Optimal simulation schemes for Lévy driven stochastic differential equations

Arturo Kohatsu-Higa [*†] Salvador Ortiz-Latorre [‡§] Peter Tankov [¶]

Abstract

We consider a general class of high order weak approximation schemes for stochastic differential equations driven by Lévy processes with infinite activity. These schemes combine a compound Poisson approximation for the jump part of the Lévy process with a high order scheme for the Brownian driven component, applied between the jump times. The overall approximation is analyzed using a stochastic splitting argument. The resulting error bound involves separate contributions of the compound Poisson approximation and of the discretization scheme for the Brownian part, and allows, on one hand, to balance the two contributions in order to minimize the computational time, and on the other hand, to study the optimal design of the approximating compound Poisson process. For driving processes whose Lévy measure explodes near zero in a regularly varying way, this procedure allows to construct discretization schemes with arbitrary order of convergence.

Key words: Lévy-driven stochastic differential equations, high order discretization schemes, weak approximation, regular variation

2010 Mathematics Subject Classification: 65C30, 60G51

1 Introduction

Let X_t be the unique solution of the SDE

$$X_t = x + \int_0^t b\left(X_s\right) ds + \int_0^t \sigma\left(X_s\right) dB_s + \int_0^t h\left(X_{s-}\right) dZ_s, \tag{1}$$

where b, σ and h are C^1 functions with bounded derivatives, B is a (multi-dimensional) Brownian motion and Z a one-dimensional, at least square integrable, infinite activity pure jump Lévy process with Lévy measure ν. In this paper we are interested in the weak approximation of X_t, using random partitions of the time interval.

The traditional approach, analyzed, e.g., in Jacod et al. [10] and Protter and Talay [18], consists in approximating X using the Euler scheme with a uniformly spaced time grid. It suffers from two difficulties: first, for a general Lévy measure ν, there is no available algorithm to simulate the increments of the driving Lévy process and second, a large jump of Z occurring between two discretization points can lead to a large discretization error.

With the aim of resolving these problems, Rubenthaler [19] (see also Bruti-Liberati and Platen [4] and Mordecki et al. [13] in the context of finite intensity Lévy processes) introduced the idea of replacing the driving process Z by a suitable compound Poisson approximation and placing the discretization points at the jump times of the compound Poisson process. This approach is problematic when the jump activity of the driving Lévy process Z is strong, that is, the Lévy measure has a strong singularity at zero.

[*]This author acknowledges financial support from the Japanese goverment

[†]Ritsumeikan University. Department of Mathematical Sciences and Japan Science and Technology Agency. Email: arturokohatsu@gmail.com

[‡]University of Oslo, Department of Mathematics, Email: salvador.ortiz-latorre@cma.uio.no

[§]This author acknowledges financial support from the Catalan government, BP-DGR 2009, and the the Norwegian Research Council under grant Evita/205328

[¶]LPMA, Paris-Diderot University. Email: tankov@math.univ-paris-diderot.fr

In Kohatsu-Higa and Tankov [9], the authors introduce and analyze a new approximation scheme in the case $\sigma \equiv 0$, building on the ideas of Rubenthaler and Asmussen and Rosinski [2]. The idea is to replace the driving process Z by an approximating process Z^ε, which incorporates all jumps of Z bigger than ε and approximates the jumps of Z smaller than ε with a suitable chosen Brownian motion, matching the second moment of Z. The solution to the continuous SDE between the jump times can then be approximated with a suitable high order scheme. More recently, a similar approximation was used in the context of multilevel Monte Carlo schemes for Lévy-driven SDEs [5].

Although the previous approach improves the rates of convergence obtained with Rubenthaler's scheme, there are limits on how well the small jumps of a Lévy process can be approximated by a Brownian motion (think of non-symmetric Lévy processes). In Tankov [21], the author presented a new scheme in the case $\sigma \equiv 0$ based on approximating Z by a finite intensity Lévy process, which incorporates all jumps bigger than ε and matches a given number of moments of Z with an additional compound Poisson term. The main advantages of this approach are that the schemes are very easy to implement, because the driving process is piecewise deterministic, and that one can, in specific cases, obtain arbitrarily high order of convergence by matching a sufficiently large number of moments of Z.

In this paper we are interested in two aspects of approximation schemes for Lévy driven SDE's. First, in many of the previously mentioned schemes one assumes that there is no Brownian motion component in the equation (1) (i.e. $\sigma \equiv 0$). The reason for this was that the speed of convergence of the approximating scheme for the jump component is fast and therefore it was not clear how to match this speed with the approximation of the Brownian component without wasting computing resources. Furthermore the fact that the equation does not have a Brownian component facilitates the error analysis and the implementation of the scheme because the SDE between jumps is deterministic, as in [21], or can be treated as a deterministic equation perturbed by a small noise term as in [9]. On the other hand, recent developments in the area of weak approximations for continuous SDE's [15, 14] allow for high order approximations of the Brownian component. Therefore one may expect that the right combination of these approximation techniques with suitable jump adapted approximation schemes for pure jump SDE's can be achieved.

Our second goal is a systematic study of the new moment-matching approximation schemes introduced in [21], with the objective of designing optimal compound Poisson approximations and studying their convergence in a more general setting.

In this article, we show that the mathematical framework developed in Tanaka and Kohatsu [20] is the appropriate tool in order to deal with the general situation ($\sigma \neq 0$). However, it needs to be adapted to the present setting where the partition is random while in [20], the partition is fixed. This framework is based on semigroup decompositions, which allow the study of a complex generator by decomposing it into simple components. The error estimate is obtained by a local analysis of each component.

In the resulting error bound, the contributions of the compound Poisson approximation and of the discretization scheme for the Brownian part are separate and tractable. This allows to balance the two contributions by an appropriate choice of the order of the discretization scheme for the Brownian part, in order to minimize the computational time. On the other hand, this decomposition enables us to formulate the problem of choosing the compound Poisson approximation as an optimization problem (minimizing the error bound). We characterize the optimal approximating process in the general case and provide explicit representation in specific situations. Often, the optimal solution is to keep all the jumps bigger than ε and add an additional compound Poisson process to match the moment structure of the small jumps. Under a regularity assumption on the Lévy measure, we show that this methodology can be used to construct approximations with arbitrarily high order of convergence.

An interesting consequence of our analysis is that the Asmussen-Rosinski approach is not the optimal procedure to approximate the small jumps in the setting of weak convergence. We give a better procedure, which uses Lévy measures with point masses to approximate the small jumps (see Remark 24) .

In order to correctly describe the optimality aspect, let $\bar{X}_t$ be the unique solution of (1) but using $\bar{Z}$ as driving process instead of Z. $\bar{Z}$ is a finite activity, at least square integrable, Lévy process with Lévy measure $\bar{\nu}$, which may have a Wiener component. Furthermore, let $\hat{X}_t$ be a computable approximation of $\bar{X}_t$ which shares the same jump times as $\bar{X}$. The first objective is to find an upper bound for the difference $\mathcal{D}_1 = \mathbb{E}[f(X_1)] - \mathbb{E}[f(\bar{X}_1)]$ in terms of $\bar{\lambda} = \bar{\nu}(\mathbb{R}) < \infty$ (the average number of partition intervals) and the

moments of $\nu - \bar{\nu}$ and $|\nu - \bar{\nu}|$. This part assumes then that the Brownian component can be simulated exactly.

In the second part, we approximate the Brownian component and analyze the error $\hat{\mathcal{D}}_1 = \mathbb{E}[f\left(\bar{X}_1\right)] - \mathbb{E}[f(\hat{X}_1)]$. To analyze $\hat{\mathcal{D}}_1$, we extend the operator approach developed in [20] to jump-adapted random partitions.

In conclusion, we find that we can express an upper bound for $\mathcal{D}_1$ in terms of the moments of $\nu - \bar{\nu}$ and $|\nu - \bar{\nu}|$ and an upper bound for $\hat{\mathcal{D}}_1$ in terms of $\bar{\lambda}$. Now, for fixed $\bar{\lambda}$ (and, hence, $\hat{\mathcal{D}}_1$) we consider $\bar{\nu}$ as a variable and minimize the upper bound for $\mathcal{D}_1$, obtaining an optimal Lévy measure $\bar{\nu}$ for the approximating finite intensity process $\bar{Z}$. Once the optimal error is known as a function of $\bar{\lambda}$ (this is done as a worse case analysis or in asymptotic form) one can identify the order of the approximation that is needed for the Brownian component.

The paper is structured as follows. In Section 2, we introduce the notation. In Section 3, we start introducing the assumptions in order to study the weak error of the approximations and we give the main error estimate, which will be the base for the study of optimal approximations. The expansion of the error is given in terms of $\bar{\lambda}$ and the moments of $\nu - \bar{\nu}$.

The proof of the main error estimate is given in Sections 4.1 and 4.2, which analyze, respectively, $\mathcal{D}_1$ and $\hat{\mathcal{D}}_1$. In Section 5, we formulate the problem of finding the optimal compound Poisson approximation of Z as an optimization problem, characterize its solution and prove an existence result. Explicit examples of solutions are given in Section 5.1, and Section 5.2 analyzes the convergence rates of the resulting scheme. Specific algorithms and numerical illustrations are provided in Section 6. Finally, in the appendix we gather some technical lemmas.

Throughout the article we use the Einstein notation of summation over double indices. δ_y denotes the point mass measure at $y \in \mathbb{R}$. Various positive constants are denoted by C or K with the dependence on various parameters. In particular they will depend on the constants appearing in the bounds of the coefficients of the Lévy driven SDE. We will not make this dependence explicit. The exact values of the constants may change from one line to the next without further mentioning.

2 Preliminaries and notation

Let the process $X = \{X_t\}_{t\in[0,1]}$ be the unique solution of the following d-dimensional SDE

$$X_t = x + \int_0^t b\left(X_s\right) ds + \int_0^t \sigma\left(X_s\right) dB_s + \int_0^t h\left(X_{s-}\right) dZ_s, \tag{2}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d, h : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times k}$ are $C^1\left(\mathbb{R}^d\right)$ functions with bounded derivatives, $B = \{B_t\}_{t\in[0,1]}$ is a k-dimensional standard Brownian motion and $Z = \{Z_t\}_{t\in[0,1]}$ is a one dimensional, square integrable, Lévy process (independent of B) with the following representation

$$\begin{aligned} Z_t &= \int_0^t \int_{\mathbb{R}} y\tilde{N}\left(ds, dy\right), \\ \tilde{N}\left(ds, dy\right) &= N\left(ds, dy\right) - \nu\left(dy\right) ds, \end{aligned}$$

where ν is an infinite activity Lévy measure, that is $\nu\left(\mathbb{R}\right) = +\infty$, and N is a Poisson random measure on $\mathbb{R} \times [0, \infty)$ with intensity $\nu\left(dy\right) \times dt$.

Let $\bar{X} = \{\bar{X}_t\}_{t\in[0,1]}$ be the approximating process, which is the solution of the SDE

$$\bar{X}_t = x + \int_0^t b(\bar{X}_s)ds + \int_0^t \sigma(\bar{X}_s)dB_s + \int_0^t h(\bar{X}_{s-})d\bar{Z}_s, \tag{3}$$

where $\bar{Z} = \{\bar{Z}_t\}_{t\in[0,1]}$ is a, square integrable, Lévy process (independent of B) with the following repre-

sentation

$$\bar{Z}_t = \bar{\mu}t + \bar{\sigma}W_t + \int_0^t \int_{\mathbb{R}} y\tilde{\bar{N}}\left(ds, dy\right),$$
$$\tilde{\bar{N}}\left(ds, dy\right) = \bar{N}\left(ds, dy\right) - \bar{\nu}\left(dy\right)ds,$$

where we assume that $\bar{\lambda} = \int_{\mathbb{R}} \bar{\nu}\left(dy\right) < \infty$, $\bar{\sigma}^2 \geq 0$ and $\bar{N}$ is a Poisson random measure on $\mathbb{R} \times [0, \infty)$ with intensity $\bar{\nu}\left(dy\right) \times ds$ and $W = \{W_t\}_{t\in[0,1]}$ is a standard k-dimensional Brownian motion independent of all the other processes. We assume that $(\bar{\mu}, \bar{\nu}, \bar{\sigma})$ belongs to a set of possible approximation parameters denoted by $\mathcal{A}$. Without loss of generality we may sometimes abuse the notation and write $\bar{\nu} \in \mathcal{A}$ to denote the Lévy measure for which there exists $\bar{\mu}$ and $\bar{\sigma}$ so that $(\bar{\mu}, \bar{\nu}, \bar{\sigma}) \in \mathcal{A}$.

Remark 1 *In order to simplify the notation, we write the pure jump parts of the driving Lévy processes Z_t and $\bar{Z}_t$ as compensated Poisson random measures on $\mathbb{R} \times [0, \infty)$. This can be done because they have finite expectation. Moreover, one can do all the analyses below using the standard Itô-Lévy representation, which only compensates the small jumps, and obtain equivalent results (it is a reparametrization of the drift $\bar{\mu}$).*

Note that, if we define

$$\bar{b}\left(x\right) = b\left(x\right) + h\left(x\right)\left(\bar{\mu} - \int_{\mathbb{R}} y\bar{\nu}\left(dy\right)\right),$$

then we can write

$$\bar{X}_t = x + \int_0^t \bar{b}\left(\bar{X}_s\right)ds + \int_0^t \sigma\left(\bar{X}_s\right)dB_s + \bar{\sigma}\int_0^t h\left(\bar{X}_s\right)dW_s + \int_0^t \int_{\mathbb{R}} h(\bar{X}_{s-})y\bar{N}\left(dy, ds\right).$$

Sometimes, the following flow notation will be useful

$$\begin{aligned}\bar{X}_t\left(s, x\right) = x &+ \int_s^t \bar{b}\left(\bar{X}_u\left(s, x\right)\right)du + \int_s^t \sigma\left(\bar{X}_u\left(s, x\right)\right)dB_u \\ &+ \bar{\sigma}\int_s^t h(\bar{X}_u\left(s, x\right))dW_u + \int_s^t \int_{\mathbb{R}} h(\bar{X}_{u-}\left(s, x\right))\bar{N}\left(dy, ds\right).\end{aligned}$$

Define the process

$$\bar{Y}_t(s, x) = x + \int_s^t \bar{b}(\bar{Y}_u(s, x))du + \int_s^t \sigma(\bar{Y}_u(s, x))dB_u + \bar{\sigma}\int_s^t h(\bar{Y}_u(s, x))dW_u \tag{4}$$

and the following semigroup operator

$$\left(\bar{P}_t f\right)(x) = \mathbb{E}[f(\bar{Y}_t(0, x))],$$

where $f \in C_p$, the set of functions with, at most, polynomial growth. The semigroup $\bar{P}$ corresponds to the process between jumps once the Lévy driving process is approximated by a compound Poisson process and an independent Brownian motion to approximate the small jumps as proposed in [2].

We consider the following stopping times

$$\begin{aligned}\bar{T}_i &= \inf\{t > \bar{T}_{i-1} : \bar{N}\left(\mathbb{R}, (\bar{T}_{i-1}, t]\right) \neq 0\}, \quad i \in \mathbb{N}, \\ \bar{T}_0 &= 0.\end{aligned}$$

and the associated *jump* operators

$$\begin{aligned}\left(\bar{S}^i f\right)(x) &= \mathbb{E}[f(x + h\left(x\right)\Delta\bar{Z}_{\bar{T}_i})], \quad i \in \mathbb{N} \\ \left(\bar{S}^0 f\right)(x) &= f\left(x\right),\end{aligned}$$

where $f \in C_p$.

