The Annals of Applied Probability
2011, Vol. 21, No. 1, 283–311
DOI: 10.1214/10-AAP695

MULTILEVEL MONTE CARLO ALGORITHMS FOR LÉVY-DRIVEN SDES WITH GAUSSIAN CORRECTION

BY STEFFEN DEREICH

Philipps-Universität Marburg

We introduce and analyze multilevel Monte Carlo algorithms for the computation of $\mathbb{E}f(Y)$, where $Y = (Y_t)_{t\in[0,1]}$ is the solution of a multidimensional Lévy-driven stochastic differential equation and f is a real-valued function on the path space. The algorithm relies on approximations obtained by simulating large jumps of the Lévy process individually and applying a Gaussian approximation for the small jump part. Upper bounds are provided for the worst case error over the class of all measurable real functions f that are Lipschitz continuous with respect to the supremum norm. These upper bounds are easily tractable once one knows the behavior of the Lévy measure around zero.

In particular, one can derive upper bounds from the Blumenthal–Getoor index of the Lévy process. In the case where the Blumenthal–Getoor index is larger than one, this approach is superior to algorithms that do not apply a Gaussian approximation. If the Lévy process does not incorporate a Wiener process or if the Blumenthal–Getoor index β is larger than $\frac{4}{3}$, then the upper bound is of order $\tau^{-(4-\beta)/(6\beta)}$ when the runtime τ tends to infinity. Whereas in the case, where β is in $[1, \frac{4}{3}]$ and the Lévy process has a Gaussian component, we obtain bounds of order $\tau^{-\beta/(6\beta-4)}$. In particular, the error is at most of order $\tau^{-1/6}$.

1. Introduction. Let $d_Y \in \mathbb{N}$ and denote by $D[0, 1]$ the Skorokhod space of functions mapping $[0, 1]$ to $\mathbb{R}^{d_Y}$ endowed with its Borel-σ-field. 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 (with state space $\mathbb{R}^{d_Y}$), and
- $f : D[0, 1] \to \mathbb{R}$ is a Borel measurable function that is Lipschitz continuous with respect to the *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 depen-

Received August 2009; revised January 2010.
AMS 2000 subject classifications. Primary 60H35; secondary 60H05, 60H10, 60J75.
Key words and phrases. Multilevel Monte Carlo, Komlós–Major–Tusnády coupling, weak approximation, numerical integration, Lévy-driven stochastic differential equation.

dent) option, and in the past several concepts have been employed for dealing with it.

A common stochastic approach is to perform a Monte Carlo simulation of numerical approximations to the solution Y. Typically, the Euler or Milstein schemes are used to obtain approximations. Also higher order schemes can be applied provided that samples of iterated Itô integrals are supplied and the coefficients of the equation are sufficiently regular. In general, the problem is tightly related to *weak approximation* which is, for instance, extensively studied in the monograph by Kloeden and Platen [12] for diffusions.

Essentially, one distinguishes between two cases. Either $f(Y)$ depends only on the state of Y at a fixed time or alternatively it depends on the whole trajectory of Y. In the former case, extrapolation techniques can often be applied to increase the order of convergence, see [21]. For Lévy-driven stochastic differential equations, the Euler scheme was analyzed in [17] under the assumption that the increments of the Lévy process are simulatable. Approximate simulations of the Lévy increments are considered in [11].

In this article, we consider functionals f that depend on the whole trajectory. Concerning results for diffusions, we refer the reader to the monograph [12]. For Lévy-driven stochastic differential equations, limit theorems in distribution are provided in [10] and [18] for the discrepancy between the genuine solution and Euler approximations.

Recently, Giles [7, 8] (see also [9]) introduced the so-called *multilevel Monte Carlo method* to compute $S(f)$. It is very efficient when Y is a diffusion. Indeed, it even can be shown that it is—in some sense—optimal, see [5]. For Lévy-driven stochastic differential equations, multilevel Monte Carlo algorithms are first introduced and studied in [6]. Let us explain their findings in terms of the Blumenthal–Getoor index (BG-index) of the driving Lévy process which is an index in $[0, 2]$. It measures the frequency of small jumps, see (3), where a large index corresponds to a process which has small jumps at high frequencies. In particular, all Lévy processes which have a finite number of jumps has BG-index zero. Whenever the BG-index is smaller or equal to one, the algorithms of [6] have worst case errors at most of order $\tau^{-1/2}$, when the runtime τ tends to infinity. Unfortunately, the efficiency decreases significantly for larger Blumenthal–Getoor indices.

Typically, it is not feasible to simulate the increments of the Lévy process perfectly, and one needs to work with approximations. This necessity typically worsens the performance of an algorithm, when the BG-index is larger than one due to the higher frequency of small jumps. It represents the main bottleneck in the simulation. In this article, we consider approximative Lévy increments that simulate the large jumps and approximate the small ones by a normal distribution (*Gaussian approximation*) in the spirit of Asmussen and Rosiński [2] (see also [4]). Whenever the BG-index is larger than one, this approach is superior to the approach taken in [6], which neglects small jumps in the simulation of Lévy increments.

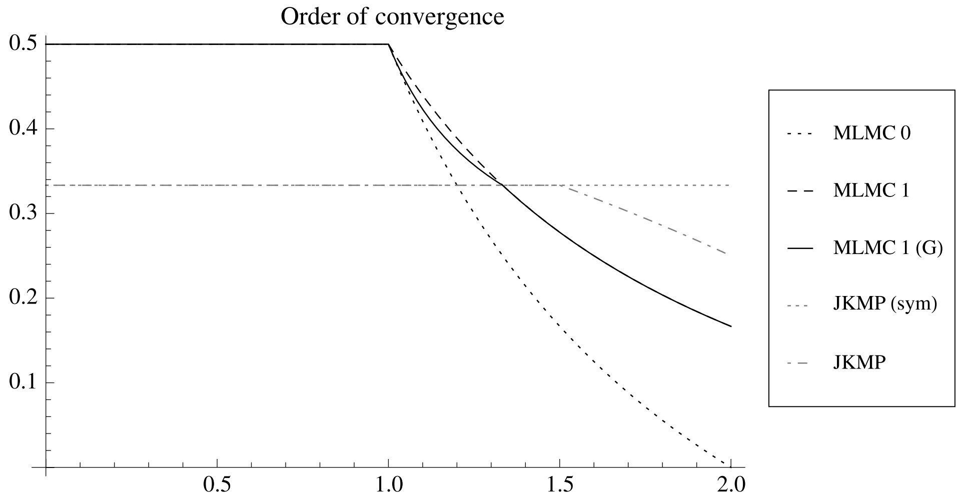

FIG. 1. *Order of convergence in dependence on the Blumenthal–Getoor index.*

To be more precise, we establish a new estimate for the Wasserstein metric between an approximative solution with Gaussian approximation and the genuine solution, see Theorem 3.1. It is based on a consequence of Zaitsev's generalization [22] of the Komlós–Major–Tusnády coupling [13, 14] which might be of its own interest itself, see Theorem 6.1. With these new estimates, we analyze a class of multilevel Monte Carlo algorithms together with a *cost* function which measures the computational complexity of the individual algorithms. We provide upper error bounds for individual algorithms and optimize the error over the parameters under a given cost constraint. When the BG-index is larger than one, appropriately adjusted algorithms lead to significantly smaller *worst case errors* over the class of Lipschitz functionals than the ones analyzed so far, see Theorem 1.1, Corollary 1.2 and Figure 1. In particular, one always obtains numerical schemes with errors at most of order $\tau^{-1/6}$ when the runtime τ of the algorithm tends to infinity.

Notation and universal assumptions. We denote by $|\cdot|$ the Euclidean norm for vectors as well as the Frobenius norm for matrices and let $\|\cdot\|$ denote the supremum norm over the interval $[0,1]$. $X=(X_t)_{t\geq 0}$ denotes an d_X-dimensional L^2-integrable Lévy process. By the Lévy–Khintchine formula, it is characterized by a square integrable Lévy-measure ν [a Borel measure on $\mathbb{R}^{d_X}\backslash\{0\}$ with $\int |x|^2\nu(\mathrm{d}x)<\infty$], a positive semi-definite matrix $\Sigma\Sigma^*$ (Σ being a $d_X\times d_X$-matrix), and a drift $b\in\mathbb{R}^{d_X}$ via

$$\mathbb{E}e^{i\langle\theta,X_t\rangle}=e^{t\psi(\theta)},$$

where

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

Briefly, we call X a $(\nu, \Sigma\Sigma^*, b)$-Lévy process, and when $b = 0$, a $(\nu, \Sigma\Sigma^*)$-Lévy martingale. All Lévy processes under consideration are assumed to be càdlàg. As is well known, we can represent X as sum of three independent processes

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

where $W = (W_t)_{t\ge 0}$ is a d_X-dimensional Wiener process and $L = (L_t)_{t\ge 0}$ is a L^2-martingale that comprises the compensated jumps of X. We consider the integral equation

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

where $y_0 \in \mathbb{R}^{d_Y}$ is a fixed deterministic initial value. We impose the standard Lipschitz assumption on the function $a : \mathbb{R}^{d_Y} \to \mathbb{R}^{d_Y \times d_X}$: for a fixed $K < \infty$, and all $y, y' \in \mathbb{R}^{d_Y}$, one has

$$|a(y) - a(y')| \le K|y - y'| \quad \text{and} \quad |a(y_0)| \le K.$$

Furthermore, we assume without further mentioning that

$$\int |x|^2 \nu(\mathrm{d}x) \le K^2, \qquad |\Sigma| \le K \quad \text{and} \quad |b| \le K.$$

We refer to the monographs [3] and [20] for details concerning Lévy processes. Moreover, a comprehensive introduction to the stochastic calculus for discontinuous semimartingales and, in particular, Lévy processes can be found in [16] and [1].

In order to approximate the small jumps of the Lévy process, we need to impose a uniform ellipticity assumption.

ASSUMPTION UE. There are $\mathfrak{h} \in (0, 1]$, $\vartheta \ge 1$ and a linear subspace $\mathcal{H}$ of $\mathbb{R}^{d_X}$ such that for all $h \in (0, \mathfrak{h}]$ the Lévy measure $\nu|_{B(0,h)}$ is supported on $\mathcal{H}$ and satisfies

$$\frac{1}{\vartheta}\int_{B(0,h)} \langle y, x\rangle^2 \nu(\mathrm{d}x) \le \int_{B(0,h)} \langle y', x\rangle^2 \nu(\mathrm{d}x) \le \vartheta \int_{B(0,h)} \langle y, x\rangle^2 \nu(\mathrm{d}x)$$

for all $y, y' \in \mathcal{H}$ with $|y| = |y'|$.

Main results. We consider a class of multilevel Monte Carlo algorithms $\mathcal{A}$ together with a cost function $\mathrm{cost} : \mathcal{A} \to [0, \infty)$ that are introduced explicitly in Section 2. For each algorithm $\widehat{S} \in \mathcal{A}$, we denote by $\widehat{S}(f)$ a real-valued random variable representing the random output of the algorithm when applied to a given measurable function $f : D[0, 1] \to \mathbb{R}$. 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, see also [15]. Our cost function represents the runtime of the algorithm reasonably well when supposing that

- one can sample from the distribution $\nu|_{B(0,h)^c}/\nu(B(0,h)^c)$ 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 less than a constant multiple of its breakpoints plus one time units.

As pointed out below, in that case, the average runtime to evaluate $\widehat{S}(f)$ is less than a constant multiple of $\mathrm{cost}(\widehat{S})$. We analyze the minimal *worst case error*

$$\mathrm{err}(\tau) = \inf_{\substack{\widehat{S}\in\mathcal{A}:\\ \mathrm{cost}(\widehat{S})\le\tau}} \sup_{f\in\mathrm{Lip}(1)} \mathbb{E}[|S(f)-\widehat{S}(f)|^2]^{1/2}, \qquad \tau \ge 1.$$

Here and elsewhere, $\mathrm{Lip}(1)$ denotes the class of measurable functions $f : D[0,1] \to \mathbb{R}$ that are Lipschitz continuous with respect to *supremum* norm with coefficient one.

