Numerical Solution of Jump-Diffusion SDEs

Alex Shkolnik* Kay Giesecke† Gerald Teng‡ Yexiang Wei§

April 6, 2017
This Version: March 6, 2020¶

Abstract

This paper formulates and analyzes a discretization scheme for a jump-diffusion process with general state-dependent drift, volatility, jump intensity, and jump size. The jump times of the process are constructed as time-changed Poisson arrival times, and the Euler method is used to generate the process between the jump epochs. Under conditions on the coefficient functions specifying the process, the scheme is proved to converge weakly with order one for functions with polynomial growth. The use of higher-order methods in between the jumps does not generally improve the weak order of convergence. Extensive numerical experiments illustrate our results.

[1]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106. EMAIL: shkolnik@ucsb.edu

[2]Department of Management Science & Engineering, Stanford University, Stanford, CA 94305. EMAIL: giesecke@stanford.edu.

[3]PDT Partners, New York, NY 10019. EMAIL: geraldteng@gmail.com

[4]Cutler Group, San Francisco, CA 94104. EMAIL: yxwei@alumni.stanford.edu.

¶Portions of this work were completed while Teng and Wei were doctoral students at the Department of Management Science & Engineering at Stanford University. We are grateful to Peter Glynn for several insightful discussions on discretization schemes for SDEs, George Papanicolaou, Tze Leung Lai, Isabelle Camilier, Philip Protter, Gustavo Schwenkler, Dmitry Smelov and to participants of the 2012 SIAM Conference on Financial Mathematics and Engineering in Minneapolis for useful comments. We also thank two anonymous referees for their insightful comments. Zihan Lin provided excellent research assistance, for which we are very grateful.

1. Introduction

Jump-diffusion processes are widely used in finance and economics. They serve as models for asset, commodity and energy prices, interest and exchange rates, and the timing of corporate and sovereign defaults. The probability distributions of jump-diffusions are rarely analytically tractable, so Monte Carlo simulation methods are often used to treat the pricing, risk management, and statistical estimation problems arising in applications of jump-diffusion models. This paper formulates a discretization scheme for multi-dimensional jump-diffusion processes with general state-dependent drift, volatility, jump intensity and jump magnitude, and analyzes its weak convergence properties.

The scheme constructs the jump times of the process as time-changed Poisson arrival times. This approach is based on a result of Meyer (1971), which states that under mild conditions, a point process is a Poisson process after a change of time defined by its cumulative intensity (or compensator). The Euler method is used to construct the compensator and the process in between the jump times. We prove that under smoothness and (hypo-) ellipticity conditions on the jump-diffusion coefficient functions, the scheme converges weakly with order one for functions with polynomial growth that need not be smooth. Using a higher-order method for discretizing the process between the jump times does generally not improve the weak order of convergence. This is because a higher-order method does not impact the order of the error associated with the timing of the jumps. Numerical tests for affine jump-diffusion models of the short rate and of default timing illustrate the convergence behavior of our scheme.

Prior work has studied the numerical solution of stochastic differential equations. Kloeden & Platen (1999) review the literature on diffusion models. Bally & Talay's (1996) work on diffusions is particularly relevant to us. They prove that under smoothness and hypoellipticity conditions on the drift and volatility functions, the Euler scheme converges weakly with order one for functions with polynomial growth. We extend their arguments to treat the case with state-dependent jumps. The key to our approach is the introduction of a certain measure tied to the diffusion process that describes the behavior of the jump-diffusion between the jumps. This measure represents a "discounted" transition law of the inter-jump diffusion, with discount rate given by the jump intensity function evaluated at the inter-jump diffusion. Results from Kusuoka & Stroock (1985) yield locally Gaussian bounds on the density of this measure, which are then used to bound the error of the discretization scheme under conditions that are natural extensions of those that Bally & Talay (1996) impose on the diffusion they study.

The treatment of jump-diffusion processes has received less attention than that of diffusions. Platen & Bruti-Liberati (2010) review the work in this area. For a jump-diffusion with constant jump intensity, Mikulevicius & Platen (1988) propose a hierarchy of schemes which are shown to have arbitrarily high weak order of convergence when the coefficient functions are sufficiently smooth. In particular, the Euler scheme is shown to converge with weak order one. For the same class of processes, Kubilius & Platen (2002) demonstrate the order one weak convergence of the Euler scheme when the coefficient functions are Hölder continuous, and Liu & Li (2000) analyze schemes of any weak order that, unlike the aforementioned methods and ours, are based on time discretizations that do not include the jump times. Maghsoodi (1998), Mordecki, Szepessy, Tempone & Zouraris (2008), and others study schemes for a jump-diffusion with time-varying but deterministic jump intensity. With a state-independent intensity, the jump times can be generated without error, independently of the diffusion. The only source of error is the discretization of the process between the jumps.

Glasserman & Merener (2003) treat a jump-diffusion with a state-dependent jump intensity. Assuming a uniformly bounded intensity function, they construct the jump times by thinning a Poisson process using state-dependent thinning probabilities. Between the jumps the process is constructed using a discretization method. Glasserman & Merener (2004) show that this scheme has the same weak order of convergence as the discretization scheme used in between the jump times. Thus, in contrast to our time-change scheme, combining thinning with higher-order methods improves the overall convergence rate. On the other hand, our time-change approach avoids the uniform bound on the jump intensity function that the thinning scheme of Glasserman & Merener (2003) requires. Therefore, our method and convergence results also apply to jump-diffusions with unbounded state-dependent intensities. Models with unbounded intensities are common in finance and other areas. Examples include the affine jump-diffusions of Duffie, Pan & Singleton (2000), the linear-quadratic jump-diffusions of Cheng & Scaillet (2007), the local volatility model of Carr & Madan (2010), the convertible bond model of Andersen & Buffum (2003), the non-parametric short rate model of Johannes (2004), the CEV jump-to-default model of Carr & Linetsky (2006), as well as many others. Protter & Talay (1997) prove that the Euler scheme for Lévy-driven jump-diffusion processes has weak order of convergence one, while Platen & Bruti-Liberati (2010) prove convergence results for a series of higher-order schemes. In this paper, we consider processes with finite-activity jumps rather than infinite-activity ones.

Casella & Roberts (2011), Giesecke & Smelov (2013), and Gonçalves &

Roberts (2014) develop exact sampling schemes for jump-diffusions with state-dependent drift, volatility, jump intensity, and jump size satisfying various sets of conditions. These schemes yield unbiased estimators for expectations of functions of a jump-diffusion, but are limited to one-dimensional processes.[1]

We highlight that our convergence result holds for (measurable) functions of a jump-diffusion that have polynomial growth. Unlike the aforementioned papers analyzing the convergence of discretization schemes for jump-diffusions, we do not require smoothness or boundedness. Therefore, our convergence result also covers a range of discontinuous or unbounded problems of interest, including the estimation of event probabilities. However, the degree of smoothness we require for the coefficient functions of the process tends to be greater than that required in the jump-diffusion discretization literature discussed above.

The time-change approach to simulating point processes has been widely used in applications. Norros (1986) and Shaked & Shanthikumar (1987) provide early discussions of this approach, focusing on the case where the underlying filtration is generated by the point process. Daley & Vere-Jones (2003) discuss the case of a general filtration. The time-change approach has also been used to construct the jump times in discretization schemes for jump-diffusions with state-independent (i.e., deterministic) jump intensities; see Mordecki et al. (2008), for example. Here, the compensator does not need to be discretized and the jump times are exact. To our knowledge, we are the first to formulate and analyze a discretization scheme employing the time-change approach to treat state-dependent intensities. Unlike for deterministic intensities, for state-dependent intensities the time-change approach requires the discretization of the compensator. With that, we only obtain approximate jump times whose error must be analyzed together with the discretization error in between the jumps.

The rest of this paper is organized as follows. Section 2 states the problem. Section 3 describes the implementation of the discretization scheme. Section 4 analyzes the weak convergence of the scheme. Section 5 discusses the convergence result and various extensions and limitations. Section 6 provides numerical results and Section 7 concludes. Technical appendices contain the proofs.

[1] Chen, Shkolnik & Giesecke (2019) develop an unbiased (but not exact) method for the simulation of a one-dimensional jump-diffusion with state-dependent coefficients. They leverage the unbiased estimators for diffusions developped in Rhee & Glynn (2015), Bally & Kohatsu-Higa (2015), Henry-Labordere, Tan & Touzi (2015) and Andersson & Kohatsu-Higa (2017).

2. Problem Formulation

2.1. Jump-diffusion. We consider a jump-diffusion process X that models the behavior of a vector of state variables such as share prices. The drift, volatility, jump arrival rate, and jump magnitude coefficients of X are state-dependent. Formally, we fix a complete probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and a right-continuous and complete filtration $(\mathcal{F}_t)_{t\geq 0}$. Let X be an adapted Markov process on the domain $\mathcal{D} \subseteq \mathbb{R}^d$ that solves the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t + \mathrm{d}J_t, \tag{1}$$

where X_0 is a constant, $\mu : \mathcal{D} \to \mathbb{R}^d$ is the drift coefficient, $\sigma : \mathcal{D} \to \mathbb{R}^{d\times m}$ is the volatility coefficient, W is a standard Brownian motion in $\mathbb{R}^m$, and

$$J_t = \sum_{n=1}^{N_t} \Delta(X_{T_n-}, Z_n), \tag{2}$$

where N is a non-explosive counting process with arrival times $(T_n)_{n\geq 1}$ and intensity $\Lambda(X)$ for some $\Lambda : \mathcal{D} \to (0, \infty)$. The function $\Delta : \mathcal{D} \times \mathcal{M} \to \mathcal{D}$ governs the jump magnitudes of X, and $(Z_n)_{n\geq 1}$ is a sequence of i.i.d. $\mathcal{F}_{T_n}$-measurable mark variables with law ν on $\mathcal{M}$, a subset of Euclidean space. We will impose conditions on the functions $\mu, \sigma, \Delta, \Lambda$ and the law ν that guarantee existence and (strong) uniqueness of solutions to the SDE (1). We remark that for suitable $f : \mathcal{D} \to \mathbb{R}$, the infinitesimal generator $\mathscr{A}$ of X takes the form

$$\begin{aligned}(\mathscr{A} f)(x) &= \mu(x)^\top \nabla f(x) + \frac{1}{2}\mathrm{tr}(\sigma(x)\sigma(x)^\top \nabla^2 f(x)) \\ &\quad + \Lambda(x) \int_{\mathcal{M}} \big(f(x + \Delta(x,z)) - f(x)\big)\, \nu(\mathrm{d}z).\end{aligned} \tag{3}$$

2.2. Examples. Stochastic models of the form (1) are ubiquitous in finance and other application areas. Below we provide a few specific choices of the parameter set $(\mu, \sigma, \Lambda, \Delta, \nu)$ representing several standard financial models.

Example 2.1. *The parameter set with $\mu(x) = \mu x$, $\sigma(x) = \sigma x$, $\Lambda(x) = \lambda$, $\Delta(x,z) = x(e^z - 1)$ for $\mu, \sigma, \lambda \in (0, \infty)$ and ν the normal distribution represents the classical jump-diffusion model of Merton (1976). Choosing ν to be the double-exponential distribution gives the jump-diffusion model of Kou (2002).*

The models in this example have a constant jump intensity which is trivially state-independent and bounded. Many standard models however have state-dependent and unbounded intensity, as the following examples illustrate.

Example 2.2. *The parameter set with* $\mu(x) = (r + \Lambda(x))x$*,* $\sigma(x) = ax^{\beta+1}$*,* $\Lambda(x) = b + ca^2x^{2\beta}$*,* $\Delta(x, z) = -x$ *for* $r, a, b > 0$ *and* $c \geq 1/2, \beta < 0$ *gives the jump-to-default extended CEV model of Carr & Linetsky (2006).*

Example 2.3. *Given* $m, d \geq 1$*, choosing the affine functions* $\mu(x) = K_0 + K_1 x$*,* $(\sigma(x)\sigma(x)^\top)_{ij} = (H_0)_{ij} + (H_1)_{ij} x$*,* $\Lambda(x) = \lambda_0 + \lambda_1 x$*,* $\Delta(x, z) = z$ *for* $(K_0, K_1) \in \mathbb{R}^d \times \mathbb{R}^{d\times d}$*,* $(H_0, H_1) \in \mathbb{R}^{d\times m} \times \mathbb{R}^{d\times d\times m}$*,* $(\lambda_0, \lambda_1) \in \mathbb{R} \times \mathbb{R}^d$ *yields the affine jump-diffusion model of Duffie et al. (2000).*

Example 2.4. *Selecting* $d = m = 1$ *and non-parametric* μ*,* σ *and* Λ *while requiring log-normal jumps of the form* $\Delta(x, z) = x(e^z - 1)$ *with* ν *the normal distribution, yields the non-parametric short rate model of Johannes (2004).*

2.3. Objective. Fix $T > 0$. Our goal is to estimate expectations of the form

$$\Pi = \mathrm{E}(g(X_T))$$

for suitable "payoff functions" $g : \mathcal{D} \to \mathbb{R}$. In financial applications, Π could represent the (undiscounted) price of a derivative paying $g(X_T)$ at time T.

Direct computation of Π has limited scope as the law of X_T is known only in special cases, such as the models in Example 2.1. Semi-analytic Fourier methods have a broader scope, and apply when the transform of X_T is known, as in the case of Example 2.3. If the transform is not tractable, one can attempt to numerically solve the partial integro-differential equation associated with Π. But, this may be computationally prohibitive, especially in higher dimensions.

Monte Carlo methods have the widest applicability, and are appropriate for high dimensional problems as well as nonparametric models (see Example 2.4). Here, Π is estimated by the average $\sum_{k=1}^{M} g(X_T^{(k)})/M$, where M is the number of simulated samples $\{X_T^{(k)}\}$ of X_T. The performance is evaluated based on the root-mean-squared error $\mathrm{RMSE} = \sqrt{\mathrm{Bias}^2 + \mathrm{SE}^2}$, for $\mathrm{Bias} = \mathrm{E}(g(X_T^{(1)})) - \Pi$ and SE, the sample standard error. If the samples $\{X_T^{(k)}\}$ of the process are exact (that is, if $X_T^{(1)}$ and X_T have the same law), then the simulation bias (Bias) is zero and the RMSE vanishes with rate $1/\sqrt{M}$. Exact sampling methods for jump-diffusions have been developed by Casella & Roberts (2011), Giesecke & Smelov (2013), Gonçalves & Roberts (2014) and Chen et al. (2019). However, these schemes are currently limited to one-dimensional stochastic processes.

