A multilevel Monte Carlo algorithm for Lévy driven stochastic differential equations

by

Steffen Dereich[1] and Felix Heidenreich[2]

[1] *Fachbereich Mathematik und Informatik*
Philipps–Universität Marburg
Hans–Meerwein Straße
35032 Marburg
Germany

[2] *Fachbereich Mathematik*
Technische Universität Kaiserslautern
Erwin-Schrödinger-Straße
67663 Kaiserslautern
Germany

Version of 11 December 2010

Summary. This article introduces and analyzes multilevel Monte Carlo schemes for the evaluation of the expectation $\mathbb{E}[f(Y)]$, where $Y = (Y_t)_{t\in[0,1]}$ is a solution of a stochastic differential equation driven by a Lévy process. Upper bounds are provided for the worst case error over the class of all path dependent measurable functions f that are Lipschitz continuous with respect to supremum norm. In the case where the Blumenthal-Getoor index of the driving process is smaller than one, one obtains convergence rates of order $1/\sqrt{n}$, when the computational cost n tends to infinity. This rate is optimal up to logarithms in the case where Y is itself a Lévy process. Furthermore, an error estimate for Blumenthal-Getoor indices larger than one is included together with results of numerical experiments.

Keywords. Multilevel Monte Carlo, numerical integration, quadrature, Lévy-driven stochastic differential equation.

2000 Mathematics Subject Classification. Primary 60G51, Secondary 60H10, 60J75

1 Introduction

In this article, we analyze numerical schemes for the evaluation of

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

where $Y = (Y_t)_{t\in[0,1]}$ is a solution to a multivariate stochastic differential equation driven by a multidimensional Lévy process, and $f : D[0,1] \to \mathbb{R}$ is a Borel measurable mapping from the Skorokhod space $D[0,1]$ of $\mathbb{R}^{d_Y}$-valued functions over the time interval $[0,1]$ that is Lipschitz continuous with respect to *supremum* norm.

This is a classical problem which appears for instance in finance, where Y models the risk neutral stock price and f denotes the payoff of a (possibly path dependent) option, and in the past several concepts have been employed for dealing with it. We refer in particular to [PT97], [Rub03], and [JKMP05] for an analysis of the Euler scheme for Lévy-driven SDEs.

Recently, Giles [Gil08b] introduced the so called *multilevel Monte Carlo method* in the context of stochastic differential equations, and this turned out to be very efficient when Y is a continuous diffusion. Indeed, it can be shown that it is optimal on the Lipschitz class [CDMGR08], see also [Gil08a, Avi09] for further recent results and [MGR09] for a survey and further references.

In this article, we analyze multilevel Monte Carlo algorithms for the computation of $S(f)$ with a focus on *path dependent* f's that are Lipschitz functions on the space $D[0,1]$ with supremum norm. In order to gain approximative solutions we decompose the Lévy process in a purely discontinuous Lévy martingale with discontinuities of size smaller than a threshold parameter and a Lévy process with discontinuities larger than the former parameter. The latter process can be efficiently simulated on an arbitrary finite time set and we apply an Euler approximation to get approximative solutions for the stochastic differential equation (see for instance [Rub03] for an analysis of such approximations).

The article is structured as follows. In the next subsection 1.1, the main notation as well as the assumptions for the SDE are stated. Furthermore, basic facts concerning the Lévy-Ito-decomposition of a Lévy process are given as a reminder. The actual algorithm, and in particular the coupled Euler schemes, are described in detail in section 2, while the main result and the right choice of the parameters of the algorithm are postponed to section 3. Both depend on $g(h) \geq \int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x)$ on $(0,\infty)$ and on whether the driving Lévy process has a Brownian component.

Let us explain our main findings in terms of the Blumenthal-Getoor index (BG-index) β of the driving Lévy process which is an index in $[0,2]$. It measures the frequency of small jumps, see (12), where a large index corresponds to a process which has small jumps at high frequencies. If the Blumenthal-Getoor index is smaller than one, appropriately adjusted algorithms achieve the same error bounds as those obtained in Giles [Gil08b] for continuous diffusions i.e., the error is of the order $n^{-1/2}(\log n)^{3/2}$ in terms of the computation time n (in the case where f depends on the whole trajectory and is Lipschitz w.r.t. supremum norm). If the driving Lévy process does not include a Wiener process one even obtains error estimtates of order $n^{-1/2}$. Unfortunately the error rates become significantly worse for larger Blumenthal-Getoor indices. In this case, a remedy would be to incorporate a Gaussian term as compensation for the disregarded discontinuous Lévy martingale (see for instance [AR01]).

Derivations of convergence rates for multilevel schemes are typically based on a *weak* and a *strong* error estimate for the approximative solutions. In this article, the main technical tool is an error estimate in the *strong sense*. We shall use as weak error estimate the one that is

induced by the strong estimate. As is well known this approach is suboptimal when the payoff $f(Y)$ actually does not depend on the whole trajectory of the process Y but on the value of Y at a particular deterministic time instance. In that case, an analysis based on the weak estimates of [JKMP05] or [TKH09] seems to be favorable.

Unfortunately, in the case where f is path dependent, one does not have better error estimates at hand. To gain better results for large BG-indices it is preferable to incorporate a Gaussian correction. In the case where $f(Y)$ depends only on the value of the process at a given time instance, one can combine the strong estimate of this article with weak estimates of [JKMP05] to improve the error estimates. However, to do so one has to impose significantly stronger assumptions on f and the Lévy process. The case where f is path dependent is analyzed in a forthcoming article [Der10]. Here the strong estimate of this article is combined with a new weak estimate for the approximative solutions with Gaussian correction.

1.1 Assumptions and Basic Facts

Let us now introduce the main notation. We denote by $|\cdot|$ the Euclidean norm for vectors as well as the Frobenius norm for matrices. For $h > 0$, we put $B_h = \{x \in \mathbb{R}^{d_X} : |x| < h\}$. Moreover, we let $D[0,1]$ denote the space of càdlàg functions from $[0,1]$ to $\mathbb{R}^{d_Y}$, where $\mathbb{R}^{d_Y}$ is the state space of Y to be defined below. The space $D[0,1]$ is endowed with the σ-field induced by the projections on the marginals (or, equivalently, the Borel-σ-field for the Skorokhod topology). Often, we consider *supremum norm* on $D[0,1]$ and we denote $\|f\| = \sup_{t\in[0,1]} |f(t)|$ for $f \in D[0,1]$. Furthermore, the set of Borel measurable functions $f : D[0,1] \to \mathbb{R}$, that are Lipschitz continuous with Lipschitz constant one with respect to the *supremum norm* is denoted by Lip(1). In general, we write $f \sim g$ iff $\lim f/g = 1$, while $f \lesssim g$ stands for $\limsup f/g \leq 1$. Finally, $f \approx g$ means $0 < \liminf f/g \leq \limsup f/g < \infty$, and $f \precsim g$ means $\limsup f/g < \infty$.

Let $X = (X_t)_{t\geq 0}$ denote a d_X-dimensional L^2-Lévy process with Lévy measure ν, drift parameter b and Gaussian covariance matrix $\Sigma\Sigma^*$, that is a process with independent and stationary increments satisfying

$$\mathbb{E}[e^{i\langle\theta,X_t\rangle}] = e^{t\psi(\theta)}, \qquad \theta \in \mathbb{R}^{d_X}$$

for

$$\psi(\theta) = -\frac{1}{2}|\Sigma^*\theta|^2 + i\langle b,\theta\rangle + \int_{\mathbb{R}^{d_X}} \left(e^{i\langle\theta,x\rangle} - 1 - i\langle\theta,x\rangle\right)\nu(\mathrm{d}x).$$

Briefly, we call X a $(\nu, \Sigma\Sigma^*, b)$-Lévy process. Note that we do not need to work with a truncation function since the marginals of the process are in $L^2(\mathbb{P})$ by assumption. We consider the stochastic integral equation

$$Y_t = y_0 + \int_0^t a(Y_{s-})\,\mathrm{d}X_s \tag{1}$$

with deterministic initial value $y_0 \in \mathbb{R}^{d_Y}$ and we impose a standard Lipschitz assumption on the function a, which implies, in particular, existence and strong uniqueness of the solution. More precisely, we require 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

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

For a general account on Lévy processes we refer the reader to the books by [Ber98] and [Sat99]. Moreover, concerning stochastic analysis, we refer the reader to the books by Protter [Pro05] and Applebaum [App04].

The distribution of the driving Lévy process X is uniquely characterized by the parameters $b \in \mathbb{R}^{d_X}, \Sigma \in \mathbb{R}^{d_X \times d_X}$ and ν. We sketch a construction of X with a view towards simulation in the multilevel setting. This construction is based on the L^2-approximation of Lévy processes as it is presented in [Pro05] and [App04].

Consider a stochastic process $(N(t,A))_{t\ge 0, A\in\mathfrak{B}(\mathbb{R}^{d_X}\setminus\{0\})}$ on some probability space $(\Omega, \mathfrak{A}, P)$ with the following properties. For every $\omega \in \Omega$ the mapping $[0,t] \times A \mapsto N(t,A)(\omega)$ induces a σ-finite counting measure on $\mathfrak{B}(\mathbb{R}_+ \times (\mathbb{R}^{d_X}\setminus\{0\}))$. For every $A \in \mathfrak{B}(\mathbb{R}^{d_X}\setminus\{0\})$ that is bounded away from zero the process $(N(t,A))_{t\ge 0}$ is a Poisson process with intensity $\nu(A)$. For pairwise disjoint sets $A_1, \dots, A_r \in \mathfrak{B}(\mathbb{R}^{d_X}\setminus\{0\})$ the stochastic processes $(N(t,A_1))_{t\ge 0}, \dots, (N(t,A_r))_{t\ge 0}$ are independent.

