bisecting the largest error contributing interval, such that the error should be at most 10^{-7} after stopping. We then choose $\varepsilon_k = 1/\nu_Z(B^c_{h_k})$. The bias and variance estimation step is performed as before on the first five levels with 1000 simulations each. The results are shown in Figure 5.4, where the decay parameter are given by $\kappa_{\text{bias}} = 1.231289$, $\varrho_{\text{bias}} = 0.4658864$, $\kappa_{\text{var}} = 9.155878$ and $\varrho_{\text{var}} = 0.3499452$. Resulting from this, we obtain the empirical order of convergence 1.1 for the bias

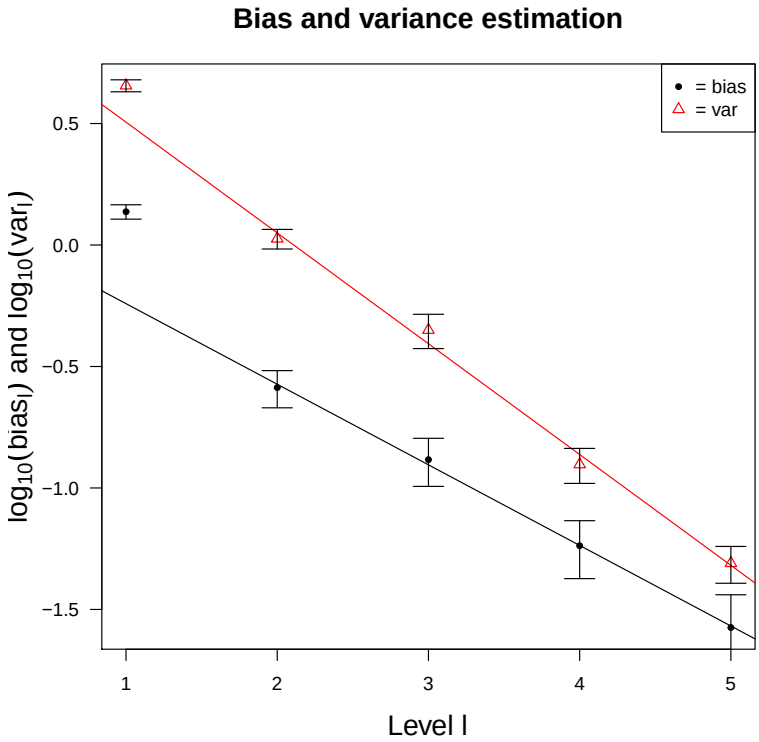

Figure 5.4: Estimates of bias_k and var_k with corresponding regression lines.

of our approximation, which yields approximately the convergence order 0.34 for the well-balanced classical Monte Carlo algorithm, while due to the Blumenthal-Getoor index of the driving process being $\beta = 1/2$, we expect a convergence order of approximately $1/2$ for our multilevel algorithm. We again perform 1000 simulations

of both the multilevel and the classical Monte Carlo scheme. As reference solution, we perform a multilevel algorithm with accuracy $\delta = 0.002$ to receive as master computation the value 1.71108 and estimate the root mean square error of the algorithms by the arithmetical means of the squared errors with respect to this master computation. The desired precisions and corresponding replication numbers for the multilevel scheme are shown in Figure 5.5. In the Monte Carlo setting, we choose the highest level with the help of the bias estimation step from the multilevel algorithm, although the classical algorithm has no direct bias information available. The unknown variance is supposed to be 1 and thus the number of replications are chosen by $n = \lceil 2/\delta^2 \rceil$. The result of the simulation experiment is shown in