An alternative to exact methods are approximate methods of generating samples of X_T based on a discretization of the time interval $[0, T]$ with step size $h > 0$. Discretization methods are relatively simple to implement and easily applicable to multi-dimensional problems. However, unlike exact methods they usually generate biased estimators of Π. In order to evaluate the performance of a discretization method it is important to understand how the simulation bias depends on h. A discretization X^h is said to have *weak order of convergence* β if

$$|\Pi - \mathrm{E}(g(X_T^h))| = O(h^\beta) \tag{4}$$

for g belonging to a suitable class of payoff functions. Here, $O(h^\beta)$ corresponds to the existence of a constant C, not depending on h, such that the left side of (4) is bounded above by Ch^β. Once the order $\beta > 0$ is known we can determine the optimal allocation of computational resources between M and h. Duffie & Glynn (1995) provide conditions guaranteeing that for a fixed computation budget B and a discretization with weak order of convergence β, setting $M \propto h^{-2\beta}$ gives the asymptotically optimal error convergence rate of $\mathrm{RMSE} \propto B^{-\beta/(1+2\beta)}$. Therefore, determining the weak order of convergence β for a discretization scheme X^h has important practical consequences for simulation efficiency.

3. Discretization scheme

We provide a method for discretizing the jump-diffusion SDE (1). The jump times are constructed from Poisson arrival times by a change of time. This approach is based on a result of Meyer (1971), which states that under mild conditions, a point process time-changed by its cumulative intensity (compensator) is a Poisson process. The Euler scheme (e.g., Kloeden & Platen (1999)) is used to construct the compensator as well as the process between the jump times.

We approximate paths of X on $[0, T]$ by an "Euler process" X^h, which is defined on a probability space $(\Omega^h, \mathcal{F}^h, \mathrm{P}^h)$ supporting a standard m-dimensional Brownian motion W^h, a standard Poisson process with arrival times $(S_n)_{n\geq 1}$, and an i.i.d. sequence of random variables $\{Z_n^h\}_{n\geq 1}$ with law ν, all mutually independent. The discretization times $0 = \tau_0 < \tau_1 < \cdots$ used to construct X^h is the superposition of the horizon T, a fixed time grid, and the random jump times $0 = T_0^h < T_1^h < \cdots$ of X^h. Formally, the discretization times $(\tau_i)_{i\in\mathbb{N}}$ are given by the ordered sequence of times in the set $\{nh \wedge T, T_n^h\}_{n\in\mathbb{N}}$, for $h > 0$.

For integers $n = 1, 2, \ldots$, we define the jump times of X^h by

$$T_n^h = \inf\left\{t > 0 : A_t^h \geq S_n\right\}, \tag{5}$$

where A^h is the piece-wise linear process given by $A_0^h = 0$ and

$$A_t^h = A_{\tau_i}^h + \Lambda(X_{\tau_i}^h)(t - \tau_i), \qquad \tau_i \leq t < \tau_{i+1}, \tag{6}$$

where $X_{\tau_i}^h$, the value of the Euler process at a discretization time, is specified in equation (8) below. The process A^h is an approximation of the compensator

$A_\cdot = \int_0^\cdot \Lambda(X_s)\,\mathrm{d}s$ of the process N counting the jumps of X in (2), obtained by taking the arrival intensity to be constant on $[\tau_i, \tau_{i+1})$ and equal to $\Lambda(X^h_{\tau_i})$. The associated process N^h counting the jumps of X^h may be constructed from the times $(T^h_n)_{n\geq 1}$. The definition in (5) constructs the jump times of X^h as time-changed Poisson arrival times, with the increasing process A^h acting as the time change. Due to (6), relation (5) can easily be inverted for the jump times.

We can now specify the process X^h. Between the τ_i, we approximate X using the continuous Euler scheme. We take $X^h_0 = X_0$ and

$$(7)\qquad X^h_t = X^h_{\tau_i} + \mu(X^h_{\tau_i})(t-\tau_i) + \sigma(X^h_{\tau_i})(W^h_t - W^h_{\tau_i}), \qquad \tau_i \leq t < \tau_{i+1}.$$

At the discretization times, we add $\Delta(X^h_{\tau_i-}, Z^h_n)$ if a jump occured, i.e,

$$(8)\qquad X^h_{\tau_i} = X^h_{\tau_i-} + \sum_{n\geq 1} \Delta(X^h_{\tau_i-}, Z^h_n) 1_{\{\tau_i = T^h_n\}}.$$

The Euler process X^h is a jump-diffusion whose drift, volatility, and jump intensity are time-invariant between the discretization times. On any $[\tau_i, \tau_{i+1})$, X^h follows a Brownian motion started at $y = X^h_{\tau_i}$, with drift $\mu(y)$ and volatility $\sigma(y)$. The process X^h is adapted to the right-continuous and complete filtration $(\mathcal{F}^h_t)_{t\geq 0}$ generated by the marked point process (T^h_n, Z^h_n) and the W^h.

Algorithm 1 describes the steps for generating a skeleton of X^h over $(\tau_i)_{i\geq 1}$. and returns a sample of $g(X^h_T)$. By convention, we take $\delta/0 = \infty$ for $\delta > 0$.

Algorithm 1 (Jump-Diffusion Discretization Scheme). Let $\{\mathcal{N}_i\}_{i\geq 1}$ and $(S_n)_{n\geq 1}$ be i.i.d. standard normal $\mathbb{R}^m$-vectors and Poisson arrival times, respectively. Let $\{Z^h_n\}_{n\geq 1}$ be an i.i.d. sequence of random variables with law ν. Initialize $h > 0$, the counters $n = i = \ell = 0$ and variables $\tau_0 = 0$, $X^h_0 = X_0$ and $A^h_0 = 0$.

1. Set $\tau_{i+1} = \big(\tau_i + (S_{n+1} - A^h_{\tau_i})/\Lambda(X^h_{\tau_i})\big) \wedge (\ell+1)h \wedge T$.
2. Set $X^h_{\tau_{i+1}-} = X^h_{\tau_i} + \mu(X^h_{\tau_i})(\tau_{i+1} - \tau_i) + \sigma(X^h_{\tau_i})\sqrt{\tau_{i+1} - \tau_i}\,\mathcal{N}_{i+1}$.
3. If $\tau_{i+1} = (\ell+1)h \wedge T$ then set $X^h_{\tau_{i+1}} = X^h_{\tau_{i+1}-}$ and $\ell \leftarrow \ell + 1$.

   Else, set $X^h_{\tau_{i+1}} = X^h_{\tau_{i+1}-} + \Delta(X^h_{\tau_{i+1}-}, Z^h_{n+1})$ and $n \leftarrow n+1$.
4. Set $A^h_{\tau_{i+1}} = A^h_{\tau_i} + \Lambda(X^h_{\tau_i})(\tau_{i+1} - \tau_i)$ and $i \leftarrow i+1$.
5. If $\tau_i = T$ then return $g(X^h_T)$. Else go to step 1.

4. Convergence analysis

We analyze the rate of convergence of the approximation X^h to the jump-diffusion process X. In particular, we provide conditions on $(\mu, \sigma, \Lambda, \Delta, \nu)$ and the payoff function g in (4) guaranteeing a weak order of convergence $\beta = 1$.

We follow a relatively standard approach to bound the error of the approximation, which for fixed $T > 0$ and $h > 0$ is given by

$$\mathcal{E}_g^h(x) = \mathrm{E}_x^h(g(X_T^h)) - \mathrm{E}_x(g(X_T)), \quad x \in \mathcal{D}. \tag{9}$$

Here, E_x denotes expectation with respect to P_x, indicating that initially $X_0 = x$. The symbols E_x^h and P_x^h are defined analogously. We introduce a function u on $\mathcal{D} \times [0, T]$ given by $u(x, t) = \mathrm{E}_x(g(X_{T-t}))$. Since $u(X_T^h, T) = g(X_T^h)$,

$$\mathcal{E}_g^h(x) = \mathrm{E}_x^h\Big(\sum_{\tau_i \le T} u(X_{\tau_i}^h, \tau_i) - u(X_{\tau_{i-1}}^h, \tau_{i-1})\Big). \tag{10}$$

We apply Itô's formula to the summands, and bound the terms. We will show that the following conditions are sufficient to execute this program.

Assumption 4.1. *The functions μ, σ and Λ are infinitely differentiable with bounded derivatives (the functions themselves need not be bounded).*

Assumption 4.2. *The function σ satisfies $\inf_{x \in \mathcal{D}} |\sigma(x)\sigma(x)^\top| > 0$.*

Assumption 4.3. *The function Δ satisfies $\inf_{(x,z) \in \mathcal{D} \times \mathcal{M}} |I + \nabla_x \Delta(x, z)| > 0$.*

Assumption 4.4. *The law ν has compact support on $\mathcal{M}$. For each $z \in \mathcal{M}$, the function $\Delta(\cdot, z)$ is infinitely differentiable. Let $V_i(x, z)$ denote any partial derivative of $\Delta_i(x, z)$ with respect to x of any order greater than one. Every $\sup_{z \in \mathcal{M}} V_i(\cdot, z)$ is measurable and $(1 \vee |\cdot|) \sup_{z \in \mathcal{M}} V_i(\cdot, z)$ is bounded on $\mathcal{D}$.*

These assumptions are typical of the conditions used in the literature studying the convergence of discretization schemes. They also guarantee existence and (strong) uniqueness of solutions to (1). Assumption 4.2 is a uniform ellipticity condition for $\sigma\sigma^\top$. It implies the significantly weaker uniform hypoellipticity condition of Bally & Talay (1996) and Kusuoka & Stroock (1985), which is stated in Appendix A.0 and which we will use in the proofs. We state the stricter condition 4.2 because of its simplicity. Under Assumption 4.1, uniform hypoellipticity can be further relaxed to the UFG condition in Kusuoka (2003) and Crisan & Delarue (2012). Together with Assumption 4.2, Assumption 4.3 is a uniform ellipticity condition for $(I + \nabla_x \Delta)\sigma\sigma^\top(I + \nabla_x \Delta)^\top$. It implies a weaker uniform hypoellipticity condition (see Appendix A.0), which we will use in the proofs. Assumption 4.4 controls the magnitude of the jumps of X and ensures that X is non-explosive and has finite moments. The compact support condition for ν is imposed for clarity in the exposition and can be relaxed (see Appendix ??).

Proposition 4.5. *Suppose Assumptions 4.1–4.4 hold and that the payoff function g is measurable and has polynomial growth. Then, for any multi-index $\alpha \in \mathbb{N}^d$ and $k \in \mathbb{N}$ the partial derivatives $\partial_t^k \partial_x^\alpha u(x,t)$ exist and are continuous for all $(x,t) \in \mathcal{D} \times [0,T)$ and have polynomial growth in x for each fixed $t \in [0,T)$. Moreover, the generator $\mathscr{A}$ of X in (3) satisfies the Kolmogorov backward equation*

$$\begin{aligned} -\frac{\partial}{\partial t} u(x,t) &= \mathscr{A} u(x,t) \\ u(x,T) &= g(x) \end{aligned} \qquad (x,t) \in \mathcal{D} \times [0,T). \tag{11}$$

The proof of this result is given in Appendix A. The smoothness and growth properties that are established in Proposition 4.5 are key to our approach. They allow us to bound the approximation error via applications of Itô's formula to (10). To see where the smoothness of u comes from, consider the process Y describing the behavior of the jump-diffusion X between its jump times. This process is governed by the SDE $\mathrm{d}Y_t = \mu(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t^*$, where W^* is a Brownian motion in $\mathbb{R}^m$ independent of W in (1) and μ and σ are as in (1). Assumption 4.1 guarantees that Y is a strong Markov process. It has several other implications. For Borel sets $\mathscr{B} \subseteq \mathcal{D}$ and $x \in \mathcal{D}$, consider the measure

$$\mathscr{P}_t(x, \mathscr{B}) = \mathrm{E}_x\left(e^{-\int_0^t \Lambda(Y_s)\,\mathrm{d}s} 1_{\{Y_t \in \mathscr{B}\}}\right). \tag{12}$$

To study the Euler scheme for a diffusion, Bally & Talay (1996) consider (12) in the special case that $\Lambda = 0$, when it is the transition law $\mathrm{P}_x(Y_t \in \mathscr{B})$. To treat the jump-diffusion case $\Lambda > 0$, we extend the transition law to the measure (12), which represents $\mathrm{P}_x(Y_t \in \mathscr{B}, \pi_t = 0)$, where π is a doubly-stochastic Poisson process with intensity $\Lambda(Y)$. The introduction of π accounts for the fact that X is a jump-diffusion rather than a diffusion as in Bally & Talay (1996). Further, Theorems 3.3 and 3.17 and Corollary 3.25 in Kusuoka & Stroock (1985) imply that, under Assumption 4.1 and the uniform hypoellipticity of $\sigma\sigma^\top$ ensured by Assumption 4.2, the measure (12) has a density with respect to Lebesgue measure whose partial derivatives exist and satisfy locally Gaussian bounds. These properties are sufficient to imply the smoothness of the function u above.

We can now apply Itô's formula to the summands in (10). The polynomial growth established in Proposition 4.5 allows us to bound each of the summands by an $O(|x|h^2)$ term, and since there are $O(1/h)$ summands in (10), the desired error bound follows. The proof of the following result is given in Appendix B.

Theorem 4.6. *Suppose that Assumptions 4.1–4.4 hold and assume that the payoff function g is measurable and has polynomial growth. We have,*

$$|\mathscr{E}_g^h(x)| \leq Ch(1 + |x|^Q)$$

for some positive constants C, Q that are independent of both h and x.