Then $N(t,.)$ P-a.s. defines a finite measure on B_h^c with values in $\mathbb{N}_0$. The integral $\int_{B_h^c} x\, N(t,\mathrm{d}x)$ thus is a random finite sum, which gives rise to a compound Poisson process. More specifically put $\lambda^{(h)} = \nu(B_h^c) < \infty$, which satisfies $\lambda^{(h)} > 0$ for sufficiently small $h > 0$. In this case $\mu^{(h)} = \nu|_{B_h^c}/\lambda^{(h)}$ defines a probability measure on $\mathbb{R}^{d_X}\setminus\{0\}$ such that

$$\int_{B_h^c} x\, N(t,\mathrm{d}x) \overset{d}{=} \sum_{i=1}^{N_t} \xi_i, \tag{2}$$

where $\overset{d}{=}$ denotes equality in distribution, $(N_t)_{t\ge 0}$ is a Poisson process with intensity $\lambda^{(h)}$ and $(\xi_i)_{i\in\mathbb{N}}$ is an i.i.d. sequence of random variables with distribution $\mu^{(h)}$ and independent of $(N_t)_{t\ge 0}$. Its expectation calculates to $\mathbb{E}\big[\int_{B_h^c} x\, N(t,\mathrm{d}x)\big] = F_0(h)t$, where we set

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

The compensated process $L^{(h)} = (L_t^{(h)})_{t\ge 0}$, given by

$$L_t^{(h)} = \int_{B_h^c} x\, N(t,\mathrm{d}x) - tF_0(h), \tag{3}$$

is an L^2-martingale, and the same holds true for its L^2-limit $L = (L_t)_{t\ge 0} = \lim_{h\downarrow 0} L^{(h)}$, see, e.g., Applebaum [App04].

With W denoting a d_X-dimensional Brownian motion independent of L, we define the Lévy process X by

$$X_t = \Sigma W_t + L_t + bt. \tag{4}$$

We add that the Lévy-Itô-decomposition guarantees that every L^2-Lévy process has a representation (4).

2 The Algorithm

2.1 A Coupled Euler Scheme

The multilevel Monte Carlo algorithm introduced in Section 2.2 is based on a hierarchy of coupled Euler schemes for the approximation of the solution process Y of the SDE (1). In every single Euler scheme we approximate the driving process X in the following way. At first we neglect all the jumps with size smaller than h. The jumps of size at least h induce a random time discretization, which, because of the Brownian component, is refined so that the step sizes are at most ε. Finally W as well as the drift component are approximated by piecewise constant functions with respect to the refined time discretization. In this way, we get an approximation $\hat{X}^{(h,\epsilon)}$ of X. The multilevel approach requires simulation of the joint distribution of $\hat{X}^{(h,\epsilon)}$ and $\hat{X}^{(h',\varepsilon')}$ for different values of $h' > h > 0$ and $\varepsilon' > \varepsilon > 0$. More precisely, we proceed as follows.

For any càdlàg process L we denote by $\Delta L_t = L_t - \lim_{s \nearrow t} L_s$ the jump-discontinuity at time t. The jump times of $L^{(h)}$ are then given by $T_0^{(h)} = 0$ and

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

for $k \geq 1$. The time differences $T_k^{(h)} - T_{k-1}^{(h)}$ form an i.i.d. sequence of random variables that are exponentially distributed with parameter $\lambda^{(h)}$. Furthermore, this sequence is independent of the sequence of jump heights $\Delta L^{(h)}_{T_k^{(h)}}$, which is i.i.d. with distribution $\mu^{(h)}$, and on every interval $[T_k^{(h)}, T_{k+1}^{(h)})$ the process $L^{(h)}$ is linear with slope $-F_0(h)$. See (2) and (3).

The 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, and therefore the jumps of the process $L^{(h')}$ can be obtained from those of $L^{(h)}$ by

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

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 $\mu^{(h)}$ and from an exponential distribution.

Because of the Brownian component we refine the time discretization by $T_0^{(h,\varepsilon)} = 0$ and

$$T_j^{(h,\varepsilon)} = \inf\{T_k^{(h)} > T_{j-1}^{(h,\varepsilon)} : k \in \mathbb{N}\} \wedge (T_{j-1}^{(h,\varepsilon)} + \varepsilon) \tag{5}$$

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 $j \geq 1$. Observe that

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

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.

For the SDE (1) the Euler scheme with the driving process $(\Sigma W_t + L_t^{(h)} + bt)_{t\geq 0}$ and the random time discretization $(T_j)_{j\in\mathbb{N}_0} = (T_j^{(h,\varepsilon)})_{j\in\mathbb{N}_0}$ is defined by $\hat{Y}_0^{(h,\epsilon)} = y_0$ and

$$\hat{Y}_{T_j}^{(h,\epsilon)} = \hat{Y}_{T_{j-1}}^{(h,\epsilon)} + a(\hat{Y}_{T_{j-1}}^{(h,\epsilon)})(\hat{X}_{T_j}^{(h,\epsilon)} - \hat{X}_{T_{j-1}}^{(h,\epsilon)}) \tag{6}$$

for $j \geq 1$. Furthermore $\hat{Y}_t^{(h,\epsilon)} = \hat{Y}_{T_j}^{(h,\epsilon)}$ for $t \in [T_j, T_{j+1})$. In the multilevel approach the solution process Y of the SDE (1) is approximated via coupled Euler schemes $(\hat{Y}^{(h,\epsilon)}, \hat{Y}^{(h',\varepsilon')})$, which are obtained by applying the Euler scheme (6) 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.

2.2 The Multilevel Monte Carlo Algorithm

We fix two positive and decreasing sequences $(\varepsilon_k)_{k\in\mathbb{N}}$ and $(h_k)_{k\in\mathbb{N}}$, and we put $Y_t^{(k)} = \hat{Y}_t^{(h_k,\epsilon_k)}$. For technical reasons we define $Y_t^{(k)}$ for all $t \geq 0$, although we are typically only interested in $Y^{(k)} := (Y_t^{(k)})_{t\in[0,1]}$. For $m \in \mathbb{N}$ and a given measurable function $f : D[0,1] \to \mathbb{R}$ with $f(Y^{(k)})$ being integrable for $k = 1, \ldots, m$, we write $\mathbb{E}[f(Y^{(m)})]$ as telescoping sum

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

In the multilevel approach each expectation on the right-hand side is approximated separately by means of independent classical Monte Carlo approximations. For $k = 2, \ldots, m$ we denote by n_k the number of replications for the approximation of $\mathbb{E}[f(Y^{(k)}) - f(Y^{(k-1)})]$ and by n_1 the number of replications for the approximation of $\mathbb{E}[f(Y^{(1)})]$. For $(Z_{j,1}^{(k)}, Z_{j,2}^{(k)})_{j=1,\ldots,n_k}$ being i.i.d. copies of the coupled Euler scheme $(Y^{(k-1)}, Y^{(k)})$ for $k = 2, \ldots, m$ and $(Z_j^{(1)})_{j=1,\ldots,n_1}$ being i.i.d. copies of $Y^{(1)}$, the corresponding multilevel Monte Carlo algorithm is given by

$$\widehat{S}(f) = \frac{1}{n_1}\sum_{j=1}^{n_1} f(Z_j^{(1)}) + \sum_{k=2}^{m} \frac{1}{n_k}\sum_{j=1}^{n_k} \big[f(Z_{j,2}^{(k)}) - f(Z_{j,1}^{(k)})\big].$$

The algorithm $\widehat{S}$ is uniquely described by the parameters m and $(n_k, h_k, \varepsilon_k)_{k=1,\ldots,m}$ so that we formally identify the algorithm $\widehat{S}$ with these parameters.

The error of the algorithm

For measurable functions $f : D[0,1] \to \mathbb{R}$ with $f(Y), f(Y^{(1)}), \ldots, f(Y^{(m)})$ being square integrable, the mean squared error of $\widehat{S}(f)$ calculates to

$$\mathbb{E}[|S(f) - \widehat{S}(f)|^2] = |\mathbb{E}[f(Y) - f(Y^{(m)})]|^2 + \mathrm{var}(\widehat{S}(f)).$$

If f is in Lip(1), then

$$\mathbb{E}[|S(f) - \widehat{S}(f)|^2] \leq \mathbb{E}\|Y - Y^{(m)}\|^2 + \sum_{k=2}^{m} \frac{1}{n_k}\mathbb{E}\|Y^{(k)} - Y^{(k-1)}\|^2 + \frac{1}{n_1}\mathbb{E}\|Y^{(1)} - y_0\|^2. \tag{7}$$

In particular, the upper bound does not depend on the choice of f. Hence (7) remains valid for the worst case error

$$e^2(\widehat{S}) = \sup_{f\in\mathrm{Lip}(1)} \mathbb{E}[|S(f) - \widehat{S}(f)|^2]. \tag{8}$$

The cost of the algorithm

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 suppose that

- one can sample from the distribution $\nu|_{B_h^c}/\nu(B_h^c)$ for sufficiently small $h > 0$ and the uniform distribution on $[0,1]$ in constant time,
- one can evaluate a at any point $y \in \mathbb{R}^{d_Y}$ in constant time, and
- f can be evaluated for piecewise constant functions in time less than a constant multiple of its breakpoints plus one.