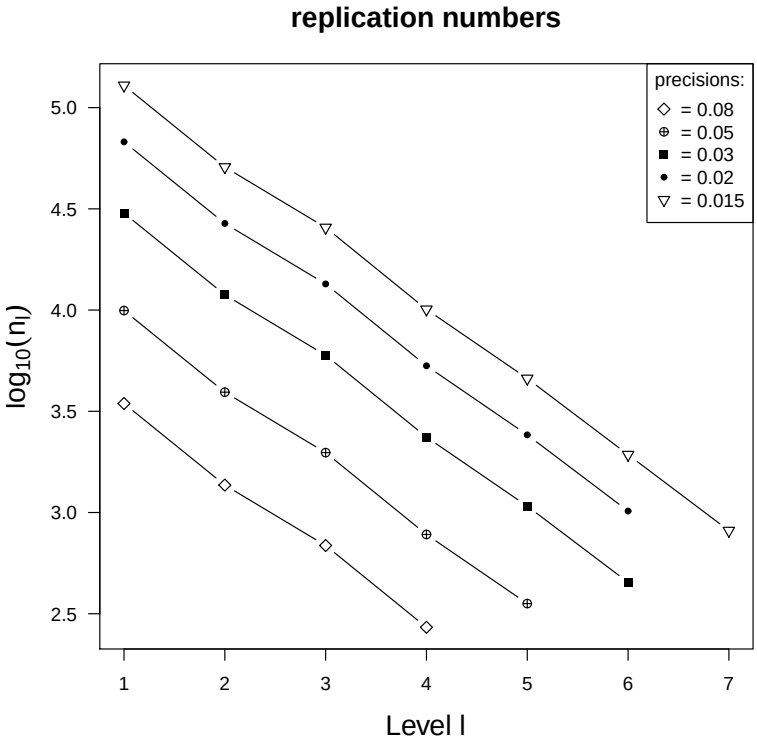

Figure 5.5: Replication numbers for different precisions δ.

a log-log plot of the estimated root mean square error against the computational cost with corresponding regression lines in Figure 5.6. As one can see, the multilevel algorithm is always superior to the Monte Carlo algorithm, although we used the bias information from the multilevel scheme for the latter. This has been done because our

approximation result is not necessarily true for this problem. Due to X only modeling the log-price, the functional is not Lipschitz continuous in general. Nonetheless, the slopes of the regression lines replicate the theoretical orders of convergence perfectly with empirical order 0.49 for the multilevel Monte Carlo algorithm and 0.34 for the standard Monte Carlo scheme.

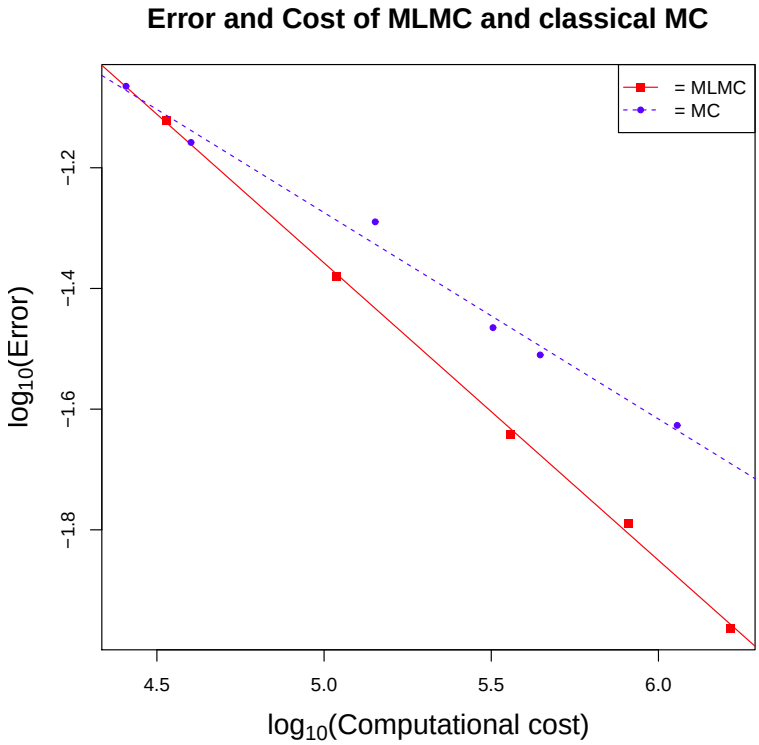

Figure 5.6: Cost and error of multilevel (MLMC) and classical (MC) Monte Carlo.

6 Appendix

Theorem 8 (Gronwall's inequality)**.** *Let* $z : [0, \infty) \to [0, \infty)$ *be integrable such that*

$$z(s) \leq \alpha_2 + \alpha_1 \cdot \int_0^s z(u)\,\mathrm{d}u < \infty$$

for $0 \leq s \leq t$ *and reals* $\alpha_1, \alpha_2 > 0$*. Then*

$$z(t) \leq \alpha_2 \cdot \exp(\alpha_1 \cdot t).$$

Proof: Define for $0 \leq s \leq t$

$$g(s) = \exp(-\alpha_1 \cdot s) \cdot \int_0^s z(u)\mathrm{d}u.$$

Differentiating yields

$$g'(s) = \exp(-\alpha_1 \cdot s) \left(z(s) - \alpha_1 \cdot \int_0^s z(u)\,\mathrm{d}u \right) \leq \exp(-\alpha_1 \cdot s)\alpha_2,$$

and thus

$$g(t) = \int_0^t g'(s)\,\mathrm{d}s \leq \int_0^t \exp(-\alpha_1 \cdot s)\alpha_2\,\mathrm{d}s = \frac{\alpha_2}{\alpha_1} - \frac{\alpha_2}{\alpha_1}\exp(-\alpha_1 \cdot t).$$