5. Discussion

5.1. Higher-order and exact methods. One could improve the approximation of X between the jump times. The Euler scheme (7) could be replaced by the second-order scheme of Milstein (1973) or one of the higher-order schemes studied by Kloeden & Platen (1999). Combining several order one schemes can often lead to a higher order method (for example, see Romberg extrapolation (Talay & Tubaro 1990))). In some cases, the Euler scheme can even be replaced by an exact scheme. For example, the SDE governing X between jumps might have an explicit solution or a known transition density, or allow the application of the exact rejection algorithms of Beskos & Roberts (2005), Chen & Huang (2013), Giesecke & Smelov (2013), or others (e.g., Blanchet & Zhang (2020)).[2]

However, the use of higher-order or exact methods does not necessarily improve the overall weak order of convergence. This is because the process A^h, which is used to sample the approximate jump times, is still generated using the Euler scheme (6). The order of the error associated with the timing of the jumps remains unchanged. In the degenerate case $\Delta = 0$, the jumps do not affect X, and X^h clearly inherits the convergence properties of the simulation method used between the jumps. While using a higher-order or exact scheme in case $\Delta > 0$ may not improve the overall order of convergence, it may still improve the simulation bias relative to the simple Euler scheme. However, any reduction in the simulation bias should be evaluated against the additional computational cost to determine whether the tradeoff is worthwhile in practice. See Section 6.

To improve the overall rate of convergence we should use a higher-order scheme to approximate the joint process (X, A), where $A_\cdot = \int_0^\cdot \Lambda(X_s)\,\mathrm{d}s$ is the compensator of the process counting the jumps of X. By recursively applying Itô-Taylor expansions, we find the second-order analogue of (6) for A^h. For

[2]See also unbiased simulation estimators for diffusions; e.g., Wagner (1989), Bally & Kohatsu-Higa (2015), Henry-Labordère, Tan & Touzi (2017) and Chen, Shkolnik & Giesecke (2020).

$\tau_i \le t < \tau_{i+1}$, it is given by

$$\begin{aligned} A_t^h = A_{\tau_i}^h &+ \Lambda(X_{\tau_i}^h)(t-\tau_i) \\ &+ \frac{1}{2}\Big(\nabla\Lambda(X_{\tau_i}^h)^\top \mu(X_{\tau_i}^h) + \frac{1}{2}\mathrm{tr}\big(\sigma(X_{\tau_i}^h)\sigma(X_{\tau_i}^h)^\top \nabla^2\Lambda(X_{\tau_i}^h)\big)\Big)(t-\tau_i)^2 \\ &+ \nabla\Lambda(X_{\tau_i}^h)^\top \sigma(X_{\tau_i}^h)\int_{\tau_i}^t W_s^h \, \mathrm{d}s\,. \end{aligned}$$

However, there are several complications that arise from using higher-order schemes to approximate the compensator. One issue is that unlike A^h in (6), the higher-order approximation specified above is no longer guaranteed to be non-decreasing due to the Brownian path integral. Although this could be resolved numerically (by imposing an absorbing or reflecting boundary on the Brownian motion, for example), the proof of Theorem 4.6 relies on this property.

Another issue involves computing T_n^h from the higher-order approximation. One approach would be to sample A^h at each point jh and determine whether the $T_n^h \in [jh, (j+1)h)$ by checking if $A_{(j+1)h}^h > S_n$. We must then determine jointly $(X_{T_n^h}^h, T_n^h)$ conditioned on $(A_{jh}^h, A_{(j+1)h}^h)$. This is equivalent to sampling the hitting time and point of a Brownian bridge, which can be very challenging (particularly in multiple dimensions; see Buchmann & Petersen (2006)).[3]

5.2. The discrete Euler scheme. Another modification of the scheme described in Section 3 is the discrete Euler approximation. This method is the same as continuous Euler except that instead of (5), we take $T_n^h = \min\{jh : A_{jh}^h \ge S_n\}$. This means we only have to generate (X^h, A^h) at the points jh, leading to a simpler algorithm that is slightly faster than the continuous Euler scheme.

The discrete Euler scheme has the same weak order of convergence as the continuous Euler scheme. But in practice, the bias of the discrete scheme may be greater than that of the continuous scheme so that the faster simulation speed may not be worthwhile. See Section 6 for numerical results illustrating this.

5.3. Point processes. Consider a vector point process L of the form

$$L_t = \sum_{n=1}^{N_t} \Gamma(X_{T_n-}, Z_n), \tag{13}$$

where $\Gamma : \mathcal{D} \times \mathcal{M} \to \mathbb{R}^k$ is a suitable function, $k \ge 1$, and N, X, and Z_n are as defined in Section 2. Definition (13) is analogous to the specification (2) of

[3]These issues are exacerbated further if we wish to simulate the X exactly conditioned on the approximate stopping time.

the point process J governing the jumps of X. The process L jumps at the same times as J does. In fact, L might even represent some of the elements of J.

We are interested in estimating expectations of the form $\mathrm{E}(q(L_T))$ for suitable functions $q : \mathbb{R}^k \to \mathbb{R}$. An expectation of this form is relevant in the modeling of portfolio credit risk, for example. In this context, L would represent the cumulative loss due to default in a portfolio of credit-sensitive positions, and $q(L_T)$ could represent the payoff of a credit derivative at maturity T.

To treat the estimation of $\mathrm{E}(q(L_T))$ in our framework, we could include the elements of L in the jump-diffusion X and select the payoff function g appropriately. Then Algorithm 1 is directly applicable to estimating $\mathrm{E}(q(L_T))$. However, Theorem 4.6 cannot be used to establish the weak order of convergence in this case. The reason is that Assumption 4.2, as well as the weaker uniform hypoellipticity condition (see Appendix $A.0$), are violated when L is an element of X. We would nevertheless expect the error to be of weak order one for suitable q. This behavior is also suggested by our numerical experiments in Section 6 below.

Giesecke & Shkolnik (2020) develop and analyze a change of probability measure that can potentially reduce or even eliminate the error of the discretization scheme when estimating expectations for point processes. They focus on a system of indicator point processes that represents a model for the timing of a finite number of distinct events. The measure change turns the compensator of the event counting process into a piecewise linear process that is easily inverted. It eliminates the need to approximate an event time on a discrete-time grid. This approach can potentially be extended to our jump-diffusion setting, which involves a non-terminating point process N that counts the jumps of X. In this setting, the change of measure would transform the compensator $A = \int_0^{\cdot} \Lambda(X_s)\mathrm{d}s$ of N into a piecewise linear process which does not need to be approximated on a discrete-time grid. The measure change would also pave the way for harnessing the improved accuracy of higher-order schemes for constructing X in between the event times. We leave the analysis of this approach for future research.

6. Numerical results

This section provides numerical results that illustrate the performance of our discretization schemes and support our convergence results. We estimate various quantities of interest for an affine model $X = (X^1, \ldots, X^d)$ governed by

$$\mathrm{d}X_t^i = \kappa_i(\theta_i - X_t^i)\mathrm{d}t + \sigma_i\sqrt{X_t^i}\mathrm{d}W_t^i + \delta_i \mathrm{d}J_t^i, \quad X_0^i > 0, \tag{14}$$

where $\kappa_i, \theta_i, \sigma_i > 0$ are constants satisfying $2\kappa_i\theta_i > \sigma_i^2$, each $\delta_i \geq 0$ and jump process $J^i = \sum_{n=1}^N Z_n^i$ for $Z_n^i \geq 0$. The law ν of $Z_n = (Z_n^1, \ldots, Z_n^d)$ is uniform on some compact set. The intensity function $\Lambda(x) = \lambda_0 + \lambda_1 \cdot x$ for constants $\lambda_0 \geq 0$ and $\lambda_1 \in \mathbb{R}^d_+$, so the frequency of the jumps can depend on all elements of X, generating a non-trivial correlation structure for the vector X.

Due to the square-root volatility term, the model in (14) violates Assumptions 4.1 and 4.2. It does not satisfy the weaker uniform hypoellipticity condition of Kusuoka & Stroock (1985) either. However, the discretization scheme can still be implemented in this case. The numerical results support our convergence Theorem 4.6 even though several of its assumptions are violated.

We test several variants of the discretization method. We consider the *continuous Euler scheme* described in Section 3, the *continuous exact scheme* described in Section 5.1, the *discrete Euler scheme* described in Section 5.2, and the *discrete exact scheme*.[4] The latter entails the exact sampling of X in between the jumps times as in Section 5.1 and the specification of the approximate jump times as in Section 5.2. The exact schemes use the fact that between the jump times, X is a vector of independent square-root (Feller) diffusions so the conditional distribution of each X^i_t is non-central chi-squared. We use the method in Section 3.4 of Glasserman (2003) to sample from that transition law.

We analyze the performance of a scheme by estimating the simulation bias (BIAS), standard error (SE), and root mean square error (RMSE) of the corresponding estimator for values of the discretization spacing h ranging between $2^{-4} = 0.0625$ and $2^{-8} \approx 0.0039$. We approximate the bias and the variance by their sample estimates using 2^{32} trials.[5] The "true" value Π required for the bias computation is estimated using the exact scheme of Giesecke & Shkolnik (2020), also with 2^{32} trials.[6] All estimates are accurate to the 5th significant digit within the 99% (Normal) confidence intervals. For each value of h, the SE is estimated by the square root of the variance over the asymptotically optimal number of trials. The asymptotically optimal number of trials is set to $100h^{-2}$ (see Duffie & Glynn (1995)). The RMSE estimates are based on those of the SE and BIAS. For each value of h, the simulation time is estimated by the product of $100h^{-2}$ and the average run time per trial. The latter is estimated from the original 2^{32} trials. The simulations are performed on a 2.5GHz Linux server. The code was written in C and compiled with `gcc` version 4.9.2 linked to GSL version 1.15.

[4]When an Euler scheme is used to generate X, then the simulated values in between jump times may become negative. For the parameter values we consider below, the probability of this event is negligible, and indeed we never observe this event in our simulations.

[5]We use the pairwise update method described in Chan, Golub & LeVeque (1983)) to estimate the sample variance with high precision. This method also circumvents the prohibitive memory requirements of that our large number of trials would entail otherwise.

[6]Alternatively, for some payoff functions the transform methods of Errais, Giesecke & Goldberg (2010) can be used to compute the true value semi-analytically.

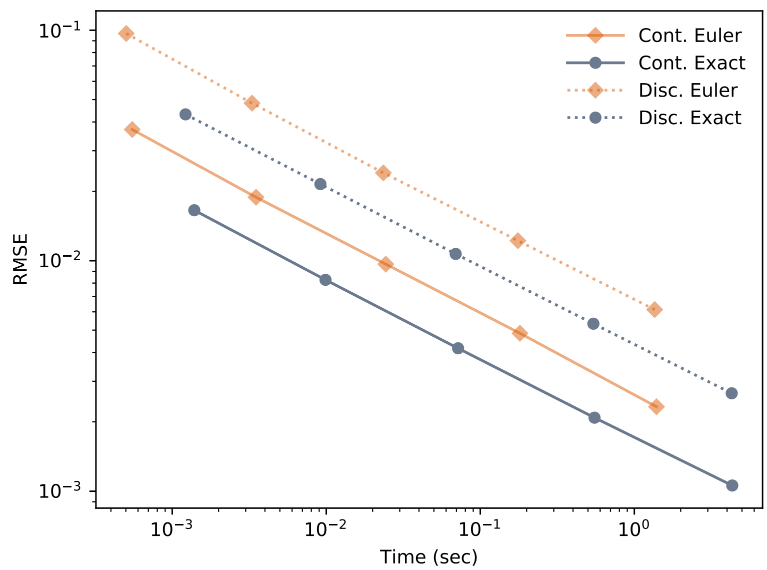

Figure 1. *Convergence of the* RMSE *of the interest rate cap problem in Table 1.*

6.1. Interest rate cap. For our first example, we take $d = 1$ and suppose that the X in (14) represents the short rate under a risk-neutral measure P. We estimate $\Pi = \mathrm{E}((X_T - K)_+)$, the undiscounted price of a cap with strike K and expiration T. The results, along with the parameter values, are in Table 1.

The convergence of the RMSE is shown as a log-log plot in Figure 1. All four variations of the discretization scheme give a RMSE roughly proportional to the inverse cube root of the simulation time, consistent with the theoretical results. The continuous exact scheme outperforms the others, and is followed by continuous Euler. The continuous exact scheme generates less bias than the continuous Euler scheme but runs significantly slower because exact sampling (in between the jumps) is computationally expensive. The discrete schemes runs about as fast as the continuous counterparts but generates significantly more bias. The discrete Euler scheme is outperformed by all other methods in terms of the RMSE. This indicates that the error from approximating the jump times by fixed grid points (see Section 5.2) can be the dominating source of bias.

6.2. Asian option. In the setting of Section 6.1, we estimate the undiscounted price of an Asian option with ℓ monitoring dates, strike K and expiration T, i.e., $\Pi = \mathrm{E}((\ell^{-1} \sum_{i=1}^{\ell}(X_{t_i} - K)_+)$. The results, and parameter values, are in Table 2. The convergence of the RMSE is shown as a log-log plot in Figure 2.

While the Asian option payoff is not directly covered by Theorem 4.6, the

Method	h	Bias ($\times 10^3$)	SE ($\times 10^3$)	RMSE ($\times 10^3$)	Time (ms)
Continuous Exact	2^{-4}	16.06	4.049	16.56	1.391
	2^{-5}	8.015	2.002	8.262	9.890
	2^{-6}	4.050	0.995	4.170	71.75
	2^{-7}	2.025	0.496	2.085	551.5
	2^{-8}	1.030	0.248	1.059	4323.
Continuous Euler	2^{-4}	−36.89	3.862	37.09	0.550
	2^{-5}	−18.76	1.952	18.86	3.500
	2^{-6}	−9.616	0.982	9.666	24.35
	2^{-7}	−4.820	0.493	4.845	180.9
	2^{-8}	−2.313	0.247	2.326	1394.
Discrete Exact	2^{-4}	−42.99	3.681	43.15	1.220
	2^{-5}	−21.42	1.909	21.50	9.163
	2^{-6}	−10.66	0.972	10.70	69.25
	2^{-7}	−5.309	0.491	5.331	542.4
	2^{-8}	−2.649	0.247	2.660	4287.
Discrete Euler	2^{-4}	−96.70	3.468	96.77	0.503
	2^{-5}	−48.21	1.856	48.25	3.300
	2^{-6}	−24.03	0.959	24.05	23.48
	2^{-7}	−12.21	0.487	12.22	175.5
	2^{-8}	−6.132	0.245	6.137	1356.