Note that the stopping times $\bar{T}_i$ are well defined because $\bar{\lambda} < \infty$ and that $\bar{S}^i$ is independent of i because the jump sizes of a compound Poisson process are identically distributed. Still, we will keep this notation as it will help to keep track of the number of jumps.

We will also assume that there exist a process $\hat{X} = \{\hat{X}_t\}_{t\in[0,1]}$ satisfying the following stochastic representation condition for a given function space $\mathcal{C}$.

Assumption 2 ($\mathcal{SR}$) *Assume that for $f \in \mathcal{C}$, $\hat{X}$ satisfies*

$$\mathbb{E}[\mathbf{1}_{\{1<\bar{T}_1\}} f(\hat{X}_1)] = \mathbb{E}[\mathbf{1}_{\{1<\bar{T}_1\}} \bar{S}^0 \hat{P}_1 f\,(x)],$$
$$\mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} f(\hat{X}_1)] = \mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \bar{S}^0 \hat{P}_{\bar{T}_1\wedge 1} \bar{S}^1 \hat{P}_{\bar{T}_2-\bar{T}_1} \cdots \bar{S}^i \hat{P}_{1-\bar{T}_i} f\,(x)],$$

for $i \in \mathbb{N}$, where $\hat{P}_t$ is a linear operator.

Remark 3 *The process $\hat{X}$ and the linear operator $\hat{P}$ correspond to the actual simulation scheme chosen to approximate the solution of equation* (3) *between jumps. That is, this semigroups represents the approximation of the Wiener and drift parts of the equation. Similarly, $\mathcal{C}$ corresponds to the domain of compositions of the operators $\bar{S}^i \hat{P}$.*

Recall that for each multi-index $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{Z}_+^d$, where $\mathbb{Z}_+ := \mathbb{N} \cup \{0\}$, we denote by $|\alpha| := \alpha_1 + \cdots + \alpha_d = m$ the order of α. We also use the following notation $f^\alpha = \prod_{i=1}^d (f^i)^{\alpha_i}$ for any function $f : \mathbb{R}^k \to \mathbb{R}^d$. For $m, p \in \mathbb{Z}_+$, we introduce the following spaces of functions:

- C_p^m : the set of C^m functions $f : \mathbb{R}^d \to \mathbb{R}$ such that for each multi-index α with $0 \le |\alpha| \le m$,

$$\left| \frac{\partial^\alpha}{\partial x^\alpha} f\,(x) \right| \le C\,(\alpha)\,(1 + \|x\|^p)$$

  for some positive constant $C\,(\alpha)$.

We will use the notation $C_p := C_p^0$. In each C_p^m we consider the norm

$$\|f\|_{C_p^m} = \inf\{C > 0 : \left| \frac{\partial^\alpha}{\partial x^\alpha} f\,(x) \right| \le C\,(1 + \|x\|^p)\,, 0 \le |\alpha| \le m, x \in \mathbb{R}\}.$$

- C_b^m : the set of C^m functions $f : \mathbb{R}^d \to \mathbb{R}$ such that for each multi-index α with $1 \le |\alpha| \le m$,

$$\left| \frac{\partial^\alpha}{\partial x^\alpha} f\,(x) \right| \le C\,(\alpha)$$

  for some positive constant $C\,(\alpha)$.

For any function space $\mathcal{C}$, we may abuse the notation and write either $f \in \mathcal{C}$ or $f(x) \in \mathcal{C}$.

For $m \in \mathbb{N}$ and $p \in \mathbb{Z}_+$, we will consider the following assumption:

Assumption 4 ($\mathcal{H}_{m,p}$) $h, b, \sigma \in C_b^m$, $\int_{\mathbb{R}} |y|^{m+p}\, \nu\,(dy) < \infty$ *and* $\sup_{\bar{\nu}\in\mathcal{A}} \int_{\mathbb{R}} |y|^{m+p}\, \bar{\nu}\,(dy) < \infty$.

In fact, all the results exposed in this article require moment conditions up to a certain order depending on the regularity of each of the coefficients and the functional of X_t to be evaluated. The case $p = 0$ will also be considered in what follows and it corresponds to the case that f and its derivatives are bounded.

3 Weak error estimate

Our next objective is to establish the main error estimate of this paper. In order to do this, we need to introduce a modification of the framework introduced in [20] in the next section. The error estimate will then be given in Section 3.2.

3.1 Framework for weak approximation of operator compositions

To simplify the notation, we define the non commutative product of operators as follows. Given a finite number of linear operators $A^1, ..., A^n$, we define

$$\prod_{i=1}^{n} A^i := A^1 A^2 \cdots A^n.$$

Suppose we are given two sequences of linear operators $\{\bar{P}_t^i\}_{i=1,..,n}$ and $\{Q_t^i\}_{i=1...,n}$, $t \in [0,1]$. Furthermore, assume that for each $i \in \{1, ..., n\}$, Q_t^i approximates $\bar{P}_t^i$ in some sense to be defined later (see Assumption 7). Given a partition $\pi = \{0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1\}$, we define its norm as $|\pi|_n := \sup_{i=1,...,n}(t_i - t_{i-1})$. Now, we would like to estimate the following quantity

$$\bar{P}_{t_1}^1 \bar{P}_{t_2-t_1}^2 \cdots \bar{P}_{1-t_{n-1}}^n f(x) - Q_{t_1}^1 Q_{t_2-t_1}^2 \cdots Q_{1-t_{n-1}}^n f(x).$$

In order to achieve this goal, we will make use of the following expansion

$$\begin{aligned}
&\prod_{i=1}^{n} \bar{P}_{t_i-t_{i-1}}^i f(x) - \prod_{i=1}^{n} Q_{t_i-t_{i-1}}^i f(x) \\
&= \sum_{k=1}^{n} \left(\prod_{i=1}^{k-1} Q_{t_i-t_{i-1}}^i (\bar{P}_{t_k-t_{k-1}}^k - Q_{t_k-t_{k-1}}^k) \prod_{i=k+1}^{n} \bar{P}_{t_i-t_{i-1}}^i \right) f(x). \qquad (5)
\end{aligned}$$

Hence, if we have an appropriate norm estimate for $\prod_{i=k+1}^{n} \bar{P}_{t_i-t_{i-1}}^i$, a norm estimate of the error $\bar{P}_{t_k-t_{k-1}}^k - Q_{t_k-t_{k-1}}^k$ and a norm estimate for $\prod_{i=1}^{k-1} Q_{t_i-t_{i-1}}^i$, then we can expect to obtain a norm estimate for the difference between $\prod_{i=1}^{n} Q_{t_i-t_{i-1}}^i f(x)$ and $\prod_{i=1}^{n} \bar{P}_{t_i-t_{i-1}}^i f(x)$.

From now on, $\bar{P}_t^i : \cup_{p\in\mathbb{Z}_+} C_p \to \cup_{p\in\mathbb{Z}_+} C_p, i \in \{1, ..., n\}$ is a linear operator for $t \in [0,1]$ and $Q_t^i : \cup_{p\in\mathbb{Z}_+} C_p \to \cup_{p\in\mathbb{Z}_+} C_p, i \in \{1, ..., n\}$ is a linear operator for $t \in [0,1]$. The set $\mathcal{A}(Q)$ is the set that characterizes the approximating operator sequences $(Q^1, ..., Q^n) \in \mathcal{A}(Q)$ for all $n \in \mathbb{N}$. In particular, the set $\mathcal{A}(Q)$ englobes all approximations for different values of n and therefore it is a universal set of approximation sequences for the problem at hand. Similarly, we also define $\mathcal{A}(\bar{P})$ as the universal set that englobes all approximations of interest using finite Lévy measures.

Assumption 5 $(\mathcal{M}_0)$ *For each $p \in \mathbb{Z}_+$, if $f \in C_p$ then $Q_t^i f \in C_p$ and*

$$\sup_{t\in[0,1]} \left\| Q_t^i f \right\|_{C_p} \leq K(\mathcal{A}(Q), p) \|f\|_{C_p},$$

for some constant $K(\mathcal{A}(Q), p) > 0$, which does not depend on $i \in \mathbb{N}$ or f. Furthermore, we assume $0 \leq Q_t^i f(x) \leq Q_t^i g(x)$ whenever $0 \leq f \leq g$ and $Q_t^i \mathbf{1}_{\mathbb{R}}(x) = \mathbf{1}_{\mathbb{R}}(x)$.

As we will always consider the biggest constant, we will write $K(\mathcal{A}, p) \equiv K(\mathcal{A}(Q), p)$. Similarly, we will simplify the notation on other constants of this type. So in general, writing $\mathcal{A}$ will mean a dependence on $\mathcal{A}(Q)$ and $\mathcal{A}(\bar{P})$.

Assumption 6 $(\mathcal{M})$ *For all $i \in \mathbb{N}$, Q_t^i satisfies $(\mathcal{M}_0)$ and for each $f_p(x) := |x|^p$ $(p \in \mathbb{N})$,*

$$Q_t^i f_p(x) \leq (1 + K(\mathcal{A}, p)\, t) f_p(x) + K(\mathcal{A}, p)\, t$$

for some positive constant $K(\mathcal{A}, p)$.

For $r \in \mathbb{N}$, we define $\delta_r(t) = t^r$ as the rate function which measures the error.

Assumption 7 $(\mathcal{R}_r)$ *For all $i \in \mathbb{N}$, define $\mathrm{Err}_t^i \equiv \mathrm{Err}_t^{\bar{\nu},i} = \bar{P}_t^i - Q_t^i$. For each $p \in \mathbb{Z}_+$, there exists a constant $q = q(p, r)$ such that if $f \in C_p^{2r+2}$ then*

$$\left\| \mathrm{Err}_t^i f \right\|_{C_q} \leq K(\mathcal{A}, p, r)\, t \delta_r(t) \|f\|_{C_p^{2r+2}},$$

for all $t \in [0,1]$.

Assumption 8 $(\mathcal{M}_P)$ *For each $r \in \mathbb{N}$ and $p \in \mathbb{Z}_+$, if $f \in C_p^r$ one has that there exists a positive constant $K(\mathcal{A}, p, r)$ such that for $k = 1, ..., n-1$*

$$\sup_{(t_{k+1},...,t_n)\in[0,1]^{n-k}} \left\| \prod_{i=k+1}^{n} \bar{P}_{t_i}^i f \right\|_{C_p^r} \leq K(\mathcal{A}, p, r) \|f\|_{C_p^r}.$$

Lemma 9 *Under assumption $(\mathcal{M})$, the operators $\{Q_t^i\}_{i\geq 1}$ satisfy*

$$\sup_n \sup_{(Q^1,...,Q^n)\in\mathcal{A}(Q)} \max_{1\leq k\leq n} \left(\prod_{i=1}^{k-1} Q_{t_i - t_{i-1}}^i \right) f(x) < \infty,$$

for any positive function $f \in C_p$, $p \in \mathbb{Z}_+$ and $|\pi|_n n \leq C$ for some positive constant C.

Proof. Let $f_p(x) = |x|^p$ for $p \in \mathbb{N}$. Using assumptions $(\mathcal{M}_0)$ and $(\mathcal{M})$, we have

$$\begin{aligned}
&\prod_{i=1}^{k-1} Q_{t_i - t_{i-1}}^i f_p(x) = \left(\prod_{i=1}^{k-2} Q_{t_i - t_{i-1}}^i \right) (Q_{t_{k-1} - t_{k-2}}^{k-1} f_p)(x) \\
&\leq (1 + K(\mathcal{A}, p)(t_{k-1} - t_{k-2})) \left(\prod_{i=1}^{k-2} Q_{t_i - t_{i-1}}^i \right) f_p(x) + K(\mathcal{A}, p)(t_{k-1} - t_{k-2}) \\
&\leq (1 + K(\mathcal{A}, p)|\pi|_n) \left(\prod_{i=1}^{k-2} Q_{t_i - t_{i-1}}^i \right) f_p(x) + K(\mathcal{A}, p)|\pi|_n,
\end{aligned}$$

Since $(1 + K(\mathcal{A}, p)|\pi|_n)^{k-1} \leq e^{CK(\mathcal{A},p)}$, by induction follows that

$$\sup_n \sup_{(Q^1,...,Q^n)\in\mathcal{A}(Q)} \max_{1\leq k\leq n} \left(\prod_{i=1}^{k-1} Q_{t_i - t_{i-1}}^i \right) f(x) \leq K(\mathcal{A}, p) e^{K(\mathcal{A},p)} (1 + |x|^p) < \infty.$$

■

Theorem 10 *Assume $(\mathcal{M})$ for $\bar{P}_t^i$ and Q_t^i as well as $(\mathcal{R}_r)$ for some $r \in \mathbb{N}$. Let $f \in C_p^{2r+2}$, for some $p \in \mathbb{Z}_+$, then there exists a positive constant $K = K(x, \mathcal{A}, p, r) \in C_q$ such that*

$$\left| \prod_{i=1}^{n} \bar{P}_{t_i - t_{i-1}}^i f(x) - \prod_{i=1}^{n} Q_{t_i - t_{i-1}}^i f(x) \right| \leq K \|f\|_{C_p^{2r+2}} \sum_{k=1}^{n} (t_k - t_{k-1}) \delta_r(t_k - t_{k-1}).$$

Proof. Using the expansion (5), we analyze the inner term, using assumptions $(\mathcal{R}_r)$ and $(\mathcal{M}_P)$, to obtain for $q \equiv q(p, r)$ (stated in $(\mathcal{R}_r)$),

$$\begin{aligned}
&\left| \left((\bar{P}_{t_k - t_{k-1}}^k - Q_{t_k - t_{k-1}}^k) \prod_{i=k+1}^{n} \bar{P}_{t_i - t_{i-1}}^i \right) f(x) \right| \\
&\leq K(\mathcal{A}, p, r)(t_k - t_{k-1}) \delta_r(t_k - t_{k-1}) (1 + |x|^q) \left\| \prod_{i=k+1}^{n} \bar{P}_{t_i - t_{i-1}}^i f \right\|_{C_p^{2r+2}} \\
&\leq K(\mathcal{A}, p, r)(t_k - t_{k-1}) \delta_r(t_k - t_{k-1}) (1 + |x|^q) \|f\|_{C_p^{2r+2}}.
\end{aligned}$$

Now, Lemma 9 yields

$$\begin{aligned}
&\left| \left(\prod_{i=1}^{k-1} Q_{t_i - t_{i-1}}^i (\bar{P}_{t_k - t_{k-1}}^k - Q_{t_k - t_{k-1}}^k) \prod_{i=k+1}^{n} \bar{P}_{t_i - t_{i-1}}^i \right) f(x) \right| \\
&\leq K(\mathcal{A}, p, r)(t_k - t_{k-1}) \delta_r(t_k - t_{k-1}) \|f\|_{C_p^{2r+2}} \prod_{i=1}^{k-1} Q_{t_i - t_{i-1}}^i ((1 + |x|^q)) \\
&\leq K(x, \mathcal{A}, p, r)(t_k - t_{k-1}) \delta_r(t_k - t_{k-1}) \|f\|_{C_p^{2r+2}}.
\end{aligned}$$