For a piecewise constant $\mathbb{R}^{d_Y}$-valued function $y = (y_t)_{t\in[0,1]}$, we denote by $\Upsilon(y)$ its number of breakpoints. Then the cost of the algorithm $\widehat{S}$ is defined by the function

$$\operatorname{cost}(\widehat{S}) = \sum_{k=1}^{m} n_k \, \mathbb{E}[\Upsilon(Y^{(k)})]. \tag{9}$$

Under the above assumptions, the average runtime to evaluate $\widehat{S}(f)$ is indeed less than a constant multiple of $\operatorname{cost}(\widehat{S})$.

3 Results and Examples

We consider algorithms $\widehat{S}$ as described in the previous section with cost function defined by (9) and with respect to the worst case error criterion (8). Our results depend on the frequency of small jumps of the driving process X. To quantify this property and to use it for the choice of the parameters of the multilevel algorithm we take a a decreasing and invertible function $g : (0,\infty) \to (0,\infty)$ such that

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

The explicit choices for the parameters m and $(n_k, \varepsilon_k, h_k)_{k=1,\dots,m}$ in the corresponding algorithms are given after the statement of the main results.

Theorem 1.

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

$$g(h) \precsim \frac{1}{h(\log 1/h)^{1+\gamma}} \tag{10}$$

as $h \downarrow 0$, then there exists a sequence of multilevel algorithms $\widehat{S}_n$ such that $\operatorname{cost}(\widehat{S}_n) \le n$ and

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

(ii) If there exists $\gamma \geq 1/2$ such that

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

as $h \downarrow 0$, then there exists a sequence of multilevel algorithms $\widehat{S}_n$ such that $\mathrm{cost}(\widehat{S}_n) \leq n$ and

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

(iii) If there exists $\gamma > 1$ such that for all sufficiently small $h > 0$

$$g(\frac{\gamma}{2} h) \geq 2g(h), \tag{11}$$

then there exists a sequence of multilevel algorithms $\widehat{S}_n$ such that $\mathrm{cost}(\widehat{S}_n) \leq n$ and

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

In terms of the Blumenthal-Getoor index

$$\beta := \inf\Big\{p > 0 : \int_{B_1} |x|^p \, \nu(\mathrm{d}x) < \infty\Big\} \in [0, 2] \tag{12}$$

we get the following corollary.

Corollary 1. *There exists a sequence of multilevel algorithms $(\widehat{S}_n)$ with $\mathrm{cost}(\widehat{S}_n) \leq n$ such that*

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

Remark 1. Parts (ii) and (iii) of Theorem 1 deal with stochastic differential equations that include a Brownian component in the driving process X, i.e. $\Sigma \neq 0$, see Section 1.1. Essentially these two results cover the case $\Sigma \neq 0$. When $\gamma = 1/2$ in (ii), the asymptotics are the same as in the classical diffusion setting analyzed in [Gil08b]. Similarly, as in our proof one can also treat different terms of lower order instead of log. Certainly, it also makes sense to consider $\gamma < 1/2$, when $\Sigma = 0$. The computations are similar and therefore omitted. Part (iii) covers, in particular, all cases where g is regularly varying at 0 with exponent strictly smaller than -1.

As the following remark shows the results of parts (i) and (ii) of Theorem 1 are close to optimal. In particular, these parts cover the case of a Blumenthal-Getoor index $\beta \leq 1$.

Remark 2. To provide a lower bound we adopt a very broad notion of algorithm that was introduced and analyzed in [CDMGR08]. Roughly it can be described as follows. An algorithm $\widehat{S}$ is allowed to carry out real number computations and evaluations of the functional f for arbitrary paths from $D[0, 1]$. The algorithm is required to end in finite time with a possibly random output $\widehat{S}(f)$. For fixed f, the mean number of f-evaluations is referred to as cost of the algorithm when applied to f. Its maximal value over the class $\mathrm{Lip}(1)$ is considered as cost of the algorithm $\widehat{S}$, that is

$$\mathrm{cost}(\widehat{S}) = \sup_{f \in \mathrm{Lip}(1)} \mathbb{E}[\#\text{ of } f\text{-evaluations used in the computation of } \widehat{S}(f)].$$

For $n \in \mathbb{N}$, we consider the optimal worst case error

$$e_n := \inf_{\widehat{S}} \sup_{f \in \mathrm{Lip}(1)} \mathbb{E}[(S(f) - \widehat{S}(f))^2]^{1/2},$$

where the infimum is taken over all algorithms $\widehat{S}$ with $\mathrm{cost}(\widehat{S}) \leq n$. Note that any reasonable implementation of such an algorithm will have a runtime that dominates the cost assigned in our considerations. Hence, the lower bound provided below prevails in more realistic models.

Suppose that $Y = X$. Furthermore assume that X includes a Wiener process in which case we set $\beta = \frac{3}{2}$ or that $\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x) \gtrsim h^{-\alpha}$ for some $\alpha \in (0,2)$ in which case we set $\beta = 1 + \frac{1}{\alpha}$. Then

$$\limsup_{n \to \infty} \sqrt{n} (\log n)^{\beta} \, e_n > 0.$$

The statement follows by combining the lower bound on so called quantization numbers of Theorem 1.5 of [AD09] with Theorem 3 of [CDMGR08].

The choice for the parameters

The algorithm $\widehat{S}$ is completely determined by the parameters m and $(n_k, \varepsilon_k, h_k)_{k=1,\dots,m}$ and we now give the parameters which achieve the error estimates provided in Theorem 1. Recall that Theorem 1 depends on an invertible and decreasing function $g : (0,\infty) \to (0,\infty)$ satisfying

$$\int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x) \leq g(h) \quad \text{for all } h > 0,$$

and we set, for $k \in \mathbb{N}$,

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

We choose the remaining parameters by

$$m = \inf\{k \in \mathbb{N} : \; C \, h_k < 1\} - 1 \qquad \text{and} \qquad n_k = \lfloor C \, h_{k-1} \rfloor \quad \text{for } k = 1, \dots, m, \tag{13}$$

where

$$\text{(i) } C = n, \text{ (ii) } C = \frac{n}{(\log n)^{\gamma+1}}, \text{ and (iii) } C = 1/g^{-1}(n)$$

in the respective cases. Here we need to assume additionally that g is such that $h^{-2/3} \precsim g(h)$ in case (i) and $h^{-1}\sqrt{\log 1/h} \precsim g(h)$ in case (ii). These assumptions do not result in a loss of generality. The parameters optimize (up to constant multiples) the error estimate induced by equation (7) together with Theorem 2 below.

Numerical results

In this section, we provide numerical examples for the new algorithms and compare our approach with the classical Monte-Carlo approach. We denote by X a *symmetric truncated α-stable Lévy process*, that is a $(\nu, 0, 0)$-Lévy process with

$$\frac{\mathrm{d}\nu}{\mathrm{d}x}(x) = \mathbb{1}_{(0,u]}(x) \frac{c}{|x|^{1+\alpha}} + \mathbb{1}_{[-u,0)}(x) \frac{c}{|x|^{1+\alpha}}, \qquad x \in \mathbb{R} \backslash \{0\},$$

where $\alpha \in (0,2)$, $c > 0$ and $u > 0$ denotes the truncation threshold. Note that

$$\int \frac{|x|^2}{h^2} \wedge 1 \; \nu(\mathrm{d}x) \leq \frac{2c}{2\alpha - \alpha^2} h^{-\alpha} = g(h).$$

so that our theoretical findings (Theorem 1) imply that there exists a sequence of algorithms $(\widehat{S}_n)$ each satisfying the cost constraint $\mathrm{cost}(\widehat{S}_n) \leq n$ such that

$$e(\widehat{S}_n) \precsim \begin{cases} n^{-\frac{1}{2}}, & \alpha < \frac{1}{2} \\ n^{-1/2} (\log n)^{3/2}, & \alpha = 1 \\ n^{-(\frac{1}{\alpha} - \frac{1}{2})}, & \alpha > 1 \end{cases}$$

In the case where $\alpha = 1$, the rates can be improved to $e(\widehat{S}_n) \precsim n^{-\frac{1}{2}} \log n$ along a similar proof since the process X does not include a Gaussian term.

In the numerical study we consider a *lookback option* with strike 1, that is

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

We treat the stochastic differential equation

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

with $y_0 = 1$ and X being a symmetric α-stable Lévy process with parameters $u = 1$, $c = 0.1$, and $\alpha \in \{0.5, 0.8, 1.2\}$.

In the aproach that is to be presented next, we allow the user to specify a desired precision δ measured in the root mean squared error. The highest level in the simulation and the iteration numbers $(n_k)_{k=1,\ldots,m}$ are then derived from simulations of the coarse levels. The remaining parameters are chosen as $\varepsilon_k = 2^{-k}$ and $h_k = \left(\frac{2^k \alpha}{2c} + u^{-\alpha} \right)^{-1/\alpha}$ which is the uniqe value with $\nu(B_{h_k}^c) = 2^k$.