Table 1. *Interest rate cap. Parameters:* $T = 1.00$, $K = 1.00$, $X_0 = 2.00$, $\kappa = 2.50$, $\theta = 1.50$, $\sigma = 0.25$, $\nu = U\{0.50, 1.00\}$, $\delta = 1.00$, $lambda_0 = 0.00$ *and* $\lambda_1 = 1.00$. *The "true" value of* Π *is* 1.11804.

Method	h	Bias ($\times 10^3$)	SE ($\times 10^3$)	RMSE ($\times 10^3$)	Time (ms)
Continuous Exact	2^{-4}	4.093	6.355	7.559	1.495
	2^{-5}	2.475	3.172	4.024	11.24
	2^{-6}	1.035	1.585	1.893	82.09
	2^{-7}	0.942	0.792	1.230	632.7
	2^{-8}	0.257	0.396	0.472	4919.
Continuous Euler	2^{-4}	1.350	6.351	6.493	0.753
	2^{-5}	0.680	3.171	3.243	3.760
	2^{-6}	0.342	1.584	1.621	26.73
	2^{-7}	0.162	0.792	0.808	199.6
	2^{-8}	0.064	0.396	0.401	1487.
Discrete Exact	2^{-4}	−62.69	5.770	62.96	1.382
	2^{-5}	−31.71	3.026	31.86	9.833
	2^{-6}	−15.45	1.548	15.53	75.67
	2^{-7}	−7.974	0.783	8.012	618.0
	2^{-8}	−3.978	0.394	3.997	4663.
Discrete Euler	2^{-4}	−65.78	5.766	66.03	0.658
	2^{-5}	−32.65	3.025	32.79	4.553
	2^{-6}	−16.58	1.548	16.66	30.94
	2^{-7}	−7.841	0.783	7.880	192.6
	2^{-8}	−4.146	0.394	4.165	1555.

Table 2. *Asian option. Parameters:* $t_i = 0.25i$ *for* $i = 1, \dots, \ell = 4$, $K = 1.00$, $X_0 = 2.50$, $\kappa = 0.25$, $\theta = 0.50$, $\sigma = 0.25$, $\nu = \mathrm{U}\{0.50, 1.00, 1.50, 2.00, 2.50\}$, $\delta = 1.00$, $\lambda_0 = 0.00$ *and* $\lambda_1 = 0.25$. *The "true" value of* Π *is* 1.786715.

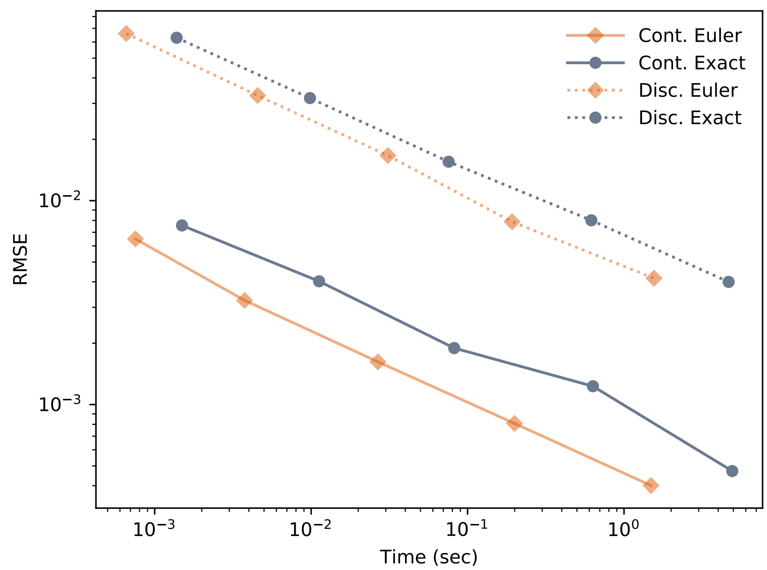

Figure 2. *Convergence of the* RMSE *of the Asian option problem in Table 2.*

numerical results suggest that the convergence of the discretization scheme is of weak order one. In contrast to the European contract, however, the Euler schemes outperform their exact schemes counterparts for the Asian contract. While the Euler schemes generate significantly more bias, their faster run time compensates to achive a superior RMSE. Additional numerical results (not reported here) highlight that the relative performance of the schemes strongly depends on the model parameter values. For some parameter values the exact schemes outperform the Euler schemes with respect to the RMSE metric.

6.3. Call on portfolio loss. To shed some insight into the behavior of the discretization method for point processes (see Section 5.3), we use (14) as a model for correlated default risk in a pool of firms. Here, the default arrival times are modeled by the jumps of X (we still take $d = 1$). In this setting, N is the default counting process, Z_n is the financial loss at the nth default, and J is the cumulative loss process. We estimate $\Pi = \mathrm{E}((J_T - K)^+)$, the undiscounted value of a call on the portfolio loss with strike K and maturity T (P continues to represent a risk-neutral measure). As shown in Errais et al. (2010), these calls are the basic building blocks of index and tranche swaps. To estimate Π, the schemes are applied to (X, J) rather than just X. Table 3 summarizes the results with the convergence of the RMSE plotted on a log-log scale in Figure 3.

All four schemes give RMSE convergence rougly proportional to the inverse

Method	h	BIAS ($\times 10^3$)	SE ($\times 10^3$)	RMSE ($\times 10^3$)	TIME (ms)
Continuous Exact	2^{-4}	-34.79	14.43	37.66	1.653
	2^{-5}	-17.14	7.231	18.60	10.79
	2^{-6}	-8.644	3.618	9.371	76.41
	2^{-7}	-4.217	1.810	4.589	571.8
	2^{-8}	-1.991	0.905	2.187	4415.
Continuous Euler	2^{-4}	-22.63	14.46	26.86	0.720
	2^{-5}	-11.33	7.238	13.44	4.192
	2^{-6}	-5.547	3.620	6.624	27.19
	2^{-7}	-2.888	1.810	3.398	192.3
	2^{-8}	-1.320	0.905	1.600	1440.
Discrete Exact	2^{-4}	-390.6	12.30	390.8	1.334
	2^{-5}	-201.9	6.682	202.0	9.614.
	2^{-6}	-102.6	3.479	102.6	71.90
	2^{-7}	-51.64	1.775	51.67	553.9
	2^{-8}	-25.94	0.986	25.96	4336.
Discrete Euler	2^{-4}	-379.5	12.33	379.7	0.617
	2^{-5}	-196.1	6.690	196.3	3.829
	2^{-6}	-99.74	3.481	99.80	25.68
	2^{-7}	-50.24	1.775	50.27	184.6
	2^{-8}	-25.11	0.896	25.12	1394.

Table 3. *Call on loss. Parameters:* $T = 1.00$, $K = 1.00$, $X_0 = 0.25$, $\kappa = 0.75$, $\theta = 2.50$, $\sigma = 0.50$, $\nu = U\{0.25, 0.50, 0.75, 1.00, 1.25\}$, $\delta = 1.00$, $\lambda_0 = 1.00$ *and* $\lambda_1 = 1.50$. *The "true" value of* Π *is* 1.93049.

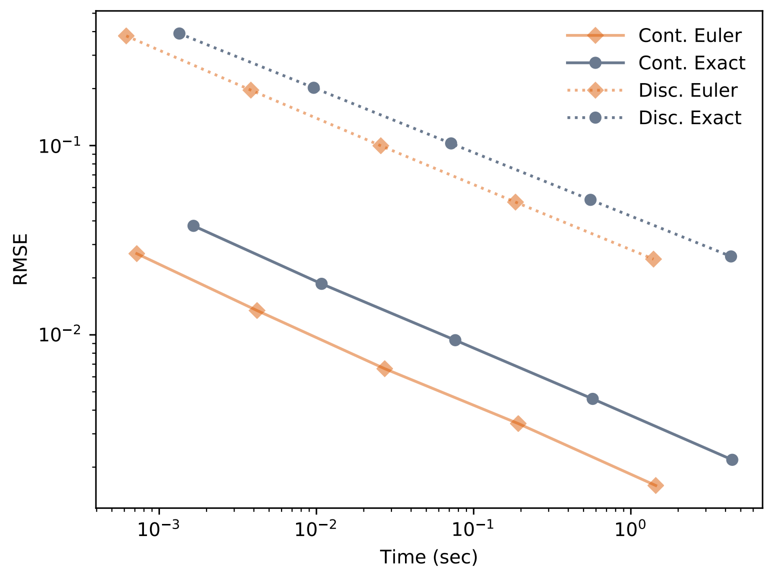

Figure 3. *Convergence of the* RMSE *of the call on loss problem in Table 3.*

cube root of the simulation time. This behavior suggests that the error converges weakly with order one for the call function of a point process driven by a jump-diffusion. Note that Theorem 4.6 does not guarantee this rate because it does not apply to point processes (see the discussion in Section 5.3 in this regard).

6.4. Loss probability. We illustrate the performance of the discretization scheme for a tail probability estimation problem. We consider $d = 4$, and let the law ν of Z_n be uniform on the set of 4-dimensional coordinate vectors. This means that at a jump time of N, we randomly select the element of X (and J) that jumps at that time. The magnitude of the jump of the ith element of X is given by the parameter δ_i (see Table 4) but the jumps of J are unit size. The model calibration takes the parameters in Sections 6.1–6.3 for the diffusions only (i.e., $X_0^i, \kappa_i, \theta_i, \sigma_i$). The 4th component parameters are specified in Table 4 and the jump parameters δ_i as well as λ_0 and $\lambda_1 = (\lambda_1^1, \ldots, \lambda_1^4)$ are also specified therein.

In a portfolio credit risk setting, N represents credit events in a portfolio of 4 firms. The process J^i represents the losses due to credit events associated with firm i. Note that credit events are allowed to influence the dynamics of X, and hence the event frequency. That is, there are self- and cross-excitation effects.

We estimate $\Pi = \mathrm{P}(\frac{1}{4}\sum_{i=1}^{4} J_T^i > K) = \mathrm{P}(N_T > 4K)$, the probability of the average loss in the portfolio exceeding the level K by time T. Note that the discontinuous payoff used here is covered by our main convergence result.

Method	h	Bias ($\times 10^3$)	SE ($\times 10^3$)	RMSE ($\times 10^3$)	Time (ms)
Continuous Exact	2^{-4}	14.29	2.430	14.49	5.3921
	2^{-5}	7.565	1.198	7.659	41.735
	2^{-6}	4.020	0.594	4.063	378.16
	2^{-7}	2.022	0.296	2.043	2437.1
	2^{-8}	1.084	0.148	1.094	18919.
Continuous Euler	2^{-4}	8.765	2.402	9.088	2.549
	2^{-5}	4.258	1.262	4.441	17.53
	2^{-6}	2.143	0.628	2.233	121.8
	2^{-7}	1.178	0.295	1.214	721.2
	2^{-8}	0.610	0.147	0.627	5824.
Discrete Exact	2^{-4}	−36.30	2.136	36.36	7.7431
	2^{-5}	−18.45	1.060	18.48	57.862
	2^{-6}	−9.072	0.536	9.087	554.42
	2^{-7}	−4.473	0.262	4.481	3627.4
	2^{-8}	−2.193	0.127	2.196	26029.
Discrete Euler	2^{-4}	−41.93	2.105	41.98	2.498
	2^{-5}	−21.12	1.052	21.14	16.69
	2^{-6}	−10.57	0.524	10.58	112.8
	2^{-7}	−5.277	0.251	5.282	701.5
	2^{-8}	−2.636	0.126	2.639	5724.

Table 4. *Loss probability. Parameters:* $d = 4$, $T = 1.00$ *and* $K = 2.00$. *The* $X_0^i, \kappa_i, \theta_i$ *and* σ_i *appear in Tables* 1–3 *for the first three components respectively, while* $X_0^4 = 1.00$, $\kappa_4 = 3.25$, $\theta_4 = 0.25$ *and* $\sigma_4 = 1.00$. *We take* $\lambda_0 = 1.0$ *and* $\lambda_1 = (1.00, 0.25, 1.50, 2.50)$ *and are* $\delta_1 = 0.75$, $\delta_2 = 1.50$, $\delta_3 = 0.75$ *and* $\delta_4 = 1.75$. *The "true" value of* Π *is* 0.172824.

Table 4 reports the results for the four schemes. The convergence of the RMSE is shown as a log-log plot in Figure 4. All four schemes give a RMSE convergence roughly proportional to the inverse cube root of the simulation time, which suggests that again the error converges weakly with order one.

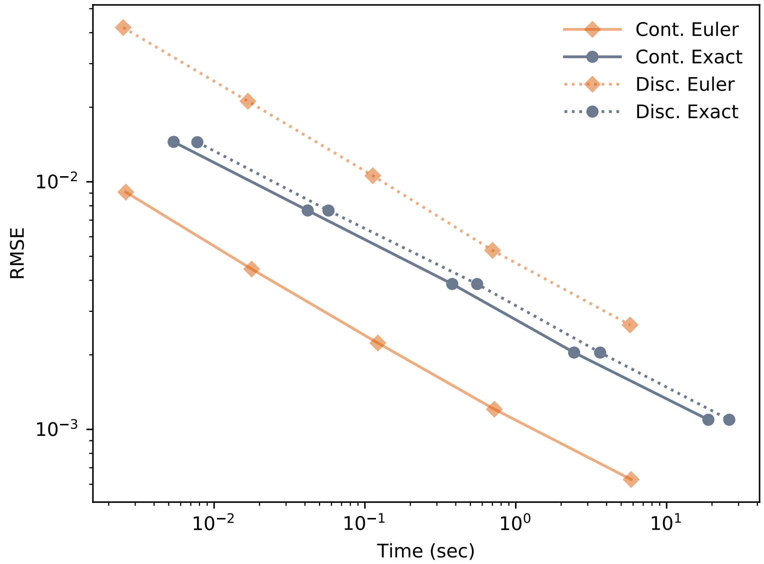

Figure 4. *Convergence of the* RMSE *of the loss probability problem in Table 4.*

7. Conclusion

We have developed, analyzed and tested a discretization scheme for simulating the solution of a multi-dimensional jump-diffusion SDE with general drift, volatility, jump intensity, and jump size functions. The jump times of the process were constructed as time-changed Poisson arrival times. The Euler method was used to generate the process between the jump epochs. We proved that under conditions on the coefficient functions, the scheme converges weakly with order one for measurable functions of polynomial growth. Numerical experiments for an affine jump-diffusion model were used to illustrate the results.