Finally, adding up the estimates

$$\left|\prod_{i=1}^{n}\bar{P}^{i}_{t_i-t_{i-1}}f\left(x\right)-\prod_{i=1}^{n}Q^{i}_{t_i-t_{i-1}}f\left(x\right)\right|$$
$$\leq K\left(x,\mathcal{A},p,r\right)\left\|f\right\|_{C_p^{2r+2}}\sum_{k=1}^{n}\left(t_k-t_{k-1}\right)\delta_r\left(t_k-t_{k-1}\right).$$

■

3.2 Main error estimate

Theorem 11 *Let $\hat{X}=\{\hat{X}_t\}_{t\in[0,1]}$ be a process satisfying assumption $(\mathcal{SR})$ with $\mathcal{C}=C_p$, for some $p\in\mathbb{Z}_+$. Assume that the operators $\bar{P}^i_t:=\bar{S}^{i-1}\bar{P}_t$ and $Q^i_t:=\bar{S}^{i-1}\hat{P}_t$, $i=1,...,n+1$ satisfy assumptions $(\mathcal{M})$ and $(\mathcal{R}_r)$, for some $r\in\mathbb{N}$. Furthermore, assume $(\mathcal{H}_{m,p})$ and $f\in C_p^{(2r+2)\vee m}$, for some integer $m\geq 3$. Then there exist positive constants $K(x,\mathcal{A},p,r)\in C_q$ and $C_i(x)\in C_{p+1}$, $i=1,...,m$ such that*

$$\begin{aligned}
&\left|\mathbb{E}[f\left(X_1\right)]-\mathbb{E}[f(\hat{X}_1)]\right|\\
&\leq C_1\left(x\right)\left\|f\right\|_{C_p^1}\left|\bar{\mu}\right|+C_2(x)\left\|f\right\|_{C_p^2}\left|\int_{\mathbb{R}}y^2(\nu-\bar{\nu})\left(dy\right)-\bar{\sigma}^2\right|\\
&+\sum_{i=3}^{m-1}C_i\left(x\right)\left\|f\right\|_{C_p^i}\left|\int_{\mathbb{R}}y^i(\nu-\bar{\nu})\left(dy\right)\right|\\
&+C_m\left(x\right)\left\|f\right\|_{C_p^m}\{\int_{\mathbb{R}}\left|y\right|^m\left|\nu-\bar{\nu}\right|\left(dy\right)+\int_{\mathbb{R}}\left|y\right|^{m+p}\left|\nu-\bar{\nu}\right|\left(dy\right)\}\\
&+K\left(x,\mathcal{A},p,r\right)\left\|f\right\|_{C_p^{2r+2}}\bar{\lambda}^{-r}.
\end{aligned}$$

Proof. The proof is developed in Section 4 and follows from Theorems 13 and 18. We recall here that $q=q(p,r)$ which appears above is the parameter defined in hypothesis $\mathcal{R}_r$. ■

We also remark that in the particular case that f is bounded, regularity hypothesis can be weakened using a method of proof similar to the one used in this article.

Example 12 *The first simple example of application of the above result is to parametrize the set $\mathcal{A}$ by a parameter $\varepsilon\in(0,1]$ so that:*

$$\begin{aligned}
\bar{\mu}\equiv\mu_\varepsilon&=0,\\
\bar{\sigma}^2\equiv\sigma_\varepsilon^2&=\int_{\mathbb{R}}y^2(\nu-\nu_\varepsilon)\left(dy\right),\\
\bar{\nu}(dy)\equiv\nu_\varepsilon(dy)&=\mathbf{1}_{\{|y|>\varepsilon\}}\nu(dy).
\end{aligned}$$

Take $\hat{P}_t\equiv\hat{P}^\varepsilon_t$ to be the operator associated with a one step Euler scheme, so that the overall approximation consists in applying the Euler scheme between the jumps of $\bar{Z}$. In this case $r=1$ and if we take $m=3$ with $p=0$, we have that the above result states that for $f\in C_0^4$ and under $(\mathcal{H}_{3,0})$, we have

$$\left|\mathbb{E}[f\left(X_1\right)]-\mathbb{E}[f(\hat{X}_1^\varepsilon)]\right|\leq C_3(x)\left\|f\right\|_{C_0^3}\int_{\{|y|\leq\varepsilon\}}\left|y\right|^3\nu\left(dy\right)+K\left(x\right)\left\|f\right\|_{C_0^4}\lambda_\varepsilon^{-1}.$$

When $\sigma\equiv 0$, this result corresponds to Theorem 2 in [9].

In the particular case of an α-stable-like Lévy process with Lévy density $\sim\frac{c}{|x|^{1+\alpha}}$ near zero, one obtains that the best convergence rate is λ_ε^{-1} for $\alpha\leq 1$ and the worse case is $\lambda_\varepsilon^{-1/2}$ for $\alpha\to 2$.

Note that we could have applied high order schemes for Wiener driven SDEs in order to improve the last term above to λ_ε^{-m}.

Additional examples, algorithms, and numerical illustrations will be given in Section 6.

4 Proof of the main error estimate

4.1 Estimation of $\mathcal{D}_1 = \mathbb{E}[f(X_1)] - \mathbb{E}[f(\bar{X}_1)]$

Throughout this section we will use the notation $u(t,x) = \mathbb{E}[f(X_1(t,x))]$. Here $X.(t,x)$ denotes the flow associated with X. Some auxiliary properties of this function $u(t,x)$ are established in Lemma 36.

Theorem 13 *Assume $(\mathcal{H}_{m,p})$ and $f \in C_p^m$, for some integers $m \geq 3$ and $p \geq 0$. Then we have the expansion*

$$\begin{aligned}\mathbb{E}[f(X_1) - f(\bar{X}_1)] = -\bar{\mu}\int_0^1 \bar{B}_t^1 dt + \int_0^1 \bar{B}_t^2 dt \left(\int_{\mathbb{R}} y^2(\nu-\bar{\nu})(dy) - \bar{\sigma}^2\right) \\ + \sum_{i=3}^{m-1}\int_0^1 \bar{B}_t^i dt \int_{\mathbb{R}} y^i(\nu-\bar{\nu})(dy) + \int_0^1 \bar{B}_t^m dt, \end{aligned} \tag{6}$$

where

$$\begin{aligned}\bar{B}_t^i &:= \mathbb{E}\left[\sum_{|\alpha|=i}\frac{1}{\alpha!}\frac{\partial^{|\alpha|}}{\partial x^\alpha}u\left(t,\bar{X}_t\right)h^\alpha\left(\bar{X}_t\right)\right], \quad i=1,...,m, \\ \bar{B}_t^m &:= \mathbb{E}\Bigg[\sum_{|\alpha|=m}\int_{\mathbb{R}}\left(\int_0^1 \frac{\partial^{|\alpha|}}{\partial x^\alpha}u\left(t,\bar{X}_t+\theta y h\left(\bar{X}_t\right)\right)\frac{(1-\theta)^{|\alpha|-1}}{(m-1)!}d\theta\right) \\ &\quad \times h^\alpha\left(\bar{X}_t\right)y^\alpha(\nu-\bar{\nu})(dy)\Bigg].\end{aligned}$$

Furthermore, one has that $|\bar{B}_t^i| \leq C_i(x)\|f\|_{C_p^i}, i=1,...,m-1$, and

$$\left|\int_0^1 \bar{B}_t^m dt\right| \leq C_m(x)\|f\|_{C_p^m}\left\{\int_{\mathbb{R}}|y|^m|\nu-\bar{\nu}|(dy) + \int_{\mathbb{R}}|y|^{m+p}|\nu-\bar{\nu}|(dy)\right\}, \tag{7}$$

where the positive constants $C_i(x) \in C_{p+1}, i=1,...,m$, do not depend on $\bar{\nu}$.

Proof. To simplify the notation we will give the proof in the case $d=k=1$. Note that $\mathbb{E}[f(X_1)] = \mathbb{E}[f(X_1(0,x))] = u(0,x)$ and

$$\mathbb{E}[f(X_1) - f(\bar{X}_1)] = \mathbb{E}[u(0,x) - u(1,\bar{X}_1)].$$

Applying Itô formula to $u(1,\bar{X}_1)$ and taking into account the equation satisfied by $u(t,x)$ (see Lemma 36), we have

$$\begin{aligned}&\mathbb{E}[u(0,x) - u(1,\bar{X}_1)] \\ &= -\bar{\mu}\mathbb{E}\left[\int_0^1 \frac{\partial u}{\partial x}\left(t,\bar{X}_t\right)h\left(\bar{X}_t\right)dt\right] \\ &+ \mathbb{E}\left[\int_0^1\int_{\mathbb{R}}\left\{u\left(t,\bar{X}_t+h\left(\bar{X}_t\right)y\right) - u\left(t,\bar{X}_t\right) - \frac{\partial u}{\partial x}\left(t,\bar{X}_t\right)h\left(\bar{X}_t\right)y\right\}(\nu-\bar{\nu})(dy)\,dt\right] \\ &- \mathbb{E}\left[\frac{\bar{\sigma}^2}{2}\int_0^1 \frac{\partial^2 u}{\partial x^2}\left(t,\bar{X}_t\right)h^2\left(\bar{X}_t\right)dt\right].\end{aligned}$$

Making a Taylor expansion of order $m \geq 3$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\int_0^1 \int_{\mathbb{R}} \left\{u\left(t, \bar{X}_t + h\left(\bar{X}_t\right) y\right) - u\left(t, \bar{X}_t\right) - \frac{\partial u}{\partial x}\left(t, \bar{X}_t\right) h\left(\bar{X}_t\right) y\right\} (\nu - \bar{\nu})(dy)\, dt\right] \\
&= \sum_{i=2}^{m-1} \mathbb{E}\left[\int_0^1 \int_{\mathbb{R}} \frac{1}{i!} \frac{\partial^i}{\partial x^i} u\left(t, \bar{X}_t\right) h^i\left(\bar{X}_t\right) y^i (\nu - \bar{\nu})(dy)\, dt\right] \\
&+ \mathbb{E}\left[\int_0^1 \int_{\mathbb{R}} \left(\int_0^1 \frac{\partial^m}{\partial x^m} u\left(t, \bar{X}_t + \theta y h\left(\bar{X}_t\right)\right) \frac{(1-\theta)^{m-1}}{(m-1)!} d\theta\right) \times h^m\left(\bar{X}_t\right) y^m (\nu - \bar{\nu})(dy)\, dt\right].
\end{aligned}$$

Hence, collecting terms, we have

$$\begin{aligned}
&\mathbb{E}[u(0,x) - u\left(1, \bar{X}_1\right)] = -\bar{\mu} \int_0^1 \mathbb{E}\left[\frac{\partial u}{\partial x}\left(t, \bar{X}_t\right) h\left(\bar{X}_t\right)\right] dt \\
&+ \left(\int_{\mathbb{R}} y^2 (\nu - \bar{\nu})(dy) - \bar{\sigma}^2\right) \mathbb{E}\left[\int_0^1 \frac{1}{2!} \frac{\partial^2}{\partial x^2} u\left(t, \bar{X}_t\right) h^2\left(\bar{X}_t\right) dt\right] \\
&+ \sum_{i=3}^{m} \int_0^1 \mathbb{E}\left[\frac{1}{i!} \frac{\partial^i}{\partial x^i} u\left(t, \bar{X}_t\right) h^i\left(\bar{X}_t\right)\right] dt \int_{\mathbb{R}} y^i (\nu - \bar{\nu})(dy) \\
&+ \int_0^1 \mathbb{E}\left[\int_{\mathbb{R}} \left(\int_0^1 \frac{\partial^m}{\partial x^m} u\left(t, \bar{X}_t + \lambda y h\left(\bar{X}_t\right)\right) \frac{(1-\theta)^{m-1}}{(m-1)!} d\theta\right) \times h^m\left(\bar{X}_t\right) y^m (\nu - \bar{\nu})(dy)\right] dt,
\end{aligned}$$

and we obtain the expansion (6) . Under the assumption $(\mathcal{H}_{m,p})$, using Lemmas 33 and 36, one obtains the bounds for $\bar{B}_t^i$ and (7). ■

4.2 Estimation of $\hat{\mathcal{D}}_1 = \mathbb{E}[f\left(\bar{X}_1\right)] - \mathbb{E}[f(\hat{X}_1)]$

Lemma 14 *For $i \in \mathbb{Z}_+$, one has that*

$$\mathbb{E}[\mathbf{1}_{\{\bar{T}_i < 1 < \bar{T}_{i+1}\}} f(\bar{X}_1)] = \mathbb{E}[\mathbf{1}_{\{\bar{T}_i < 1 < \bar{T}_{i+1}\}} \bar{P}_{1-\bar{T}_i} f(\bar{X}_{\bar{T}_i})].$$

Proof. Define $\bar{\mathcal{H}}^{i,j} := \sigma(\bar{X}_{\bar{T}_j}, \bar{T}_1, ..., \bar{T}_{i+1}), i \in \mathbb{N} \cup \{0\}, j = 1, .., i$. Then, on the set $\left\{\bar{T}_i < 1 < \bar{T}_{i+1}\right\}$

$$\begin{aligned}
&\mathbb{E}[f(\bar{X}_1(\bar{T}_i, \bar{X}_{\bar{T}_i}))|\bar{\mathcal{H}}^{i,i}] \\
&= \mathbb{E}\Bigg[f(x + \int_t^1 \bar{b}(\bar{X}_s(t,x))ds \\
&+ \int_t^1 \sigma(\bar{X}_s(t,x))dB_s + \bar{\sigma} \int_t^1 h(\bar{X}_s(t,x))dW_s)\Bigg|\bar{\mathcal{H}}^{i,i}\Bigg]\Bigg|_{t=\bar{T}_i, x=\bar{X}_{\bar{T}_i}} \\
&= \mathbb{E}[f\left(\bar{Y}_1(t,x)\right)]|_{t=\bar{T}_i, x=\bar{X}_{\bar{T}_i}},
\end{aligned}$$

where in the last equality we have used that $\bar{X}_s(t,x)$ satisfies the same SDE as $\bar{Y}_s(t,x)$ on $\bar{T}_i \leq t < 1 < \bar{T}_{i+1}$. Now applying Lemma 35 and the definition of $(\bar{P}_t f)(x)$ we obtain the result. ■

Remark 15 *Applying the previous lemma with $i = 0$ and using that $\bar{S}^0$ is the identity operator we obtain that*

$$\mathbb{E}[\mathbf{1}_{\{1 < \bar{T}_1\}} f(\bar{X}_1)] = \mathbb{E}[\mathbf{1}_{\{1 < \bar{T}_1\}} \bar{S}^0 \bar{P}_1 f(x)].$$

Proposition 16 *For $i \in \mathbb{N}$, the following equality holds.*

$$\mathbb{E}[\mathbf{1}_{\{\bar{T}_i < 1 < \bar{T}_{i+1}\}} f\left(\bar{X}_1\right)] = \mathbb{E}[\mathbf{1}_{\{\bar{T}_i < 1 < \bar{T}_{i+1}\}} \bar{S}^0 \bar{P}_{\bar{T}_1} \bar{S}^1 \bar{P}_{\bar{T}_2 - \bar{T}_1} \cdots \bar{S}^i \bar{P}_{1-\bar{T}_i} f(x)].$$