In this article, we use asymptotic comparisons. We write $f \approx g$ for $0 < \liminf \frac{f}{g} \le \limsup \frac{f}{g} < \infty$, and $f \precsim g$ or, equivalently $g \succsim f$, for $\limsup \frac{f}{g} < \infty$. Our main findings are summarized in the following theorem.

THEOREM 1.1. *Assume that Assumption* UE *is valid and let* $g : (0,\infty) \to (0,\infty)$ *be a decreasing and invertible function such that for all* $h > 0$

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

and, for a fixed $\gamma > 1$,

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

for all sufficiently small $h > 0$.

(I) *If* $\Sigma = 0$ *or*

$$g^{-1}(x) \succsim x^{-3/4} \qquad \text{as } x \to \infty,$$

then

$$\mathrm{err}(\tau) \precsim g^{-1}((\tau\log\tau)^{2/3})\tau^{1/6}(\log\tau)^{2/3} \qquad \text{as } \tau \to \infty.$$

(II) *If*

$$g^{-1}(x) \precsim x^{-3/4} \qquad \text{as } x \to \infty,$$

then

$$\mathrm{err}(\tau) \precsim \sqrt{\frac{\log\tau}{g^*(\tau)}} \qquad \text{as } \tau \to \infty,$$

where $g^*(\tau) = \inf\{x > 1 : x^3 g^{-1}(x)^2(\log x)^{-1} \ge \tau\}$.

The class of algorithms $\mathcal{A}$ together with appropriate parameters which establish the error estimates above are stated explicitly in Section 2.

In terms of the Blumenthal–Getoor index

$$\beta := \inf\left\{p > 0 : \int_{B(0,1)} |x|^p \nu(\mathrm{d}x) < \infty\right\} \in [0, 2] \tag{3}$$

we get the following corollary.

COROLLARY 1.2. *Assume that Assumption UE is valid and that the BG-index satisfies $\beta \ge 1$. If $\Sigma = 0$ or $\beta \ge \frac{4}{3}$, then*

$$\sup\{\gamma \ge 0 : \mathrm{err}(\tau) \precsim \tau^{-\gamma}\} \ge \frac{4-\beta}{6\beta},$$

and, if $\Sigma \ne 0$ and $\beta < \frac{4}{3}$,

$$\sup\{\gamma \ge 0 : \mathrm{err}(\tau) \precsim \tau^{-\gamma}\} \ge \frac{\beta}{6\beta - 4}.$$

Visualization of the results and relationship to other work. Figure 1 illustrates our findings and related results. The x-axis and y-axis represent the Blumenthal–Getoor index and the order of convergence, respectively. Note that MLMC 0 stands for the multilevel Monte Carlo algorithm which does not apply a Gaussian approximation, see [6]. Both lines marked as MLMC 1 illustrate Corollary 1.2, where the additional (G) refers to the case where the SDE comprises a Wiener process.

These results are to be compared with the results of Jacod et al. [11]. Here an approximate Euler method is analyzed by means of weak approximation. In contrast to our investigation, the object of that article is to compute $\mathbb{E} f(X_T)$ for a fixed time $T > 0$. Under quite strong assumptions (for instance, a and f have to be four times continuously differentiable and the eights moment of the Lévy process needs to be finite), they provide error bounds for a numerical scheme which is based on Monte Carlo simulation of one approximative solution. In the figure, the two lines quoted as JKMP represent the order of convergence for general, respectively pseudo symmetrical, Lévy processes. Additionally to the illustrated schemes, [11] provide an expansion which admits a Romberg extrapolation under additional assumptions.

We stress the fact that our analysis is applicable to general path dependent functionals and that our error criterion is the worst case error over the class of Lipschitz continuous functionals with respect to supremum norm. In particular, our class contains most of the continuous payoffs appearing in finance.

We remark that our results provide upper bounds for the inferred error and so far no lower bounds are known. The worst exponent appearing in our estimates is $\frac{1}{6}$ which we obtain for Lévy processes with Blumenthal–Getoor index 2. Interestingly, this is also the worst exponent appearing in [19] in the context of strong approximation of SDEs driven by subordinated Lévy processes.

Agenda. The article is organized as follows. In Section 2, we introduce a class of multilevel Monte Carlo algorithms together with a cost function. Here, we also provide the crucial estimate for the mean squared error which motivates the consideration of the Wasserstein distance between an approximative and the genuine solution, see (6). Section 3 states the central estimate for the former Wasserstein distance, see Theorem 3.1. In this section, we explain the strategy of the proof and the structure of the remaining article in detail. For the proof, we couple the driving Lévy process with a Lévy process constituted by the large jumps plus a Gaussian compensation of the small jumps and we write the difference between the approximative and the genuine solution as a telescoping sum including further auxiliary processes, see (9) and (10). The individual errors are then controlled in Sections 4 and 5 for the terms which do not depend on the particular choice of the coupling and in Section 7 for the error terms that do depend on the particular choice. In between, in Section 6, we establish the crucial KMT like coupling result for the Lévy process. Finally, in Section 8, we combine the approximation result for the Wasserstein metric (Theorem 3.1) with estimates for strong approximation of stochastic differential equations from [6] to prove the main results stated above.

2. Multilevel Monte Carlo. Based on a number of parameters, we define a multilevel Monte Carlo algorithm $\widehat{S}$: We denote by m and $n_1, \dots, n_m$ natural numbers and let $\varepsilon_1, \dots, \varepsilon_m$ and $h_1, \dots, h_m$ denote decreasing sequences of positive reals. Formally, the algorithm $\widehat{S}$ can be represented as a tuple constituted by these parameters, and we denote by $\mathcal{A}$ the set of all possible choices for $\widehat{S}$. We continue with defining processes that depend on the latter parameters. For ease of notation, the parameters are omitted in the definitions below.

We choose a square matrix $\Sigma^{(m)}$ such that $(\Sigma^{(m)}(\Sigma^{(m)})^*)_{i,j} = \int_{B(0,h_m)} x_i x_j \times \nu(\mathrm{d}x)$. Moreover, for $k = 1, \dots, m$, we let $L^{(k)} = (L_t^{(k)})_{t \ge 0}$ denote the $(\nu|_{B(0,h_k)^c}, 0)$-Lévy martingale which comprises the compensated jumps of L that are larger than h_k, that is

$$L_t^{(k)} = \sum_{s \le t} \mathbb{1}_{\{|\Delta L_s| \ge h_k\}} \Delta L_s - t \int_{B(0,h_k)^c} x \nu(\mathrm{d}x).$$

Here and elsewhere, we denote $\Delta L_t = L_t - L_{t-}$. We let $B = (B_t)_{t \ge 0}$ be an independent Wiener process (independent of W and $L^{(k)}$), and consider, for $k = 1, \dots, m$, the processes $\mathcal{X}^{(k)} = (\Sigma W_t + \Sigma^{(m)} B_t + L_t^{(k)} + bt)_{t \ge 0}$ as driving processes. Let $\Upsilon^{(k)}$ denote the solution to

$$\Upsilon_t^{(k)} = y_0 + \int_0^t a(\Upsilon_{s-}^{(k)}) \,\mathrm{d}\mathcal{X}_{\iota^{(k)}(s)},$$

where $(\iota^{(k)}(t))_{t \ge 0}$ is given via $\iota^{(k)}(t) = \max(\mathbb{I}^{(k)} \cap [0, t])$ and the set $\mathbb{I}^{(k)}$ is constituted by the random times $(T_j^{(k)})_{j \in \mathbb{Z}_+}$ that are inductively defined via $T_0^{(k)} = 0$ and

$$T_{j+1}^{(k)} = \inf\{t \in (T_j^{(k)}, \infty) : |\Delta L_t| \ge h_k \text{ or } t = T_j^{(k)} + \varepsilon_k\}.$$

Clearly, $\Upsilon^{(k)}$ is constant on each interval $[T_j^{(k)}, T_{j+1}^{(k)})$ and one has

$$\Upsilon_{T_{j+1}^{(k)}}^{(k)} = \Upsilon_{T_j^{(k)}}^{(k)} + a(\Upsilon_{T_j^{(k)}}^{(k)})(\mathcal{X}_{T_{j+1}^{(k)}} - \mathcal{X}_{T_j^{(k)}}). \tag{4}$$

Note that we can write

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

The multilevel Monte Carlo algorithm—identified with $\widehat{S}$—estimates each expectation $\mathbb{E}[f(\Upsilon^{(k)}) - f(\Upsilon^{(k-1)})]$ (resp., $\mathbb{E}[f(\Upsilon^{(1)})]$) individually by sampling independently n_k (resp., n_1) versions of $f(\Upsilon^{(k)}) - f(\Upsilon^{(k-1)})$ $[f(\Upsilon^{(1)})]$ and taking the average. The output of the algorithm is then the sum of the individual estimates. We denote by $\widehat{S}(f)$ a random variable that models the random output of the algorithm when applied to f.

The mean squared error of an algorithm. The Monte Carlo algorithm introduced above induces the mean squared error

$$\begin{aligned} \operatorname{mse}(\widehat{S}, f) = {}& \left|\mathbb{E}[f(Y)] - \mathbb{E}[f(\Upsilon^{(m)})]\right|^2 + \sum_{k=2}^{m} \frac{1}{n_k} \operatorname{var}\bigl(f(\Upsilon^{(k)}) - f(\Upsilon^{(k-1)})\bigr) \\ & + \frac{1}{n_1} \operatorname{var}\bigl(f(\Upsilon^{(1)})\bigr), \end{aligned}$$

when applied to f. For two $D[0,1]$-valued random elements $Z^{(1)}$ and $Z^{(2)}$, we denote by $\mathcal{W}(Z^{(1)}, Z^{(2)})$ the Wasserstein metric of second-order with respect to supremum norm, that is

$$\mathcal{W}(Z^{(1)}, Z^{(2)}) = \inf_{\xi} \biggl(\int \|z^{(1)} - z^{(2)}\|^2 \, \mathrm{d}\xi(z^{(1)}, z^{(2)}) \biggr)^{1/2}, \tag{5}$$

where the infimum is taken over all probability measures ξ on $D[0,1] \times D[0,1]$ having first marginal $\mathbb{P}_{Z^{(1)}}$ and second marginal $\mathbb{P}_{Z^{(2)}}$. Clearly, the Wasserstein distance depends only on the distributions of $Z^{(1)}$ and $Z^{(2)}$. Now, we get for $f \in \operatorname{Lip}(1)$, that

$$\begin{aligned} \operatorname{mse}(\widehat{S}, f) \le {}& \mathcal{W}(Y, \Upsilon^{(m)})^2 + \sum_{k=2}^{m} \frac{1}{n_k} \mathbb{E}[\|\Upsilon^{(k)} - \Upsilon^{(k-1)}\|^2] \\ & + \frac{1}{n_1} \mathbb{E}[\|\Upsilon^{(1)} - y_0\|^2]. \end{aligned} \tag{6}$$

We set

$$\operatorname{mse}(\widehat{S}) = \sup_{f \in \operatorname{Lip}(1)} \operatorname{mse}(\widehat{S}, f),$$

and remark that estimate (6) remains valid for the worst case error $\mathrm{mse}(\widehat{S})$.

The main task of this article is to provide good estimates for the Wasserstein metric $\mathcal{W}(Y, \Upsilon^{(m)})$. The remaining terms on the right-hand side of (6) are controlled with estimates from [6].