Altogether we have

$$z(t) \leq \alpha_2 + \alpha_1 \exp(\alpha_1 \cdot t) g(t) \leq \alpha_2 \cdot \exp(\alpha_1 \cdot t).$$

□

Lemma 2. *A pure jump Lévy process with Lévy measure*

$$\nu(\mathrm{d}x) = \frac{ab^2}{2\sqrt{2\pi}} \cdot x^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}b^2 x\right) \cdot \mathbb{1}_{x>0} \cdot \mathrm{d}x.$$

is of compound Poisson type. It is of the form

$$Z_t = \frac{1}{b^2} \cdot \sum_{n=1}^{N_t} V_n^2,$$

where $(V_n)_{n\geq 1}$ is an i.i.d. sequence of standard normal random variables, independent of the Poisson process $(N_t)_{t\geq 0}$ with parameter $ab/2$.

Proof: First, we show, that the Lévy measure is finite and equal to $ab/2$, which already yields the compound Poisson structure.

$$\begin{aligned}
\nu(\mathbb{R}) &= \int_0^\infty \frac{ab^2}{2\sqrt{2\pi}} \cdot x^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}b^2 x\right) \mathrm{d}x \\
&= \frac{ab^2}{2\sqrt{2\pi}} \int_0^\infty \frac{\exp\left(-\frac{1}{2}b^2 y^2\right)}{y} \cdot 2y \,\mathrm{d}y \\
&= \frac{ab^2}{\sqrt{2\pi}} \int_0^\infty \exp\left(-\frac{1}{2}b^2 y^2\right) \mathrm{d}y = \frac{ab^2}{\sqrt{2\pi}} \cdot \frac{\sqrt{2\pi}}{2b} = ab/2,
\end{aligned}$$

What remains to show is that

$$\nu(\mathbb{R})^{-1} \cdot \nu(\mathrm{d}x) = \frac{b}{\sqrt{2\pi}} \cdot \frac{\exp\left(-\frac{1}{2}b^2 x\right)}{x^{\frac{1}{2}}} \cdot \mathbb{1}_{x>0} \,\mathrm{d}x$$

is the distribution of V^2/b^2 for a standard normal variable $V \sim \mathcal{N}(0,1)$. We show that the distributions coincide on the system of half-open intervals. As this system is stable w.r.t. intersections and generates the Borel σ-field $\mathcal{B}(\mathbb{R})$ the distributions then coincide on the whole $\mathcal{B}(\mathbb{R})$. Therefore consider sets of the form $]c.d]$ with

$0 \leq c \leq d < \infty$ and observe that

$$\begin{aligned} P(\{V^2/b^2 \in]c,d]\}) &= 2 \cdot P(\{V \in]b\sqrt{c}, b\sqrt{d}]\}) \\ &= 2 \cdot \int_{b\sqrt{c}}^{b\sqrt{d}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \mathrm{d}x \\ &= 2 \cdot \int_{\sqrt{c}}^{\sqrt{d}} \frac{b}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}b^2x^2\right) \mathrm{d}x \\ &= \int_c^d \frac{b}{\sqrt{2\pi}} \cdot \frac{\exp\left(-\frac{1}{2}b^2x\right)}{x^{\frac{1}{2}}} \mathrm{d}x. \end{aligned}$$

□

Lemma 3. *Let $r \in \mathbb{N}$ and $(\mathcal{G}_j)_{j=0,1,\ldots,r}$ denote a filtration. Moreover, let, for $j = 0,\ldots,r-1$, 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*

$$\mathrm{E}\big[\max_{j=0,\ldots,r-1} U_jV_j\big] \leq \mathrm{E}\big[\max_{j=0,\ldots,r-1} U_j\big] \cdot \mathrm{E}\big[\max_{j=0,\ldots,r-1} V_j\big].$$

Proof: See [13]. □

Bibliography

[1] David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge Stud. Adv. Math. Cambridge University Press, second edition, 2009.