To obtain mean squared error at most δ^2, we want to choose the highest level m and the iteration numbers $n_1, \ldots, n_m$ such that

$$|\mathrm{bias}(\widehat{S}_\delta(f))|^2 := |\mathbb{E}[f(Y) - f(\hat{Y}^{(m)})]|^2 \leq \frac{\delta^2}{2} \tag{14}$$

and

$$\mathrm{var}(\widehat{S}_\delta(f)) \leq \frac{\delta^2}{2}. \tag{15}$$

A direct estimate for $\mathrm{bias}(\widehat{S}_\delta(f))$ is not available, and therefore we proceed as follows. Put $\mathrm{bias}_k(f) := \mathbb{E}[f(\hat{Y}^{(k)}) - f(\hat{Y}^{(k-1)})]$ for $k \geq 2$, to obtain $\mathrm{bias}(\widehat{S}_\delta(f)) = \sum_{k=m+1}^{\infty} \mathrm{bias}_k(f)$, and note that $\mathrm{bias}_k(f)$ can be estimated for small levels k at a small computational cost. This suggests to use extrapolation to large levels k. In particular, if there is an exponential decay

$$|\mathrm{bias}_k(f)| \lesssim \gamma \cdot \rho^k, \tag{16}$$

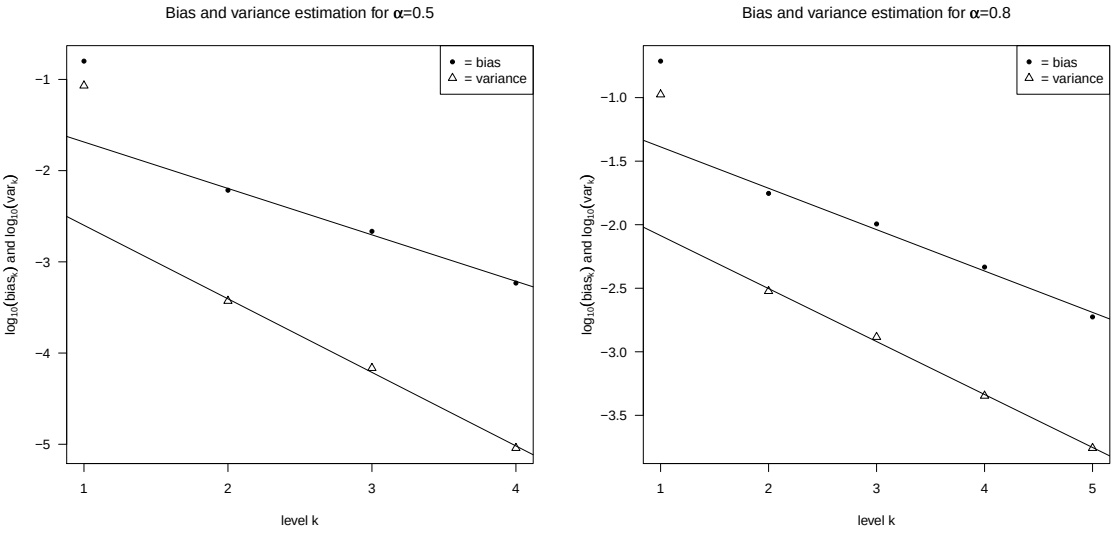

Figure 1: Estimates of $\mathrm{bias}_k(f)$ and $\mathrm{var}_k(f)$ with corresponding regression lines.

with $\gamma > 0$ and $\rho \in]0,1[$, we have $|\mathrm{bias}(\widehat{S}_\delta(f))| \lesssim \gamma/(1-\rho) \cdot \rho^{m+1}$ and thus (14) holds for

$$m = \left\lceil \frac{\log((1-\rho)\delta) - \log(\sqrt{2}\cdot\gamma)}{\log(\rho)} - 1 \right\rceil .$$

To check the validity of assumption (16), we estimate $\mathrm{bias}_k(f)$ for small levels k (here $k \le 4$ or $k \le 5$) by the empirical mean of 1000 simulations. It turns out that the estimates for $\mathrm{bias}_k(f)$ fit very well to the assumption (16), see Figure 1.

For the variance of $\widehat{S}_\delta(f)$ we proceed in the same way. Put $\mathrm{var}_k(f) := \mathrm{var}(f(\hat{Y}^{(k)}) - f(\hat{Y}^{(k-1)}))$ for $k \ge 2$ and $\mathrm{var}_1(f) = \mathrm{var}(f(\hat{Y}^{(1)}))$ to obtain $\mathrm{var}(\widehat{S}_\delta(f)) = \sum_{k=1}^m \mathrm{var}_k(f)/n_k$. Estimates of $\mathrm{var}_k(f)$ for small levels fit very well to an exponential decay, see Figure 1, and we use extrapolation to large levels again. Finally the replication numbers $n_1, \dots, n_m$ are determined by minimizing the cost of $\widehat{S}_\delta(f)$, which is given by $\sum_{k=1}^m n_k 2^k$, up to a constant, subject to the constraint (15). For several values of δ, the corresponding choice of replication numbers is shown in Figure 2.

We now study the relation between $(\mathbb{E}[S(f) - \widehat{S}_\delta(f)]^2)^{1/2}$ and $\mathrm{cost}(\widehat{S}_\delta(f))$ by means of a simulation experiment. Here, different precisions δ and the respective replication numbers and maximal levels are chosen according to Figure 2. We use 1000 Monte Carlo simulations of $\widehat{S}_\delta(f)$ to estimate the root mean-squared error and the cost, where the unknown value of $S(f)$ is replaced by the output of a master computation. The results are presented in a log-log plot in Figure 3. In this experiment, the empirical orders of convergence, i.e. the slopes of the regression lines, are close to the asymptotic results from Theorem 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 $1/2$. For stability index $\alpha = 1.2$, the orders are 0.38 empirically and $1/3$ according to Theorem 1.

Let us compare the multilevel algorithm $\widehat{S}_\delta$ 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

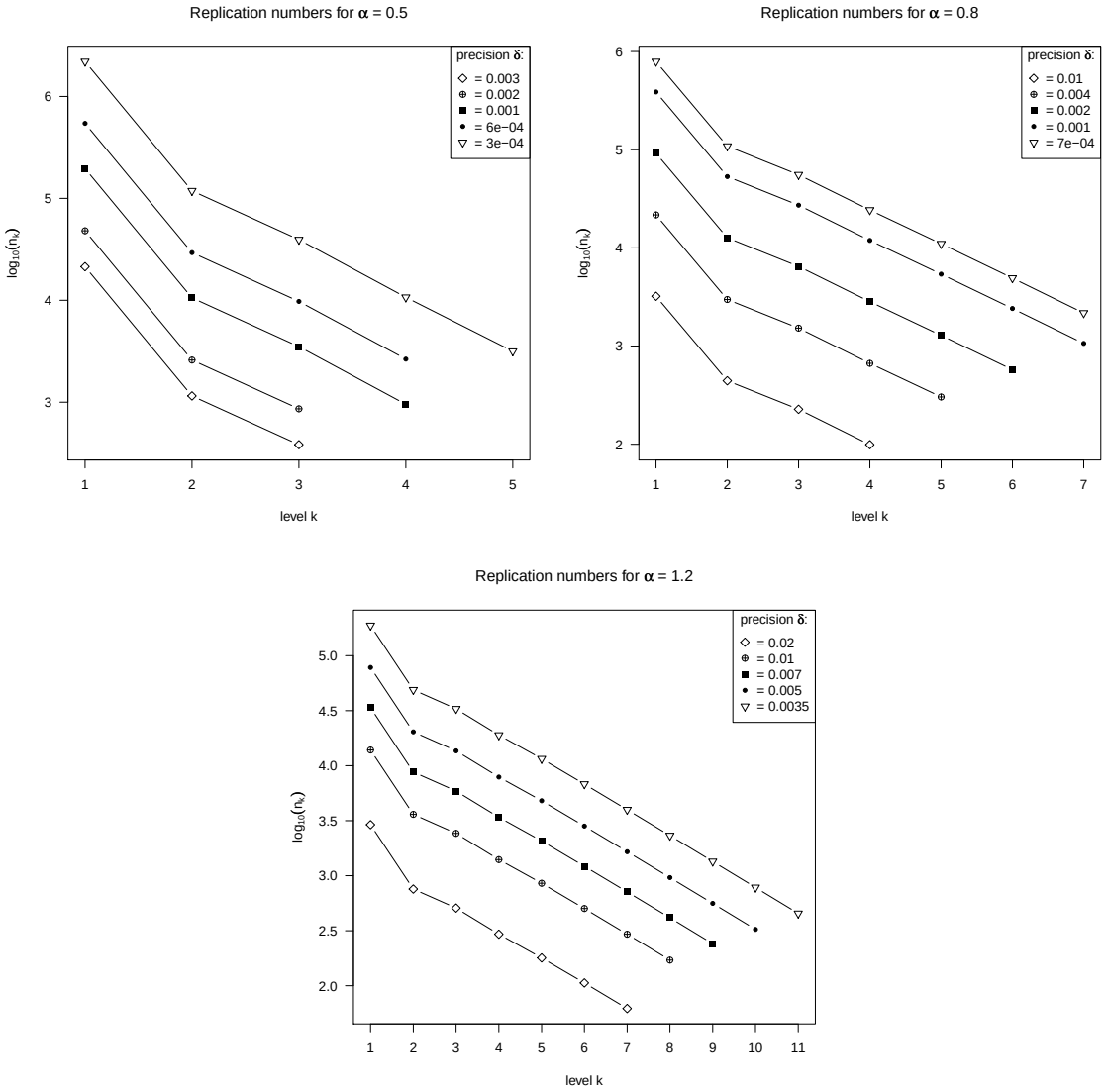

Figure 2: Replication numbers for different precisions δ.