The cost function. In order to simulate one pair $(\Upsilon^{(k-1)}, \Upsilon^{(k)})$, we need to simulate all displacements of L of size larger or equal to h_k on the time interval $[0, 1]$. Moreover, we need the increments of the Wiener process on the time skeleton $(\mathbb{I}^{(k-1)} \cup \mathbb{I}^{(k)}) \cap [0, 1]$. Then we can construct our approximation via (4). In the real number model of computation (under the assumptions described in the Introduction), this can be performed with runtime less than a multiple of the number of entries in $\mathbb{I}^{(k)} \cap [0, 1]$, see [6] for a detailed description of an implementation of a similar scheme. Since

$$\mathbb{E}\big[\#\big(\mathbb{I}^{(k)} \cap [0, 1]\big)\big] \le 1 + \frac{1}{\varepsilon_k} + \mathbb{E}\bigg[\sum_{t\in[0,1]} \mathbb{1}_{\{|\Delta L_t| \ge h_k\}}\bigg] = \nu(B(0, h_k)^c) + \frac{1}{\varepsilon_k} + 1,$$

we define, for $\widehat{S} \in \mathcal{A}$,

$$\mathrm{cost}(\widehat{S}) = \sum_{k=1}^{m} n_k \bigg[\nu(B(0, h_k)^c) + \frac{1}{\varepsilon_k} + 1\bigg].$$

Then supposing that $\varepsilon_1 \le 1$ and $\nu(B(0, h_k)^c) \le \frac{1}{\varepsilon_k}$ for $k = 1, \dots, m$, yields that

$$\mathrm{cost}(\widehat{S}) \le 3 \sum_{k=1}^{m} n_k \frac{1}{\varepsilon_k}. \tag{7}$$

Algorithms achieving the error rates of Theorem 1.1. Let us now quote the choice of parameters which establish the error rates of Theorem 1.1. In general, one chooses $\varepsilon_k = 2^{-k}$ and $h_k = g^{-1}(2^k)$ for $k \in \mathbb{Z}_+$. Moreover, in case (I), for sufficiently large τ, one picks

$$m = \lfloor \log_2 C_1 (\tau \log \tau)^{2/3} \rfloor \quad \text{and} \quad n_k = \bigg\lfloor C_2 \tau^{1/3} (\log \tau)^{-2/3} \frac{g^{-1}(2^k)}{g^{-1}(2^m)} \bigg\rfloor \quad \text{for } k = 1, \dots, m,$$

where C_1 and C_2 are appropriate constants that do not depend on τ. In case (II), one chooses

$$m = \lfloor \log_2 C_1 g^*(\tau) \rfloor \quad \text{and} \quad n_k = \bigg\lfloor C_2 \frac{g^*(\tau)^2}{\log g^*(\tau)} \frac{g^{-1}(2^k)}{g^{-1}(2^m)} \bigg\rfloor \quad \text{for } k = 1, \dots, m,$$

where again C_1 and C_2 are appropriate constants. We refer the reader to the proof of Theorem 1.1 for the error estimates of this choice.

3. Weak approximation. In this section, we provide the central estimate for the Wasserstein metric appearing in (6). For ease of notation, we denote by ε and h two positive parameters which correspond to $h^{(m)}$ and $\varepsilon^{(m)}$ above. We denote by Σ' a square matrix with $\Sigma'(\Sigma')^* = (\int_{B(0,h)} x_i x_j \nu(\mathrm{d}x))_{i,j\in\{1,\dots,d_X\}}$. Moreover, we let L' denote the process constituted by the compensated jumps of L of size larger than h, and let $B = (B_t)_{t\ge 0}$ be a d_X-dimensional Wiener process that is independent of W and L'. Then we consider the solution $\Upsilon = (\Upsilon_t)_{t\ge 0}$ of the integral equation

$$\Upsilon_t = y_0 + \int_0^t a\big(\Upsilon_{\iota(s-)}\big)\,\mathrm{d}\mathcal{X}_s,$$

where $\mathcal{X} = (\mathcal{X}_t)_{t\ge 0}$ is given as $\mathcal{X}_t = \Sigma W_t + \Sigma' B_t + L'_t + bt$ and $\iota(t) = \max(\mathbb{I} \cap [0,t])$, where $\mathbb{I}$ is, in analogy to above, the set of random times $(T'_j)_{j\in\mathbb{Z}_+}$ defined inductively via $T'_0 = 0$ and

$$T'_{j+1} = \inf\{t \in (T'_j,\infty) : |\Delta L_t| \ge h \text{ or } t = T'_j + \varepsilon\} \qquad \text{for } j \in \mathbb{Z}_+.$$

The process Υ is closely related to $\Upsilon^{(m)}$ from Section 2 and choosing $\varepsilon = \varepsilon_m$ and $h = h_m$, implies that $(\Upsilon_{\iota(t)})_{t\ge 0}$ and $\Upsilon^{(m)}$ are identically distributed.

We need to introduce two further crucial quantities: for $h > 0$, let $F(h) = \int_{B(0,h)} |x|^2 \nu(\mathrm{d}x)$ and $F_0(h) = \int_{B(0,h)^c} x\,\nu(\mathrm{d}x)$.

THEOREM 3.1. *Suppose that Assumption UE is valid. There exists a finite constant κ that depends only on K, d_X and ϑ such that for $\varepsilon \in (0,\frac12]$, $\varepsilon' \in [2\varepsilon, 1]$, and $h \in (0,\mathfrak{h}]$ with $\nu(B(0,h)^c) \le \frac{1}{\varepsilon}$, one has*

$$\mathcal{W}(Y, \Upsilon_{\iota(\cdot)})^2 \le \kappa\biggl[F(h)\varepsilon' + \frac{h^2}{\varepsilon'}\log\biggl(\frac{\varepsilon' F(h)}{h^2} \vee e\biggr)^2 + \varepsilon \log\frac{e}{\varepsilon}\biggr],$$

and, if $\Sigma = 0$, one has

$$\mathcal{W}(Y, \Upsilon_{\iota(\cdot)})^2 \le \kappa\biggl[F(h)\biggl(\varepsilon' + \varepsilon\log\frac{e}{\varepsilon}\biggr) + \frac{h^2}{\varepsilon'}\log\biggl(\frac{\varepsilon' F(h)}{h^2} \vee e\biggr)^2 + |b - F_0(h)|^2\varepsilon^2\biggr].$$

COROLLARY 3.2. *Under Assumption UE, there exists a constant $\kappa = \kappa(K, d_X, \vartheta)$ such that for all $\varepsilon \in (0,\frac14]$ and $h \in (0,\mathfrak{h}]$ with $\nu(B(0,h)^c) \vee \frac{F(h)}{h^2} \le \frac{1}{\varepsilon}$, one has*

$$\mathcal{W}(Y, \Upsilon_{\iota(\cdot)})^2 \le \kappa\biggl(h^2\frac{1}{\sqrt{\varepsilon}} + \varepsilon\biggr)\log\frac{e}{\varepsilon},$$

and, in the case where $\Sigma = 0$,

$$\mathcal{W}(Y, \Upsilon_{\iota(\cdot)})^2 \le \kappa\biggl(h^2\frac{1}{\sqrt{\varepsilon}}\log\frac{e}{\varepsilon} + |b - F_0(h)|^2\varepsilon^2\biggr).$$

PROOF. Choose $\varepsilon' = \sqrt{\varepsilon}\log 1/\varepsilon$ and observe that $\varepsilon' \ge 2\varepsilon$ since $\varepsilon \le \frac14$. Using that $\frac{F(h)}{h^2} \le g(h) \le \frac{1}{\varepsilon}$, it is straight forward to verify the estimate with Theorem 3.1. $\square$

3.1. *Strategy of the proof of Theorem 3.1 and main notation.* We represent X as

$$X_t = \Sigma W_t + L'_t + L''_t + bt,$$

where $L'' = (L''_t)_{t\geq 0} = L - L'$ is the process which comprises the compensated jumps of L of size smaller than h. Based on an additional parameter $\varepsilon' \in [2\varepsilon, 1]$, we couple L'' with ΣB. The introduction of the explicit coupling is deferred to Section 7. Let us roughly explain the idea behind the parameter ε'. In classical Euler schemes, the coefficients of the SDE are updated in either a deterministic or a random number of steps of a given (typical) length. Our approximation updates the coefficients at steps of order ε as the classical Euler method. However, in our case the Lévy process that comprises the small jumps is ignored for most of the time steps. It is only considered on steps of order of size ε'.

On the one hand, a large ε' reduces the accuracy of the approximation. On the other hand, the part of the small jumps has to be approximated by a Wiener process and the error inferred from the coupling decreases in ε'. This explains the increasing and decreasing terms in Theorem 3.1. Balancing ε' and ε then leads to Corollary 3.2.

We need some auxiliary processes. Analogously to $\mathbb{I}$ and ι, we let $\mathbb{J}$ denote the set of random times $(T_j)_{j\in\mathbb{Z}_+}$ defined inductively by $T_0 = 0$ and

$$T_{j+1} = \min\bigl(\mathbb{I} \cap (T_j + \varepsilon' - \varepsilon, \infty)\bigr)$$

so that the mesh-size of $\mathbb{J}$ is less than or equal to ε'. Moreover, we set $\eta(t) = \max(\mathbb{J} \cap [0, t])$.

Let us now introduce the first auxiliary processes. We set $X' = (X_t - L''_t)_{t\geq 0}$ and we consider the solution $\bar{Y}' = (\bar{Y}'_t)_{t\geq 0}$ to the integral equation

$$\bar{Y}'_t = y_0 + \int_0^t a\bigl(\bar{Y}'_{\iota(s-)}\bigr)\,\mathrm{d}X'_s + \int_0^t a\bigl(\bar{Y}'_{\eta(s-)}\bigr)\,\mathrm{d}L''_{\eta(s)} \tag{8}$$

and the process $\bar{Y} = (\bar{Y}_t)_{t\geq 0}$ given by

$$\bar{Y}_t = \bar{Y}'_t + a\bigl(\bar{Y}'_{\eta(t)}\bigr)\bigl(L''_t - L''_{\eta(t)}\bigr).$$

It coincides with $\bar{Y}'$ for all times in $\mathbb{J}$ and satisfies

$$\bar{Y}_t = y_0 + \int_0^t a\bigl(\bar{Y}'_{\iota(s-)}\bigr)\,\mathrm{d}X'_s + \int_0^t a\bigl(\bar{Y}_{\eta(s-)}\bigr)\,\mathrm{d}L''_s.$$

Next, we replace the term L'' by the Gaussian term $\Sigma' B$ in the above integral equations and obtain analogs of $\bar{Y}'$ and $\bar{Y}$ which are denoted by $\bar{\Upsilon}'$ and $\bar{\Upsilon}$. To be more precise, $\Upsilon' = (\bar{\Upsilon}'_t)_{t\geq 0}$ is the solution to the stochastic integral equation

$$\bar{\Upsilon}'_t = y_0 + \int_0^t a\bigl(\bar{\Upsilon}'_{\iota(s-)}\bigr)\,\mathrm{d}X'_s + \int_0^t a\bigl(\bar{\Upsilon}'_{\eta(s-)}\bigr)\Sigma'\,\mathrm{d}B_{\eta(s)},$$

and $\bar{\Upsilon} = (\bar{\Upsilon}_t)_{t\ge 0}$ is given via

$$\bar{\Upsilon}_t = \bar{\Upsilon}'_t + a(\bar{\Upsilon}'_{\eta(t)})\Sigma'(B_t - B_{\eta(t)}).$$

We now focus on the discrepancy of Y and $\Upsilon_{\iota(\cdot)}$. By the triangle inequality, one has

$$\|Y - \Upsilon_{\iota(\cdot)}\| \le \|Y - \bar{Y}\| + \|\bar{Y} - \bar{\Upsilon}\| + \|\bar{\Upsilon} - \Upsilon\| + \|\Upsilon - \Upsilon_{\iota(\cdot)}\|. \tag{9}$$

Moreover, the second term on the right satisfies

$$\|\bar{Y} - \bar{\Upsilon}\| \le \|\bar{Y}' - \bar{\Upsilon}'\| + \|\bar{Y} - \bar{Y}' - (\bar{\Upsilon} - \bar{\Upsilon}')\|. \tag{10}$$

In order to prove Theorem 3.1, we control the error terms individually. The first term on the right-hand side of (9) is considered in Proposition 4.1. The third and fourth term are treated in Propositions 5.1 and 5.2, respectively. The terms on the right-hand side of (10) are investigated in Propositions 7.1 and 7.2, respectively. Note that only the latter two expressions depend on the particular choice of the coupling of L'' and $\Sigma' B$. Once the above-mentioned propositions are proved, the statement of Theorem 3.1 follows immediately by combining these estimates and identifying the dominant terms.