A. Proof of Proposition 4.5

For measurable $f : \mathcal{D} \to \mathbb{R}$ of polynomial growth (see Appendix A.0), let

$$v(x,t) = \mathrm{E}_x(f(X_t)) \tag{15}$$

for $(x,t) \in \mathcal{D} \times [0,T]$ so that the u of Section (4) satisfies $u(x,t) = v(x,T-t)$ when f is taken to be the payoff function g. The subscript $x \in \mathcal{D}$ denotes the starting position of X. We decompose (15) as $v = \sum_{n\in\mathbb{N}} v_n$ where

$$v_n(x,t) = \mathrm{E}_x(f(X_t)1_{\{N_t=n\}}) \tag{16}$$

and then prove that v inherits the required properties from those of $\{v_n\}_{n\in\mathbb{N}}$. The proof proceeds in the three steps found in subsections A.1, A.2, and A.3.

We begin by discussing some preliminary results.

A.0. Preliminaries. Define a d-dimensional process Y, that describes the dynamics of X in (1) in between the jump times, as the solution to the SDE

$$\mathrm{d}Y_t = \mu(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t^*, \qquad Y_0 \in \mathbb{R}^d, \tag{17}$$

where W^* is a standard m-dimensional Brownian motion independent of W in (1). Assumption 4.1 on the coefficients μ and σ guarantees Y is a strong Markov process (Kusuoka & Stroock 1985, Theorem 3.3). Let $\mathbb{G} = \{\mathcal{G}_t\}_{t\ge 0}$ be the right-continuous and complete filtration generated by Y. Our analysis makes extensive use of a measure $\mathscr{P}$ analyzed in Kusuoka & Stroock (1985).

The subscript x of E_x and P_x refers to the initial condition of Y. Let

$$\mathscr{P}_t(x,\mathscr{B}) = \mathrm{E}_x\big(e^{-\int_0^t \Lambda(Y_s)\,\mathrm{d}s} 1_{\{Y_t\in\mathscr{B}\}}\big) \quad x\in\mathbb{R}^d,\, t\ge 0. \tag{18}$$

for Borel $\mathscr{B}\subseteq\mathbb{R}^d$, The right side of (18) equals $\mathrm{P}_x(\theta > t, Y_t\in\mathscr{B})$ where the θ denotes the first jump time of a doubly-stochastic Poisson process with intensity $\Lambda(Y)$ (see also Giesecke & Zhu (2013)), i.e., condition on $\mathcal{G}_t$ and use the fact that, for a standard exponential $\mathcal{E}$ independent of $\mathcal{G}_\infty$, θ admits the representation

$$\theta = \inf\big\{t>0 : B_t \ge \mathcal{E}\big\}; \qquad B_t = \int_0^t \Lambda(Y_s)\mathrm{d}s. \tag{19}$$

We say $\psi : \mathbb{R}^d\times[0,T]\to\mathbb{R}$ is *smooth* if it is infinitely differentiable, i.e., the partial derivatives $\partial_t^k\partial_x^\alpha\psi(x,t) = \partial_t^k\partial_{x_1}^{\alpha_1}\cdots\partial_{x_d}^{\alpha_d}\psi(x,t)$ exist for every $\alpha\in\mathbb{N}^d, k\in\mathbb{N}$ and $(x,t)\in\mathbb{R}^d\times[0,T]$. The multi-index $\alpha = (\alpha_1,\dots,\alpha_d)$ and the k indicate the order of the derivatives. A 0th order derivative is just the function itself, e.g., $\partial_t^0\psi(x,t) = \psi(x,t)$. Let $|\alpha| = \sum_{i=1}^d \alpha_i$. We denote by ∇ and ∇^2 the gradient and Hessian operators in the spacial coordinate.

For ψ smooth, we further say that ψ is *tempered* if all its derivatives are slowly increasing, i.e., for any multi-index $\alpha\in\mathbb{N}^d$ and $k\in\mathbb{N}$ there exist a constant $Q_{k,\alpha}\in\mathbb{R}_+$ and $t\to K_t^{k,\alpha}\in\mathbb{R}_+$ continuous on $(0,T]$ such that

$$|\partial_t^k\partial_x^\alpha\psi(x,t)| \le K_t^{k,\alpha}(1+|x|^{Q_{k,\alpha}}). \tag{20}$$

We allow for the possibility that $K_0^{k,\alpha} = \infty$ and call ψ *uniformly tempered* if every $\sup_{0\le t\le T} K_t^{k,\alpha} < \infty$. When any measurable $\psi : \mathbb{R}^d\times[0,T]\to\mathbb{R}$ (not necessarily smooth) satisfies (20) with $\alpha = 0_d$ and $k=0$, we say ψ has *polynomial growth*. It is of *uniform polynomial growth* if $\sup_{0\le t\le T} K_t^{0,0_d} < \infty$.

For measurable ψ of polynomial growth, we adopt the notation

$$\mathscr{P}_t(\psi)(x,s) = \int_{\mathbb{R}^d} \psi(y,s)\mathscr{P}_t(x,\mathrm{d}y).$$

Observe that $\mathscr{P}_0\psi = \psi$. If $\psi(x,s) = \varphi(x)$, then $\mathscr{P}_t(\psi)(x,s) = \mathscr{P}_t(\varphi)(x)$. By Kusuoka & Stroock (1985, Theorem 3.11), for any tempered $\varphi : \mathbb{R}^d \to \mathbb{R}$,

$$\partial_t \mathscr{P}_t(\varphi) = (\mathscr{L} - \Lambda)\mathscr{P}_t(\varphi) \qquad 0 < t < T, \tag{21}$$

where $\mathscr{L}$ is the generator of Y. Here, $(\Lambda\psi)(x,t) = \Lambda(x)\psi(x,t)$ and $\mathscr{L}\psi_t$ denotes $\mathscr{L}$ applied to $\psi_t(\cdot) = \psi(\cdot,t)$ for suitable $\psi : \mathbb{R}^d \times [0,T] \to \mathbb{R}$, i.e.,

$$\mathscr{L}\psi_t = \mu \cdot \nabla\psi_t + \frac{1}{2}\mathrm{tr}(\sigma\sigma^\top \nabla^2\psi_t). \tag{22}$$

We now strengthen (21) (see Lemma *A.1* below) under the assumption of uniform hypoellipticity for diffusion Y (Kusuoka & Stroock 1985). As discussed in Section 4, this condition is implied by the uniform ellipticity of Assumption 4.2. Uniform hypoellipticity is essentially the requirement that Hörmander's condition holds everywhere in the domain. For any smooth functions[7] $U_1, U_2 : \mathbb{R}^d \to \mathbb{R}^d$ we define $[U_1, U_2] = \mathrm{J}_{U_1}U_2 - \mathrm{J}_{U_2}U_1$, the Lie bracket; where the J_{U_i} is the Jacobian of U_i (cf. pg. 8 Kusuoka & Stroock (1985)).

Let $V_0, V_1, \ldots, V_m : \mathbb{R}^d \to \mathbb{R}^d$ be smooth. For $k \geq 1$ and a multi-index χ with each entry $\chi_i \in \{0,1,\ldots,d\}$ and $i = 1,2,\ldots,\|\chi\|$ (the length of the index with $\|\emptyset\| = 0$), we define (Kusuoka & Stroock 1985, (2.11))

$$V_k^\chi = \begin{cases} [V_{\chi_\ell}, V_k^{(\chi_1,\ldots,\chi_{\ell-1})}] & \|\chi\| = \ell \geq 1, \\ V_k & \text{otherwise.} \end{cases} \tag{23}$$

Let $\widehat{\mu} = V_0 + \frac{1}{2}\sum_{k=1}^m \mathrm{J}_{V_k}V_k$ and $\widehat{\sigma} = (V_1,\ldots,V_m)$ and assume $\widehat{\mu}$ and $\widehat{\sigma}$ have bounded partial derivatives. We say that a diffusion with drift $\widehat{\mu}$ and volatility $\widehat{\sigma}$ satisfies *uniform hypoellipticity* (UH) if there is an integer $L \geq 1$ with

$$\mathscr{V}_L = 1 \wedge \inf_{\substack{x,\eta\in\mathbb{R}^d \\ |\eta|=1}} \sum_{k=1}^d \sum_{\|\chi\|_0 \leq L-1} \left(V_k^\chi(x)\cdot\eta\right)^2 > 0 \tag{24}$$

where $\|\chi\|_0 = \|\chi\| - \mathrm{card}\{k : \chi_k = 0\}$ (cf. Bally & Talay (1996)).

It suffices (for our claims) to weaken Assumption 4.2 (uniform ellipticity) to require only that Y satisfy UH. Similarly, we may relax Assumption 4.3 to the requirement that $Y + \Delta(Y,z)$ satisfy UH (see Remark A.3 for the definitions of $\widehat{\mu}$

[7] That is, the $\partial^\alpha U_i = (\partial^\alpha U_i^1, \ldots, \partial^\alpha U_i^d)$ exist for all $\alpha \in \mathbb{N}^d$.

and $\widehat{\sigma}$) uniformly in $z \in \mathcal{M}$. The latter, uniform property is true by Assumption 4.4. Note also that Assumption 4.4 implies every $\sup_{z\in\mathcal{M}} \Delta_i(x,z)$ is bounded.

As Y satisfies UH, $\mathscr{P}_t$ has a density p_t, i.e., $\mathscr{P}_t(x,\mathscr{B}) = \int_{\mathscr{B}} \mathrm{p}_t(x,y)\mathrm{d}y$ for all $(x,t) \in \mathbb{R}^d \times [0,T]$ (see Kusuoka & Stroock (1985, Theorem 3.17)). Moreover, by Kusuoka & Stroock (1985, Corollary 3.25), for any $k \in \mathbb{N}$ and multi-index $\alpha \in \mathbb{N}^d$, $\partial_t^k \partial_x^\alpha p_t(x,y)$ exists for $(t,x,y) \in (0,T] \times \mathbb{R}^d \times \mathbb{R}^d$ and,

$$(25)\qquad |\partial_t^k \partial_x^\alpha \mathrm{p}_t(x,y)| \le \frac{(1+|x|^Q)}{(1+|x-y|^2)^{\ell/2}} \left(\frac{\mathscr{K}_{T,\ell}^{\mathscr{V}_L}}{t^{q_\ell}}\right) e^{-\frac{(|x-y|\wedge 1)^2}{t c_\ell}}$$

for every integer $\ell \ge 2k + \sum_{i=1}^d \alpha_i$ and constants $Q, \mathscr{K}_{T,\ell}^{\mathscr{V}_L}, c_\ell, q_\ell > 0$. Here, the constants $\mathscr{K}_{T,\ell}^{\mathscr{V}_L}, c_\ell, q_\ell$ depend on ℓ (and on L in the definition of UH) while Q does not (see comments above (3.27) in Kusuoka & Stroock (1985)). The constant $\mathscr{K}_{T,\ell}^{\mathscr{V}_L}$ depends inversely on $\mathscr{V}_L$ in (24) and is increasing in T on $[0,\infty)$.

Lemma A.1. *Suppose that Y satisfies UH and a measurable $\varphi : \mathbb{R}^d \to \mathbb{R}$ has polynomial growth. Then, (21) holds. Moreover, the function ψ, defined by $\psi(x,t) = \mathscr{P}_t(\varphi)(x)$, is tempered and has uniform polynomial growth.*

This is a direct consequence of (25). Note, $\mathscr{P}\varphi$ is not uniformly tempered. Indeed, for $x = y$, as $t \downarrow 0$ the bound (25) becomes degenerate as $\mathrm{p}_t(x,\cdot) = \delta_x$, the Dirac delta-function. To help address this degeneracy in the proofs, we state a variant of Lemma 4.1 in Bally & Talay (1996). Its proof is in Appendix C.

Lemma A.2. *Suppose that $\gamma : \mathbb{R}^d \to \mathbb{R}^d$ is smooth and for $\alpha \in \mathbb{N}^d$ with $\|\alpha\| \ge d$ satisfies $|\partial_x^\alpha \gamma(x)||x| \to 0$ as $|x| \to \infty$. Assume $\gamma(Y)$ satisfies UH. Let $\psi : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ and $\varphi : \mathbb{R}^d \to \mathbb{R}$ be tempered. Suppose ψ has uniform polynomial growth, and $\psi_s(x) = \psi(x,s)$. Then, $(x,s) \to \mathscr{P}_s(\varphi\, \partial_x^\alpha \psi_{t-s} \circ \gamma)(x)$ has uniform polynomial growth when restricted to any $[0,t] \subseteq [0,T]$.*

Remark A.3. *When the coefficients of Y satisfy Assumption 4.1, $\gamma(Y)$ is a diffusion with smooth coefficients having bounded derivatives. In particular, $\gamma_i(Y)$ has drift $(\nabla\gamma_i)^\top \mu + \frac{1}{2}\mathrm{tr}((\nabla^2\gamma_i)(\sigma\sigma^\top))$ and volatility $\nabla\gamma_i\sigma$. Thus, the stronger uniform ellipticity condition holds if $\inf_{x\in\mathcal{D}} \left|(\nabla\gamma)\sigma\sigma^\top(\nabla\gamma)^\top(x)\right| > 0$. When Assumption 4.2 holds, it suffices to require only that $\inf_{x\in\mathcal{D}} |(\nabla\gamma)(x)| > 0$.*

We show how to construct a weak solution to the SDE (1) using Y. Fix a probability space $(\Omega^*, \mathcal{F}^*, \mathrm{P}^*)$ supporting a collection $\{W_n^*\}_{n\in\mathbb{N}}$ of independent standard Brownian motions in $\mathbb{R}^m$, an i.i.d sequence $\{\mathcal{E}_n\}_{n\in\mathbb{N}}$ of standard exponential random variables, and and i.i.d. sequence $\{Z_n^*\}_{n\in\mathbb{N}}$ of random variables with law ν. The three sequences are mutually independent. Let Y^n be a solution of (17) with $W^* = W_n^*$. Given $X_0^* = X_0$ and $T_0^* = 0$, we recursively define a jump-diffusion process X^* with jump times T_n^* as follows. For every $n \in \mathbb{N}$,

$$
\begin{aligned}
B_n(t) &= \int_0^t \Lambda(Y_s^n)\,\mathrm{d}s \qquad Y_0^n = X_{T_n^*}^* \\
T_{n+1}^* &= T_n^* + B_n^{-1}(\mathcal{E}_n) \\
X_t^* &= Y_{t-T_n^*}^n \qquad\qquad T_n^* \le t < T_{n+1}^* \\
X_{T_{n+1}^*}^* &= X_{T_{n+1}^*-}^* + \Delta(X_{T_{n+1}^*-}^*, Z_{n+1}^*)
\end{aligned}
\tag{26}
$$

where the functions Λ and Δ satisfy Assumptions 4.1 and 4.4. The process X^* is adapted to the right-continuous and complete filtration $\mathbb{F}^* = (\mathcal{F}_t^*)_{t\ge 0}$ generated by the marked point process $(T_n^*, Z_n^*)_{n\in\mathbb{N}}$ and the processes $\{W_n^*\}_{n\in\mathbb{N}}$.