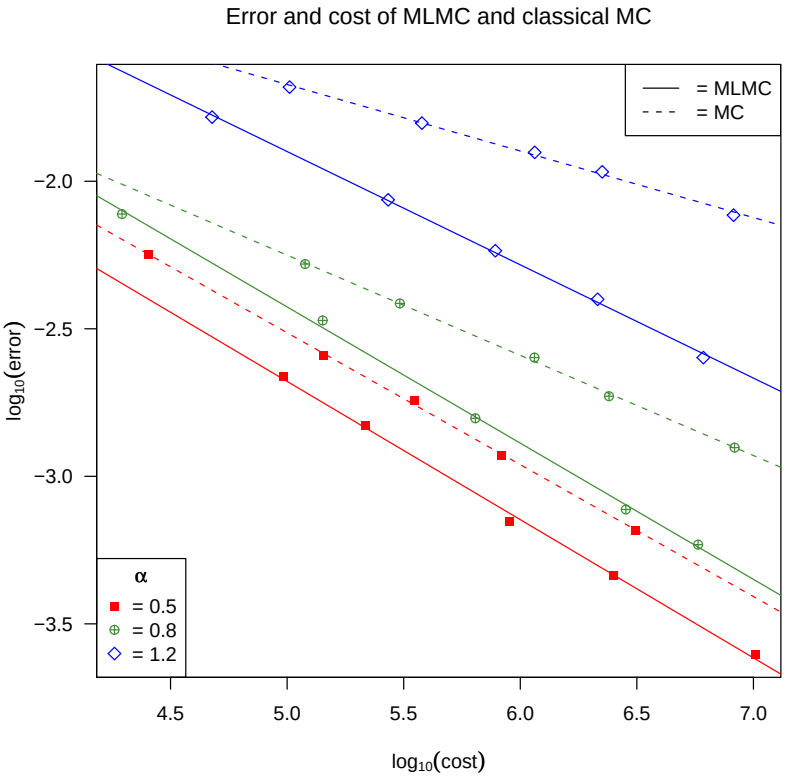

Figure 3: Cost and error of multilevel (MLMC) and classical (MC) Monte Carlo.

as well as the unknown variance are assumed to be 1. Then we determine the parameters h, ε and the replication number n in order to reach precision $\delta > 0$ for the root mean squared 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 in the classical Monte Carlo algorithm, which are $0.45, 0.34$ and 0.23 for $\alpha = 0.5, 0.8$ and 1.2, are always worse than those of the corresponding multilevel algorithm. Furthermore, 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.

Concluding, the multilevel scheme provides an algorithm $\widehat{S}_\delta(f)$ which achieves a superior order of convergence together with a heuristic control of the error, which is an important feature in applications.

4 Proofs

Proving the main result requires asymptotic error bounds for the strong approximation of Y by $\hat{Y}^{(h,\epsilon)}$ for given $\varepsilon, h > 0$. We derive an error estimate in terms of the function

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

for $h > 0$.

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

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

in the general case and

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

in the case without *Wiener process, i.e.* $\Sigma = 0$.

Remark 3. A similar Euler scheme is analyzed in [Rub03]. There it is shown that the appropriately scaled error process (the discrepancy between approximative and real solution) converges in distribution to a stochastic process. In the cases where this limit theorem is applicable, it is straight-forward to verify that the estimate provided in Theorem 2 is of the right order.

Remark 4. In the case without Wiener process the term $|b - F_0(h)|^2\varepsilon^2$ is typically of lower order than $F(h)$, so that we have in most cases that

$$\mathbb{E}\Big[\sup_{t\in[0,1]} |Y_t - \hat{Y}_t^{(h,\epsilon)}|^2\Big] \lesssim \kappa\, F(h).$$

The proof of Theorem 2 is based on the analysis of an auxiliary process $(\bar{Y}_t)$: We decompose the Lévy martingale L into a sum of the Lévy martingale $L' = L^{(h)}$ constituted by the sum of compensated jumps of size bigger than h and the remaining part $L'' = (L_t - L'_t)_{t\geq 0}$. We denote $\bar{X} = (\Sigma W_t + L'_t + tb)_{t\geq 0}$, and let $\bar{Y} = (\bar{Y}_t)_{t\geq 0}$ be the solution to the integral equation

$$\bar{Y}_t = y_0 + \int_0^t a(\bar{Y}_{\iota(s-)})\,\mathrm{d}\bar{X}_s, \tag{17}$$

where $\iota(t) = \sup[0,t]\cap\mathbb{T}$, and $\mathbb{T}$ is the set of random times $(T_j)_{j\in\mathbb{Z}_+}$ defined by $T_j = T_j^{(h,\varepsilon)}$ as in (5) in Section 2.1.

Proposition 1. *Under Assumption A, there exists a constant κ depending only on K such that for any $\varepsilon \in (0,1]$ and $h > 0$ with $\nu(B_h^c) \leq 1/\varepsilon$ we have*

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

in the general case and

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

in the case without *Wiener process, i.e. $\Sigma = 0$.*

The proof of the proposition relies on the following lemma.

Lemma 1. *Under Assumption A, we have*

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

where κ is a finite constant depending on K only.

The proof of the lemma can be achieved along the standard argument for proving bounds for second moments. Indeed, the standard combination of Gronwall's lemma together with Doob's inequality yields the result.

Proof of Proposition 1. For $t \geq 0$, we consider $Z_t = Y_t - \bar{Y}_t$ and $Z'_t = Y_t - \bar{Y}_{\iota(t)}$. We fix a stopping time τ and let $z_\tau(t) = \mathbb{E}[\sup_{s\in[0,t\wedge\tau]} |Z_s|^2]$. To indicate that a process is stopped at time τ we put τ in its superscript. The main task of the proof is to establish an estimate of the form

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

with values $\alpha_1, \alpha_2 > 0$ that do not depend on the choice of τ. Then by using a localizing sequence of stopping times (τ_n) with finite $z_{\tau_n}(1)$, we deduce from Gronwall's inequality that

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

We analyze

$$Z_t = \underbrace{\int_0^t (a(Y_{s-}) - a(\bar{Y}_{\iota(s-)}))\,\mathrm{d}(\Sigma W_s + L'_s) + \int_0^t a(Y_{s-})\,\mathrm{d}L''_s}_{=:M_t} + \int_0^t (a(Y_{s-}) - a(\bar{Y}_{\iota(s-)}))\,b\,\mathrm{d}s \tag{18}$$

with $M = (M_t)_{t\geq 0}$ being a local L^2-martingale. By Doob and Lemma 3, we get

$$\mathbb{E}\sup_{s\in[0,t\wedge\tau]}|M_s|^2 \leq 4\,\mathbb{E}\Big[\int_0^{t\wedge\tau}|a(Y_{s-}) - a(\bar{Y}_{\iota(s-)})|^2\,\mathrm{d}\langle \Sigma W + L'\rangle_s + \int_0^{t\wedge\tau}|a(Y_{s-})|^2\,\mathrm{d}\langle L''\rangle_s\Big],$$

where in general for a local L^2-martingale A we set $\langle A\rangle_t = \sum_j \langle A^{(j)}\rangle_t$, where $\langle A^{(j)}\rangle$ denotes the predictable compensator of the classical bracket process of the j-th coordinate of A. Note that $\mathrm{d}\langle \Sigma W + L'\rangle_t = (|\Sigma|^2 + \int_{B_h^c}|x|^2\,\nu(\mathrm{d}x))\,\mathrm{d}t \leq 2K^2\,\mathrm{d}t$ and similarly $\mathrm{d}\langle L''\rangle_t = F(h)\,\mathrm{d}t$. Consequently, by Assumption A and Fubini's theorem, we get

$$\mathbb{E}\sup_{s\in[0,t\wedge\tau]}|M_s|^2 \leq 8K^4\int_0^t \mathbb{E}[\mathbb{1}_{\{s\leq\tau\}}|Z'^{\tau}_{s-}|^2]\,\mathrm{d}s + 4K^2F(h)\int_0^t \mathbb{E}[(|Y_{s-} - y_0| + 1)^2]\,\mathrm{d}s.$$

Conversely, by the Cauchy-Schwarz inequality and Fubini's theorem, one has for $t \in [0,1]$ that

$$\mathbb{E}\Big[\Big|\int_0^{t\wedge\tau}(a(Y_{s-}) - a(\bar{Y}_{\iota(s-)}))\,b\,\mathrm{d}s\Big|^2\Big] \leq K^4\int_0^t \mathbb{E}[\mathbb{1}_{\{s\leq\tau\}}|Z'^{\tau}_{s-}|^2]\,\mathrm{d}s.$$

Using that for $a, b \in \mathbb{R}$, $(a+b)^2 \leq 2a^2 + 2b^2$, we deduce with (18) that

$$\mathbb{E}\sup_{s\in[0,t\wedge\tau]}|Z_s|^2 \leq 18K^4\int_0^t \mathbb{E}[\mathbb{1}_{\{s\leq\tau\}}|Z'^{\tau}_{s-}|^2]\,\mathrm{d}s + 8K^2F(h)\int_0^t \mathbb{E}[(|Y_{s-} - y_0| + 1)^2]\,\mathrm{d}s.$$

Since $Z'_t = Z_t + \bar{Y}_t - \bar{Y}_{\iota(t)}$ we conclude that

$$\mathbb{E}\sup_{s\in[0,t\wedge\tau]}|Z_s|^2 \leq 36K^4\int_0^t \big[\mathbb{E}[|Z^{\tau}_{s-}|^2] + \mathbb{E}[\mathbb{1}_{\{s\leq\tau\}}|\bar{Y}_{s-} - \bar{Y}_{\iota(s-)}|^2]\big]\,\mathrm{d}s + 8K^2F(h)\int_0^t \mathbb{E}[(|Y_{s-} - y_0| + 1)^2]\,\mathrm{d}s.$$

By Lemma 1, $\mathbb{E}[\sup_{s\in[0,1]}(|Y_s - y_0| + 1)^2]$ is bounded by a constant, and we get, for $t \in [0,1]$,

$$z_\tau(t) \leq \kappa_1\Big[\int_0^t \big[z_\tau(s) + \mathbb{E}[\mathbb{1}_{\{s\leq\tau\}}|\bar{Y}_{s-} - \bar{Y}_{\iota(s-)}|^2]\big]\,\mathrm{d}s + F(h)\Big], \tag{19}$$

where κ_1 is a constant that depends only on K.