4. Approximation of Y by $\bar{Y}$.

PROPOSITION 4.1. *There exists a constant $\kappa > 0$ depending on K only such that, for $\varepsilon \in (0, \frac{1}{2}]$, $\varepsilon' \in [2\varepsilon, 1]$ and $h > 0$ with $\nu(B(0,h)^c) \le \frac{1}{\varepsilon}$, one has*

$$\mathbb{E}\Big[\sup_{t\in[0,1]} |Y_t - \bar{Y}_t|^2\Big] \le \kappa[F(h)\varepsilon' + |b - F_0(h)|^2\varepsilon^2],$$

if $\Sigma = 0$, and

$$\mathbb{E}\Big[\sup_{t\in[0,1]} |Y_t - \bar{Y}_t|^2\Big] \le \kappa(\varepsilon + F(h)\varepsilon') \tag{11}$$

for general Σ.

PROOF. For $t \ge 0$, we consider $Z_t = Y_t - \bar{Y}_t$, $Z'_t = Y_t - \bar{Y}'_{\iota(t)}$, $Z''_t = Y_t - \bar{Y}_{\eta(t)}$ and $z(t) = \mathbb{E}[\sup_{s\in[0,t]} |Z_s|^2]$. The main task of the proof is to establish an estimate of the form

$$z(t) \le \alpha_1 \int_0^t z(s)\,\mathrm{d}s + \alpha_2$$

for appropriate values $\alpha_1, \alpha_2 > 0$. Since z is finite (see, for instance, [6]), then Gronwall's inequality implies as upper bound:

$$\mathbb{E}\Big[\sup_{s\in[0,1]} |Y_s - \bar{Y}_s|^2\Big] \le \alpha_2 \exp(\alpha_1).$$

We proceed in two steps.

1st step. Note that

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

$$+ \int_0^t \big(a(Y_{s-}) - a(\bar{Y}'_{\iota(s-)})\big) b\,\mathrm{d}s,$$

so that

$$|Z_t|^2 \le 2|M_t|^2 + 2\left|\int_0^t \big(a(Y_{s-}) - a(\bar{Y}_{\iota(s-)})\big) b\,\mathrm{d}s\right|^2. \tag{12}$$

For $t \in [0, 1]$, we conclude with the Cauchy–Schwarz inequality that the second term on the right-hand side is bounded by $2K^4 \int_0^t |Z'_{s-}|^2\,\mathrm{d}s$.

Certainly, (M_t) is a (local) martingale with respect to the canonical filtration, and we apply the Doob inequality together with Lemma A.1 to deduce that

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

$$+ \int_0^t \big|a(Y_{s-}) - a(\bar{Y}_{\eta(s-)})\big|^2\,\mathrm{d}\langle L''\rangle_s\Big].$$

Here and elsewhere, for a multivariate local L^2-martingale $S = (S_t)_{t\ge 0}$, we denote $\langle S\rangle = \sum_j \langle S^{(j)}\rangle$ and $\langle S^{(j)}\rangle$ denotes the predictable compensator of the classical bracket process of the jth coordinate $S^{(j)}$ of S. Note that $\mathrm{d}\langle \Sigma W + L'\rangle_t = (|\Sigma|^2 + \int_{B(0,h)^c} |x|^2 \nu(\mathrm{d}x))\,\mathrm{d}t \le 2K^2\,\mathrm{d}t$ and $\mathrm{d}\langle L''\rangle_t = F(h)\,\mathrm{d}t$. Consequently,

$$\mathbb{E}\Big[\sup_{s\in[0,t]} |M_s|^2\Big] \le 4\mathbb{E}\Big[2K^4 \int_0^t |Z'_s|^2\,\mathrm{d}s + K^2 F(h) \int_0^t |Z''_s|^2\,\mathrm{d}s\Big].$$

Hence, by (12) and Fubini's theorem, one has

$$\mathbb{E}\Big[\sup_{s\in[0,t]} |Z_s|^2\Big] \le \kappa_1 \int_0^t \big[z(s) + \mathbb{E}[|Z'_s|^2] + F(h)\mathbb{E}[|Z''_s|^2]\big]\,\mathrm{d}s$$

for a constant κ_1 that depends only on K. Since $Z'_t = Z_t + \bar{Y}_t - \bar{Y}'_{\iota(t)}$ and $Z''_t = Z_t + \bar{Y}_t - \bar{Y}_{\eta(t)}$, we get

$$z(t) \le \kappa_2 \int_0^t \big[z(s) + \mathbb{E}\big[|\bar{Y}_s - \bar{Y}'_{\iota(s)}|^2\big] + F(h)\mathbb{E}\big[|\bar{Y}_s - \bar{Y}_{\eta(s)}|^2\big]\big]\,\mathrm{d}s \tag{13}$$

for an appropriate constant $\kappa_2 = \kappa_2(K)$.

2nd step. In the second step we provide appropriate estimates for $\mathbb{E}[|\bar{Y}_t - \bar{Y}'_{\iota(t)}|^2]$ and $\mathbb{E}[|\bar{Y}_t - \bar{Y}_{\eta(t)}|^2]$. The processes W and L'' are independent of the random time

$\iota(t)$. Moreover, L' has no jumps in $(\iota(t), t)$, and we obtain

$$\begin{aligned}\bar{Y}_t - \bar{Y}'_{\iota(t)} &= \bar{Y}'_t - \bar{Y}'_{\iota(t)} + a(\bar{Y}_{\eta(t)})(L''_t - L''_{\eta(t)}) \\ &= a(\bar{Y}'_{\iota(t)})\big(\Sigma(W_t - W_{\iota(t)}) + (b - F_0(h))(t - \iota(t))\big) \\ &\quad + a(\bar{Y}_{\eta(t)})(L''_t - L''_{\eta(t)})\end{aligned}$$

so that

$$\begin{aligned}\mathbb{E}\big[|\bar{Y}_t - \bar{Y}'_{\iota(t)}|^2\big] \le 3K^2\big[&\mathbb{E}\big[(|\bar{Y}'_{\iota(t)} - y_0| + 1)^2\big](|\Sigma|^2\varepsilon + |b - F_0(h)|^2\varepsilon^2) \\ &+ \mathbb{E}\big[(|\bar{Y}_{\eta(t)} - y_0| + 1)^2\big]F(h)\varepsilon'\big].\end{aligned}$$

By Lemma A.2, there exists a constant $\kappa_3 = \kappa_3(K)$ such that

$$\mathbb{E}\big[|\bar{Y}_t - \bar{Y}'_{\iota(t)}|^2\big] \le \kappa_3[|\Sigma|^2\varepsilon + |b - F_0(h)|^2\varepsilon^2 + F(h)\varepsilon']. \tag{14}$$

Similarly, we estimate $\mathbb{E}[|\bar{Y}_t - \bar{Y}_{\eta(t)}|^2]$. Given $\eta(t)$, $(L'_{\eta(t)+u} - L'_{\eta(t)})_{u\in[0,(\varepsilon'-\varepsilon)\wedge(t-\eta(t))]}$ is distributed as the unconditioned Lévy process L' on the time interval $[0, (\varepsilon' - \varepsilon) \wedge (t - \eta(t))]$. Moreover, we have $\mathrm{d}L'_u = -F_0(h)\,\mathrm{d}u$ on $(\eta(t) + \varepsilon' - \varepsilon, t]$. Consequently,

$$\begin{aligned}\bar{Y}_t - \bar{Y}_{\eta(t)} &= \int_{\eta(t)}^t \mathbb{1}_{\{s-\eta(t)\le\varepsilon'-\varepsilon\}} a(\bar{Y}_{\iota(s-)})\,\mathrm{d}(\Sigma W_s + L'_s + bs) \\ &\quad + \int_{\eta(t)}^t \mathbb{1}_{\{s-\eta(t)>\varepsilon'-\varepsilon\}} a(\bar{Y}_{\iota(s-)})\,\mathrm{d}\big(\Sigma W_s + (b - F_0(h))s\big) \\ &\quad + a(\bar{Y}_{\eta(t)})(L''_t - L''_{\eta(t)}),\end{aligned}$$

and analogously as we obtained (14) we get now that

$$\mathbb{E}\big[|\bar{Y}_t - \bar{Y}_{\eta(t)}|^2\big] \le \kappa_4[\varepsilon' + |b - F_0(h)|^2\varepsilon^2]$$

for a constant $\kappa_4 = \kappa_4(K)$. Next, note that, by the Cauchy–Schwarz inequality, $|F_0(h)|^2 \le \int_{B(0,h)^c} |x|^2\nu(\mathrm{d}x) \cdot \nu(B(0,h)^c) \le \frac{K^2}{\varepsilon}$ so that we arrive at

$$\mathbb{E}\big[|\bar{Y}_t - \bar{Y}_{\eta(t)}|^2\big] \le \kappa_5\varepsilon'.$$

Combining this estimate with (13) and (14), we obtain

$$z(t) \le \kappa_2 \int_0^t z(s)\,\mathrm{d}s + \kappa_6[|\Sigma|^2\varepsilon + F(h)\varepsilon' + |b - F_0(h)|^2\varepsilon^2].$$

In the case where $\Sigma = 0$, the statement of the proposition follows immediately via Gronwall's inequality. For general Σ, we obtain the result by recalling that $|F_0(h)|^2 \le \frac{K^2}{\varepsilon}$. $\square$

5. Approximation of $\bar{\Upsilon}$ by $\Upsilon_{\iota(\cdot)}$.

PROPOSITION 5.1. *Under the assumptions of Proposition 4.1, one has*

$$\mathbb{E}[\|\bar{\Upsilon} - \Upsilon\|^2] \le \kappa \varepsilon' F(h)$$

for a constant κ depending only on K.

PROOF. The proposition can be proved as Proposition 4.1. Therefore, we only provide a sketch of the proof. The arguments from the first step give, for $t \in [0, 1]$,

$$z(t) \le \kappa_1 \int_0^t \big[z(s) + \mathbb{E}\big[\big|\bar{\Upsilon}_{\iota(s)} - \bar{\Upsilon}'_{\iota(s)}\big|^2\big] + F(h)\mathbb{E}\big[\big|\bar{\Upsilon}_{\iota(s)} - \bar{\Upsilon}_{\eta(s)}\big|^2\big]\big]\,\mathrm{d}s,$$

where $z(t) = \mathbb{E}[\sup_{s\in[0,t]} |\Upsilon_s - \bar{\Upsilon}_s|^2]$ and $\kappa_1 = \kappa_1(K)$ is an appropriate constant.