Lemma A.4. *Suppose Assumptions 4.1 and 4.4. Then the process X^* constructed per (26) is a weak solution to the* SDE (1).

The proof of Lemma A.4 is in Appendix C. Lemma A.4 allows us to work with the solution X^* of (1) for the remainder of Appendix A. To ease the notation, we write (X, T_n, Z_n) and $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathrm{P})$ to denote that solution.

Lemma A.5. *Suppose Assumptions 4.1 and 4.4. For $k \in \mathbb{N}$ there is $C_k \in \mathbb{R}_+$ such that $\mathrm{E}_x(|X_\tau|^k) \le C_k(1 + |x|^k)$ for all stopping times $\tau \le T$ and $x \in \mathcal{D}$.*

Remark A.6. *This statement also holds for the Euler process X^h of Section 3. The proof is identical as the properties of the coefficients $\mu, \sigma, \Lambda, \Delta$ used as part of Assumptions 4.1 and 4.4 are related to their (linear) growth, not smoothness.*

The proof of Lemma A.5 is also in Appendix C. Lemma A.5 guarantees that $v(x, t)$ in (15) is well-defined for any measurable f with polynomial growth.

A.1. Proof of step 1. We construct a representation of v_n in (16) in terms of the diffusion Y solving (17). In particular, we prove that $v_n = \phi_n$ on $\mathcal{D} \times [0, T]$ for all $n \in \mathbb{N}$ for the following functions $\{\phi_n\}_{n\in\mathbb{N}}$ that are defined recursively.[8]

$$
\begin{aligned}
\phi_{n+1}(x, t) &= \mathrm{E}_x(1_{\{\theta\le t\}}\,\phi_n^\Delta(Y_\theta, t-\theta)) \qquad \forall n \in \mathbb{N}, \\
\phi_0(x, t) &= \mathrm{E}_x(1_{\{\theta> t\}}\,f(Y_t))
\end{aligned}
\tag{27}
$$

for θ in (19) and where for any $\psi : \mathbb{R}^d \times [0, T] \to \mathbb{R}$ we have defined,

$$
\psi^\Delta(x, t) = \int_{\mathcal{M}} \psi(x + \Delta(x, z), t)\,\nu(\mathrm{d}z). \tag{28}
$$

[8]Note that the indicator of $\{\theta \le t\}$ prevents the evaluation of ϕ_n^Δ at a negative argument. The argument $(t - \theta)$ may be replaced with $(t - \theta)_+$, but we proceed otherwise to ease the notation.

The ϕ_n results from iterated expectations of $v_n(x,t) = \mathrm{E}_x(f(X_t)1_{\{N_t=n\}})$ over the $(n+1)$ intervals of the form $[T_k, T_{k+1} \wedge t)$ on which $X_{T_k+\cdot}$ equals $Y^k_\cdot$ per construction of X in (26). The functions $\{\phi_n\}$ facilitate the analysis of the v_n in terms of the measure $\mathscr{P}_t(x,\cdot)$ (induced by Y) that has nice properties.

We let $(x,t) \in \mathcal{D} \times [0,T]$. Below, we use the notation $\theta_n = T_n - T_{n-1}$ to denote the inter-arrival time. We first show the case $n = 0$ by using the construction in (26). Since $Y^0_0 = X_0$ and $\theta_1 = T_1$, as required,

$$v_0(x,t) = \mathrm{E}_x(f(Y^0_t)1_{\{T_1>t\}}) = \phi_0(x,t).$$

Fix $n \geq 1$ and apply (27) together with the construction (26). As (Y^k, θ_{k+1}) and T_k are $\mathrm{P}_{Y^k_0}$-independent, T_k is $\mathcal{F}_{T_k}$-measurable and $Y^k_0 = X_{T_k}$,

$$\begin{aligned}\mathrm{E}(1_{\{\theta_{k+1}\leq t-T_k\}}\phi^\Delta_{n-k-1}(Y^k_{\theta_{k+1}}, t-T_k-\theta_{k+1}) \mid \mathcal{F}_{T_k})& \\ = \phi_{n-k}(X_{T_k}, t-T_k) \quad \text{on} \quad \{T_k \leq t\},\, 1 \leq k < n.&\end{aligned} \tag{29}$$

Furthermore, since T_k is $\mathcal{F}_{T_k-}$-measurable (it is a stopping time),

$$\begin{aligned}\mathrm{E}_x(1_{\{T_k\leq t\}}\phi_{n-k}(X_{T_k}, t-T_k)) &= \mathrm{E}_x\big(1_{\{T_k\leq t\}}\mathrm{E}(\phi_{n-k}(X_{T_k}, t-T_k) \mid \mathcal{F}_{T_k-})\big) \\ &= \mathrm{E}_x(1_{\{T_k\leq t\}}\phi^\Delta_{n-k}(X_{T_k-}, t-T_k))\end{aligned} \tag{30}$$

by the definition in (28). Consequently, applying the construction in (26) and taking the expectation of both sides of (29), we obtain

$$\mathrm{E}_x(1_{\{T_{k+1}\leq t\}}\phi^\Delta_{n-k-1}(X_{T_{k+1}-}, t-T_{k+1})) = \mathrm{E}_x(1_{\{T_k\leq t\}}\phi^\Delta_{n-k}(X_{T_k-}, t-T_k)).$$

Iterating this identity over all $1 \leq k < n$ yields

$$\mathrm{E}_x(1_{\{T_1\leq t\}}\phi^\Delta_{n-1}(X_{T_1}, t-T_1)) = \mathrm{E}_x(1_{\{T_n\leq t\}}\phi^\Delta_0(X_{T_n-}, t-T_n)).$$

The left side is simply $\phi_n(x,t)$ since $X_{T_1} = Y^0_{\theta_1}$ and $T_1 = \theta_1$ as prescribed by (26). The right side, applying the argument from (30), simplifies to

$$\begin{aligned}\mathrm{E}_x\big(1_{\{T_n\leq t\}}\phi^\Delta_0(X_{T_n-}, t-T_n)\big) &= \mathrm{E}_x\big(1_{\{T_n\leq t\}}\phi_0(X_{T_n}, t-T_n)\big) \\ &= \mathrm{E}_x\big(\mathrm{E}(f(Y^n_{t-T_n})1_{\{T_{n+1}-T_n>t-T_n\geq 0\}} \mid \mathcal{F}_{T_n})\big) \\ &= \mathrm{E}_x(f(X_t)1_{\{T_n\leq t<T_{n+1}\}}) \\ &= v_n(x,t)\end{aligned}$$

which concludes the proof of this step.

A.2. Proof of step 2. We prove that for every v_n (and of course ϕ_n) satisfies[9]

$$\begin{aligned} \partial_t v_n(x,t) &= ((\mathscr{L}-\Lambda)v_n)(x,t) + (\Lambda v_{n-1}^{\Delta})(x,t) \qquad && 0 < t \le T \\ v_n(x,0) &= (1-n)_+ f(x) && x \in \mathbb{R}^d . \end{aligned} \tag{31}$$

We adopt the convention $v_{-1} = 0$. Equation (31) is shown by direct differentiation of the ϕ_n in (27). For f smooth, (31) also follows from Itô's formula.

We proceed by strong induction on n. Assume (31) holds for all $\{v_k\}_{k<n}$, each v_k is tempered and is in the domain of $\mathscr{L}$. For v_0, these are consequences of Lemma A.1 and that Y is Feller (Kusuoka & Stroock 1985, Theorem 3.3).

We first express the ϕ_n in (27) as integrals with respect to $\mathscr{P}$. By (19),

$$v_0(x,t) = \phi_0(x,t) = \mathrm{E}_x\big(f(Y_t)\mathrm{P}(\theta > t \,|\, \mathcal{G}_t)\big) = \mathscr{P}_t(f)(x) \tag{32}$$

Let $n \ge 1$. Again by (19), $\mathrm{P}(\theta > t \,|\, \mathcal{G}_\infty) = \mathrm{P}(\mathcal{E} > B_t \,|\, \mathcal{G}_\infty) = e^{-B_t}$ almost surely and we obtain, after conditioning $v_n = \phi_n$ in (27) on $\mathcal{G}_\infty$, that

$$\begin{aligned} v_n(x,t) &= \mathrm{E}_x \int_0^t v_{n-1}^{\Delta}(Y_s, t-s)\mathrm{P}(\theta \in \mathrm{d}s \,|\, \mathcal{G}_\infty) \\ &= \mathrm{E}_x \int_0^t e^{-B_s}\Lambda(Y_s) v_{n-1}^{\Delta}(Y_s, t-s)\,\mathrm{d}s . \end{aligned}$$

Let $\phi_t^n(x,s) = \mathscr{P}_s\big(\Lambda v_{n-1}^{\Delta}\big)(x, t-s)$, finite by the IH and Lemma A.2. Also,

$$\phi_t^n(x,s) = \mathrm{E}_x\big(e^{-B_s}\Lambda(Y_s) v_{n-1}^{\Delta}(Y_s, t-s)\big) \tag{33}$$

and, by Tonelli's theorem (Bogachev 2007, Theorem 3.4.5), $v_n(x,t) < \infty$, so

$$v_n(x,t) = \int_0^t \phi_t^n(x,s)\,\mathrm{d}s . \tag{34}$$

Fix $n \ge 1$ and $t \in (0,T]$. We proceed by differentiating (34). To this end, we must justify the interchanges of $\int_0^t$ and the partial derivatives. In particular,

$$\partial_t v_n(x,t) = \phi_t^n(x,t) + \int_0^t \partial_t \phi_t^n(x,s)\,\mathrm{d}s \tag{35a}$$

$$(\mathscr{L}-\Lambda)v_n(x,t) = \int_0^t (\mathscr{L}-\Lambda)\phi_t^n(x,s)\,\mathrm{d}s . \tag{35b}$$

We assume (35*a*)–(35*b*), deferring their proof. By Bogachev (2007, Corollary 2.8.7) applied with Assumption 4.4 and that v_{n-1} is tempered (by the IH)

$$\partial_t v_{n-1}^{\Delta}(x,t) = \int_{\mathcal{M}} \partial_t v_{n-1}(x + \Delta(x,z), t)\,\nu(\mathrm{d}z) . \tag{36}$$

[9] Here, $(\Lambda v_n)(x,t) = \Lambda(x) v_n(x,t)$ and $\mathscr{L} v_n(x,t)$ is the operator $\mathscr{L}$ applied to $v_n(\cdot,t)$ at x.

The same argument reveals that $\sup_{t\in(a,b]} \partial_t v_{n-1}^{\Delta}(x,t)$ is $\mathscr{P}_s(x,\cdot)$-integrable for every $(a,b] \subseteq [0,T]$. Again by Bogachev (2007, Corollary 2.8.7), we have

$$\partial_t \phi_t^n(x,s) = \mathscr{P}_s\big(\Lambda \partial_t v_{n-1}^{\Delta}\big)(x,t-s) \qquad 0 \le s < t\,. \tag{37}$$

Applying Fubuni's theorem to (35a), followed by integration by parts, we obtain

$$\begin{aligned}\partial_t v_n(x,t) &= \Lambda(x) v_{n-1}^{\Delta}(x,t) + \int_{\mathcal{D}} \int_0^t v_{n-1}^{\Delta}(y,t-s)\Lambda(y)\partial_s \mathscr{P}_s(x,\mathrm{d}y)\mathrm{d}s \\ &= \Lambda(x) v_{n-1}^{\Delta}(x,t) + \int_0^t (\mathscr{L}-\Lambda)\phi_t^n(x,s)\mathrm{d}s \end{aligned}\tag{38}$$

In the first equality, we used the fact that $\phi_t^n(x,t) = \mathscr{P}_t(\Lambda v_{n-1}^{\Delta})(x,0)$ and $\partial_t \phi_{n-1}^{\Delta}(y,t-s) = -\partial_s \phi_{n-1}^{\Delta}(y,t-s)$. The first term in sum on the right side of (38) arises from the integration against $\delta_x(\mathrm{d}y)$ over $\mathcal{D}$. In the second equality, Fubini's theorem is again used to exchange $\int_0^t$ and $\int_{\mathcal{D}}$, and Bogachev (2007, Corollary 2.8.7) is applied with the bound (25) to take ∂_s outside the integral $\mathscr{P}_s(\Lambda v_{n-1}^{\Delta})(x,t-s)$. Lemma A.1 is applied with $\varphi(y) = \Lambda(y) v_{n-1}^{\Delta}(y,t-s)$ to introduce the $(\mathscr{L}-\Lambda)$ operator. By (35b), we have that (31) holds.

By induction, (31) extends to all $n \in \mathbb{N}$. We conclude the proof with a justification of (35a) and (35b) keeping in force our inductive hypotheses. Note that the conclusion that each v_n is tempered follows from the bounds proved in Sections A.2.1 and A.2.2 below. That fact that v_n is in the domain of $\mathscr{L}$ follows from representation (34) and Ethier & Kurtz (2009, Chapter 1, Proposition 1.5).