Next, we provide an appropriate estimate for $\mathbb{E}[\mathbb{1}_{\{s\leq\tau\}}|\bar{Y}_t - \bar{Y}_{\iota(t)}|^2]$. Since L' has no jumps on $(\iota(t), t)$ we have

$$\bar{Y}_t - \bar{Y}_{\iota(t)} = a(\bar{Y}_{\iota(t)})\big[\Sigma(W_t - W_{\iota(t)}) + \big(b - F_0(h)\big)(t - \iota(t))\big]$$

so that

$$\mathbb{E}[\mathbb{1}_{\{t\leq\tau\}}|\bar{Y}_t - \bar{Y}_{\iota(t)}|^2] \leq 2K^2\mathbb{E}[(|\bar{Y}^{\tau}_{\iota(t)} - y_0| + 1)^2]\,\big[|\Sigma|^2\varepsilon + |b - F_0(h)|^2\varepsilon^2\big].$$

Here we used the independence of the Wiener process and the random times in $\mathbb{T}$. Next, we use that $|\bar{Y}_{\iota(t)} - y_0| \le |Y_{\iota(t)} - y_0| + |Z_{\iota(t)}|$ to deduce that

$$\mathbb{E}[\mathbb{1}_{\{t\le\tau\}}|\bar{Y}_t - \bar{Y}_{\iota(t)}|^2] \le 4K^2\big[\mathbb{E}[(|Y^\tau_{\iota(t)} - y_0| + 1)^2] + \mathbb{E}[|Z^\tau_{\iota(t)}|^2]\big]\,\big[|\Sigma|^2\varepsilon + |b - F_0(h)|^2\varepsilon^2\big].$$

Recall that $\mathbb{E}[\sup_{s\in[0,1]}(|Y_s - y_0| + 1)^2]$ is uniformly bounded. Moreover, by the Cauchy-Schwarz inequality, $|F_0(h)|^2 \le K^2\nu(B_h^c) \le K^2/\varepsilon$ so that the right bracket in the latter equation is uniformly bounded. Consequently, for $t \in [0, 1]$,

$$\mathbb{E}[\mathbb{1}_{\{t\le\tau\}}|\bar{Y}_t - \bar{Y}_{\iota(t)}|^2] \le \kappa_2\big[|\Sigma|^2\varepsilon + |b - F_0(h)|^2\varepsilon^2 + z_\tau(t)\big]$$

where κ_2 is an appropriate constant that depends only on K.

Inserting this estimate into (19) gives

$$z_\tau(t) \le \kappa_3\Big[\int_0^t z_\tau(s)\,\mathrm{d}s + |\Sigma|^2\varepsilon + |b - F_0(h)|^2\varepsilon^2 + F(h)\Big].$$

for a constant κ_3 that depends only on K. If $\Sigma = 0$, then the statement of the proposition follows from the Gronwall inequality. The general case is an immediate consequence of the estimates

$$|\Sigma|^2\varepsilon \le K^2\varepsilon \quad\text{and}\quad |b - F_0(h)|^2\varepsilon^2 \le 2K^2(\varepsilon^2 + \varepsilon) \le 4K^2\varepsilon,$$

where we used again that $|F_0(h)|^2 \le K^2/\varepsilon$. □

For the proof of Theorem 2, we need a further lemma.

Lemma 2. *Let $r \in \mathbb{N}$ and $(\mathcal{G}_j)_{j=0,1,\dots,r}$ denote a filtration. Moreover, let, for $j = 0, \dots, r-1$, 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*

$$\mathbb{E}\big[\max_{j=0,\dots,r-1} U_jV_j\big] \le \mathbb{E}\big[\max_{j=0,\dots,r-1} U_j\big]\cdot\mathbb{E}\big[\max_{j=0,\dots,r-1} V_j\big].$$

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

We proceed by induction. For $r = 1$ the statement is trivial, since U_0 and V_0 are independent random variables. Next, we let $r \ge 1$ arbitrary and note that

$$\mathbb{E}[\max_{j=0,\dots,r} U_jV_j] = \mathbb{E}[\max_{j=1,\dots,r} U_jV_j] + \mathbb{E}[(U_0V_0 - \max_{j=1,\dots,r} U_jV_j)^+].$$

Using the monotonicity of (U_j), we get that

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

Next, we use the induction hypothesis for $\mathbb{E}[\max_{j=1,\dots,r} U_jV_j]$ to deduce that

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

□

Proof of Theorem 2. By Proposition 1, it remains to find an appropriate upper bound for $\mathbb{E}[\sup_{t\in[0,1]}|\bar{Y}_t-\hat{Y}_t|^2]$ where we denote $\hat{Y}_t=\hat{Y}_t^{(h,\epsilon)}$. First note that for all $j\in\mathbb{N}$, one has $\Delta L_{T_j}=\Delta L'_{T_j}$ and

$$L'_{T_{j+1}}=L'_{T_j}+\Delta L'_{T_{j+1}}-(T_{j+1}-T_j)F_0(h)$$

so that by definition, the processes $(\bar{Y}_t)$ and $(\hat{Y}_t)$ coincide almost surely for all times in $\mathbb{T}$ (see (17) and (6)). Hence,

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

Since two neighboring points in $\mathbb{T}$ are at most ε units apart we get

$$\mathbb{E}\big[\sup_{t\in[0,1]}|B_t|^2\big]\le K^2\,\mathbb{E}[(\|\bar{Y}-y_0\|+1)^2]|b-F_0(h)|^2\varepsilon^2. \tag{20}$$

It remains to analyze

$$\mathbb{E}\big[\sup_{t\in[0,1]}|A_t|^2\big]\le K^2|\Sigma|^2\,\mathbb{E}\big[\sup_{t\in[0,1]}(|\bar{Y}_{\iota(t)}-y_0|+1)^2|W_t-W_{\iota(t)}|^2\big].$$

We apply Lemma 2 with $U_j=\mathbb{1}_{\{T_j<1\}}(|\bar{Y}_{T_j}-y_0|+1)^2$, $V_j=\sup_{t\in[T_j,T_{j+1}\wedge 1)}|W_t-W_{T_j}|^2$ and $\mathcal{G}_j=\mathcal{F}_{T_j}$ $(j=0,1,\dots)$. For $r\in\mathbb{N}$ we obtain

$$\mathbb{E}\big[\sup_{t\in[0,1\wedge T_r]}|A_t|^2\big]\le K^2|\Sigma|^2\,\mathbb{E}\big[\sup_{t\in[0,1\wedge T_r]}(|\bar{Y}_{\iota(t)}-y_0|+1)^2\big]\,\mathbb{E}\big[\sup_{j=0,1,\dots,r-1}\sup_{t\in[T_j,T_{j+1}\wedge 1)}|W_t-W_{T_j}|^2\big].$$

By Lévy's modulus of continuity, the term

$$\|W\|_\varphi:=\sup_{0\le s<t\le 1}\frac{|W_t-W_s|}{\varphi(t-s)}$$

is almost surely finite for $\varphi:[0,1]\to[0,\infty),\delta\mapsto\sqrt{\delta\log(e/\delta)}$ and, by Fernique's theorem, $\mathbb{E}\|W\|_\varphi^2$ is finite. Recalling that neighboring points in $\mathbb{T}$ are at most ε units apart, we conclude with the monotonicity of φ on $[0,1]$ that

$$\mathbb{E}\big[\sup_{j=0,\dots,r-1}\sup_{t\in[T_j,T_{j+1}\wedge 1)}|W_t-W_{T_j}|^2\big]\le\mathbb{E}\big[\|W\|_\varphi^2\big]\,\varphi(\varepsilon)^2$$

Letting r to infinity we arrive at

$$\mathbb{E}\big[\sup_{t\in[0,1]}|A_t|^2\big]\le K^2|\Sigma|^2\,\mathbb{E}[\|W\|_\varphi^2]\,\mathbb{E}\big[\sup_{t\in[0,1]}(|\bar{Y}_{\iota(t)}-y_0|+1)^2\big]\,\varphi(\varepsilon)^2, \tag{21}$$

Combining (20) and (21), we get

$$\mathbb{E}[\|\bar{Y}-\hat{Y}\|^2]\le 2K^2\,\mathbb{E}[(\|\bar{Y}-y_0\|+1)^2]\big(|\Sigma|^2\,\mathbb{E}\big[\|W\|_\varphi^2\big]\,\varphi(\varepsilon)^2+|b-F_0(h)|^2\varepsilon^2\big).$$

Next, note that by, Proposition 1 and Lemma 1, $\mathbb{E}[(\|\bar{Y}-y_0\|+1)^2]$ is bounded from above by some constant depending on K only. Consequently, there exists a constant κ with

$$\mathbb{E}[\|\bar{Y}-\hat{Y}\|^2]\le\kappa\big(|\Sigma|^2\,\varphi(\varepsilon)^2+|b-F_0(h)|^2\varepsilon^2\big).$$

Together with Proposition 1, one immediately obtains the statement for the case without Wiener process. In order to obtain the statement of the general case, we again use that $|\int_{B_h^c}x\,\nu(\mathrm{d}x)|^2\le K^2/\varepsilon$ due to the Cauchy-Schwarz inequality. □