[2] S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.*, 38(2):482–493, 2001.

[3] F. Aurzada and S. Dereich. The coding complexity of Lévy processes. *Found. Comput. Math.*, 9:359–390, 2009.

[4] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance Stoch.*, 13(3):381–401, 2009.

[5] O.E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Statist. Soc. B*, 63:167–241, 2001.

[6] A. Barth, A. Lang, and C. Schwab. Multi-level Monte Carlo finite element method for parabolic stochastic partial differential equations. Research Report No. 2011-30, ETH Zürich, 2011.

[7] J. Bertoin. *Lévy processes*. Cambridge University Press, 1998.

[8] R.M. Blumenthal and R.K. Getoor. Sample functions of stochastic processes with stationary independent increments. *J. Math. Mech.*, 10:493–516, 1961.

[9] S. Burgos and M.B. Giles. Computing Greeks using multilevel path simulation. In L. Plaskota and H. Wozniakowski, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2010*, volume 23 of *Springer Proceedings in Mathematics & Statistics*, pages 281–296. Springer Berlin Heidelberg, 2012.

[10] R. Cont and P. Tankov. *Financial modelling with jump processes*. Financ. Math. Ser. Chapman & Hall/CRC, London, 2004.

[11] J. Creutzig, S. Dereich, Th. Müller-Gronbach, and K. Ritter. Infinite-dimensional quadrature and approximation of distributions. *Found. Comput. Math.*, 9:391–429, 2009.

[12] S. Dereich. Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Probab.*, 21(1):283–311, 2011.

[13] S. Dereich and F. Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Process. Appl.*, 121(7):1565–1587, 2011.

[14] M. Fischer and G. Nappo. On the moments of the modulus of continuity of Itô processes. *Stoch. Anal. Appl.*, 28:103–122, 2010.

[15] N. Fournier. Simulation and approximation of Lévy-driven stochastic differential equations. *ESAIM Probab. Stat.*, 15:233–248, 2011.

[16] M.B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[17] M.B. Giles and B.J. Waterhouse. Multilevel quasi-Monte Carlo path simulation. In H. Albrecher, W.J. Runggaldier, and W. Schachermayer, editors, *Advanced Financial Modelling*, Radon Ser. Comput. Appl. Math., pages 165–181. de Gruyter, 2009.

[18] S. Heinrich. Monte Carlo complexity of global solution of integral equations. *J. Complexity*, 14:151–175, 1998.

[19] S. Heinrich. Multilevel Monte Carlo methods. In P. Yalamov S. Argenov, J. Wasniewski, editor, *Large Scale Scientific Computing*, pages 58–67. Lecture Notes in Comput. Sci. 2179, Springer, Berlin, 2001.

[20] F.J. Hickernell, Th. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo algorithms for infinite-dimensional integration on $\mathbb{R}^{\mathbb{N}}$. *J. Complexity*, 26(3):229–254, 2010.

[21] D.J. Higham, X. Mao, M. Roj, Q. Song, and G. Yin. Mean exit times and the multi-level Monte Carlo method. Technical Report 5, Department of Mathematics and Statistics, University of Strathclyde, 2011.

[22] F. Hubalek and C. Sgarra. On the explicit evaluation of the geometric Asian options in stochastic volatility models with jumps. *J. Comput. Appl. Math.*, 235(11):3355–3365, 2011.

[23] J. Jacod, Th. G. Kurtz, S. Méléard, and Ph. Protter. The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 41(3):523–558, 2005.

[24] J. Jacod and P. Protter. *Probability Essentials*. Springer, Berlin, second edition, 2003.

[25] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Grad. Texts in Math.* Springer, Berlin, second edition, 1991.

[26] A. Kebaier. Statistical Romberg extrapolation: A new variance reduction method and applications to option pricing. *Ann. Appl. Probab.*, 15(4):2681–2705, 2005.

[27] A. Kohatsu-Higa and P. Protter. The Euler scheme for SDE's driven by semimartingales. In H. Kunita and H. Kuo, editors, *Stochastic analysis on infinite dimensional spaces*, volume 310 of *Pitman Research Notes in Mathematics Series*, pages 141–151. Longman Scientific & Technical, Harlow, 1994.

[28] S.G. Kou. A jump-diffusion model for option pricing. *Manage. Sci.*, 48(8):1086–1101, 2002.