Proof. Define $\bar{\mathcal{G}}^{i,j} := \sigma(\bar{X}_{\bar{T}_j-}, \bar{T}_1, ..., \bar{T}_{i+1}), i \in \mathbb{N}, j = 1, .., i$. By Lemma 14 and the definition of the operator $\bar{S}^i$ we have that

$$\begin{aligned}
&\mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} f\left(\bar{X}_1\right)] \\
&= \mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \bar{P}_{1-\bar{T}_i} f(\bar{X}_{\bar{T}_i})] \\
&= \mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \mathbb{E}[\bar{P}_{1-\bar{T}_i} f(\bar{X}_{\bar{T}_i-} + h(\bar{X}_{\bar{T}_i-})\Delta\bar{Z}_{\bar{T}_i})|\bar{\mathcal{G}}^{i,i}]] \\
&= \mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \bar{S}^i \bar{P}_{1-t} f(x)|_{t=\bar{T}_i, x=\bar{X}_{\bar{T}_i-}}] \\
&= \mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \bar{S}^i \bar{P}_{1-\bar{T}_i} f(\bar{X}_{\bar{T}_i-}(\bar{T}_{i-1}, \bar{X}_{\bar{T}_{i-1}}))] \\
&= \mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \bar{S}^i \bar{P}_{1-\bar{T}_i} f(\bar{Y}_{\bar{T}_i}(\bar{T}_{i-1}, \bar{X}_{\bar{T}_{i-1}}))].
\end{aligned}$$

Where in the last equality we have used that

$$\int_{\bar{T}_{i-1}}^{\bar{T}_i-} \int_{\mathbb{R}} h(\bar{X}_s(\bar{T}_{i-1}, \bar{X}_{\bar{T}_{i-1}}))y\bar{N}\left(dy, ds\right) = 0.$$

Reasoning analogously to the proof of Lemma 14, one has that

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \bar{S}^i \bar{P}_{1-\bar{T}_i} f(\bar{Y}_{\bar{T}_i}(\bar{T}_{i-1}, \bar{X}_{\bar{T}_{i-1}}))\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \mathbb{E}[\bar{S}^i \bar{P}_{1-\bar{T}_i} f(\bar{Y}_{\bar{T}_i}(\bar{T}_{i-1}, \bar{X}_{\bar{T}_{i-1}}))|\bar{\mathcal{H}}^{i,i-1}]\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \mathbb{E}[\bar{S}^i \bar{P}_{1-t_i} f(\bar{Y}_{t_i}\left(t_{i-1}, x\right))]|_{t_i=\bar{T}_i, t_{i-1}=\bar{T}_{i-1}, x=\bar{X}_{\bar{T}_{i-1}}}\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \bar{P}_{\bar{T}_i-\bar{T}_{i-1}} \bar{S}^i \bar{P}_{1-\bar{T}_i} f(X_{\bar{T}_{i-1}})]\right].
\end{aligned}$$

Iterating this procedure the result follows. ■

Now we need the following technical result.

Proposition 17 *We have for $r \in \mathbb{N}$,*

$$\sum_{i=0}^{\infty}\sum_{k=1}^{i+1} \mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \left(\bar{T}_k \wedge 1 - \bar{T}_{k-1}\right)^{r+1}\right] \leq C\left(r\right)\bar{\lambda}^{-r}.$$

Proof. From Lemma 11 in [9], one has that

$$\mathbb{E}\left[\int_0^1 \left(t - \eta\left(t\right)\right)^m dt\right] \leq C\left(m\right)\bar{\lambda}^{-m},$$

where $\eta\left(t\right) = \sup\{\bar{T}_i : \bar{T}_i \leq t\}$ and $C\left(r\right)$ is a constant that only depends on r. We can write

$$\begin{aligned}
\mathbb{E}\left[\int_0^1 \left(t - \eta\left(t\right)\right)^r dt\right] &= \sum_{i=0}^{\infty} \mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \int_0^1 \left(t - \eta\left(t\right)\right)^r dt\right] \\
&= \sum_{i=0}^{\infty}\sum_{k=1}^{i+1} \mathbb{E}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}} \int_{\bar{T}_{k-1}}^{\bar{T}_k\wedge 1} \left(t - \eta\left(t\right)\right)^r dt],
\end{aligned}$$

and the result follows by integration. ■

The main result of this section is the following.

Theorem 18 *Let $\{\bar{X}_t\}_{t\in[0,1]}$ be the process defined in* (3) *and $\{\hat{X}_t\}_{t\in[0,1]}$ a process satisfying assumption* $(\mathcal{SR})$ *with $\mathcal{C} = C_p$ for some $p \in \mathbb{Z}_+$. If the operators $\bar{P}_t^i := \bar{S}^{i-1}\bar{P}_t$ and $Q_t^i := \bar{S}^{i-1}\hat{P}_t$ associated to these processes satisfy assumptions* $(\mathcal{M})$ *and* $(\mathcal{R}_r)$ *with $\delta_r(t) = t^r$ for some $r \in \mathbb{N}$. Then for any $f \in C_p^{2r+2}$ there exists a positive constant $K = K\left(x, \mathcal{A}, p, r\right) \in C_q$ such that*

$$\left|\mathbb{E}[f(\bar{X}_1)] - \mathbb{E}[f(\hat{X}_1)]\right| \leq K\left(x, \mathcal{A}, p, r\right)\|f\|_{C_p^{2r+2}}\bar{\lambda}^{-r}$$

Proof. We can write

$$\mathbb{E}[f(\bar{X}_1)] - \mathbb{E}[f(\hat{X}_1)] = \mathbb{E}\left[\sum_{i=0}^{\infty}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}}(f(\bar{X}_1) - f(\hat{X}_1))\right].$$

By Proposition 16 and assumption $(\mathcal{SR})$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=0}^{\infty}[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}}(f(\bar{X}_1) - f(\hat{X}_1))\right] \\
&= \sum_{i=0}^{\infty}\mathbb{E}\Bigg[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}}(\bar{S}^0\bar{P}_{\bar{T}_1}\bar{S}^1\bar{P}_{\bar{T}_2-\bar{T}_1}\cdots\bar{S}^i\bar{P}_{1-\bar{T}_i} \\
&- \bar{S}^0\hat{P}_{\bar{T}_1}\bar{S}^1\hat{P}_{\bar{T}_2-\bar{T}_1}\cdots\bar{S}^i\hat{P}_{1-\bar{T}_i})f(x)\Bigg] \\
&= \sum_{i=0}^{\infty}\mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}}\left(\prod_{k=1}^{i+1}\bar{P}^k_{\bar{T}_k\wedge 1-\bar{T}_{k-1}} - \prod_{k=1}^{i+1}Q^k_{\bar{T}_k\wedge 1-\bar{T}_{k-1}}\right)f(x)\right].
\end{aligned}$$

Then, by Theorem 10, we obtain that

$$\begin{aligned}
&\left|\mathbb{E}[f(\bar{X}_1)] - \mathbb{E}[f(\widehat{X}_1)]\right| \\
&\leq \sum_{i=0}^{\infty}\left|\mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}}\left(\prod_{k=1}^{i+1}\bar{P}^k_{\bar{T}_k\wedge 1-\bar{T}_{k-1}} - \prod_{k=1}^{i+1}Q^k_{\bar{T}_k\wedge 1-\bar{T}_{k-1}}\right)f(x)\right]\right| \\
&\leq K(x,\mathcal{A},p,r)\|f\|_{C_p^{2r+2}}\sum_{i=0}^{\infty}\sum_{k=1}^{i+1}\mathbb{E}\left[\mathbf{1}_{\{\bar{T}_i<1<\bar{T}_{i+1}\}}\left(\bar{T}_k\wedge 1-\bar{T}_{k-1}\right)\delta_r\left(\bar{T}_k\wedge 1-\bar{T}_{k-1}\right)\right],
\end{aligned}$$

Then the result follows by Proposition 17. ■

5 Optimal approximation of Lévy measures

In this section, we discuss the optimization of the error bound in Theorem 11, i) with respect to the choice of the approximating Lévy process $\bar{Z}$. We would like to choose the parameters $\bar{\mu}$ and $\bar{\sigma}$ and the Lévy measure $\bar{\nu}$ in order to make the first four terms in the expansion small, that is, we concentrate on

$$\begin{aligned}
&C_1(x)|\bar{\mu}| + C_2(x)\left|\int_{\mathbb{R}}y^2(\nu-\bar{\nu})(dy) - \bar{\sigma}^2\right| \\
&+ \sum_{i=3}^{m-1}C_i(x)\left|\int_{\mathbb{R}}y^i(\nu-\bar{\nu})(dy)\right| + C_m(x)\int_{\mathbb{R}}|y|^m|\nu-\bar{\nu}|(dy).
\end{aligned} \tag{8}$$

Our approach is to take $\bar{\mu}=0$ and $\bar{\sigma}=0$, so that the expansion becomes

$$\sum_{i=2}^{m-1}C_i(x)\left|\int_{\mathbb{R}}y^i(\nu-\bar{\nu})(dy)\right| + C_m(x)\int_{\mathbb{R}}|y|^m|\nu-\bar{\nu}|(dy),$$

(see Remark 24 for an alternative choice of $\bar{\sigma}$).

Next, we choose the Lévy measure $\bar{\nu}$ in the class of measures for which the first sum is equal to zero and then optimize over $\bar{\nu}$ in this class with fixed intensity $\Lambda = \bar{\nu}(\mathbb{R}) < \infty$ in order to make the last term as small as possible. We will denote by $\mathcal{M}$ the set of all positive finite measures on $\mathbb{R}$. The problem of finding the optimal approximating Lévy measure then takes the following form.

Problem 19 ($\Omega_{m,\Lambda}$) *Let ν be a Lévy measure on $\mathbb{R}$ with finite moment of order m, where $m \geq 2$, and define $\mu_k = \int_{\mathbb{R}} y^k \nu(dy), 2 \leq k \leq m$. For any $\bar{\nu} \in \mathcal{M}$ define the functional*

$$J(\bar{\nu}) := \int_{\mathbb{R}} |y|^m |\nu - \bar{\nu}|(dy).$$

The problem $\Omega_{m,\Lambda}, m \geq 2$, consists in finding

$$\mathcal{E}_m(\Lambda) := \min_{\bar{\nu} \in \mathcal{M}} J(\bar{\nu}), \tag{9}$$

under the constraints

$$\int_{\mathbb{R}} \bar{\nu}(dy) = \Lambda \quad \text{and} \quad \int_{\mathbb{R}} y^k \bar{\nu}(dy) = \mu_k, \ k = 2, \ldots, m-1, \tag{10}$$

where $\Lambda \geq \min_{\bar{\nu} \in \mathcal{M}_{m-1}} \bar{\nu}(\mathbb{R})$, where for $m \geq 2$ we set

$$\mathcal{M}_m := \{\bar{\nu} \in \mathcal{M} : \int_{\mathbb{R}} y^k \bar{\nu}(dy) = \mu_k, k = 2, \ldots, m\}, \tag{11}$$

and $\mathcal{M}_1 := \mathcal{M}$.

The computation of $\min_{\bar{\nu} \in \mathcal{M}_m} \bar{\nu}(\mathbb{R})$ for $m \geq 2$ is a classical problem, known as the Hamburger problem. A summary of known results on this problem is provided in Appendix A.

Proposition 20 *The problem $\Omega_{m,\Lambda}, m \geq 2$, admits a solution.*

Proof. By Corollary 31, there exist at least one measure satisfying the constraints (10). For $m \geq 3$, we define by M_m^Λ the set of all such measures. For $m = 2$, we define by M_2^Λ the set of all measures $\bar{\nu} \in \mathcal{M}$ satisfying $\bar{\nu}(\mathbb{R}) = \Lambda$ and $\int_{\mathbb{R}} y^2 \bar{\nu}(dy) \leq C$, where

$$C = 2 \int_{\mathbb{R}} y^2 \nu(dy).$$

It is clear that the minimum in (9) is the same as the minimum over the set M_m^Λ for any $m \geq 2$.

Define

$$K_a := \{y \in \mathbb{R} : |y| \leq a\}, \quad a > 0.$$

By Chebyshev's inequality we have that

$$\bar{\nu}(\mathbb{R} \backslash K_a) = \int_{\{|y| > a\}} \bar{\nu}(dy) \leq \frac{1}{a^2} \int_{\mathbb{R}} y^2 \bar{\nu}(dy), \quad \forall \nu \in M_m^\Lambda,$$

which yields the tightness of M_m^Λ. By Prokhorov's theorem, we have that the set M_m^Λ is relatively sequentially compact but, as M_m^Λ is closed (see e.g., Chapter VII in Doob [6]), we also have that is sequentially compact. The set $\{J(\bar{\nu}) : \bar{\nu} \in M_m^\Lambda\}$ is bounded from below and, hence, it has an infimum, say $\mathcal{E}_m(\Lambda)$. Then, by the basic properties of the the infimum, we can find a sequence of real numbers of the form $\{J(\bar{\nu}_k)\}_{k \geq 1}$ converging to $\mathcal{E}_m(\Lambda)$. As M_m^Λ is sequentially compact we can always find a sequence $\{\bar{\nu}_{k_l}\}_{l \geq 1}$ that converges weakly to some $\bar{\nu}^* \in M_m^\Lambda$. But $\{J(\bar{\nu}_{k_l})\}_{l \geq 1}$, being a subsequence of the convergent sequence $\{J(\bar{\nu}_k)\}_{k \geq 1}$, must converge to $\mathcal{E}_m(\Lambda)$. Hence, we only need to prove the lower semicontinuity of the functional J, that is, if $\bar{\nu}_k$ converges weakly to $\bar{\nu}$ then $\liminf_{k \to \infty} J(\bar{\nu}_k) \geq J(\bar{\nu})$.

Let $\bar{\nu} \in M_m^\Lambda$. By the Hahn decomposition theorem, there exist disjoint measurable sets S^+ and S^- such that $S^+ \cup S^- = \mathbb{R}$, $\nu - \bar{\nu}$ is nonnegative on S^+ and nonpositive on S^-. The functional $J(\bar{\nu})$ can be alternatively written as

$$\begin{aligned} J(\bar{\nu}) &= \sup_{f \in L^\infty, \|f\| \leq 1} \int_{\mathbb{R}} |y|^m f(y)(\nu - \bar{\nu})(dy), \\ &= \int_{\mathbb{R}} |y|^m f^*(y)(\nu - \bar{\nu})(dy), \quad \text{with} \quad f^*(y) = \mathbf{1}_{S^+}(y) - \mathbf{1}_{S^-}(y), \end{aligned}$$

where L^∞ is the space of bounded measurable functions endowed with the essential supremum norm. This implies that

$$J(\bar{\nu}) \geq \sup_{f \in C_0, \|f\| \leq 1} \int_{\mathbb{R}} |y|^m f(y)(\nu - \bar{\nu})(dy), \tag{12}$$

where C_0 is the space of continuous functions with compact support.