Proof of part (i) of Theorem 1

We can assume without loss of generality that

$$\frac{1}{h^{2/3}} \precsim g(h) \precsim \frac{1}{h(\log 1/h)^{1+\gamma}}, \tag{22}$$

since otherwise, we can modify g in such a way that the new g is larger than the old one and enjoys the wanted properties. We consider a multilevel Monte Carlo algorithm $\widehat{S}$ (as introduced in Section 2.2) with $h_k = g^{-1}(2^k)$ and $\varepsilon_k = 2^{-k}$ for $k \in \mathbb{Z}_+$. For technical reasons, we also define ε_0 and h_0 although they do not appear in the algorithm. The parameters $m \in \mathbb{N}$ and $n_1, \dots, n_m \in \mathbb{N}$ are specified below. Note that

$$\nu(B_{h_k}^c) \le g(h_k) = 1/\varepsilon_k \quad \text{and} \quad \varepsilon_k \le 1, \tag{23}$$

so that by Theorem 2, we have

$$\begin{aligned}\mathbb{E}[\|Y^{(k)} - Y^{(k-1)}\|^2] &\le 2\mathbb{E}[\|Y - Y^{(k)}\|^2] + 2\mathbb{E}[\|Y - Y^{(k-1)}\|^2] \\ &\le \kappa_1 \big[F(h_{k-1}) + |b - F_0(h_k)|^2 \varepsilon_{k-1}^2\big].\end{aligned}$$

for some constant $\kappa_1 > 0$. By Lemma 1 and Theorem 2, $\mathbb{E}[\|Y^{(1)} - y_0\|^2]$ is bounded from above by some constant depending on K only. Hence, equation (7) together with Theorem 2 imply the existence of a constant κ_2 such that

$$e^2(\widehat{S}) \le \kappa_2 \sum_{k=1}^{m+1} \frac{1}{n_k} \big[F(h_{k-1}) + |b - F_0(h_k)|^2 \varepsilon_{k-1}^2\big], \tag{24}$$

where we set $n_{m+1} = 1$. Next, we analyze the terms in (24). Using assumption (10), we have, for a sufficiently small $v \in (0,1)$ and an appropriate constant κ_3,

$$\begin{aligned}|F_0(h_k)| &\le \int |x|\, \nu(\mathrm{d}x) \le \frac{1}{v} \int |x|(v \vee |x|)\, \nu(\mathrm{d}x) \le \frac{1}{v} \int_{B_v^c} |x|^2\, \nu(\mathrm{d}x) + \int_0^v \nu(B_u^c)\, \mathrm{d}u \\ &\le \frac{1}{v} \int |x|^2\, \nu(\mathrm{d}x) + \kappa_3 \int_0^v \frac{1}{u(\log 1/u)^{1+\gamma}}\, \mathrm{d}u.\end{aligned}$$

Both integrals are finite. Moreover, we have

$$F(h_k) \le g(h_k)\, h_k^2 = 2^k g^{-1}(2^k)^2, \tag{25}$$

and using (22) we get that

$$\frac{1}{y^{3/2}} \precsim g^{-1}(y) \precsim \frac{1}{y(\log y)^{1+\gamma}} \qquad \text{as } y \to \infty. \tag{26}$$

Hence, we have $2^k g^{-1}(2^k)^2 \succsim 2^{-2k} = \varepsilon_k^2$ as k tends to infinity, and there exists a constant κ_4 such that

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

We shall now fix m and $n_1, \dots, n_m$. For a given parameter $Z \geq 1/g^{-1}(1)$, we choose $m = m(Z) = \inf\{k \in \mathbb{N} : Zg^{-1}(2^k) < 1\} - 1$. Moreover, we set $n_k = n_k(Z) = \lfloor Zg^{-1}(2^{k-1}) \rfloor$ for $k = 1, \dots, m$, and set again $n_{m+1} = 1$. Then $1/n_k \leq 2/(Zg^{-1}(2^{k-1}))$ for $k = 1, \dots, m+1$, so that

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

By (26), $2^k g^{-1}(2^k) \precsim k^{-(1+\gamma)}$ and the latter sum is uniformly bounded for all m. Hence, there exists a constant κ_5 depending only on g and K such that

$$e^2(\widehat{S}) \leq \kappa_5 \frac{1}{Z}.$$

It remains to analyze the cost of the algorithm. The expected number of breakpoints of $Y^{(k)}$ is less than $1/\varepsilon_k + \nu(B_{h_k}^c) \leq 2^{k+1}$ (see (23)) so that

$$\operatorname{cost}(\widehat{S}) \leq \sum_{k=1}^{m} 2^{k+1} n_k. \tag{27}$$

Hence,

$$\operatorname{cost}(\widehat{S}) \leq 4Z \sum_{k=1}^{m} 2^{k-1} g^{-1}(2^{k-1}) \leq \kappa_6 Z,$$

where $\kappa_6 > 0$ is an appropriate constant.

Proof of part (ii) of Theorem 1

We proceed similarly as in the proof of part (i). We assume without loss of generality that g satisfies

$$\frac{\sqrt{\log 1/h}}{h} \precsim g(h) \precsim \frac{(\log 1/h)^\gamma}{h}, \tag{28}$$

since, otherwise, we can enlarge g appropriately. In analogy to the proof of part (i), we choose $h_k = g^{-1}(2^k)$ and $\varepsilon_k = 2^{-k}$ for $k \in \mathbb{Z}_+$, and we note that estimates (23) and (25) remain valid. Next, we deduce with equation (7) and Theorem 2 that, for some constant κ_1,

$$e^2(\widehat{S}) \leq \kappa_1 \sum_{k=1}^{m+1} \frac{1}{n_k} \Big[F(h_{k-1}) + \varepsilon_{k-1} \log \frac{e}{\varepsilon_{k-1}} \Big] \leq \kappa_1 \sum_{k=1}^{m+1} \frac{1}{n_k} \Big[2^{k-1} g^{-1}(2^{k-1})^2 + 2^{-(k-1)} \log(e 2^{k-1}) \Big],$$

where again $n_{m+1} = 1$. Note that (28) implies that

$$\frac{\sqrt{\log y}}{y} \precsim g^{-1}(y) \precsim \frac{(\log y)^\gamma}{y} \quad \text{as } y \to \infty, \tag{29}$$

so that, in particular, $2^k g^{-1}(2^k)^2 \succsim 2^{-k} \log(e 2^k)$ as k tends to infinity. Hence, we find a constant κ_2 such that

$$e^2(\widehat{S}) \leq \kappa_2 \sum_{k=1}^{m+1} \frac{1}{n_k} 2^{k-1} g^{-1}(2^{k-1})^2.$$

For a parameter $Z \geq e \vee (1/g^{-1}(1))$, we choose $m = m(Z) = \inf\{k \in \mathbb{N} : Zg^{-1}(2^k) < 1\} - 1$, and we set $n_k = n_k(Z) = \lfloor Zg^{-1}(2^{k-1}) \rfloor$ for $k = 1, \ldots, m$. Then we get with (29) that

$$e^2(\widehat{S}) \leq 2\kappa_2 \frac{1}{Z} \sum_{k=1}^{m+1} 2^{k-1} g^{-1}(2^{k-1}) \leq \kappa_3 \frac{1}{Z}\, m^{\gamma+1}$$

for an appropriate constant κ_3. By definition, $Zg^{-1}(2^m) \geq 1$ so that, by equation (28),

$$m \leq \log g(\frac{1}{Z}) / \log 2 \precsim \log Z \qquad \text{as } Z \to \infty.$$

Consequently, there exists a constant κ_4 such that

$$e^2(\widehat{S}) \leq \kappa_4 \frac{(\log Z)^{\gamma+1}}{Z}.$$

Similarly, we get for the cost function

$$\text{cost}(\widehat{S}) \leq \sum_{k=1}^{m} 2^{k+1} n_k \leq 4Z \sum_{k=1}^{m} 2^{k-1} g^{-1}(2^{k-1}) \leq \kappa_5 Z (\log Z)^{\gamma+1}.$$

Next, we choose

$$Z = Z(n) = \frac{1}{2\kappa_5} \frac{n}{(\log n)^{\gamma+1}}$$

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

$$\text{cost}(\widehat{S}) \leq \kappa_5 Z (\log Z)^{\gamma+1} \sim \frac{1}{2} n,$$

and we have, for all sufficiently large n, $\text{cost}(\widehat{S}) \leq n$. Conversely, we find

$$e^2(\widehat{S}) \leq \kappa_4 \frac{(\log Z)^{\gamma+1}}{Z} \approx \frac{(\log n)^{2(\gamma+1)}}{n}.$$

Proof of part (iii) of Theorem 1

First note that property (11) is equivalent to

$$\frac{\gamma}{2} g^{-1}(u) \leq g^{-1}(2u) \tag{30}$$

for all sufficiently large $u > 0$. This implies that there exists a finite constant κ_1 depending only on g such that for all $k, l \in \mathbb{Z}_+$ with $k \leq l$ one has

$$g^{-1}(2^k) \leq \kappa_1 (\frac{2}{\gamma})^{l-k} g^{-1}(2^l). \tag{31}$$

We proceed similar as in the proof of part (i) and consider $\widehat{S} \in \mathcal{A}$ with $h_k = g^{-1}(2^k)$ and $\varepsilon_k = 2^{-k}$ for $k \in \mathbb{Z}_+$. The maximal index m and the number of iterations n_k are fixed later in the discussion. Again estimates (23) and (25) remain valid and we get with Theorem 2

$$e^2(\widehat{S}) \leq \kappa_2 \sum_{k=1}^{m+1} \frac{1}{n_k} \Big[F(h_{k-1}) + \varepsilon_{k-1} \log \frac{e}{\varepsilon_{k-1}} \Big] \leq \kappa_2 \sum_{k=1}^{m+1} \frac{1}{n_k} \Big[g^{-1}(2^{k-1}) 2^{k-1} + 2^{-(k-1)} \log(e 2^{k-1}) \Big],$$

for a constant κ_2 and $n_{m+1} = 1$ as before. By (30), we have $g^{-1}(2^k) \gtrsim (\gamma/2)^k$ and recalling that $\gamma > 1$ we conclude that $2^{-k}\log(e2^k) \precsim g^{-1}(2^k)^2 2^k$ as k tends to infinity. Hence, there exists a constant κ_3 with