A.2.1 Proof of identity (35a)

Write $v_n(x,t) = \Phi_x(t,t)$ where $\Phi_x(t,a) = \int_0^t \phi_a^n(x,s)\mathrm{d}s$. By the chain rule,

$$\partial_t v_n(x,t) = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_x(t,t) = \partial_a \Phi_x(a,t)|_{a=t} + \partial_a \Phi_x(t,a)|_{a=t}\,.$$

If $\partial_a \int_0^t \phi_a^n(x,s)\mathrm{d}s = \int_0^t \partial_a \phi_a^n(x,s)\mathrm{d}s$ for all $|a-t| < \epsilon$, then (35a) holds. The former holds if $\sup_{|a-t|<\epsilon} \partial_a \phi_a^n(x,\cdot)$ is integrable on $[0,t]$ by Bogachev (2007, Corollary 2.8.7). We show this is true next.

Let $\gamma_z(y) = y + \Delta(y,z)$. Also write $\psi_s^1(x) = v_{n-1}(x,s)$ and $\psi_s^2(x) = \Lambda(x) v_{n-2}^{\Delta}(x,s)$. From (36) and the IH (in particular (31)) on v_{n-1} we obtain[10]

$$\partial_r v_{n-1}^{\Delta}(y,r) = \int_{\mathcal{M}} ((\mathscr{L}-\Lambda)\psi_r^1) \circ \gamma_z(y) + \psi_r^2 \circ \gamma_z(y)\nu(\mathrm{d}z)$$

[10]Here, $(\mathscr{L}-\Lambda)v_{n-1}(y+\Delta(y,z),a-s)$ is the result of applying the operator $(\mathscr{L}-\Lambda)$ to the function $v_{n-1}(\cdot,a-s)$ and then evaluating at $y+\Delta(y,z)$. See footnote 9 for related notes.

Using (37), the above identity and Fubini's theorem, we obtain

$$\partial_a \phi_a^n(x,s) = \int_{\mathcal{M}} \mathscr{P}_s\big(\Lambda((\mathscr{L}-\Lambda)\psi_{a-s}^1)\circ\gamma_z + \psi_{a-s}^2\circ\gamma_z\big)(x)\nu(\mathrm{d}z)\,.$$

The second term is bounded unifomly over $|a-t| < \epsilon$ by Lemma A.1 (applying Assumption 4.4 to γ_z and the IH on v_{n-2}). We appeal to Lemma A.2 to conclude the first term (and the IH on v_{n-1}) is also uniformly bounded over $|a-t| < \epsilon$.

A.2.2 Proof of identity (35*b*)

The interchange of $(\mathscr{L}-\Lambda)$ and the Lebesgue integral in (35*a*) is justified by Ethier & Kurtz (2009, Chapter 1, Lemma 1.4) provided $\phi_t^n(x,s)$ is in the domain of $(\mathscr{L}-\Lambda)$ for each $s \in [0,t]$ and $(\mathscr{L}-\Lambda)\phi_t^n(x,\cdot)$ is integrable over $[0,t]$.

We proceed by defining $\widehat{\phi}_a^n(x,s) = \phi_{a+s}^n(x,s)$ for $a \in (0,T]$ (cf. (33)), i.e.,

$$\widehat{\phi}_a^n(x,s) = \mathrm{E}_x\big(e^{-B_s}\Lambda(Y_s)v_{n-1}^{\Delta}(Y_s,a)\big)\,.$$

Observe that $\widehat{\phi}_a^n(x,s) = \mathscr{P}_s(\psi_a)(x)$ where $\psi_a(y) = v_{n-1}^{\Delta}(y,a)\Lambda(y)$. By the IH on v_{n-1}^{Δ} we may apply Itô's formula to (Y,B) and $(y,b) \to \psi_a(y)e^{-b}$. Then,

$$\psi_a(Y_r)e^{-B_r} - \psi_a(Y_0) = \int_0^r e^{-B_s}(\mathscr{L}-\Lambda)\psi_a(Y_s)\,\mathrm{d}s + e^{-B_s}(\nabla\psi_a^{\top}\sigma)(Y_s)\mathrm{d}W_s\,.$$

Note, $\mathrm{E}_x\big(\nabla\psi_a^{\top}\sigma\sigma^{\top}\nabla\psi_a(Y_s)\big) < \infty$ for all $0 < s < T$ by (25) (Assumption 4.1 imposes linear growth on σ) and Lemma A.5. Thus,

$$\widehat{\phi}_a^n(x,r+\delta) - \widehat{\phi}_a^n(x,r) = \int_r^{r+\delta} \mathscr{P}_s((\mathscr{L}-\Lambda)\psi_a)(x)\,\mathrm{d}s\,. \tag{39}$$

By the IH on v_{n-1} and Assumption 4.4, ψ_a is of polynomial growth. Thus by Lemma A.1, it follows that $\partial_r\widehat{\phi}_a^n(x,r) = \partial_r\mathscr{P}_r\psi_a = (\mathscr{L}-\Lambda)(\mathscr{P}_r\psi_a)$ for all $0 < r \le T$. Dividing both sides of (39) by δ and taking $\delta \downarrow 0$ yields the standard result that the infintesimal generator commutes in the sense that $(\mathscr{L}-\Lambda)(\mathscr{P}_s\psi_a) = \mathscr{P}_s((\mathscr{L}-\Lambda)\psi_a)$. Now, by Lemma A.2 with γ the identity map and $a = t-s$, there exist finite constants $C_t, Q > 0$ such that for all $x \in \mathbb{R}$

$$|(\mathscr{L}-\Lambda)\widehat{\phi}_{t-s}^n(x,s)| = |\mathscr{P}_s((\mathscr{L}-\Lambda)\psi_{t-s})(x)| \le C_t(1+|x|^Q)\,. \tag{40}$$

As $\widehat{\phi}_{t-s}^n(x,s) = \phi_t^n(x,s)$, $(\mathscr{L}-\Lambda)\phi_t^n(x,\cdot)$ is Lebesgue integrable on $[0,t]$.

A.3. Proof of step 3. We complete the proof of Proposition 4.5. In particular, we prove for $x \in \mathcal{D}$ that $\partial_t v(x,t) = \mathscr{A} v(x,t)$ for $0 < t \le T$ and $v(x,0) = f(x)$ where $\mathscr{A}$ is the operator in (3). We then show v is a tempered map. Then by Clairaut's theorem, $\partial_t^2 v(x,t) = \partial_t \mathscr{A} v(x,t) = \mathscr{A} \partial_t v(x,t) = \mathscr{A}^2 v(x,t)$ and so on. Writing $\widehat{v}_k = \sum_{n=0}^k v_n$, we have by (31) that

$$\begin{aligned}\partial_t \widehat{v}_k(x,t) &= \sum_{n=0}^{k} (\mathscr{L} v_n)(x,t) - \Lambda(x) v_n(x,t) + \Lambda(x) v_{n-1}^{\Delta}(x,t) \\ &= (\mathscr{L} - \Lambda)\widehat{v}_k(x,t) + \Lambda(x)\widehat{v}_{k-1}^{\Delta}(x,t)\end{aligned}$$

From this we obtain $(-\partial_t + \mathscr{L} - \Lambda)\widehat{v}_k + \Lambda \widehat{v}_k^{\Delta} = \Lambda v_k^{\Delta}$. Fix a compact set $\mathscr{K} \subset \mathcal{D}$ and $[0,b] \subseteq (0,T]$. Since N_t is almost surely finite, v_k^{Δ} converges to zero uniformly on $\mathscr{K} \times [a,b]$. Then it is easy to see that $\widehat{v}_k$ and $\widehat{v}_k^{\Delta}$ are Cauchy sequences and thus converge uniformly to v and v^{Δ} on $\mathscr{K} \times [a,b]$. Therefore, $(\partial_t - \mathscr{L} - \Lambda)\widehat{v}_k$ converges uniformly to Λv^{Δ} on $\mathscr{K} \times [a,b]$. Moreover, summing over n in (34) and applying Ethier & Kurtz (2009, Chapter 1, Proposition 1.5) we have that every $\widehat{v}_k$ is in the domain of $(\partial_t - \mathscr{L} - \Lambda)$. The latter is the generator of the joint process $(Y_t, B_t, t)_{t \ge 0}$ which is Feller since (Y, B) is Feller (Böttcher 2014, Theorem 3.2). Consequently, we have by Ethier & Kurtz (2009, Chapter 1, Proposition 1.6) that the operator $(\partial_t - \mathscr{L} - \Lambda)$ is closed so that $(\partial_t - \mathscr{L} - \Lambda)\widehat{v}_k$ coverges uniformly to $(\partial_t - \mathscr{L} - \Lambda)v$ on $\mathscr{K} \times [a,b]$. Therefore,

$$\partial_t v(x,t) = \mathscr{L} v(x,t) + \Lambda(x)(v^{\Delta} - v)(x,t) \tag{41}$$

$$= \mathscr{A} v(x,t) \qquad (x,t) \in \mathscr{K} \times [a,b]. \tag{42}$$

The claim extends to the desired domain since $\mathscr{K} \times [a,b]$ was arbitrary.

That v is tempered follows from the results obtained in Section A.2. In particular, since each v_n is shown to be tempered, we may take any number of derivatives of (31) with respect to time. Consequently, any $\partial_t^k \widehat{v}_k(x,t)$ may be expressed as a sum of spacial derivatives of the functions $\{v_n\}_{n \le k}$. Therefore, it suffices to show only bounds of the type (20) on $\partial_x^{\alpha} \widehat{v}_k(x,t)$. This is achieved by the same arguments as in Section A.2.2, applying Itô's formula iteratively and concluding by Lemma A.2. Finally, all of these bounds are extended to v by the same arguments as above; specifically, we again apply Ethier & Kurtz (2009, Chapter 1, Proposition 1.6) with the closed operator $(\mathscr{L} - \Lambda)$.

B. Proof of Theorem 4.6

We write $\mathscr{E}_g^h(x) = \mathrm{E}_x^h(g(X_T^h)) - \mathrm{E}_x^h(g(X_T))$ as in (10) so that $\mathscr{E}_g^h(x) = \mathrm{E}_x^h \sum_{k=1}^\infty \mathscr{D}_h(\tau_k, u) 1_{\{\tau_k \le T\}}$ where

$$\mathscr{D}_k^h(s, v) = v(X_s^h, s) - v(X_{\tau_{k-1}}^h, \tau_{k-1}), \tag{43}$$

for a function $v : \mathcal{D} \times [0, T] \to \mathbb{R}$.

$$\mathrm{E}_x(u(X_T^h, T) - u(X_\tau^h, \tau)) = \mathrm{E}_x(g(X_T^h) - \mathrm{E}_{X_\tau^h}(g(X_{T-\tau})))$$

Fix $h > 0$ and $K = \lceil T/h \rceil$. Let $\mathscr{L}_y^h$ denote the generator of X^h over the interval (τ_{k-1}, τ_k), i.e., for $X_{\tau_{k-1}} = y$ and $\phi : \mathcal{D} \times [0, T] \to \mathbb{R}$ smooth, let

$$\mathscr{L}_k^h \phi(x, t) = \sum_{i=1}^d \mu_i(X_{\tau_{k-1}}^h) \partial_{x_i} \phi(x, t) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d (\sigma\sigma^\top)_{ij}(X_{\tau_{k-1}}) \partial_{x_i} \partial_{x_j} \phi(x, t)$$

the generator of X^h on (τ_{k-1}, τ_k) so that its generator on $(\tau_{k-1}, \tau_k]$ is

$$\mathscr{A}_k^h \phi(x, t) = \mathscr{L}_k^h \phi(x, t) + \Lambda(X_{\tau_{k-1}}^h) \int_{\mathcal{M}} \big(\phi(x + \Delta(x, z), t) - \phi(x, t)\big) \nu(\mathrm{d}z).$$

Recall $u(x, t) = \mathrm{E}_x(g(X_{T-t}))$. We write $\mathscr{E}_g^h(x) = \mathrm{E}_x^h(g(X_T^h)) - \mathrm{E}_x^h(g(X_T))$ as in (10) so that $\mathscr{E}_g^h(x) = \mathrm{E}_x^h \sum_{k=1}^\infty \mathscr{D}_k^h(\tau_k, u) 1_{\{\tau_k \le T\}}$ where

$$\mathscr{D}_k^h(s, v) = v(X_s^h, s) - v(X_{\tau_{k-1}}^h, \tau_{k-1}), \tag{44}$$

for a function $v : \mathcal{D} \times [0, T] \to \mathbb{R}$. Since $\tau_k \le T$ implies $N_T^h \ge K$ for $k > 2K$

(45)

$$|\mathscr{E}_g^h(x)| \le T^2 C(1 + \|x\|^Q) \mathrm{P}_x(N_T^h \ge K) + \Big| \sum_{k=1}^{2K} \mathrm{E}_x\big(\mathscr{D}_k^h(\tau_k, u) 1_{\{\tau_k \le T\}}\big) \Big|$$

by Lemma A.5. By the latter it also holds that $\mathrm{P}_x(N_T^h > K) \le C T^2/K^2 (1 + \|x\|^2)$. Thus, it suffices to consider the second term in (45) only. By Itô's formula,

$$\mathscr{D}_k^h(t, u) = \int_{(\tau_{k-1}, t]} \Big(\frac{\partial}{\partial s} u(X_s^h, s) + \mathscr{L}_k^h u(X_s^h, s)\Big) \mathrm{d}s + \sum_{\tau_{k-1} < s \le t} \Delta u(X_s^h, s)$$
$$+ \int_{(\tau_{k-1}, t]} \nabla_x u(X_{s-}^h, s) \cdot \sigma(X_{\tau_k}^h) \mathrm{d}W_s^h.$$

Taking an expectation, since $\mathrm{E}_x^h\big(\nabla_x u^\top(X_{s-}^h, s)\sigma\sigma^\top(X_{\tau_k}^h)\nabla_x u^\top(X_{s-}^h, s)\big) < \infty$ by Lemma A.5 we obtain

$$\big|\mathrm{E}_x^h\big(\mathscr{D}_k^h(\tau_k, u)\big)1_{\{\tau_k \le T\}}\big| = \bigg|\mathrm{E}_x^h\bigg(1_{\{\tau_k \le T\}}\int_{(\tau_{k-1},\tau_k]}\Big(\frac{\partial}{\partial s}u(X_s^h, s) + \mathscr{A}_k^h u(X_s^h, s)\Big)\mathrm{d}s\bigg)\bigg|.$$