Fix $\varepsilon > 0$. By the monotone convergence theorem there exists $A \in (1, \infty)$ such that

$$J(\bar{\nu}) - \int_{-A}^{A} |y|^m f^*(y)(\nu - \bar{\nu})(dy) \leq \varepsilon.$$

Since the measure $\mu := |y|^m(\nu - \bar{\nu})$ is a finite measure on $\mathbb{R}$, both measures in its Jordan decomposition are also finite and hence inner regular (see e.g. V.16 in [6]). Therefore, we can find two closed sets $B^+ \subseteq S^+ \cap (-A, A)$ and $B^- \subseteq S^- \cap (-A, A)$ such that μ is positive on B^+, negative on B^- and $\mu(\mathbb{R} \setminus (B^+ \cup B^-)) \leq 2\varepsilon$. By Lusin's theorem, we can find an interpolation between 1_{B^+} and 1_{B^-}. That is, a function $f \in C_0$ with $\|f\| \leq 1$ such that $f(y) = 1$ for $y \in B^+$, $f(y) = -1$ for $y \in B^-$ and $f(y) = 0$ for $y \notin (-A, A)$ with

$$\mu\{y \in \mathbb{R};\ |f - 1_{B^-} + 1_{B^+}|(y) > \varepsilon\} < \varepsilon.$$

Therefore, finally

$$J(\bar{\nu}) - \int_{\mathbb{R}} |y|^m f(y)(\nu - \bar{\nu})(dy) \leq \varepsilon + \int_{-A}^{A} |y|^m (f^*(y) - f(y))(\bar{\nu} - \nu)(dy) \leq 3\varepsilon,$$

which, together with (12) means that

$$J(\bar{\nu}) = \sup_{f \in C_0, \|f\| \leq 1} \int_{\mathbb{R}} |y|^m f(y)(\nu - \bar{\nu})(dy),$$

because the choice of ε was arbitrary.

For a sequence $(\bar{\nu}_k)$ which converges weakly to $\bar{\nu}$, we have, for every $f \in C_0$ with $\|f\| \leq 1$:

$$\begin{aligned}
\int_{\mathbb{R}} |y|^m f(y)(\nu - \bar{\nu})(dy) &= \liminf_k \int_{\mathbb{R}} |y|^m f(y)(\nu - \bar{\nu}_k)(dy) \\
&\leq \liminf_k \sup_{f \in C_0, \|f\| \leq 1} \int_{\mathbb{R}} |y|^m f(y)(\nu - \bar{\nu}_k)(dy) \\
&= \liminf_k J(\bar{\nu}_k).
\end{aligned}$$

Now, taking the sup with respect to f in the left-hand side, we obtain the desired result. ■

The following result provides a characterization of the solutions of $\Omega_{m,\Lambda}$, which will be useful in finding explicit representations for small m.

Proposition 21 *The measure $\bar{\nu}$ is a solution of (9) if and only if it satisfies the constraints (10), and there exists a piecewise polynomial function $P(y) = a_0 + \sum_{i=2}^{m-1} a_i y^i + |y|^m$ such that $P(y) \geq 0$ for all $y \in \mathbb{R}$, a function $\alpha : \mathbb{R} \mapsto [0, 1]$ and a positive measure τ on $\mathbb{R}$ such that*

$$\bar{\nu}(dy) = \nu(dy)\mathbf{1}_{\{P(y) < 2|y|^m\}} + \alpha(y)\nu(dy)\mathbf{1}_{\{P(y) = 2|y|^m\}} + (\tau(dy) + \nu(dy))\mathbf{1}_{\{P(y)=0\}}. \tag{13}$$

Remark 22 *If the measure ν is absolutely continuous with respect to Lebesgue's measure, the expression (13) simplifies to*

$$\bar{\nu}(dy) = \nu(dy)\mathbf{1}_{\{P(y) < 2|y|^m\}} + \tau(dy)\mathbf{1}_{\{P(y)=0\}}.$$

Moreover, in the case $m = 2q, q \in \mathbb{N}$, $P(y)$ is a polynomial and the measure τ may always be taken to be an atomic measure with at most q atoms (because a positive polynomial of degree $m = 2q$ has at most q distinct roots).

Proof. Since the functional J is convex, a measure $\bar{\nu}^*$ which satisfies the constraints (10) is a solution of (9) if and only if there exists a vector of Lagrange multipliers $(p_0, p_2, \ldots, p_m)$ such that $\bar{\nu}^*$ minimizes the Lagrangian $\mathcal{L}(\bar{\nu}, p)$ over all measures $\bar{\nu} \in \mathcal{M}$, and $\mathcal{L}(\bar{\nu}^*, p) > -\infty$. The Lagrangian for this problem takes the form (dropping the terms which do not depend on $\bar{\nu}$):

$$\mathcal{L}(\bar{\nu}, p) = \int_{\mathbb{R}} |y|^m |\bar{\nu} - \nu|(dy) + \int_{\mathbb{R}} \bar{\nu}(dy)(p_0 + \sum_{i=2}^{m-1} p_i y^i).$$

Set $P(y) = p_0 + \sum_{i=2}^{m-1} p_i y^i + |y|^m$. Let $y_0 \in \mathbb{R}$ be such that $P(y_0) < 0$, and consider the family of measures $\bar{\nu}_a(dy) = a\delta_{y_0}$, where $a > 0$. Then, for any $(p_0, ..., p_m)$,

$$\mathcal{L}(\bar{\nu}_a, p) = \int_{\mathbb{R}\setminus\{y_0\}} |y|^m \nu(dy) + |a - a_0||y_0|^m + a\left(p_0 + \sum_{i=2}^{m-1} p_i y_0^i\right),$$

where $a_0 = \nu(\{y_0\})$. For $a > a_0$, we have that

$$\mathcal{L}(\bar{\nu}_a, p) = \int_{\mathbb{R}\setminus\{y_0\}} |y|^m \nu(dy) - a_0|y_0|^m + aP(y_0) \underset{a\to+\infty}{\to} -\infty.$$

Therefore, necessarily $P(y) \geq 0$ for all $y \in \mathbb{R}$. Now, as before, let the Jordan decomposition of $\bar{\nu} - \nu$ be given by $\bar{\nu} - \nu = \mu^+ - \mu^-$, where μ^+ and μ^- are supported on disjoint measurable sets. Then,

$$\mathcal{L}(\bar{\nu}, p) = \int_{\mathbb{R}} P(y)\mu^+(dy) + \int_{\mathbb{R}} (2|y|^m - P(y))(\mu^-(dy) - \nu(dy)) + \int_{\mathbb{R}} |y|^m \nu(dy).$$

Then, it is clear that at optimum,

- μ^+ should be equal to a measure with support $\{y : P(y) = 0\}$. Therefore in general, there will be no uniqueness.
- $\mu^- \equiv 0$ on $\{y : 2|y|^m - P(y) > 0\}$.
- $\mu^- \equiv \nu$ on $\{y : 2|y|^m - P(y) < 0\}$. This follows because μ^+ and μ^- are supported on disjoint measurable sets and $\mu^- \leq \nu$.
- μ^- satisfies $\nu - \mu^- \geq 0$ on $\{y : 2|y|^m - P(y) = 0\}$.

Combining these observations, we complete the proof. ■

Example 23 *Let $m = 2q, q \in \mathbb{N}$, and ν be absolutely continuous. Since there are $m-1$ constraints* (10)*, the polynomial P must, in general, have exactly q distinct roots, otherwise there are not enough degrees of freedom to satisfy all the constraints. Therefore, we shall restrict our attention to polynomials of the form*

$$P(y) = (y - a_1)^2 \cdots (y - a_q)^2,$$

and the measure τ will be of the form

$$\tau(dy) = \sum_{i=1}^{q} \alpha_i \delta_{a_i}.$$

The roots $\{a_i\}_{i=1}^q$ and the masses $\{\alpha_i\}_{i=1}^q$: are found from the following system of m nonlinear equations:

$$\sum_{j=1}^{q} a_j \prod_{i\neq j}^{q} a_i^2 = 0,$$

$$\int_{\{(y-a_1)^2\cdots(y-a_q)^2 > 2y^{2q}\}} \nu(dy) + \sum_{i=1}^{q} \alpha_i = \Lambda,$$

$$\int_{\{(y-a_1)^2\cdots(y-a_q)^2 > 2y^{2q}\}} y^k \nu(dy) = \sum_{i=1}^{q} \alpha_i a_i^k, \quad k = 2, ..., 2q-1.$$

Obviously, in general, the solution to this system can only be approximated numerically and this does not seem an easy task. For $m \leq 4$, the solutions are quite explicit; they are discussed in the following section.

To complete the analysis we need to quantify the dependence of the optimal value of the error $\mathcal{E}_m(\Lambda)$ on Λ when Λ tends to infinity. This is achieved in the following section for small values of m and in Section 5.2 for general m, under a regularity assumption on the Lévy measure.

5.1 Explicit examples for small values of m

Throughout this section we assume that the measure ν is absolutely continuous with respect to the Lebesgue measure.

The case $m = 2$. We use the characterization of Proposition 21 (see also Remark 22). The function $P(y)$ is necessarily of the form $P(y) = a_0 + y^2$ for some $a_0 \geq 0$, and therefore the optimal solution is given by

$$\bar{\nu}_\varepsilon(dy) = \mathbf{1}_{\{y^2>\varepsilon\}}\nu(dy),$$

where $\varepsilon = \varepsilon(\Lambda)$ solves $\nu(\{y^2 > \varepsilon\}) = \Lambda$. The approximation error $\mathcal{E}_2(\Lambda)$ is given by

$$\mathcal{E}_2(\Lambda) = J(\bar{\nu}_{\varepsilon(\Lambda)}) = \int_{\{y^2\leq\varepsilon(\Lambda)\}} y^2\nu(dy),$$

which can go to zero at an arbitrarily slow rate as $\Lambda \to \infty$.

The case $m = 3$. The function $P(y)$ is now of the form $P(y) = a_0 + a_2y^2 + |y|^3$ and, as $P(y) = P(-y)$, it suffices to study $P(y), y \geq 0$. To be able to satisfy the constraints (10), the optimal measure must have nontrivial atomic part. Hence, $P(y)$ has to have at least one positive root. The nonnegativity restriction on $P(y)$ yields that $P(y)$ must have exactly one positive root of multiplicity 2. Therefore, $P(y)$ must be of the form

$$P(y) = (y+\varepsilon)(y-\delta)^2, \quad , y \geq 0 \quad \varepsilon > 0, \quad \delta > 0.$$

Since the coefficient in front of the first power of y is zero, we conclude that $\delta = 2\varepsilon$ and $P(y)$ is necessarily of the form

$$\begin{aligned} P(y) &= (y+\varepsilon)(y-2\varepsilon)^2, \quad y \geq 0 \\ P(y) &= -(y+2\varepsilon)^2(y-\varepsilon), \quad y < 0, \end{aligned}$$

or, in other words, $P(y) = |y|^3 - 3\varepsilon y^2 + 4\varepsilon^3$, for some $\varepsilon > 0$. It is now easy to see that an optimal solution is given by

$$\bar{\nu}_\varepsilon(dy) = \mathbf{1}_{\{|y|>\varepsilon\}}\nu(dy) + \alpha_1\delta_{-2\varepsilon} + \alpha_2\delta_{2\varepsilon},$$

where $\varepsilon = \varepsilon(\Lambda)$ solves

$$\int_{\{|y|>\varepsilon\}} \nu(dy) + \frac{1}{4\varepsilon^2}\int_{\{|y|\leq\varepsilon\}} y^2\nu(dy) = \Lambda,$$

and

$$\alpha_1 + \alpha_2 = \frac{1}{4\varepsilon^2}\int_{\{|y|\leq\varepsilon\}} y^2\nu(dy).$$

The approximation error $\mathcal{E}_3(\Lambda)$ satisfies $\mathcal{E}_3(\Lambda) = o(\Lambda^{-1/2})$ as $\Lambda \to \infty$, since

$$\mathcal{E}_3(\Lambda) = \int_{\{|y|\leq\varepsilon(\Lambda)\}} |y|^3\nu(dy) + 2\varepsilon(\Lambda)\int_{\{|y|\leq\varepsilon(\Lambda)\}} y^2\nu(dy) \leq 3\varepsilon(\Lambda)\int_{\{|y|\leq\varepsilon(\Lambda)\}} y^2\nu(dy) = o(\varepsilon(\Lambda))$$

and

$$\lim_{\Lambda\to\infty} \varepsilon(\Lambda)^2\Lambda = \lim_{\varepsilon\downarrow 0}\varepsilon^2\int_{\{|y|>\varepsilon\}}\nu(dy) + \lim_{\varepsilon\downarrow 0}\frac{1}{4}\int_{\{|y|\leq\varepsilon\}} y^2\nu(dy) \leq \lim_{c\downarrow 0}\int_{\{|y|\leq c\}} y^2\nu(dy) = 0.$$

However, the scheme with $m = 4$ achieves a better rate with the same computational cost.

The case $m = 4$. The function $P(y)$ is now of the form $P(y) = a_0 + a_2y^2 + a_3y^3 + y^4$. Once again, for P to be nonnegative, the roots must have multiplicity 2 or 4, and to be able to satisfy the constraints (10), we require that P have exactly 2 roots. Using the fact that the coefficient in front of the first power of y is zero, we deduce that

$$P(y) = (y-\varepsilon)^2(y+\varepsilon)^2 = y^4 - 2y^2\varepsilon^2 + \varepsilon^4$$

for some $\varepsilon > 0$. Analyzing the function $2y^4 - P(y) = y^4 - \varepsilon^4 + 2\varepsilon^2y^2$ it is easy to check that $\{2y^4 - P(y) > 0\} = \{|y| > \varepsilon\sqrt{\sqrt{2}-1}\}$. Hence, the optimal solution is of the form

$$\bar{\nu}_\varepsilon(dy) = \nu(dy)\mathbf{1}_{\{|y|>\varepsilon\sqrt{\sqrt{2}-1}\}} + \alpha_1\delta_{-\varepsilon} + \alpha_2\delta_\varepsilon,$$

where the constants α_1 and α_2 are determined from the moment constraints and satisfy

$$\alpha_1 = \frac{1}{2\varepsilon^3}\left(-\int_{\{|y|\le\varepsilon\sqrt{\sqrt{2}-1}\}} y^3\nu(dy) + \varepsilon\int_{\{|y|\le\varepsilon\sqrt{\sqrt{2}-1}\}} y^2\nu(dy)\right),$$
$$\alpha_2 = \frac{1}{2\varepsilon^3}\left(\int_{\{|y|\le\varepsilon\sqrt{\sqrt{2}-1}\}} y^3\nu(dy) + \varepsilon\int_{\{|y|\le\varepsilon\sqrt{\sqrt{2}-1}\}} y^2\nu(dy)\right),$$

and $\varepsilon = \varepsilon(\Lambda)$ is found from the intensity constraint $F(\varepsilon) = \Lambda$, where

$$F(\varepsilon) = \int_{\{|y|>\varepsilon\sqrt{\sqrt{2}-1}\}} \nu(dy) + \frac{1}{\varepsilon^2}\int_{\{|y|\le\varepsilon\sqrt{\sqrt{2}-1}\}} y^2\nu(dy).$$

Note that F is strictly decreasing, continuous, and satisfies $\lim_{\varepsilon\downarrow 0} F(\varepsilon) = +\infty$ and $\lim_{\varepsilon\uparrow+\infty} F(\varepsilon) = 0$, which ensures the existence of a unique solution for $F(\varepsilon) = \Lambda$. Also note that

$$\left|\int_{\{|y|\le\varepsilon\sqrt{\sqrt{2}-1}\}} y^3\nu(dy)\right| \le \varepsilon\sqrt{\sqrt{2}-1}\int_{\{|y|\le\varepsilon\sqrt{\sqrt{2}-1}\}} y^2\nu(dy)$$
$$\le \varepsilon\int_{\{|y|\le\varepsilon\sqrt{\sqrt{2}-1}\}} y^2\nu(dy),$$

which ensures the non negativity of α_1, α_2.