Moreover, based on Lemma A.2 the second step leads to

$$\mathbb{E}\big[\big|\bar{\Upsilon}_{\iota(t)} - \bar{\Upsilon}'_{\iota(t)}\big|^2\big] \le \kappa_2 \varepsilon' F(h) \quad \text{and} \quad \mathbb{E}\big[\big|\bar{\Upsilon}_{\iota(t)} - \bar{\Upsilon}_{\eta(t)}\big|^2\big] \le \kappa_3 \varepsilon'$$

for appropriate constants $\kappa_2 = \kappa_2(K)$ and $\kappa_3 = \kappa_3(K)$. Then Gronwall's lemma implies again the statement of the proposition. □

PROPOSITION 5.2. *Under the assumptions of Proposition 4.1, there exists a constant κ depending only on K and d_X such that, if $\Sigma = 0$,*

$$\mathbb{E}\Big[\sup_{t\in[0,1]} \big|\Upsilon_t - \Upsilon_{\iota(t)}\big|^2\Big] \le \kappa \Big[F(h)\varepsilon \log \frac{e}{\varepsilon} + |b - F_0(h)|^2 \varepsilon^2\Big]$$

and, in the general case,

$$\mathbb{E}\Big[\sup_{t\in[0,1]} \big|\Upsilon_t - \Upsilon_{\iota(t)}\big|^2\Big] \le \kappa \varepsilon \log \frac{e}{\varepsilon}.$$

PROOF. Recall that by definition

$$\Upsilon_t - \Upsilon_{\iota(t)} = \int_{\iota(t)}^t a\big(\Upsilon_{\iota(s-)}\big)\,\mathrm{d}\mathcal{X}_s$$

so that

$$\big|\Upsilon_t - \Upsilon_{\iota(t)}\big|^2 \le K^2\big(\big|\Upsilon_{\iota(t)} - y_0\big| + 1\big)^2 \big|\mathcal{X}_t - \mathcal{X}_{\iota(t)}\big|^2.$$

Next, we apply Lemma A.4. For $j \in \mathbb{Z}_+$, we choose

$$U_j = |\Upsilon_{T'_j \wedge 1} - y_0|^2 \quad \text{and} \quad V_j = \sup_{s\in[T'_j, T'_{j+1}\wedge 1)} |\mathcal{X}_t - \mathcal{X}_{\iota(t)}|^2$$

with the convention that the supremum of the empty set is zero. Then

$$\mathbb{E}\Big[\sup_{t\in[0,1]}|\Upsilon_t-\Upsilon_{\iota(t)}|^2\Big]\le\mathbb{E}\Big[\sup_{j\in\mathbb{Z}_+}U_j\Big]\cdot\mathbb{E}\Big[\sup_{j\in\mathbb{Z}_+}V_j\Big]$$

$$\le\mathbb{E}\Big[\sup_{t\in[0,1]}(|\Upsilon_t-y_0|+1)^2\Big]\cdot\mathbb{E}\Big[\sup_{\substack{0\le s<t\le 1\\ t-s\le\varepsilon}}|\mathcal{X}_t-\mathcal{X}_s|^2\Big].$$

By Proposition 5.1 and Lemma A.2, $\mathbb{E}[\sup_{t\in[0,1]}(|\Upsilon_t-y_0|+1)^2]$ is bounded by a constant that depends only on K.

Consider $\varphi:[0,1]\to[0,\infty)$, $\delta\mapsto\sqrt{\delta\log(e/\delta)}$. By Lévy's modulus of continuity,

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

is finite almost surely, so that Fernique's theorem implies that $\mathbb{E}[\|W\|_\varphi^2]$ is finite too. Consequently,

$$\begin{aligned}&\mathbb{E}\Big[\sup_{s\in[0,t]}|\mathcal{X}_s-\mathcal{X}_{\iota(s)}|^2\Big]\\ &\quad\le 3\Big[\big(|\Sigma|^2+F(h)\big)\mathbb{E}[\|W\|_\varphi^2]\varepsilon\log\frac{e}{\varepsilon}+|b-F_0(h)|^2\varepsilon^2\Big].\end{aligned}\tag{15}$$

The result follows immediately by using that $|F_0(h)|^2\le\frac{K^2}{\varepsilon}$ and ruling out the asymptotically negligible terms. $\square$

6. Gaussian approximation via Komlós, Major and Tusnády. In this section, we prove the following theorem.

THEOREM 6.1. *Let $h>0$ and $L=(L_t)_{t\ge 0}$ be a d-dimensional $(\nu,0)$-Lévy martingale whose Lévy measure ν is supported on $B(0,h)$. Moreover, we suppose that for $\vartheta\ge 1$, one has*

$$\int\langle y',x\rangle^2\nu(\mathrm{d}x)\le\vartheta\int\langle y,x\rangle^2\nu(\mathrm{d}x)$$

for any $y,y'\in\mathbb{R}^d$ with $|y|=|y'|$, and set $\sigma^2=\int|x|^2\nu(\mathrm{d}x)$.

There exist constants $c_1,c_2>0$ depending only on d such that the following statement is true. For every $T\ge 0$, one can couple the process $(L_t)_{t\in[0,T]}$ with a Wiener process $(B_t)_{t\in[0,T]}$ such that

$$\mathbb{E}\exp\Big\{\frac{c_1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}|L_t-\Sigma B_t|\Big\}\le\exp\Big\{c_2\log\Big(\frac{\sigma^2T}{h^2}\vee e\Big)\Big\},$$

where Σ is a square matrix with $\Sigma\Sigma^*=\mathrm{cov}_{L_1}$ and $\sigma^2=\int|x|^2\nu(\mathrm{d}x)$.

The proof of the theorem is based on Zaitsev's generalization [22] of the Komlós–Major–Tusnády coupling. In this context, a key quantity is the *Zaitsev parameter*: Let Z be a d-dimensional random variable with finite exponential moments in a neighborhood of zero and set

$$\Lambda(\theta) = \log \mathbb{E} \exp\{\langle \theta, Z\rangle\}$$

for all $\theta \in \mathbb{C}$ with integrable expectation. Then the parameter is defined as

$$\tau(Z) = \inf\{\tau > 0 : |\partial_w \partial_v^2 \Lambda(\theta)| \le \tau \langle \mathrm{cov}_Z\, v, v\rangle \text{ for all } \theta \in \mathbb{C}^d, v, w \in \mathbb{R}^d \text{ with } |\theta| \le \tau^{-1} \text{ and } |w| = |v| = 1\}.$$

In the latter set, we implicitly only consider τ's for which Λ is finite on a neighborhood of $\{x \in \mathbb{C}^d : |x| \le 1/\tau\}$. Moreover, cov_Z denotes the covariance matrix of Z.

PROOF OF THEOREM 6.1. *1st step:* First, consider a d-dimensional infinitely divisible random variable Z with

$$\Lambda(\theta) := \log \mathbb{E} e^{\langle \theta, Z\rangle} = \int \bigl(e^{\langle \theta, x\rangle} - \langle \theta, x\rangle - 1\bigr)\nu'(\mathrm{d}x),$$

where the Lévy measure ν' is supported on the ball $B(0, h')$ for a fixed $h' > 0$. Then

$$\partial_w \partial_v^2 \Lambda(\theta) = \int_{B(0,h')} \langle w, x\rangle \langle v, x\rangle^2 e^{\langle \theta, x\rangle} \nu(\mathrm{d}x)$$

and

$$\langle \mathrm{cov}_Z\, v, v\rangle = \mathrm{var}\langle v, Z\rangle = \partial_v^2 \Lambda_Z(0) = \int_{B(0,h')} \langle v, x\rangle^2 \nu(\mathrm{d}x).$$

We choose $\zeta > 0$ with $e^{\zeta} = 1/\zeta$, and observe that for any $\theta \in \mathbb{C}^d$, $v, w \in \mathbb{R}^d$ with $|\theta| \le \zeta/h'$ and $|w| = |v| = 1$,

$$|\partial_w \partial_v^2 \Lambda(\theta)| \le h' e^{|\theta| h'} \langle \mathrm{cov}_Z\, v, v\rangle \le \frac{h'}{\zeta} \langle \mathrm{cov}_Z\, v, v\rangle.$$

Hence,

$$\tau(Z) \le \frac{h'}{\zeta}.$$

2nd step: In the next step, we apply Zaitsev's coupling to piecewise constant interpolations of (L_t). Fix $m \in \mathbb{N}$ and consider $L^{(m)} = (L_t^{(m)})_{t\in[0,T]}$ given via

$$L_t^{(m)} = L_{\lfloor 2^m t/T\rfloor 2^{-m} T}.$$

Moreover, we consider a d-dimensional Wiener process $B = (B_t)_{t\ge 0}$ and its piecewise constant interpolation $\Sigma B^{(m)}$ given by $B^{(m)} = (B_{\lfloor 2^m t/T\rfloor 2^{-m} T})_{t\in[0,T]}$.

Since cov_{L_1} is self-adjoint, we find a representation $\mathrm{cov}_{L_t} = tUDU^*$ with D diagonal and U orthogonal. Hence, for $A_t := (tD)^{-1/2}U^*$ we get $\mathrm{cov}_{A_tL_t} = I_d$. We denote by λ_1 the leading and by λ_2 the minimal eigenvalue of D (or cov_{L_1}). Then A_tL_t is again infinitely divisible and the corresponding Lévy measure is supported on $B(0, h/\sqrt{\lambda_2 t})$. By part one, we conclude that

$$\tau(A_tL_t) \le \frac{h}{\zeta\sqrt{\lambda_2 t}}.$$

Now the discontinuities of $A_{2^{-m}}L^{(m)}$ are i.i.d. with unit covariance and Zaitsev parameter less than or equal to $\frac{h2^{m/2}}{\zeta\sqrt{T\lambda_2}}$. By [22], Theorem 1.3, one can couple the processes L and ΣB on an appropriate probability space such that

$$\mathbb{E}\exp\left\{\kappa_1\frac{\sqrt{T\lambda_2}}{2^{m/2}h}\sup_{t\in[0,T]}\left|A_{2^{-m}}L_t^{(m)} - A_{2^{-m}}\Sigma B_t^{(m)}\right|\right\} \le \exp\left\{\kappa_2\log\left(\frac{\zeta^2T\lambda_2}{h^2}\vee e\right)\right\},$$

where $\kappa_1, \kappa_2 > 0$ are constants only depending on the dimension d. The smallest eigenvalue of $A_{2^{-m}}$ is $2^{m/2}(T\lambda_1)^{-1/2}$ and, by assumption, $\lambda_1 \le \vartheta\lambda_2$. Since $\lambda_2 \le \sigma^2$, we get

$$\mathbb{E}\exp\left\{\kappa_1\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}\left|L_t^{(m)} - \Sigma B_t^{(m)}\right|\right\} \le \exp\left\{\kappa_2\log\left(\frac{\zeta^2T\sigma^2}{h^2}\vee e\right)\right\}.$$

3rd step: The general result follows by approximation. First, note that $\sup_{t\in[0,T]}|L_t - L_t^{(m)}|$ converges as $m\to\infty$ to $\sup_{t\in[0,T]}|L_t - L_{t-}|$ so that by dominated convergence

$$\begin{aligned}\lim_{m\to\infty}\mathbb{E}\exp\left\{\kappa_1\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}\left|L_t - L_t^{(m)}\right|\right\}\\ = \mathbb{E}\exp\left\{\kappa_1\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}|L_t - L_{t-}|\right\} \le e^{\kappa_1}.\end{aligned}$$

Analogously, $\lim_{m\to\infty}\mathbb{E}\exp\{\kappa_1\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}|\Sigma B_t - \Sigma B_t^{(m)}|\} = 1$. Next, we choose $\kappa_3 \ge 1$ with $e^{\kappa_1} + 1 \le e^{\kappa_2+\kappa_3}$ and we fix $m\in\mathbb{N}$ such that

$$\begin{aligned}&\mathbb{E}\exp\left\{\frac{\kappa_1}{3}\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}|L_t - \Sigma B_t|\right\} + \mathbb{E}\exp\left\{\kappa_1\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}\left|\Sigma B_t - \Sigma B_t^{(m)}\right|\right\}\\ &\quad\le e^{\kappa_2+\kappa_3}.\end{aligned}$$

We apply the coupling introduced in step 2 and estimate

$$\begin{aligned}\mathbb{E}\exp\left\{\frac{\kappa_1}{3}\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}|L_t - \Sigma B_t|\right\} &\le \mathbb{E}\exp\left\{\kappa_1\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}\left|L_t - L_t^{(m)}\right|\right\}\\ &\quad + \mathbb{E}\exp\left\{\kappa_1\frac{1}{\sqrt{\vartheta}h}\sup_{t\in[0,T]}\left|L_t^{(m)} - \Sigma B_t^{(m)}\right|\right\}\end{aligned}$$

$$+\mathbb{E}\exp\left\{\kappa_1\frac{1}{\sqrt{\vartheta h}}\sup_{t\in[0,T]}\left|\Sigma B_t^{(m)}-\Sigma B_t\right|\right\}$$

$$\leq \exp\left\{\kappa_2\log\left(\frac{T\sigma^2}{h^2}\vee e\right)\right\}+e^{\kappa_2+\kappa_3}.$$

Straightforwardly, one obtains the assertion of the theorem for $c_1=\kappa_1/3$ and $c_2=\kappa_2+2\kappa_3$. $\square$

COROLLARY 6.2. *The coupling introduced in Theorem 6.1 satisfies*

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|L_t-\Sigma B_t|^2\Big]^{1/2}\leq\frac{\sqrt{\vartheta}h}{c_1}\left(c_2\log\left(\frac{\sigma^2T}{h^2}\vee e\right)+2\right),$$

where c_1 and c_2 are as in the theorem.