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

Conversely, we again estimate the cost by

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

Now we specify m and $n_1, \dots, n_m$. For a given parameter $Z \ge 2/g^{-1}(1)$, we let $m = m(Z) = \inf\{k \in \mathbb{N} : Z g^{-1}(2^k) < 2\} - 1$, and set $n_k = n_k(Z) = \lfloor Z g^{-1}(2^{k-1}) \rfloor$ for $k = 1, \dots, m$. Then we get with (31) that there exists a constant κ_4 with

$$\begin{aligned} e^2(\widehat{S}) &\le 2\kappa_3 \frac{1}{Z} \sum_{k=1}^{m+1} 2^{k-1} g^{-1}(2^{k-1}) \le 2\kappa_1\kappa_3 \frac{1}{Z} \sum_{k=1}^{m+1} 2^{k-1} g^{-1}(2^{m+1}) \big(\frac{2}{\gamma}\big)^{m+1-(k-1)} \\ &\le 2\kappa_1\kappa_3 \frac{1}{Z} 2^{m+1} g^{-1}(2^{m+1}) \sum_{k=1}^{m+1} \gamma^{-(m+1-(k-1))} \le \kappa_4 \frac{1}{Z} 2^{m+1} g^{-1}(2^{m+1}). \end{aligned}$$

Moreover, by (30) one has for sufficiently large m (or, equivalently, for sufficiently large Z) that $Z g^{-1}(2^{m+1}) \ge \frac{\gamma}{2} Z g^{-1}(2^m) \ge \gamma > 1$ so that

$$e^2(\widehat{S}) \le \kappa_4 2^{m+1} g^{-1}(2^{m+1})^2$$

Similarly, one obtains

$$\mathrm{cost}(\widehat{S}) \le \kappa_5 Z 2^{m+1} g^{-1}(2^{m+1}) < 2\kappa_5 2^{m+1}.$$

Next, we choose, for given $n \ge 2\kappa_5$, $Z > 0$ such that $m = \lfloor \log_2 \frac{n}{2\kappa_5} \rfloor - 1$. Then, clearly, $\mathrm{cost}(\widehat{S}) \le n$ and for sufficiently large n we have

$$e^2(\widehat{S}) \le \kappa_4 \frac{n}{2\kappa_5}\, g^{-1}\big(\frac{n}{4\kappa_5}\big)^2 \precsim n\, g^{-1}(n)^2 \ \text{ as } n \to \infty.$$

In the last step, we again used property (30).

Proof of Corollary 1

We assume without loss of generality that $\beta < 2$ since otherwise the statement of the corollary is trivial. We fix $\beta' \in (\beta, 2]$ and recall that

$$C := \int_{B_1} |x|^{\beta'} \, \nu(\mathrm{d}x)$$

is finite. We consider $\bar{g} : (0,\infty) \to (0,\infty), h \mapsto \int \frac{|x|^2}{h^2} \wedge 1 \, \nu(\mathrm{d}x)$, whose integral we split for $h \in (0,1]$ into three parts:

$$\bar{g}(h) = \int_{B_h} \frac{|x|^2}{h^2} \, \nu(\mathrm{d}x) + \int_{B_1 \setminus B_h} 1 \, \nu(\mathrm{d}x) + \int_{B_1^c} 1 \, \nu(\mathrm{d}x) =: I_1 + I_2 + I_3.$$

The last term does not depend on h and we estimate the first two terms by

$$I_1 \leq h^{-\beta'} \int_{B_h} |x|^{\beta'} \, \nu(\mathrm{d}x) \leq C \, h^{-\beta'} \quad \text{and} \quad I_2 \leq h^{-\beta'} \int_{B_1 \setminus B_h} |x|^{\beta'} \, \nu(\mathrm{d}x) \leq C \, h^{-\beta'}.$$

Hence, we can choose $\beta'' \in ((\beta' \vee 1), 2]$ arbitrarily, and a decreasing and invertible function $g : (0, \infty) \to (0, \infty)$ that dominates $\bar{g}$ and satisfies $g(h) = h^{-\beta''}$ for all sufficiently small $h > 0$. Then for $\gamma = 2^{1-1/\beta''}$, one has $g(\frac{\gamma}{2} h) = 2g(h)$ for all sufficiently small $h > 0$ and, by part (iii) of Theorem 1 there exists a sequence of multilevel algorithms $\widehat{S}_n$ such that

$$e(\widehat{S}_n) \precsim n^{\frac{1}{2} - \frac{1}{\beta''}}.$$

The general statement follows via a diagonal argument. In the case where $\beta < 1$, one can choose $\beta' = 1$ and $\beta'' > 1$ arbitrarily close to one and gets the result. Whereas for $\beta \geq 1$, one can choose for any $\beta'' > \beta$ an appropriate β' and thus retrieve the statement.

Appendix

We shall use the following consequence of the Itô isometry for Lévy processes.

Lemma 3. *Let (A_t) be a previsible process with state space $\mathbb{R}^{d_Y \times d_X}$, let (L_t) be a square integrable $\mathbb{R}^{d_X}$-valued Lévy martingale and denote by $\langle L \rangle$ the process given via*

$$\langle L \rangle_t = \sum_{j=1}^{d_X} \langle L^{(j)} \rangle_t,$$

where $\langle L^{(j)} \rangle$ denotes the predictable compensator of the classical bracket process for the j-th coordinate of L. One has, for any stopping time τ with finite expectation $\mathbb{E} \int_0^\tau |A_s|^2 \, \mathrm{d}\langle L \rangle_s$, that $(\int_0^{t \wedge \tau} A_s \, \mathrm{d}L_s)_{t \geq 0}$ is a uniformly square integrable martingale which satisfies

$$\mathbb{E} \Big| \int_0^\tau A_s \, \mathrm{d}L_s \Big|^2 \leq \mathbb{E} \int_0^\tau |A_s|^2 \, \mathrm{d}\langle L \rangle_s.$$

In general the estimate can be strengthened by replacing the Frobenius norm by the matrix norm induced by the Euclidean norm. For the convenience of the reader we provide a proof of the lemma.

Proof. Let ν denote the Lévy measure of L and $\Sigma\Sigma^*$ the covariance of its Wiener component. Let $Q : \mathbb{R}^{d_X} \to \mathbb{R}^{d_X}$ denote the self-adjoint operator given by

$$Qx = \Sigma\Sigma^* x + \int \langle x, y \rangle \, y \, \nu(\mathrm{d}y).$$

We recall the Itô isometry for Lévy processes. One has for a previsible process (A_s) with

$$\mathbb{E} \int_0^\tau |A_s Q^{1/2}|^2 \, \mathrm{d}s < \infty,$$

that $(\int_0^{t\wedge\tau} A_s \,\mathrm{d}L_s)_{t\geq 0}$ is a uniformly square integrable martingale with

$$\mathbb{E}\Big|\int_0^\tau A_s \,\mathrm{d}L_s\Big|^2 = \mathbb{E}\int_0^\tau |A_s Q^{1/2}|^2 \,\mathrm{d}s.$$

The statement now follows immediately by noticing that

$$|A_s Q^{1/2}|^2 \leq |A_s|^2 \operatorname{tr}(Q) \quad \text{and} \quad \int_0^\tau |A_s|^2 \operatorname{tr}(Q)\,\mathrm{d}s = \int_0^\tau |A_s|^2 \,\mathrm{d}\langle L\rangle_s.$$

□

Acknowledgments. The authors would like to thank Michael Scheutzow for the proof of Lemma 2. Felix Heidenreich was supported by the DFG-Schwerpunktprogramm 1324.

References

[AD09] F. Aurzada and S. Dereich. The coding complexity of Lévy processes. *Found. Comput. Math.*, 9(3):359–390, 2009.

[App04] D. Applebaum. *Lévy processes and stochastic calculus*, volume 93 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2004.

[AR01] 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.

[Avi09] R. Avikainen. On irregular functionals of SDEs and the euler scheme. *Finance Stoch.*, 13(3):381–401, 2009.

[Ber98] J. Bertoin. *Lévy processes*. Cambridge Univ. Press, 1998.

[CDMGR08] J. Creutzig, S. Dereich, T. Müller-Gronbach, and K. Ritter. Infinite-dimensional quadrature and approximation of distributions. *Found. Comput. Math.*, 2008.

[Der10] S. Dereich. Multilevel Monte Carlo for Lévy-driven SDE's with Gaussian correction. To appear in Ann. Appl. Probab., 2010.

[Gil08a] M. B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, Berlin, 2008.

[Gil08b] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3), 2008.

[JKMP05] J. Jacod, T. G. Kurtz, S. Méléard, and P. Protter. The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(3):523–558, 2005.

[MGR09] T. Müller-Gronbach and K. Ritter. Variable subspace sampling and multi-level algorithms. In A. Owen P. L'Ecuyer, editor, *Monte Carlo and Quasi-Monte Carlo Methods 2008*. Springer-Verlag, Berlin, 2009.

[Pro05] P. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. 2nd edition.

[PT97] P. Protter and D. Talay. The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, 25(1):393–423, 1997.

[Rub03] 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.

[Sat99] K. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.

[TKH09] 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.