The worst case convergence rate can be estimated similarly to the case $m = 3$ and satisfies $\mathcal{E}_4(\Lambda) = o(\Lambda^{-1})$ as $\Lambda \to \infty$. As we shall see in the next section, in the presence of a more detailed information about the explosion of the Lévy measure at zero, this convergence rate can be refined.

Remark 24

1. *The calculations of this section make it clear that as far as weak approximations are concerned, the Asmussen-Rosinski approach of approximating the small jumps of a Lévy process with a Brownian motion is not necessarily the only answer. In fact, the case $m = 3$ studied above leads to an approximation which is asymptotically equivalent to the Asmussen-Rosinski method and the case $m = 4$ leads to a scheme which converges at a faster rate, for the same computational cost.*

2. *Instead of taking $\bar{\sigma} = 0$, one may choose $\bar{\sigma}$ which makes the second term in (8) equal to zero, which leads, for $m \ge 3$, to the following optimization problem for $\bar{\nu}$:*

$$\mathcal{E}'_m(\Lambda) := \min_{\bar{\nu}\in\mathcal{M}} J(\bar{\nu})$$

under the constraints

$$\int_{\mathbb{R}} \bar{\nu}(dy) = \Lambda \quad \text{and} \quad \int_{\mathbb{R}} y^k\bar{\nu}(dy) = \mu_k,\ k = 3,\dots,m-1.$$

This problem assumes the use of the Asmussen-Rosinski approach to match the second moment of ν. The analysis of this problem can be carried out using the same tools described above and leads to similar results.

5.2 Convergence rates for regularly varying Lévy measures

The notion of regular variation provides a convenient tool to study the convergence of our moment matching schemes even in the cases when m is large and an explicit solution of $(\Omega_{m,\Lambda})$ is not available. We refer to [3] for background on regular variation.

As usual, we denote by R_α the class of regularly varying functions with index α (at zero or at infinity depending on the context). The following assumption, which is satisfied by many parametric Lévy models used in practice (stable, tempered stable/CGMY, normal inverse Gaussian, generalized hyperbolic etc.) may be used to quantify the rate of explosion of the Lévy measure near zero.

Assumption 25 *There exists $\alpha \in (0,2)$, positive constants c_+ and c_- with $c_+ + c_- > 0$ and a function $g \in R_{-\alpha}$ (at zero) such that the Lévy measure ν satisfies*

$$\nu((y,\infty)) \sim c_+ g(y) \quad \text{and} \quad \nu((-\infty,-y)) \sim c_- g(y) \quad \text{as } y \downarrow 0, \tag{R_α}$$

Theorem 26 *Let m be even and let the Lévy measure ν satisfy the assumption (R_α). Then there exists a function $f(\Lambda)$ with $f \in R_{1-m/\alpha}$ as $\Lambda \to \infty$ such that the error bound $\mathcal{E}_m(\Lambda)$ defined by (9) satisfies*

$$\underline{c} f(\Lambda) \le \mathcal{E}_m(\Lambda) \le \overline{c} f(\Lambda)$$

for all Λ sufficiently large, and for some constants $\underline{c}, \overline{c}$ with $0 < \underline{c} \le \overline{c} < \infty$. The function f is given explicitly by $f(\Lambda) = (g^{\leftarrow}(\Lambda))^m \Lambda$, where $g^{\leftarrow}$ is a generalized inverse of the function g appearing in Assumption (R_α).

Remark 27

1. *The regular variation implies that as $\Lambda \to \infty$, the error goes to zero as $\Lambda^{1-\frac{m}{\alpha}}$ times a slowly varying factor (such as logarithm). To compute the explicit convergence rate, the exact form of the regularly varying function g must be known. For example, if $g(y) = y^{-\alpha}$ then*

$$f(\Lambda) \sim C\Lambda^{1-\frac{m}{\alpha}}$$

   for some strictly positive constant C.

2. *In the case $m = 4$ it can be shown using similar methods that $\mathcal{E}_m(\Lambda) \sim C f(\Lambda)$ for some strictly positive constant C.*

Proof. Throughout the proof, we let $q = \frac{m}{2}$. To obtain an upper bound on the error, we construct a, possibly suboptimal, measure satisfying the constraints which attains the desired rate. Let $\varepsilon > 0$, and define

$$\nu_\varepsilon(dy) = \nu(dy)\mathbf{1}_{\{|y|>\varepsilon\}} + \bar{\nu}_\varepsilon(dy), \tag{14}$$

where $\bar{\nu}_\varepsilon(dy)$ is the solution (minimizer) of the moment problem

$$\bar{\Lambda}_\varepsilon := \min\{\bar{\nu}(\mathbb{R}) : \bar{\nu} \in \mathcal{M}, \int_{\mathbb{R}} y^k \bar{\nu}(dy) = \mu_k^\varepsilon, k = 2, \dots, m\},$$

where we define $\mu_k^\varepsilon := \int_{\{|y|\le\varepsilon\}} y^k \nu(dy)$. Then,

$$\begin{aligned}\mathcal{E}_m(\Lambda_\varepsilon) &\le J(\nu_\varepsilon) := \int_{\mathbb{R}} y^m |\nu - \nu_\varepsilon|(dy) \\ &\le \int_{\{|y|\le\varepsilon\}} y^m \nu(dy) + \int_{\mathbb{R}} y^m \bar{\nu}_\varepsilon(dy) = 2\int_{\{|y|\le\varepsilon\}} y^m \nu(dy),\end{aligned} \tag{15}$$

where

$$\Lambda_\varepsilon := \bar{\nu}_\varepsilon(\mathbb{R}) = \int_{\{|y|>\varepsilon\}} \nu(dy) + \bar{\Lambda}_\varepsilon.$$

By Proposition 29,

$$\bar{\Lambda}_\varepsilon = \inf\{\mu_0^\varepsilon : \{\mu_{i+j}^\varepsilon\}_{i,j=0}^q \geq 0 \quad \text{for some} \quad \mu_1^\varepsilon\}.$$

On the other hand, the matrix $\{\mu_{i+j}^\varepsilon\}_{i,j=1}^q$ is (nonnegative) positive definite, because it is a moment matrix of a measure. Therefore, by Sylvester's criterion (applied here to lower right minors) we can write

$$\bar{\Lambda}_\varepsilon = \inf\{\mu_0^\varepsilon : \det(\{\mu_{i+j}^\varepsilon\}_{i,j=0}^q) \geq 0 \quad \text{for some} \quad \mu_1^\varepsilon\}$$

and also

$$\bar{\Lambda}_\varepsilon \leq \inf\{\mu_0^\varepsilon : \det(\{\mu_{i+j}^\varepsilon \mathbf{1}_{i+j\neq 1}\}_{i,j=0}^q) \geq 0\}$$

But

$$\det(\{\mu_{i+j}^\varepsilon \mathbf{1}_{i+j\neq 1}\}_{i,j=0}^q) = \mu_0^\varepsilon \det(\{\mu_{i+j}^\varepsilon\}_{i,j=1}^q) + \det(\{\mu_{i+j}^\varepsilon \mathbf{1}_{i+j>1}\}_{i,j=0}^q)$$

and therefore

$$\bar{\Lambda}_\varepsilon \leq \frac{\left|\det(\{\mu_{i+j}^\varepsilon \mathbf{1}_{i+j>1}\}_{i,j=0}^q)\right|}{\det(\{\mu_{i+j}^\varepsilon\}_{i,j=1}^q)}$$

By integration by parts and Karamata's theorem (Theorem 1.5.11 in [3]), we show that

$$\lim_{\varepsilon\downarrow 0} \frac{\int_{(0,\varepsilon]} |y|^p \nu(dy)}{\varepsilon^p \int_{(\varepsilon,\infty)} \nu(dz)} = \frac{\alpha}{p-\alpha}, \quad \text{for all } p > \alpha.$$

and so

$$\limsup_{\varepsilon\downarrow 0} \frac{\bar{\Lambda}_\varepsilon}{\int_{|y|>\varepsilon} \nu(dy)} \leq \frac{\left|\det(\{\frac{\alpha}{i+j-\alpha}\mathbf{1}_{i+j>1}\}_{i,j=0}^q)\right|}{\det(\{\frac{\alpha}{i+j-\alpha}\}_{i,j=1}^q)}.$$

The matrix $(A_{ij})_{i,j=1}^m = (\frac{\alpha}{i+j-\alpha})_{i,j=1}^q$ is positive definite because

$$\langle z, Az\rangle = \int_0^1 x^{-\alpha-1}\Big(\sum_{i=1}^q z_i x^i\Big)^2 dx.$$

Therefore, $\det A > 0$ and there exits a constant $C < \infty$ such that

$$\bar{\Lambda}_\varepsilon \leq C \int_{|z|>\varepsilon} \nu(dz)$$

for ε sufficiently small.

To sum up, we have found that there exist two positive constants C_1 and C_2 such that for ε sufficiently small,

$$\begin{aligned} \mathcal{E}_m(\Lambda_\varepsilon) &\leq 2\int_{\{|y|\leq\varepsilon\}} y^m \nu(dy) \leq C_1 \varepsilon^m \int_{|y|>\varepsilon} \nu(dy) \\ \Lambda_\varepsilon &= \int_{|y|>\varepsilon} \nu(dy) + \bar{\Lambda}_\varepsilon \leq C_2 \int_{|y|>\varepsilon} \nu(dy). \end{aligned} \tag{16}$$

Let $\Lambda(\varepsilon) := \int_{|y|>\varepsilon} \nu(dy)$ and $\varepsilon(\Lambda) := \inf\{\varepsilon : \Lambda(\varepsilon) < \Lambda\}$. Since $\Lambda(\varepsilon) \in R_{-\alpha}$, as $\varepsilon \downarrow 0$, by Theorem 1.5.12 in [3], we also get that $\varepsilon(\Lambda) \in R_{-1/\alpha}$ as $\Lambda \to \infty$.

Now, for a given Λ, consider the measure 14 with $\varepsilon = \varepsilon(\Lambda/C_2)$, and possibly an additional atom at 0 to satisfy the intensity constraint. This measure satisfies the constraints of Problem $(\Omega_{m,\Lambda})$ and, by (16), has error bounded by

$$C_1 \varepsilon^m(\Lambda/C_2)\frac{\Lambda}{C_2} \sim C_1 C_2^{m/\alpha-1} \Lambda \varepsilon^m(\Lambda),$$

so that the upper bound of the theorem holds with $f(\Lambda) = \Lambda\varepsilon^m(\Lambda) \in R_{1-m/\alpha}$.

To compute the lower bound, observe that

$$\mathcal{E}_m(\Lambda) \geq \min_{\hat{\nu}\in\mathcal{M},\, \hat{\nu}(\mathbb{R})=\Lambda} J\left(\hat{\nu}\right),$$

and the explicit optimal solution for the problem in the right-hand side is given by

$$\nu_\varepsilon(dy) = \nu(dy)\mathbf{1}_{|y|>\varepsilon} + \xi\nu(dy)\mathbf{1}_{|y|=\varepsilon},$$

where ε and $\xi \in [0,1]$ are such that $\int_{|y|>\varepsilon} \nu(dy) + \xi\nu(\{|y| = \varepsilon\}) = \Lambda$. This is justified by an argument similar to the one in the proof of Proposition 21, with $P(y) = a_0 + |y|^m$ since there are no constraints except $\hat{\nu}(\mathbb{R}) = \Lambda$. In particular, ε coincides with $\varepsilon(\Lambda)$ introduced above. On the other hand, the error functional associated to this solution satisfies

$$J(\nu_\varepsilon) = \int_{\mathbb{R}} |y|^m |\nu - \nu_\varepsilon|(dy) \geq \int_{|y|<\varepsilon} |y|^m \nu(dy) \sim \frac{\alpha}{m-\alpha}\Lambda\varepsilon^m(\Lambda),$$

which proves the lower bound. ■

6 Description of the algorithm and numerical results

According to Section 5, our approach to find an optimal approximation for the Lévy measure starts by setting $\bar{\mu} = 0$ and $\bar{\sigma} = 0$. Hence, the solution of equation (3) between jumps satisfies the following equation

$$\bar{Y}_t(x) = x + \int_0^t \bar{b}(\bar{Y}_s(x))ds + \int_0^t \sigma(\bar{Y}_s(x))dB_s, \tag{17}$$

where

$$\bar{b}(x) = b(x) + \bar{\gamma}h(x),$$
$$\bar{\gamma} = -\int_{\mathbb{R}} y\bar{\nu}(dy).$$

This implies that the drift term of the continuous part will depend on $\bar{\nu}$ through the parameter $\bar{\gamma}$. Therefore, once we have fixed $\bar{\nu}$ the optimal approximation of the Lévy measure ν, we need to choose a weak approximation method to solve equation (17) . We will consider the following approaches:

- **Weak Taylor approximations**: These methods are based on the Itô-Taylor expansion of the solution of (17). This expansion is the stochastic analogue of the classical Taylor expansion, where the role of polynomials is played by multiple iterated stochastic integrals. Truncating the expansion at a certain degree of the iterated integrals we obtain an approximation method with global order of convergence related to that degree, see Proposition 5.11.1 in [8]. We will consider the weak Taylor approximations with global order of convergence 1,2 and 3, which we will denote by WT1, WT2 and WT3. Although the method is conceptually simple to understand, it presents some difficulties in the implementation as we need to sample from the joint law of multiple stochastic integrals of different orders. This makes the method less appealing from a practical point of view, especially when the driving Brownian motion is multi-dimensional.

- **Kusuoka-Lyons-Victoir methods**: These methods are also based on stochastic Taylor expansions. The idea is to approximate the expectation under the Wiener measure by the expectation under a probability measure supported on a finite number of paths of finite variation. By construction, the expectations of the iterated Stratonovich integrals, up to a certain degree, under this new measure match the expectations of the corresponding iterated integrals under the Wiener measure. Using the Stratonovich-Taylor formula one can deduce that the approximations obtained have global order of convergence depending on the degree of the iterated integrals taken into account, see [12]. In particular we will consider the approximation schemes of degree 3 and 5, denoted by KLV3 and KLV5, which give, respectively, global order of convergence 1 and 2. Deriving and implementing these methods is not straightforward, see [7] for an account on these issues.

- **Ninomiya-Victoir method**: The Ninomiya-Victoir method can be seen as a stochastic splitting method. The idea is to find suitable small time approximations of the semigroup associated to the solution of equation (17) . These approximations are written in terms of weighted products (compositions) of simpler semigroups associated to the so called coordinate processes and are deduced using formal Taylor expansions of the semigroups involved. The main difference with respect to the classical splitting methods is that, in the stochastic case, we need to find appropriate stochastic representations of the semigroups in order to implement the Monte Carlo method. These representations involve solving or approximating ODEs with random coefficients. We will consider the algorithm given by Ninomiya and Victoir in [14], which has global order of convergence 2.