PROOF. We set $Z=\sup_{t\in[0,T]}|L_t-\Sigma B_t|$ and $t_0=\frac{\sqrt{\vartheta}h}{c_1}c_2\log(\frac{\sigma^2T}{h^2}\vee e)$, and use that

$$\mathbb{E}[Z^2]=2\int_0^\infty t\mathbb{P}(Z\geq t)\,\mathrm{d}t\leq t_0^2+2\int_{t_0}^\infty t\mathbb{P}(Z\geq t)\,\mathrm{d}t. \tag{16}$$

By the Markov inequality and Theorem 6.1, one has for $s\geq 0$

$$\mathbb{P}(Z\geq s+t_0)\leq\frac{\mathbb{E}[\exp\{c_1/(\sqrt{\vartheta}h)Z\}]}{\exp\{c_1/(\sqrt{\vartheta}h)(s+t_0)\}}\leq\exp\left\{-\frac{c_1}{\sqrt{\vartheta}h}s\right\}.$$

We set $\alpha=\sqrt{\vartheta}h/c_1$, and deduce together with (16) that

$$\mathbb{E}[Z^2]\leq t_0^2+2\int_0^\infty(s+t_0)\exp\left\{-\frac{1}{\alpha}s\right\}\mathrm{d}s=t_0^2+2t_0\alpha+2\alpha^2\leq(t_0+2\alpha)^2. \quad\square$$

7. Coupling the Gaussian approximation. We are now in the position to couple the processes L'' and $\Sigma'B$ introduced in Section 3.1. We adopt again the notation of Section 3.1.

To introduce the coupling, we need to assume that Assumption UE is valid, and that $\varepsilon\in(0,\frac{1}{2}]$, $\varepsilon'\in[2\varepsilon,1]$ and $h\in(0,\mathfrak{h}]$ are such that $\nu(B(0,h)^c)\leq\frac{1}{\varepsilon}$. Recall that L'' is independent of W and L'. In particular, it is independent of the times in $\mathbb{J}$, and given W and L' we couple the Wiener process B with L'' on each interval $[T_i,T_{i+1}]$ according to the coupling provided by Theorem 6.1.

More explicitly, the coupling is established in such a way that, given $\mathbb{J}$, each pair of processes $(B_{t+T_j}-B_{T_j})_{t\in[0,T_{j+1}-T_j]}$ and $(L''_{t+T_j}-L''_{T_j})_{t\in[0,T_{j+1}-T_j]}$ is independent of W, L' and the other pairings, and satisfies

$$\begin{aligned}&\mathbb{E}\left[\exp\left\{\frac{c_1}{\sqrt{\vartheta}h}\sup_{t\in[T_j,T_{j+1}]}|L''_t-L''_{T_j}-(\Sigma'B_t-\Sigma'B_{T_j})|\right\}\middle|\mathbb{J}\right]\\&\quad\leq\exp\left\{c_2\log\left(\frac{F(h)(T_{j+1}-T_j)}{h^2}\vee e\right)\right\}\end{aligned} \tag{17}$$

for positive constants c_1 and c_2 depending only on d_X, see Theorem 6.1. In particular, by Corollary 6.2, one has

$$\mathbb{E}\Big[\sup_{t\in[T_j,T_{j+1}]}|L''_t - L''_{T_j} - (\Sigma' B_t - \Sigma' B_{T_j})|^2|\mathbb{J}\Big]^{1/2} \le c_3 h \log\Big(\frac{F(h)(T_{j+1}-T_j)}{h^2} \vee e\Big) \tag{18}$$

for a constant $c_3 = c_3(d_X, \vartheta)$.

PROPOSITION 7.1. *Under Assumption UE, there exists a constant κ depending only on K, ϑ and d_X such that for any $\varepsilon \in (0, \frac{1}{2}]$, $\varepsilon' \in [2\varepsilon, 1]$ and $h \in (0, \mathfrak{h}]$ with $\nu(B(0,h)^c) \le \frac{1}{\varepsilon}$, one has*

$$\mathbb{E}\Big[\sup_{[0,1]}|\bar Y'_t - \bar\Upsilon'_t|^2\Big] \le \kappa \frac{1}{\varepsilon'} h^2 \log\Big(\frac{\varepsilon' F(h)}{h^2} \vee e\Big)^2.$$

PROOF. For ease of notation, we write

$$A_t = L''_{\eta(t)} \quad \text{and} \quad A'_t = \Sigma' B_{\eta(t)}.$$

By construction, (A_t) and (A'_t) are martingales with respect to the filtration $(\mathcal{F}_t)$ induced by the processes (W_t), (L'_t), (A_t) and (A'_t). Let $Z_t = \bar Y'_t - \bar\Upsilon'_t$, $Z'_t = \bar Y'_{\iota(t)} - \bar\Upsilon'_{\iota(t)}$, $Z''_t = \bar Y'_{\eta(t)} - \bar\Upsilon'_{\eta(t)}$ and $z(t) = \mathbb{E}[\sup_{s\in[0,t]}|Z_s|^2]$. The proof is similar to the proof of Proposition 4.1.

Again, we write

$$Z_t = \underbrace{\int_0^t \big(a(\bar Y'_{\iota(s-)}) - a(\bar\Upsilon'_{\iota(s-)})\big)\,\mathrm{d}(\Sigma W_s + L'_s) + \int_0^t a(\bar Y'_{\eta(s-)})\,\mathrm{d}A_s - \int_0^t a(\bar\Upsilon'_{\eta(s-)})\,\mathrm{d}A'_s}_{=:M_t \text{ (localmartingale)}} + \int_0^t \big(a(\bar Y'_{\iota(s)}) - a(\bar\Upsilon'_{\iota(s)})\big) b\,\mathrm{d}s. \tag{19}$$

Denoting $M' = \Sigma W + L'$, we get

$$\begin{aligned}\mathrm{d}M_t &= \big(a(\bar Y'_{\iota(t-)}) - a(\bar\Upsilon'_{\iota(t-)})\big)\,\mathrm{d}M'_t + a(\bar Y'_{\eta(t-)})\,\mathrm{d}(A_t - A'_t)\\ &\quad + \big(a(\bar Y'_{\eta(t-)}) - a(\bar\Upsilon'_{\eta(t-)})\big)\,\mathrm{d}A'_t\end{aligned}$$

and, by Doob's inequality and Lemma A.1, we have

$$\begin{aligned}\mathbb{E}\Big[\sup_{s\in[0,t]}|M_s|^2\Big] &\le \kappa_1\Big[\mathbb{E}\Big[\int_0^t |Z'_{s-}|^2\,\mathrm{d}\langle M'\rangle_s\Big] + \mathbb{E}\Big[\int_0^t |Z''_{s-}|^2\,\mathrm{d}\langle A'\rangle_s\Big]\\ &\qquad + \mathbb{E}\Big[\int_0^t (|\bar Y'_{\eta(s-)}|+1)^2\,\mathrm{d}\langle A - A'\rangle_s\Big]\Big].\end{aligned} \tag{20}$$

Each bracket $\langle\cdot\rangle$ in the latter formula can be chosen with respect to a (possibly different) filtration such that the integrand is predictable and the integrator is a local L^2-martingale. As noticed before, with respect to the canonical filtration $(\mathcal{F}_t)$ one has $\mathrm{d}\langle M'\rangle_t = (|\Sigma|^2 + \int_{B(0,h)^c} |x|^2 \nu(\mathrm{d}x))\,\mathrm{d}t \le 2K^2\,\mathrm{d}t$. Moreover, we have with respect to the enlarged filtration $(\mathcal{F}_t \vee \sigma(\mathbb{J}))_{t\ge 0}$,

$$\langle A'\rangle_t = \sum_{\{j\in\mathbb{N}:T_j\le t\}} (T_j - T_{j-1})F(h) = \max(\mathbb{J}\cap[0,t])\cdot F(h),$$

and, by (18), for $j\in\mathbb{N}$,

$$\Delta\langle A - A'\rangle_{T_j} = \mathbb{E}[|L''_{T_j} - L''_{T_{j-1}} - (\Sigma' B_{T_j} - \Sigma' B_{T_{j-1}})|^2|\mathbb{J}] \le c_3^2\xi^2,$$

where $\xi := h\log(\frac{\varepsilon' F(h)}{h^2}\vee e)$. Note that two discontinuities of $\langle A - A'\rangle$ are at least $\varepsilon'/2$ units apart and the integrands of the last two integrals in (20) are constant on $(T_{j-1}, T_j]$ so that altogether

$$\mathbb{E}\Big[\sup_{s\in[0,t]} |M_s|^2\Big] \le \kappa_1\Big[2K^2\mathbb{E}\Big[\int_0^t |Z'_s|^2\,\mathrm{d}s\Big] + F(h)\mathbb{E}\Big[\int_0^t |Z''_s|^2\,\mathrm{d}s\Big] + c_3^2\xi^2\frac{2}{\varepsilon'}\mathbb{E}\Big[\int_0^t \big(|\bar{Y}'_{\eta(s-)}|+1\big)^2\,\mathrm{d}s\Big]\Big].$$

With Lemma A.2 and Fubini's theorem, we arrive at

$$\mathbb{E}\Big[\sup_{s\in[0,t]} |M_s|^2\Big] \le \kappa_2\Big[\int_0^t z(s)\,\mathrm{d}s + \xi^2\frac{1}{\varepsilon'}\Big].$$

Moreover, by Jensen's inequality, one has

$$\mathbb{E}\Big[\sup_{s\in[0,t]}\Big|\int_0^s \big(a(\bar{Y}'_{\iota(u-)}) - a(\bar{\Upsilon}'_{\iota(u-)})\big)b\,\mathrm{d}u\Big|^2\Big] \le K^4\int_0^t \mathbb{E}[|Z'_{s-}|^2]\,\mathrm{d}s.$$

Combining the latter two estimates with (19) and applying Gronwall's inequality yields the statement of the proposition. $\square$

PROPOSITION 7.2. *There exists a constant κ depending only on K and d_X such that*

$$\mathbb{E}[\|\bar{Y} - \bar{Y}' - (\bar{\Upsilon} - \bar{\Upsilon}')\|^2]^{1/2} \le \kappa\Big[h\Big[\log\Big(1+\frac{2}{\varepsilon'}\Big) + \log\Big(\frac{F(h)\varepsilon'}{h^2}\vee e\Big)\Big] + \sqrt{F(h)\varepsilon'\log\frac{e}{\varepsilon'}}\,\mathbb{E}[\|\bar{Y}' - \bar{\Upsilon}'\|^2]^{1/2}\Big].$$