[29] Andreas E. Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Universitext. Springer, Berlin, 2006.

[30] C.B. Madan and E. Senata. The Variance Gamma (v.g.) model for share market returns. *J. Bus.*, 63(4):511–524, 1990.

[31] H. Marxen. The Multilevel Monte Carlo method used on a Lévy driven SDE. *Monte Carlo Methods Appl.*, 16(2):167–190, 2010.

[32] R.C. Merton. Option pricing when underlying stock returns are discontinuous. *J. Financ. Econ.*, 3:125–144, 1976.

[33] Th. Müller-Gronbach, E. Novak, and K. Ritter. *Monte Carlo-Algorithmen*. Springer, Berlin, 2012.

[34] Th. Müller-Gronbach and K. Ritter. Variable subspace sampling and multi-level algorithms. In P. L' Ecuyer and A.B. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pages 131–156. Springer Berlin Heidelberg, 2009.

[35] E. Nicolato and E. Venardos. Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Math. Finance*, 13(4):445–466, 2003.

[36] Ph. Protter and D. Talay. The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, 25(1):393–423, 1997.

[37] Philip E. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Applications of Mathematics (New York), Stochastic Modelling and Applied Probability*. Springer, Berlin, 2004.

[38] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, third edition, 1999.

[39] S. Rubenthaler. Numerical simulation of the solution of a stochastic differential equation driven by a Lévy process. *Stochastic Process. Appl.*, 103(2):311–349, 2003.

[40] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian random processes: Stochastic models with infinite variance*. Chapman & Hall/CRC, New York, 1994.

[41] Ken-iti Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Stud. Adv. Math.* Cambridge University Press, Cambridge, 1999.

[42] W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives*. John Wiley & Sons, Ltd., 2003.

[43] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.*, 8(4):483–509, 1990.

[44] H. Tanaka and A. Kohatsu-Higa. An operator approach for Markov chain weak approximations with an application to infinite activity Lévy driven SDEs. *Ann. Appl. Probab.*, 19(3):1026–1062, 2009.

[45] P. Tankov. High order weak approximation schemes for Lévy-driven SDEs. In L. Plaskota and H. Wozniakowski, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2010*, volume 23 of *Springer Proceedings in Mathematics & Statistics*, pages 667–683. Springer Berlin Heidelberg, 2012.

[46] P. Tankov and A. Kohatsu-Higa. Jump-adapted discretization schemes for Lévy-driven SDEs. *Stochastic Process. Appl.*, 120:2258–2285, 2010.

[47] D. Werner. *Funktionalanalysis*. Springer, Berlin, 6th edition, 2007.

[48] Y. Xia and M.B. Giles. Multilevel path simulation for Jump-diffusion SDEs. In L. Plaskota and H. Wozniakowski, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2010*, volume 23 of *Springer Proceedings in Mathematics & Statistics*, pages 695–708. Springer Berlin Heidelberg, 2012.

Scientific career

June 2001	Abitur at the Max-Planck-Gymnasium Rüsselsheim
October 2002	Beginning to study Mathematics at the TU Darmstadt
August 2004	Vordiplom in Mathematics at the TU Darmstadt
March 2008	Diploma in Mathematics at the TU Darmstadt. Diploma Thesis: "The Coding Complexity of Lévy Processes"
April 2008 - April 2010	Scientific assistant at the TU Darmstadt in the Stochastic research group, partially supported by the DFG within the priority program SPP 1324
May 2010 - September 2012	Scientific assistant at the TU Kaiserslautern in the research group Computational Stochastics, partially supported by the DFG within the priority program SPP 1324

Wissenschaftlicher Werdegang

Juni 2001	Abitur am Max-Planck-Gymnasium Rüsselsheim
Oktober 2002	Beginn des Studiums der Mathematik mit Nebenfach Informatik an der Technischen Universität Darmstadt
August 2004	Vordiplom in Mathematik an der TU Darmstadt
März 2008	Diplom in Mathematik an der TU Darmstadt. Diplomarbeit: "Die Kodierungskomplexität von Lévyprozessen"
April 2008 - April 2010	Wissenschaftlicher Mitarbeiter an der TU Darmstadt in der AG Stochastik, unterstützt durch die DFG im Schwerpunktprogramm SPP 1324
Mai 2010 - September 2012	Wissenschaftlicher Mitarbeiter an der TU Kaiserslautern in der AG Computational Stochastics, unterstützt durch die DFG im Schwerpunktprogramm SPP 1324