By using similar techniques as in [20], one can show that the operators induced by the weak approximation schemes above satisfy assumptions $(\mathcal{M}_0)$, $(\mathcal{M})$ and $(\mathcal{R}_r)$. Having fixed an optimal Lévy measure $\bar{\nu}$ and a weak approximation scheme $\hat{Y}_t$ for the continuous part we can apply the following algorithm to obtain a sample of $\hat{X}_1$.

Algorithm to generate a weak approximation of $\hat{X}_1$

Requires:
The initial condition $x_0 \in \mathbb{R}^d$.
The optimal Lévy measure $\bar{\nu}$.
The weak approximation method $\hat{Y}_t(x)$, to solve $\bar{Y}_t(x), t \in (0,1], x \in \mathbb{R}^d$
Compute $\bar{\lambda} = \bar{\nu}(\mathbb{R})$ and $\bar{\gamma} = -\int_{\mathbb{R}} y\bar{\nu}(dy)$
Set $T_{last} = 0, x_{new} = x_0$
Simulate the next jump time $T \sim \text{Exp}(\bar{\lambda})$
While $(T < 1 - T_{last})$ **do**
{
Compute $\hat{Y}_T(x_{new})$
Simulate Δ, a jump from the Poisson random measure
with Lévy measure $\bar{\nu}$
Set $x_{new} = \hat{Y}_T(x_{new}) + h(\hat{Y}_T(x_{new}))\Delta$
Set $T_{last} = T$
Simulate the next jump time $T \sim \text{Exp}(\bar{\lambda})$
}
Compute $\hat{Y}_{1-T_{last}}(x_{new})$
Set $\hat{X}_1 = \hat{Y}_{1-T_{last}}(x_{new})$
Return $\hat{X}_1$

Applying, independently, the previous algorithm M times we obtain a sequence $\{\hat{X}_1^i\}_{i=1,\dots,M}$ and the Monte Carlo estimator of $\mathbb{E}[f(X_1)]$ is given by

$$\frac{1}{M}\sum_{i=1}^{M} f(\hat{X}_1^i).$$

We end this section with some numerical examples. We evaluate $\mathbb{E}[f(X_1)]$, where X is the solution of equation (1) with $b(x) \equiv \gamma_0 h(x)$ and $\sigma(x) \equiv \sigma_0 h(x)$. To approximate the Lévy process, we use the optimal schemes presented in section 5.1 with $m = 2$, $m = 3$ and $m = 4$, and denoted, respectively, by OA2, OA3 and OA4 in the examples below (here "OA stands for optimal approximation"). For solving the continuous SDE between the times of jumps, we use the schemes WT1, WT2, WT3, KLV3, KLV5 and NV mentioned above. Finally, the process Z is taken to be a CGMY process, which is a Lévy process with no diffusion component and Lévy density of the form

$$\nu(x) = C\frac{e^{-\lambda_-|x|}1_{x<0} + e^{-\lambda_+|x|}1_{x>0}}{|x|^{1+\alpha}}.$$

The third component of the characteristic triplet is chosen in such way that Z becomes a martingale. An algorithm for simulating the increments of Z is available [16], which makes it possible to compare our methods to the traditional Euler scheme. Also, this process satisfies the assumption (R_α) of the previous section, and allows us to illustrate the dependence of the convergence rates on the parameter α. Actually, combining Theorems 11 and 26 we have the following result.

Theorem 28 *Assume that the Lévy measure ν satisfies the condition (R_α), and that the approximation scheme is obtained by choosing $\bar{\nu}$ to be a solution of Problem $(\Omega_{m,\Lambda})$ for m even, $\bar{\sigma}^2 = 0$ and $\bar{\mu} = 0$, and that it satisfies the hypotheses in Theorem 11. Then, for $f \in C_p^{(2r+2)\vee m}$, there exist positive constants $K(x, \mathcal{A}, p, r) \in C_q$, $C(x) \in C_{p+1}$ and a slowly varying function l such that*

$$|\mathbb{E}[f(X_1)] - \mathbb{E}[f(\widehat{X}_1)]| \leq C(x) \|f\|_{C_p^m} l(\Lambda)\Lambda^{1-\frac{m}{\alpha}} + K(x, \mathcal{A}, p, r) \|f\|_{C_p^{2r+2}} \Lambda^{-r},$$

where $\Lambda = \bar{\nu}(\mathbb{R})$.

We use 10^6 simulation paths in all examples. For the Euler scheme, all values are computed using the same set of paths with $1, 2, 4, 8, 16, 32, 64, 128$ and 256 discretization intervals. For the optimal schemes, different paths are used for each point on the graph, and the different points are obtained by choosing the values of the parameter ε which correspond to the values of $\lambda_\varepsilon := \int_{\{|x|>\varepsilon\}} \nu(dx)$ in the range $[0.5, 1, 2, 4, 8, 16, 32]$. Also, the computing time for each point has been normalized by the standard deviation of the MC estimate, so that the times for all points correspond to the time required to get a standard deviation of 0.001. The variance of the MC estimate is about the same for all values computed with the optimal schemes. For the Euler scheme, the variance may be different, because, on one hand, the simulation method from [16] makes use of a probability change which increases variance, and on the other hand, we use a variance reduction technique for the Euler scheme (by taking $\mathbb{E}[f(x + h(x)Z_1)]$ as control variate) but not for the other schemes. In all the numerical examples below we take $\gamma_0 = 0.5$, $\sigma_0 = 0.3$, $\lambda_+ = 3.5$ and $\lambda_- = 2$. Furthermore, for data set I, we take $C = 0.5$ and $\alpha = 0.5$ (finite variation jumps) and for data set II we take $C = 0.1$ and $\alpha = 1.5$ (infinite variation jumps). These two choices yield approximately the same variance of X_1 and allow us to understand the effect of α on the convergence rate.

For our first example, we take $h(x) = x$ and $f(x) = x$. In this case, X is simply the stochastic exponential of $\gamma_0 t + \sigma_0 W_t + Z_t$, and the exact value of $\mathbb{E}[f(X_1)]$ can be computed explicitly: $E[f(X_1)] = e^{\gamma_0}$. Figure 1 plots the errors of the KLV schemes of different degrees and the NV scheme on a log-log scale for data sets I and II. In this case, the three approximations of the Lévy measure, OA2, OA3 and OA4, have very similar performance and we only plot the results for OA2. This happens because with the choice $f(x) = h(x) = x$, we have $\mathbb{E}[f(\bar{X}_1)] = \mathbb{E}[f(X_1)]$ as soon as the approximation scheme for the Lévy measure preserves the expectation of the Lévy process, which is the case for all three approximation schemes OA1, OA2 and OA3. In other words, for this choice of f and h, the approximation of the Lévy measure does not introduce any error. The error is therefore exclusively determined by the approximation scheme which is used between the jump times. However, in this case, the KLV and NV methods perform so well that all the errors are below the statistical error due to the Monte Carlo method and it is not even possible to identify the actual order of convergence.

In our second example, we take $h(x) = x$ still and $f(x) = x^2$. The exact value of $\mathbb{E}[f(X_1)]$ can also be computed explicitly and is now equal to

$$\begin{aligned}
\mathbb{E}[X_T^2] &= \mathbb{E}[\mathcal{E}(2Z + [Z, Z])_T] = \exp\{E[2Z_T] + E[[Z, Z]_T]\} \\
&= \exp\left\{2\gamma_0 T + \sigma^2 T + T\int_{\mathbb{R}} y^2 \nu(dy)\right\} \\
&= \exp\left\{2\gamma_0 T + \sigma^2 T + TC\Gamma(2-\alpha)(\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2})\right\}.
\end{aligned}$$

Figure 2 plots the errors of the weak Taylor schemes of different orders on a log-log scale for data sets I and II, together with the theoretical error rates. In this case, one can clearly see the difference between the three schemes for approximating the Lévy measure (OA2, OA3 and OA4) as well as the effect of the parameter α.

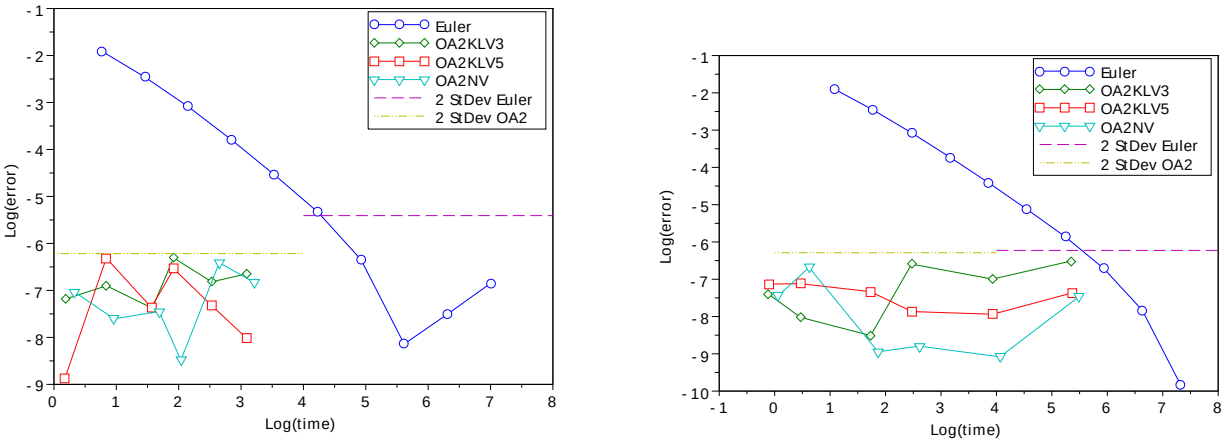

Figure 1: Errors of the cubature-based schemes for $h(x) = x$ and $f(x) = x$. Left: parameters from data set I. Right: parameters from data set II.

For $\alpha = 0.5$ (upper three graphs), the error of approximating the Lévy measure is of order of $\Lambda^{1-\frac{m}{\alpha}} = \Lambda^{-3}$ for OA2, Λ^{-5} for OA3 and Λ^{-7} for OA4. Therefore, in these graphs, the global error is dominated by the one of approximating the diffusion part: we observe a clear improvement going from WT1 to WT2 and WT3, and no visible change going from OA2 to OA3 and OA4.

On the other hand, in the lower left graph, which corresponds to $\alpha = 1.5$ and $m = 2$, the error of approximating the Lévy measure is of order of $\Lambda^{1-\frac{m}{\alpha}} = \Lambda^{-\frac{1}{3}}$, which dominates the error of approximating the continuous SDE for any of the three weak Taylor schemes, and determines the slope of the curves in this graph. In this context, using the optimal scheme with $m = 3$ (lower middle graph) or $m = 4$ (lower right graph) leads to an substantial improvement of performance. In this case, we observe similar behavior for $m = 3$ and $m = 4$ because the Lévy measure of Z is locally symmetric near zero, which means that 3-moment scheme and 4-moment scheme actually have the same convergence rate.

The theoretical error rate of the Euler scheme is always $\frac{1}{n}$, which corresponds to the straight solid line on the graphs. The observed convergence rates appears slower than the theoretical prediction due to our variance reduction method, which has better performance when the number of discretization dates is small.

Finally, for comparison, we have repeated the same simulation experiments taking the diffusion component equal to zero: $\sigma_0 = 0$. The results, shown in Figure 3, are very similar to those obtained with nonzero σ_0, emphasizing the robustness of our approach.

A A moment matching problem

In this section we present an auxiliary problem related with the moment matching of finite measures.

We want to compute $\inf_{\bar{\nu}\in\mathcal{M}_m} \bar{\nu}(\mathbb{R})$, with $\mathcal{M}_m$ defined in (11), i.e., the smallest intensity for which the moment constraints are feasible. This problem is very similar to the classical 'truncated Hamburger moment problem' which goes back to the works of Chebyshev, Markov and Stieltjes, the only difference being that we do not impose a constraint on the first moment. Known results for the Hamburger problem on an infinite interval can be summarized as follows [11]:

Proposition 29 *Let $m = 2q, q \in \mathbb{N}$ and let $\{\mu_k\}_{k=0}^m$ be given . There exists a measure $\bar{\nu} \in \mathcal{M}_m$ with $\bar{\nu}(\mathbb{R}) = \mu_0$ if and only if the matrix $\{\mu_{i+j}\}_{i,j=0}^q$ is nonnegative definite.*

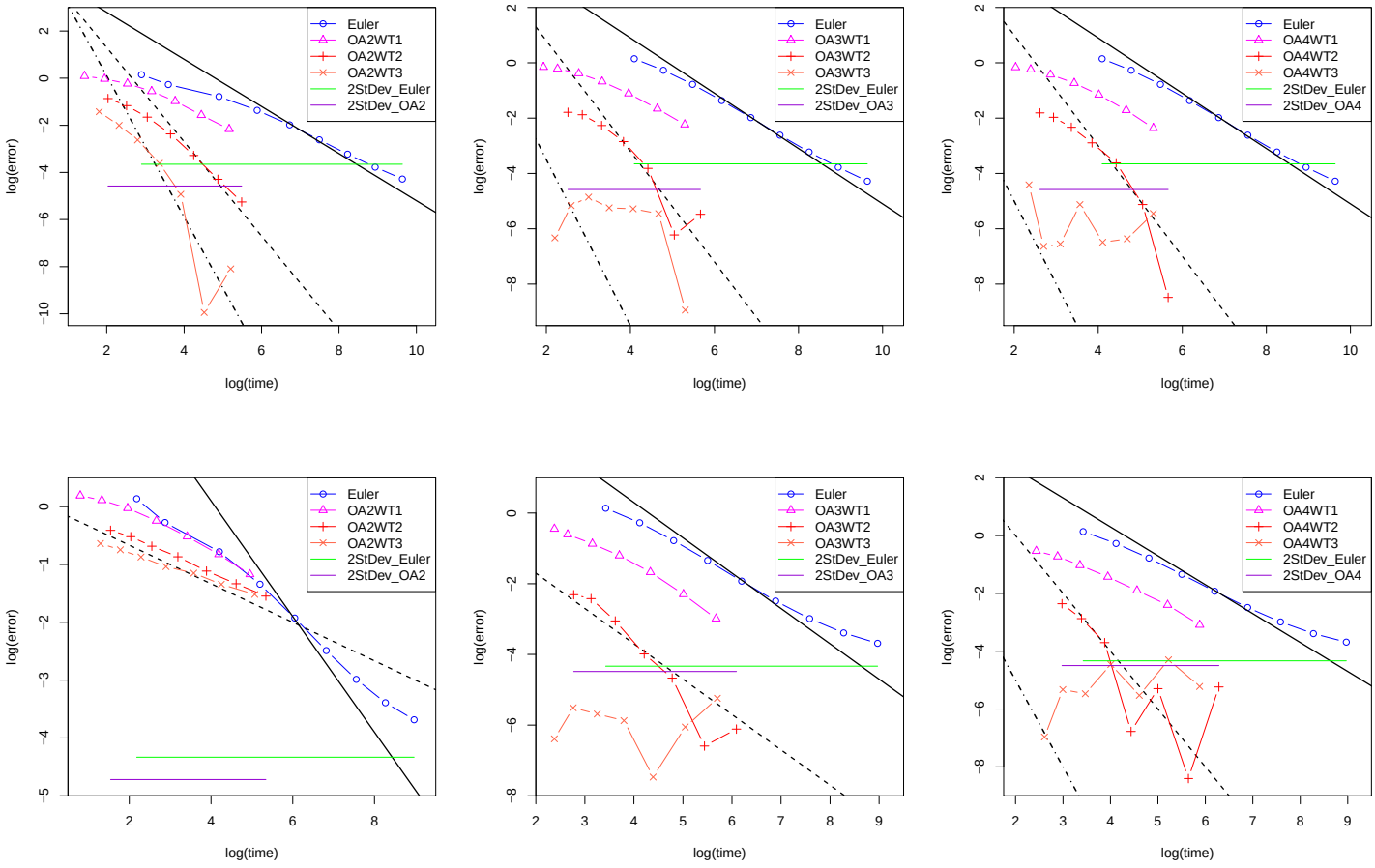

Figure 2: Errors of the weak Taylor schemes for $h(x) = x$ and $f(x) = x^2$. Top: parameters from data set I. Bottom: parameters from data set II.