PROOF. Note that

$$\begin{aligned}\bar{Y}_t - \bar{Y}'_t - (\bar{\Upsilon}_t - \bar{\Upsilon}'_t) &= a(\bar{Y}'_{\eta(t)})(L''_t - L''_{\eta(t)}) - a(\bar{\Upsilon}'_{\eta(t)})(\Sigma' B_t - \Sigma' B_{\eta(t)})\\ &= a(\bar{Y}'_{\eta(t)})(L''_t - L''_{\eta(t)} - (\Sigma' B_t - \Sigma' B_{\eta(t)}))\\ &\quad + \big(a(\bar{Y}'_{\eta(t)}) - a(\bar{\Upsilon}'_{\eta(t)})\big)(\Sigma' B_t - \Sigma' B_{\eta(t)}).\end{aligned}$$

Similar as in the proof of Proposition 5.2, we apply Lemma A.4 to deduce that

$$\begin{aligned}
&\mathbb{E}[\|\bar{Y}-\bar{Y}'-(\bar{\Upsilon}-\bar{\Upsilon}')\|^2]^{1/2} \\
&\quad \le K\mathbb{E}[(\|\bar{Y}'\|+1)^2]^{1/2}\mathbb{E}\Big[\sup_{t\in[0,1]}\big|L''_t-L''_{\eta(t)}-\big(\Sigma'B_t-\Sigma'B_{\eta(t)}\big)\big|^2\Big]^{1/2} \qquad (21)\\
&\qquad + K\mathbb{E}[\|\bar{Y}'-\bar{\Upsilon}'\|^2]^{1/2}\mathbb{E}\Big[\sup_{t\in[0,1]}\big|\Sigma'B_t-\Sigma'B_{\eta(t)}\big|^2\Big]^{1/2}.
\end{aligned}$$

Next, we estimate $\mathbb{E}[\sup_{t\in[0,1]}|L''_t-L''_{\eta(t)}-(\Sigma'B_t-\Sigma'B_{\eta(t)})|^2]$. Recall that conditional on $\mathbb{J}$, each pairing of $(L''_{t+T_j}-L''_{T_j})_{t\in[0,T_{j+1}-T_j]}$ and $(B_{t+T_j}-B_{T_j})_{t\in[0,T_{j+1}-T_j]}$ is coupled according to Theorem 6.1, and individual pairs are independent of each other.

Let us first assume that the times in $\mathbb{J}$ are deterministic with mesh smaller or equal to ε'. We denote by n the number of entries of $\mathbb{J}$ which fall into $[0,1]$, and we denote, for $j=1,\dots,n$, $\Delta_j=\sup_{t\in[T_{j-1},T_j]}|L''_t-L''_{T_{j-1}}-(\Sigma'B_t-\Sigma'B_{T_{j-1}})|$. By (17) and the Markov inequality, one has, for $u\ge 0$,

$$\mathbb{P}\Big(\sup_{j=1,\dots,n}\Delta_j\ge u\Big)\le\sum_{j=1}^n\mathbb{P}(\Delta_j\ge u)\le n\exp\Big\{c_2\log\Big(\frac{F(h)\varepsilon'}{h^2}\vee e\Big)-\frac{c_1}{\sqrt{\vartheta}h}u\Big\}.$$

Let now $\alpha=\frac{c_1}{\sqrt{\vartheta}h}$, $\beta=\frac{F(h)}{h^2}$ and $u_0=\frac{1}{\alpha}(\log n+c_2\log(\beta\varepsilon'\vee e))$. Then for $u\ge 0$

$$\mathbb{P}\Big(\sup_{j=1,\dots,n}\Delta_j\ge u\Big)\le e^{-\alpha(u-u_0)}$$

so that

$$\begin{aligned}
\mathbb{E}\Big[\sup_{j=1,\dots,n}\Delta_j^2\Big]&=2\int_0^\infty u\mathbb{P}\Big(\sup_{j=1,\dots,n}\Delta_j\ge u\Big)\,\mathrm{d}u\\
&\le u_0^2+2\int_{u_0}^\infty e^{-\alpha(u-u_0)}\,\mathrm{d}u=u_0^2+2\frac{1}{\alpha}u_0+2\frac{1}{\alpha^2}\le\Big(u_0+\frac{2}{\alpha}\Big)^2.
\end{aligned}$$

Note that the upper bound depends only on the number of entries in $\mathbb{J}\cap[0,1]$, and, since $\#(\mathbb{J}\cap[0,1])$ is uniformly bounded by $\frac{2}{\varepsilon'}+1$, we thus get in the general random setting that

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,1]}\big|L''_t-L''_{\eta(t)}-\big(\Sigma'B_t-\Sigma'B_{\eta(t)}\big)\big|^2\Big]^{1/2}\\
&\qquad\le\frac{\sqrt{\vartheta}h}{c_1}\Big[\log\Big(1+\frac{2}{\varepsilon'}\Big)+c_2\log\Big(\frac{F(h)\varepsilon'}{h^2}\vee e\Big)+2\Big].
\end{aligned}$$

Together with Lemma A.2, this gives the appropriate upper bound for the first summand in (21).

By the argument preceding (15), one has

$$\mathbb{E}\Big[\sup_{t\in[0,1]}\big|\Sigma' B_t-\Sigma' B_{\eta(t)}\big|^2\Big]^{1/2}\le \kappa_1|\Sigma'|\sqrt{\varepsilon'\log\frac{e}{\varepsilon'}}=\kappa_1\sqrt{F(h)\varepsilon'\log\frac{e}{\varepsilon'}},$$

where κ_1 is a constant that depends only on d_X. This estimate is used for the second summand in (21) and putting everything together yields the statement. □

8. Proof of the main results.

Proof of Theorem 1.1. We consider a multilevel Monte Carlo algorithm $\widehat{S}\in\mathcal{A}$ partially specified by $\varepsilon_k:=2^{-k}$ and $h_k:=g^{-1}(2^k)$ for $k\in\mathbb{Z}_+$. The maximal index $m\in\mathbb{N}$ and the number of iterations $n_1,\dots,n_m\in\mathbb{N}$ are fixed explicitly below in such a way that $h_m\le\mathfrak{h}$ and $m\ge 2$. Recall that

$$\operatorname{mse}(\widehat{S})\le \mathcal{W}\big(Y,\Upsilon^{(m)}\big)^2+\sum_{k=2}^m\frac{1}{n_k}\mathbb{E}\big[\|\Upsilon^{(k)}-\Upsilon^{(k-1)}\|^2\big]+\frac{1}{n_1}\mathbb{E}\big[\|\Upsilon^{(1)}-y_0\|^2\big];$$

see (6). We control the Wasserstein metric via Corollary 3.2. Moreover, we deduce from [6], Theorem 2, that there exists a constant κ_0 that depends only on K and d_X such that, for $k=2,\dots,m$,

$$\mathbb{E}\big[\|\Upsilon^{(k)}-\Upsilon^{(k-1)}\|^2\big]\le\kappa_0\big(\varepsilon_{k-1}\log(e/\varepsilon_{k-1})+F(h_{k-1})\big)$$

and

$$\mathbb{E}\big[\|\Upsilon^{(1)}-y_0\|^2\big]\le\kappa_0\big(\varepsilon_0\log(e/\varepsilon_0)+F(h_0)\big).$$

Consequently, one has

$$\operatorname{mse}(\widehat{S})\le\kappa_1\bigg[\bigg(h_m^2\frac{1}{\sqrt{\varepsilon_m}}+\varepsilon_m\bigg)\log\frac{e}{\varepsilon_m}+\sum_{k=0}^{m-1}\frac{1}{n_{k+1}}\bigg[F(h_k)+\varepsilon_k\log\frac{e}{\varepsilon_k}\bigg]\bigg]\tag{22}$$

in the general case, and

$$\begin{aligned}\operatorname{mse}(\widehat{S})\le\kappa_2\bigg[&h_m^2\frac{1}{\sqrt{\varepsilon_m}}\log\frac{e}{\varepsilon_m}+|b-F_0(h)|^2\varepsilon_m^2\\&+\sum_{k=0}^{m-1}\frac{1}{n_{k+1}}\bigg[F(h_k)+\varepsilon_k\log\frac{e}{\varepsilon_k}\bigg]\bigg]\end{aligned}\tag{23}$$

in the case where $\Sigma=0$. Note that $F(h_k)\le h_k^2 g(h_k)=g^{-1}(2^k)^2 2^k$. With Lemma A.3, we conclude that $h_k=g^{-1}(2^k)\succsim(\gamma/2)^k$ so that $\varepsilon_k\log\frac{e}{\varepsilon_k}=2^{-k}\log(e2^k)\precsim g^{-1}(2^k)^2 2^k$. Hence, we can bound $F(h_k)+\varepsilon_k\log\frac{e}{\varepsilon_k}$ from above by a multiple of $h_k^2 g(h_k)$ in (22) and (23).

By Lemma A.3, we have $|F_0(h_m)| \precsim h_m/\varepsilon_m$ as $m \to \infty$. Moreover, in the case with general Σ and $g^{-1}(x) \succsim x^{-3/4}$, we have $h_m^2 \frac{1}{\sqrt{\varepsilon_m}} \succsim \varepsilon_m$. Hence, in case (I), there exists a constant κ_3 such that

$$\mathrm{mse}(\widehat{S}) \le \kappa_3 \biggl[h_m^2 \frac{1}{\sqrt{\varepsilon_m}} \log \frac{e}{\varepsilon_m} + \sum_{k=0}^{m-1} \frac{1}{n_{k+1}} h_k^2 g(h_k) \biggr]. \tag{24}$$

Conversely, in case (II), i.e. $g^{-1}(x) \precsim x^{-3/4}$, the term $h_m^2 \frac{1}{\sqrt{\varepsilon_m}}$ is negligible in (22), and we get

$$\mathrm{mse}(\widehat{S}) \le \kappa_4 \biggl[\varepsilon_m \log \frac{e}{\varepsilon_m} + \sum_{k=0}^{m-1} \frac{1}{n_{k+1}} h_k^2 g(h_k) \biggr] \tag{25}$$

for an appropriate constant κ_4.

Now, we specify $n_1, \dots, n_m$ in dependence on a positive parameter Z with $Z \ge 1/g^{-1}(2^m)$. We set $n_{k+1} = n_{k+1}(Z) = \lfloor Z g^{-1}(2^k) \rfloor \ge \frac{1}{2} Z g^{-1}(2^k)$ for $k = 0, \dots, m-1$ and conclude that, by (30),

$$\begin{aligned} \sum_{k=0}^{m-1} \frac{1}{n_{k+1}} h_k^2 g(h_k) &= \sum_{k=0}^{m-1} \frac{1}{n_{k+1}} 2^k g^{-1}(2^k)^2 \le \kappa_5 \frac{1}{Z} \sum_{k=0}^{m-1} 2^k g^{-1}(2^m) \Bigl(\frac{2}{\gamma}\Bigr)^{m-k} \\ &= \kappa_5 \frac{1}{Z} 2^m g^{-1}(2^m) \sum_{k=0}^{m-1} \gamma^{-(m-k)} \\ &\le \kappa_5 \frac{1}{1-\gamma^{-1}} \frac{1}{Z} 2^m g^{-1}(2^m). \end{aligned} \tag{26}$$

Similarly, we get with (7)

$$\mathrm{cost}(\widehat{S}) \le 3 \sum_{k=0}^{m-1} 2^{k+1} n_k \le \kappa_6 Z 2^m g^{-1}(2^m). \tag{27}$$

We proceed with case (I). By (24) and (26),

$$\mathrm{mse}(\widehat{S}) \le \kappa_7 \biggl[g^{-1}(2^m)^2 2^{m/2} m + \frac{1}{Z} 2^m g^{-1}(2^m) \biggr] \tag{28}$$

so that, for $Z := 2^{m/2}/(m g^{-1}(2^m))$,

$$\mathrm{mse}(\widehat{S}) \le 2\kappa_7 g^{-1}(2^m)^2 2^{m/2} m$$

and, by (27),

$$\mathrm{cost}(\widehat{S}) \le \kappa_6 \frac{2^{(3/2)m}}{m}.$$

For a positive parameter τ, we choose $m = m(\tau) \in \mathbb{N}$ as the maximal integer with $\kappa_6 2^{(3/2)m}/m \le \tau$. Here, we suppose that τ is sufficiently large to ensure the existence of such a m and the property $h_m \le \mathfrak{h}$. Then $\mathrm{cost}(\widehat{S}) \le \tau$. Since $2^m \approx (\tau \log \tau)^{2/3}$, we conclude that

$$\mathrm{mse}(\widehat{S}) \precsim g^{-1}((\tau \log \tau)^{2/3})^2 \tau^{1/3} (\log \tau)^{4/3}.$$