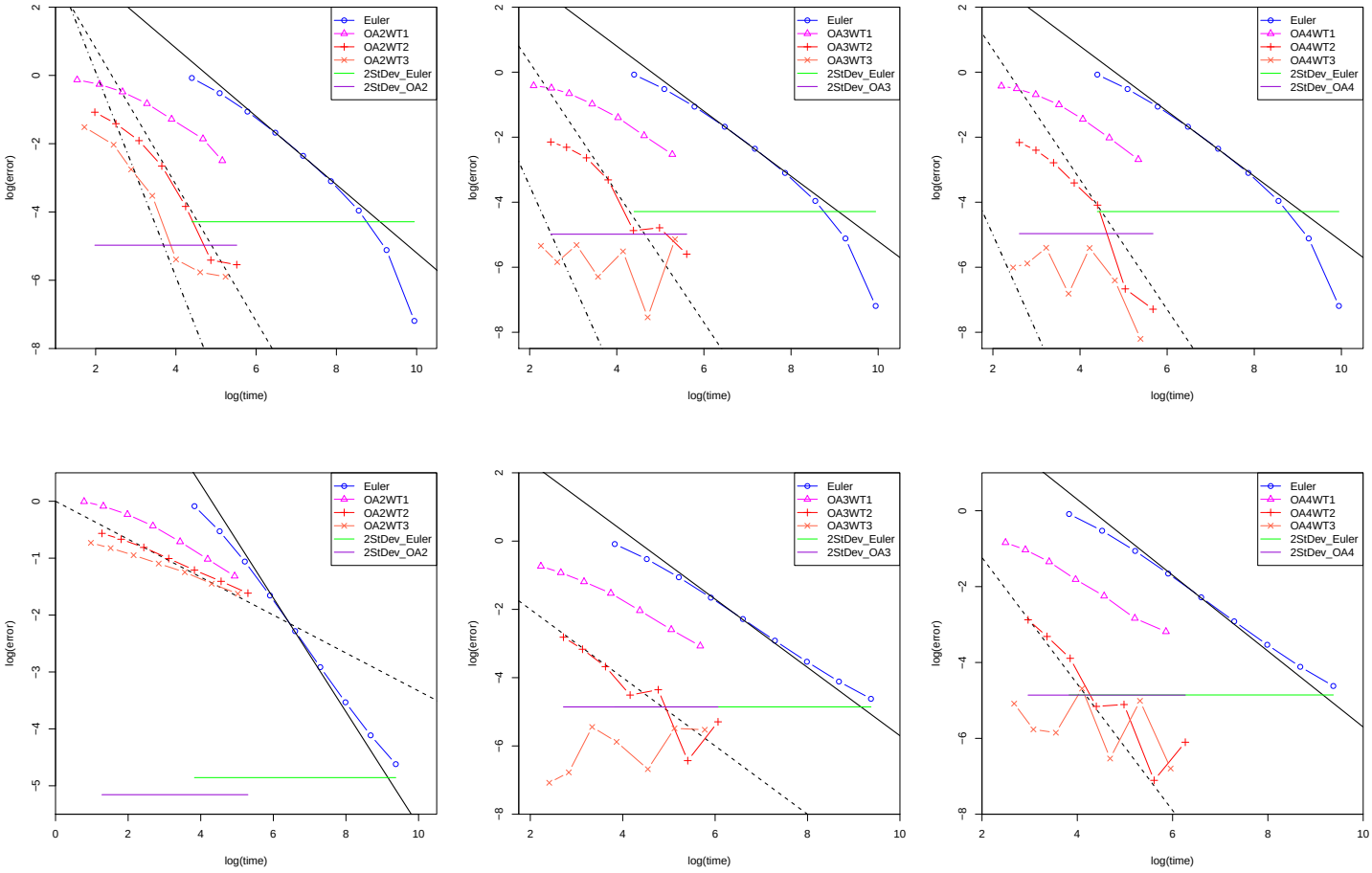

Figure 3: Errors of the weak Taylor schemes for $h(x) = x$ and $f(x) = x^2$. Top: parameters from data set I. Bottom: parameters from data set II. The diffusion coefficient σ_0 has been set equal to zero.

Corollary 30 *Let $m = 2q, q \in \mathbb{N}$, and let $\{\mu_k\}_{k=0}^m$ be given such that $\mu_k = \int_{\mathbb{R}} y^k \nu(dy)$, $2 \leq k \leq n$ for some nonnegative measure ν. Then there exists a measure $\bar{\nu} \in \mathcal{M}_m$ with $\bar{\nu}(\mathbb{R}) = \mu_0$ if and only if $\det(\{\mu_{i+j}\}_{i,j=0}^q) \geq 0$ for some $\mu_1 \in \mathbb{R}$.*

Proof. Using Proposition 29, it is enough to check that the matrix $\{\mu_{i+j}\}_{i,j=0}^q$ is nonnegative definite. By the definition of μ_k for $k = 2, ..., n$ we have that the matrix $\{\mu_{i+j}\}_{i,j=1}^q$ is nonnegative definite. Hence, by the Sylvester's criterion applied to the lower right corner minors of the matrix $\{\mu_{i+j}\}_{i,j=0}^q$, we have that in order for it to be nonnegative definite it is sufficient that $\det(\{\mu_{i+j}\}_{i,j=0}^q) \geq 0$. ■

Corollary 31 *For $(\mu_k)_{k=2}^m$ as in Corollary 30, the set of values μ_0 for which there exists a measure $\bar{\nu} \in \mathcal{M}_m$ with $\bar{\nu}(\mathbb{R}) = \mu_0$ is of the form $[\mu_0^*, \infty)$.*

The case when m is odd can be deduced from the previous one.

Corollary 32 *Let $m = 2q + 1, q \in \mathbb{N}$. There exists a measure $\bar{\nu} \in \mathcal{M}_m$ with $\bar{\nu}(\mathbb{R}) = \mu_0$ if and only if the matrix $\{\mu_{i+j}\}_{i,j=0}^{q+1}$ is nonnegative definite for some $\mu_1 \in \mathbb{R}$ and $\mu_{m+1} \in \mathbb{R}_+$.*

A simple matrix algebra computation then yields the following solutions for small m:

m	2	3	4	5
$\min_{\bar{\nu} \in \mathcal{M}_m} \bar{\nu}(\mathbb{R})$	0	0	$\frac{\mu_2^2}{\mu_4}$	$\frac{\mu_2^2}{\mu_4}$

B Some useful lemmas on the solutions of SDEs

In this section we will assume the notation established in the first section. In addition, $X_\cdot(t, x)$ will denote the flow associated with X, the solution of equation (1) .

Lemma 33 *Assume that, for some integer $p \geq 1$,*

$$\int_{\mathbb{R}} |y|^{p \vee 2} \nu(dy) < \infty, \quad \sup_{\bar{\nu} \in \mathcal{A}} \int_{\mathbb{R}} |y|^{p \vee 2} \bar{\nu}(dy) < \infty,$$

$h, b, \sigma \in C_b^1(\mathbb{R})$. Then, there exists a constant $C > 0$, which does not depend on $\bar{\nu}$, such that

$$\mathbb{E}\left[\sup_{0 \leq t \leq 1} |X_t|^p\right] \leq C(1 + |x|^p),$$
$$\mathbb{E}\left[\sup_{0 \leq t \leq 1} |\bar{X}_t|^p\right] \leq C(1 + |x|^p).$$

The proof of the this lemma is a standard generalization of the proof for continuous SDE's if one uses Kunita's second inequality (see Corollary 4.4.24 in Applebaum [1]).

Lemma 34 *For integers $p \geq 2$ and $m \geq 1$ assume*

$$\int_{\mathbb{R}} |y|^{mp} \nu(dy) < \infty,$$

$h, b, \sigma \in C_b^m(\mathbb{R})$. Then for any multi-index α with $0 < |\alpha| \leq m$ we have

$$\sup_{x \in \mathbb{R}^d} \mathbb{E}\left[\sup_{t \in [0,1]} \left|\frac{\partial^\alpha}{\partial x^\alpha} X_1(t, x)\right|^p\right] < \infty.$$

Proof. Follows from Theorem 70, Ch. V in [17]. ■

Using the time invariance of Lévy processes one obtains the following result.

Lemma 35 *1. For $0 \leq t \leq s \leq 1, X_s(t,x)$ and $X_{s-t}(0,x)$ have the same law.*

2. For $0 \leq t \leq s \leq 1, \bar{Y}_s(t,x)$ and $\bar{Y}_{s-t}(0,x)$ have the same law, where $\bar{Y}_s(t,x)$ is the process defined in (4).

Lemma 36 *Let $u(t,x) = \mathbb{E}[f(X_1(t,x))]$, $t \in [0,1]$, $x \in \mathbb{R}^d$. Assume $(\mathcal{H}_{m,p})$ and $f \in C_p^m$, for some integers $m \geq 2$ and $p \geq 0$. Then $u \in C^{1,m}([0,1] \times \mathbb{R}), u$ is a solution of the following partial differential equation*

$$\begin{aligned} &\frac{\partial u}{\partial t}(t,x) + b_i(x)\frac{\partial u}{\partial x_i}(t,x) + \frac{1}{2}\sigma_{ik}\sigma_{jk}(x)\frac{\partial^2 u}{\partial x_i \partial x_j}(t,x) \\ &+ \int_{|y|\leq 1}\{u(t,x+h(x)y) - u(t,x) - \frac{\partial u}{\partial x_i}(t,x)h_i(x)y\}\nu(dy) \\ &+ \int_{|y|>1}\{u(t,x+h(x)y) - u(t,x)\}\nu(dy) = 0 \qquad (18) \\ u(1,x) &= f(x) \end{aligned}$$

Furthermore, there exists $C < \infty$ with

$$\left|\frac{\partial^\alpha u}{\partial x^\alpha}(t,x)\right| \leq C\|f\|_{C_p^m}(1+|x|^p)$$

for all $t \in [0,1], x \in \mathbb{R}^d$ and $|\alpha| \leq m$.

Proof. For $d = 1$, the derivative $\frac{\partial u}{\partial x}$ satisfies

$$\frac{\partial u}{\partial x}(t,x) = \mathbb{E}\left[\frac{\partial f}{\partial x}(X_1(t,x))\frac{\partial}{\partial x}X_1(t,x)\right].$$

The interchange of the derivative and the expectation is justified using Lemma 34. The other derivatives with respect to x are obtained by successive differentiation under the expectation and the derivative with respect to t is obtained from Itô's formula applied to $f(X_1(t,x))$ using Lemma 35. Furthermore, one obtains by a direct estimation the polynomial growth under $(\mathcal{H}_{m,p})$ using lemmas 33 and 34. This estimation involves the use of Hölder's inequality with multiple exponents and the solution of an optimization problem for these exponents in order to find the appropriate Hölder exponents allowing the use of the above mentioned lemmas under the hypothesis $(\mathcal{H}_{m,p})$. For the general case, $d \geq 2$, one uses the multivariate chain rule formula, known as Faà di Bruno's formula, and similar reasonings apply. ■

Remark 37 *For the integro-differential equation in Lemma 36 to make sense we have to impose the first order moment of ν to be finite. The finiteness of higher order moments of ν allow us to prove existence and boundedness properties of higher order derivatives of $u(t,x)$, using Lemmas 33 and 34. These properties of the derivatives of $u(t,x)$ are crucial to obtain the main result in this paper, the error expansion in Theorem 10.*

References

[1] D. Applebaum. *Lévy Processes and Stochastic Calculus.* Cambridge University Press, 2009.

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

[3] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular Variation.* Cambridge University Press, 1987.

[4] N. Bruti-Liberati and E. Platen. *Strong approximations of stochastic differential equations with jumps.* J. Comput. Appl. Math. **205**, 982–1001 (2007).

[5] S. Dereich. *Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction.* Ann. App. Proba. **21**, 283–331 (2011).

[6] J. Doob. *Measure Theory.* Graduate Texts in Mathematics. Springer-Verlag, 1994.

[7] L.G. Gyurkó and T.J. Lyons. *Efficient and practical implementations of cubature on Wiener space.* In: Stochastic Analysis 2010, D. Crisan (ed.), Springer, 2011.

[8] P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1992.

[9] A. Kohatsu-Higa and P. Tankov. *Jump-adapted discretization schemes for Lévy-driven SDEs.* Stoch. Proc. Appl. **120**, 2258–2285 (2010).

[10] J. Jacod, T. Kurtz, P. Protter and S. Méléard. *The Approximate Euler method for Lévy driven stochastic differential equations.* Ann. Inst. H. Poincaré. **41**, 523–558 (2005).

[11] M. G. Krein and A. A. Nudelman. *The Markov Moment Problem and Extremal Problems.* American Mathematical Society, 1977.

[12] T. Lyons and N. Victoir. *Cubature on Wiener space.* Stochastic analysis with applications to mathematical finance. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **460**, 169-198 (2004)

[13] E. Mordecki, A. Szepessy, R. Tempone and G.E. Zouraris. *Adaptive weak approximations of diffusions with jumps.* SIAM J. Numer. Anal. **46**, 1732–1768 (2008).

[14] S. Ninomiya and N. Victoir. *Weak approximation of stochastic differential equations and application to derivative pricing*, Appl. Math. Finance **15**, 107–121 (2008).

[15] K. Oshima, J. Teichmann and D.Veluscek. *A new extrapolation method for weak approximation schemes with applications*, Ann. Appl. Probab., **22**, 1008–1045 (2012).

[16] J. Poirot and P. Tankov. *Monte Carlo option pricing for tempered stable (CGMY) processes.* Asia-Pacific Financial Markets, **13**, 327–344, 2006.

[17] P. Protter. *Stochastic Integration and Differential Equations: A New Approach.* Springer, 2005.

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

[19] S. Rubenthaler. *Numerical simulation of the solution of a stochastic differential equation driven by a Lévy process.* Stoch. Proc. Appl. **103**, 311–349 (2003).

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

[21] P. Tankov. *High order weak approximation schemes for Lévy-driven SDEs.* In: Proceedings of the 9^{th} International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Springer (to appear).