It remains to consider case (II). Here, (25) and (26) yield

$$\mathrm{mse}(\widehat{S}) \le \kappa_8 \left[2^{-m} m + \frac{1}{Z} 2^m g^{-1}(2^m) \right]$$

so that, for $Z := \frac{1}{m} 2^{2m} g^{-1}(2^m)$,

$$\mathrm{mse}(\widehat{S}) \le 2\kappa_8 2^{-m} m$$

and, by (27),

$$\mathrm{cost}(\widehat{S}) \le \kappa_6 \frac{1}{m} 2^{3m} g^{-1}(2^m)^2.$$

Next, let $l \in \mathbb{N}$ such that $2\kappa_6 2^{-l} \gamma^{-2l} \le 1$. Again we let τ be a positive parameter which is assumed to be sufficiently large so that we can pick $m = m(Z)$ as the maximal natural number larger than l and satisfying $2^{m+l} \le g^*(\tau)$. Then, by (29),

$$\mathrm{cost}(\widehat{S}) \le \kappa_6 \frac{1}{m} 2^{3m} g^{-1}(2^m)^2 \le 2\kappa_6 2^{-3l} \left(\frac{2}{\gamma}\right)^{2l} \frac{1}{m+l} 2^{3(m+l)} g^{-1}(2^{m+l})^2 \le \tau.$$

Conversely, since $2^{-m} \le 2^{l+1} g^*(\tau)$,

$$\mathrm{mse}(\widehat{S}) \le 2\kappa_8 2^{l+1} g^*(\tau)^{-1} \log_2 g^*(\tau).$$

Moreover, $g^{-1}(x) \succsim x^{-1}$ so that $x^3 g^{-1}(x)^2 / \log x \succsim x / \log x$, as $x \to \infty$. This implies that $\log g^*(\tau) \precsim \log \tau$.

Proof of Corollary 1.2. We fix $\beta' \in (\beta, 2]$ or $\beta' = 2$ in the case where $\beta = 2$, and note that, by definition of β,

$$\kappa_1 := \int_{B(0,1)} |x|^{\beta'} \nu(\mathrm{d}x)$$

is finite. We consider $\bar{g} : (0, \infty) \to (0, \infty), h \mapsto \int \frac{|x|^2}{h^2} \wedge 1 \nu(\mathrm{d}x)$. For $h \in (0, 1]$, one has

$$\begin{aligned}
\bar{g}(h) &= \int_{B(0,1)} \frac{|x|^2}{h^2} \wedge 1 \nu(\mathrm{d}x) + \int_{B(0,1)^c} \frac{|x|^2}{h^2} \wedge 1 \nu(\mathrm{d}x) \\
&\le \int_{B(0,1)} \frac{|x|^{\beta'}}{h^{\beta'}} \nu(\mathrm{d}x) + \int_{B(0,1)^c} 1 \nu(\mathrm{d}x) \le \kappa_2 h^{-\beta'},
\end{aligned}$$

where $\kappa_2 = \kappa_1 + \nu(B(0, 1)^c)$. Hence, we find a decreasing and invertible function $g : (0, \infty) \to (0, \infty)$ that dominates $\bar{g}$ and satisfies $g(h) = \kappa_2 h^{-\beta'}$ for $h \in (0, 1]$.

Then for $\gamma = 2^{1-1/\beta'}$, one has $g(\frac{\gamma}{2}h) = 2g(h)$ for $h \in (0,1]$ and we are in the position to apply Theorem 1.1: In the first case, we get

$$\operatorname{err}(\tau) \precsim \tau^{-(4-\beta')/(6\beta')}(\log\tau)^{(2/3)(1-1/\beta')}.$$

In the second case, we assume that $\beta' \le \frac{4}{3}$ and obtain $g^*(\tau) \approx (\tau\log\tau)^{-\beta'/(3\beta'-2)}$ so that

$$\operatorname{err}(\tau) \precsim \tau^{-\beta'/(6\beta'-4)}(\log\tau)^{(\beta'-1)/(3\beta'-2)}.$$

These estimates yield immediately the statement of the corollary.

APPENDIX

LEMMA A.1. *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 jth 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 $(\int_0^{t\wedge\tau} A_s\,\mathrm{d}L_s)_{t\ge 0}$ is a uniformly square integrable martingale which satisfies

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

The statement of the lemma follows from the Itô isometry for Lévy driven stochastic differential equations. See, for instance, [6], Lemma 3, for a proof.

LEMMA A.2. *The processes $\bar Y'$ and Υ introduced in Section 3.1 satisfy*

$$\mathbb{E}\Big[\sup_{s\in[0,1]}|\bar Y_s' - y_0|\Big] \le \kappa \quad \textit{and} \quad \mathbb{E}\Big[\sup_{s\in[0,1]}|\bar\Upsilon_s - y_0|\Big] \le \kappa,$$

where κ is a constant that depends only on K.

PROOF. The result is proven via a standard Gronwall inequality type argument that is similar to the proofs of the above propositions. It is therefore omitted. $\square$

LEMMA A.3. *Let $\bar h > 0$, $\gamma \in (1,2)$ and $g:(0,\infty)\to(0,\infty)$ be an invertible and decreasing function such that, for $h \in (0,\bar h]$,*

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

Then

$$\frac{\gamma}{2} g^{-1}(u) \le g^{-1}(2u) \tag{29}$$

for all $u \ge g(\bar{h})$. *Moreover, there exists a finite constant* κ_1 *depending only on* g *such that for all* $k, l \in \mathbb{Z}_+$ *with* $k \le l$ *one has*

$$g^{-1}(2^k) \le \kappa_1 \left(\frac{2}{\gamma}\right)^{l-k} g^{-1}(2^l). \tag{30}$$

If $\nu(B(0,h)^c) \le g(h)$ *for all* $h > 0$, *and* ν *has a second moment*, *then*

$$\int_{B(0,h)^c} |x| \nu(\mathrm{d}x) \le \kappa_2 \big(hg(h) + 1\big),$$

where κ_2 *is a constant that depends only on* g *and* $\int |x|^2 \nu(\mathrm{d}x)$.

PROOF. First, note that property (2) is equivalent to

$$\frac{\gamma}{2} g^{-1}(u) \le g^{-1}(2u)$$

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 $k \le l$ one has

$$g^{-1}(2^k) \le \kappa_1 \left(\frac{2}{\gamma}\right)^{l-k} g^{-1}(2^l).$$

For general, $h > 0$ one has

$$\int_{B(0,h)^c} |x| \nu(\mathrm{d}x) \le \int_{B(0,h)^c \cap B(0,\bar{h})} |x| \nu(\mathrm{d}x) + \frac{1}{\bar{h}} \int |x|^2 \nu(\mathrm{d}x).$$

Moreover,

$$\begin{aligned}\int_{B(0,h)^c \cap B(0,\bar{h})} |x| \nu(\mathrm{d}x) &\le \sum_{n=0}^{\infty} \nu\left(B\left(0, h\left(\frac{2}{\gamma}\right)^n\right)^c \cap B(0,\bar{h})\right) h \left(\frac{2}{\gamma}\right)^{n+1} \\ &\le \sum_{n=0}^{\infty} \mathbb{1}_{\{h(2/\gamma)^n \le \bar{h}\}} \underbrace{g\left(h\left(\frac{2}{\gamma}\right)^n\right)}_{\le 2^{-n} g(h)} h \left(\frac{2}{\gamma}\right)^{n+1} \\ &\le 2hg(h) \sum_{n=0}^{\infty} \gamma^{-(n+1)}.\end{aligned}$$

□

LEMMA A.4. *Let* $n \in \mathbb{N}$ *and* $(\mathcal{G}_j)_{j=0,1,\ldots,n}$ *denote a filtration. Moreover, let, for* $j = 0, \ldots, n-1$, U_j *and* V_j *denote nonnegative 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*

$$\mathbb{E}\Big[\max_{j=0,\ldots,n-1} U_j V_j\Big] \le \mathbb{E}\Big[\max_{j=0,\ldots,n-1} U_j\Big] \cdot \mathbb{E}\Big[\max_{j=0,\ldots,n-1} V_j\Big].$$

PROOF. See [6]. □

REFERENCES

[1] APPLEBAUM, D. (2004). *Lévy Processes and Stochastic Calculus*. *Cambridge Studies in Advanced Mathematics* **93**. Cambridge Univ. Press, Cambridge. MR2072890

[2] ASMUSSEN, S. and ROSIŃSKI, J. (2001). Approximations of small jumps of Lévy Processes with a view towards simulation. *J. Appl. Probab.* **38** 482–493. MR1834755

[3] BERTOIN, J. (1998). *Lévy Processes*. Cambridge Univ. Press. MR1406564

[4] COHEN, S. and ROSIŃSKI, J. (2007). Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes. *Bernoulli* **13** 195–210. MR2307403

[5] CREUTZIG, J., DEREICH, S., MÜLLER-GRONBACH, T. and RITTER, K. (2009). Infinite-dimensional quadrature and approximation of distributions. *Found. Comput. Math.* **9** 391–429. MR2519865

[6] DEREICH, S. and HEIDENREICH, F. (2009). A multilevel Monte Carlo algorithm for Lévy driven stochastic differential equations. Preprint.

[7] GILES, M. B. (2008). Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and Quasi-Monte Carlo Methods 2006* 343–358. Springer, Berlin. MR2479233

[8] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Oper. Res.* **56** 607–617. MR2436856

[9] HEINRICH, S. (1998). Monte Carlo complexity of global solution of integral equations. *J. Complexity* **14** 151–175. MR1629093

[10] JACOD, J. (2004). The Euler scheme for Lévy driven stochastic differential equations: Limit theorems. *Ann. Probab.* **32** 1830–1872. MR2073179

[11] JACOD, J., KURTZ, T. G., MÉLÉARD, S. and PROTTER, P. (2005). The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.* **41** 523–558. MR2139032

[12] KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations*. *Applications of Mathematics* (*New York*) **23**. Springer, Berlin. MR1214374

[13] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412

[14] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883

[15] NOVAK, E. (1995). The real number model in numerical analysis. *J. Complexity* **11** 57–73. MR1319050

[16] PROTTER, P. (2005). *Stochastic Integration and Differential Equations*, 2nd ed. *Stochastic Modelling and Applied Probability* **21**. Springer, Berlin. MR2273672

[17] PROTTER, P. and TALAY, D. (1997). The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.* **25** 393–423. MR1428514

[18] RUBENTHALER, S. (2003). Numerical simulation of the solution of a stochastic differential equation driven by a Lévy process. *Stochastic Process. Appl.* **103** 311–349. MR1950769

[19] RUBENTHALER, S. and WIKTORSSON, M. (2003). Improved convergence rate for the simulation of stochastic differential equations driven by subordinated Lévy processes. *Stochastic Process. Appl.* **108** 1–26. MR2008599

[20] SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. MR1739520

[21] TALAY, D. and TUBARO, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.* **8** 483–509. MR1091544

[22] ZAITSEV, A. Y. (1998). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM Probab. Statist*. **2** 41–108. MR1616527

PHILIPPS-UNIVERSITÄT MARBURG
FB. 12, MATHEMATIK UND INFORMATIK
HANS-MEERWEIN-STRASSE
D-35032 MARBURG
GERMANY
E-MAIL: dereich@mathematik.uni-marburg.de
URL: http://www.mathematik.uni-marburg.de/~dereich