Proposition 4.5 states $\frac{\partial}{\partial s}u = -\mathscr{A}u$. Therefore, we have $\frac{\partial}{\partial s}u(X_s^h, s)+\mathscr{A}_k^h u(X_s^h, s) = \mathscr{A}_k^h u(X_s^h, s) - \mathscr{A}_k^h u(X_{\tau_{k-1}}^h, \tau_{k-1}) - \big(\mathscr{A}u(X_s^h, s) - \mathscr{A}u(X_{\tau_{k-1}}^h, \tau_{k-1})\big)$ since the generators $\mathscr{A}$ and $\mathscr{A}^h$ coincide at the the $\{\tau_k\}$. This yields the identity

$$\big|\mathrm{E}_x^h\big(\mathscr{D}_k^h(\tau_k, u)1_{\{\tau_k \le T\}}\big)\big| = \bigg|\mathrm{E}_x^h\bigg(1_{\{\tau_k \le T\}}\int_{(\tau_{k-1},\tau_k]}\big(\mathscr{D}_k^h(s, \mathscr{A}_k^h u) - \mathscr{D}_k^h(s, \mathscr{A}u)\big)\,ds\bigg)\bigg|.$$

Applying Itô's formula to each of the terms yields, for $V_k(s) = (\tau_{k-1}, \tau_k] \times (\tau_{k-1}, s]$, the above is bounded above by

$$\bigg|\mathrm{E}_x^h 1_{\{\tau_k \le T\}}\iint_{V_k(s)}\big(\mathscr{A}_k^h\mathscr{A}u(X_r^h, r) - \mathscr{A}^2u(X_r^h, r)\big)\mathrm{d}r\,\mathrm{d}s\bigg| \\ + \bigg|\mathrm{E}_x^h 1_{\{\tau_k \le T\}}\iint_{V_k(s)}\big(\mathscr{A}\mathscr{A}_k^h u(X_r^h, r) - (\mathscr{A}_k^h)^2u(X_r^h, r)\big)\mathrm{d}r\,\mathrm{d}s\bigg|.$$

To conclude, consider the filtration $(\widehat{\mathcal{F}}_t^h)_{t\ge 0}$ defined as $(\mathcal{F}_t^h)_{t\ge 0}$ but with the sequence $\{E_n\}_{n\in\mathbb{N}}$ added to $\widehat{\mathcal{F}}_0^h$. By definition of the compensator A^h, both times τ_{k-1} and τ_k are measurable with respect to $\widehat{\mathcal{F}}_{\tau_{k-1}}^h$. Observe that Lemma A.2 also holds for a continuous Euler approximation Y^h for the process Y of Appendix A. This is proved exactly as in Bally & Talay (1996, Lemma 4.1 and Lemma 5.2). Since the process Y^h is equal to X^h in between the jumps, we have for $r > \tau_{k-1}$ that

$$\big|\mathrm{E}^h\big(\mathscr{A}_k^h\mathscr{A}u(X_r^h, r) - \mathscr{A}^2u(X_r^h, r)\,|\,\widehat{\mathcal{F}}_{\tau_{k-1}}^h\big)\big| \le C_1(1 + \|X_{\tau_{k-1}}^h\|^{Q_1}) \tag{46}$$

$$\big|\mathrm{E}^h\big(\mathscr{A}\mathscr{A}_k^h u(X_r^h, r) - (\mathscr{A}_k^h)^2u(X_r^h, r)\,|\,\widehat{\mathcal{F}}_{\tau_{k-1}}^h\big)\big| \le C_2(1 + \|X_{\tau_{k-1}}^h\|^{Q_2}) \tag{47}$$

for constants $C_i, Q_i > 0$. Since $\tau_k - \tau_{k-1} \le h$, it now follows by Fubini's theorem, the bounds above and Lemma A.5 that $\big|\mathrm{E}_x^h\big(\mathscr{D}_k^h(\tau_k, u)1_{\{\tau_k \le T\}}\big)\big| \le Ch^2(1 + \|x\|^Q)$ for constant $C, Q > 0$. The conclusion now follows by summing over k as in (45).

C. Auxiliary proofs

Proof of Lemma A.2. The proof is essentially that of Lemma 4.1 of Bally & Talay (1996) and involves Malliavin calculus. We point out only the modifications required for our version of the statement and provide references to the required definitions. We work on a fixed interval $[0, t] \subseteq [0, T]$. Note, for $s \le t/2$

the desired bound holds by the hypothesis that ψ is tempered because $K^{k,\alpha}_{t-s}$ (the coefficient in (20) for ψ_{t-s}) is bounded for $s \leq t/2$. We now let $s \in (t/2, t]$.

Let D_∞ be the space of smooth Wiener functionals in the Malliavin sense (Ikeda & Watanabe 1989, Chapter V.8). Since each γ_i is a tempered function, $\gamma_i(Y) \in \mathrm{D}_\infty$. In particular, applying Itô's formula to $\gamma_i(Y)$ yields

$$\gamma_i(Y_t) - \gamma_i(Y_0) = \int_0^t \nabla\gamma_i(Y_s)\cdot \mathrm{d}Y_s + \frac{1}{2}\int_0^t \operatorname{tr}\big((\nabla^2\gamma_i)(\sigma\sigma^\top)\big)(Y_s)\,\mathrm{d}s$$

revealing that $\gamma_i(Y)$ is a diffusion with drift coefficient $\nabla\gamma_i^\top\mu + \frac{1}{2}\operatorname{tr}\big((\nabla^2\gamma_i)(\sigma\sigma^\top)\big)$ and volatility coefficients $\nabla\gamma_i\sigma$. Note that by our hypotheses these coefficients satisfy all requirements of uniform hypoellipticity (cf. Remark A.3).

Since φ is tempered, $e^{-B}\varphi(Y) \in \mathrm{D}_\infty$, and the diffusion $\gamma(Y)$ satisfies UH, we apply (integration by parts) Bally & Talay (1996, Proposition 2.3). We have

$$\begin{aligned} \mathscr{P}_s(\varphi\,\partial_x^\alpha\psi_{t-s}\circ\gamma)(x) &= \mathrm{E}_x\big(e^{-B_s}\varphi(Y_s)(\partial_x^\alpha\psi_{t-s}\circ\gamma)(Y_s)\big) \\ &= \mathrm{E}_x\big(\Phi_Y^\alpha(s)(\psi_{t-s}\circ\gamma)(Y_s)\big) \end{aligned} \tag{48}$$

where $\Phi_\alpha(s)$ is a random variable such that for any $p \in (1,\infty)$ there are integers C, q, k, a, k', a' (depending on α, p) with (Bally & Talay 1996, Proposition 2.4),

$$(\mathrm{E}_x(\Phi_\alpha(s)^p))^{1/p} \leq C(\mathrm{E}_x(|\Gamma_s|^q))^{1/q}\|e^{-B_s}\varphi(Y_s)\|_{W^{k,a}(\Omega_x)}\|\gamma(Y_s)\|_{W^{k',a'}(\Omega_x)}$$

where $\|\cdot\|_{W(\Omega_x)}$ denotes the Sobolev norms with Malliavin (weak) derivatives (Bally & Talay 1996, Section 2) on the spaces $\mathrm{D}_\infty, \mathrm{D}_\infty^d$ with Y started at $x \in \mathbb{R}^d$. Here, Γ_s denotes the inverse of the Malliavin covariance matrix of $\gamma(Y_s)$ started at $\gamma(x)$ and $|\Gamma_s|$ its determinant. Since $\gamma(Y)$ satisfies UH (with some integer $L \geq 1$ in (24)) and $t/2 \leq s \leq t$ we have $(\mathrm{E}_x(|\Gamma_s|^q))^{1/q} \leq \frac{K}{t^{dL}}(1+|x|^Q)$ (Kusuoka & Stroock 1985, Corollary 3.25) for constants K, Q (depending on q). The two Sobolev norms obey similar bounds (Bally & Talay 1996, Lemma 5.1). The claim now follows by applying Hölder's inequality to (48). In this step, we use that φ is time independent and ψ has uniform polynomial growth. ∎

Proof of Lemma A.4. From (17) it is clear that Y^n, and thus X^*, solves the SDE (1) in between the jump times $(T_n^*)_{n\geq 1}$, temporarily assuming $T_n^* = T_n$. Definition (26) implies that X^* also solves (1) at the jump times. Thus, it suffices to show that X^* jumps at the same frequency as X. To this end we show that the counting process N^* given by $N_t^* = \sum_{n\geq 1} 1_{\{T_n^* \leq t\}}$ has $(\mathbb{F}^*, \mathrm{P}^*)$-compensator

$$A_t^* = \int_0^t \Lambda(X_s^*)\,\mathrm{d}s.$$

Clearly A^* starts at the origin, is $\mathbb{F}^*$-adapted, and has continuous and strictly increasing paths. We verify that $N^* - A^*$ is a $(\mathbb{F}^*, \mathrm{P}^*)$-local martingale.

Letting $S_n = \sum_{k \leq n} \mathcal{E}_k$, we can consider $(S_n, Z_n^*)_{n \geq 1}$ as a unit-intensity P^*-marked Poisson process with respect to a filtration $(\mathcal{H}_t)_{t \geq 0}$ with the property that $\mathcal{H}_0$ contains $\sigma(\{W_n^*, Z_n^*\}_{n \in \mathbb{N}})$. Let $N_t^0 = \sum_{n \geq 1} 1_{\{S_n \leq t\}}$ be the associated counting process. From (26) we deduce that $\{T_n^* \leq t\} = \{S_n \leq A_t^*\}$. Then,

$$N_t^* = \sum_{n \geq 1} 1_{\{T_n^* \leq t\}} = \sum_{n \geq 1} 1_{\{S_n \leq A_t^*\}} = N_{A_t^*}^0. \tag{49}$$

Given the time-change representation (49), $N^* - A^*$ is a P^*-local martingale with respect to the filtration $(\mathcal{H}_{A_t^*})_{t \geq 0}$ by Proposition A.1 of Giesecke & Tomecek (2005). Proposition A.10 of Giesecke & Tomecek (2005) proves $\mathbb{F}^*$ is a subfiltration of $(\mathcal{H}_{A_t^*})_{t \geq 0}$. Thus, $N^* - A^*$ is also a $(\mathbb{F}^*, \mathrm{P}^*)$-local martingale. ■

Proof of Lemma A.5. Proofs of the finiteness of moments have a standard approach using Itô's formula and Grönwall's inequality. We will prove the result for X solving (1). The claim then also follows for X^* in (26) via Lemma A.4.

Fix $k \in \mathbb{N}$. Let $\eta_c = \inf\{t > 0 : |X_t| \geq c\}$ for any number $c \in \mathbb{N}$ and $\tau_c = \tau \wedge \eta_c$, both η_c, τ_c, stopping times. We apply Itô's formula with $x \to |x|^{2k}$. The gradient G and the Hessian matrix H of $x \to |x|^{2k}$ simplify as follows.

$$\mathrm{G}_i(x) = \frac{\partial}{\partial x_i}|x|^{2k} = 2|x|^{2(k-1)}x_i$$

$$\mathrm{H}_{ij}(x) = \frac{\partial^2}{\partial x_i \partial x_j}|x|^{2k} = 4k(k-1)|x|^{2(k-2)}xx^\top + 2k|x|^{2(k-1)}\mathrm{I}$$

Define the $d \times d$ matrix D by $\mathrm{D}_{ij}(X) = [X^i, X^j]^{\mathrm{c}}$, the continuous part of the quadratic covariation of X^i and X^j. Its entries simplify as

$$\begin{aligned}\mathrm{D}_{ij}(X_t) &= \Big[\sum_{k=1}^m \int_0^t \sigma_{ik}(X_{s-})\,\mathrm{d}W_s^k, \sum_{\ell=1}^m \int_0^t \sigma_{j\ell}(X_{s-})\,\mathrm{d}W_s^\ell\Big] \\ &= \sum_{k=1}^m \int_0^t \sigma_{ik}(X_s)\sigma_{jk}(X_s)\,\mathrm{d}s.\end{aligned}$$

Then, $\mathrm{dD}(X_s) = \sigma(X_s)\sigma(X_s)^\top \mathrm{d}s$, and consequently $\sum_{i,j} \mathrm{H}_{ij}(X_s)\,\mathrm{dD}_{ij}(X_s) = \mathrm{tr}((\mathrm{H}\sigma\sigma^\top)(X_s))\,\mathrm{d}s$. Since ν is a finite measure (the jumps of X have finite activity), we obtain the following form of Itô's formula for X.

$$\begin{aligned}|X_{T\wedge\tau_c}|^{2k} &= |X_0|^{2k} + \int_0^{T\wedge\tau_c} \mathrm{G}(X_{s-})^\top \mathrm{d}X_s^{\mathrm{c}} + \frac{1}{2}\int_0^{T\wedge\tau_c} \mathrm{tr}((\mathrm{H}\sigma\sigma^\top)(X_{s-}))\,\mathrm{d}s \\ &\quad + \sum_{n=1}^\infty 1_{\{T_n \leq T\wedge\tau_c\}}\big(|X_{T_n-} + \Delta(X_{T_n-}, Z_n)|^{2k} - |X_{T_n-}|^{2k}\big).\end{aligned}$$

All integrands are bounded (due to $1_{\{s\le\tau_c\}}$). Taking the expectation and using that $\Lambda(X)\nu(\mathrm{d}z)$ is the intensity kernel for the marked point process $(T_n, Z_n)_{n\in\mathbb{N}}$,

$$\begin{aligned}\mathrm{E}_x|X_{T\wedge\tau_c}|^{2k} &= |x|^{2k} + \int_0^T \mathrm{E}_x\Big(1_{\{s\le\tau_c\}}\big(\mathrm{G}^\top\mu + \frac{1}{2}\mathrm{tr}(\mathrm{H}\sigma\sigma^\top)\big)(X_s)\Big)\,\mathrm{d}s \\ &\quad + \int_0^T \mathrm{E}_x\Big(1_{\{s\le\tau_c\}}\int_{\mathcal{M}}\big(|X_s+\Delta(X_s,z)|^{2k} - |X_s|^{2k}\big)\,\nu(\mathrm{d}z)\Lambda(X_s)\Big)\mathrm{d}s\,.\end{aligned}$$

We have, for a set of multinomial coefficients $\{\mathscr{C}_\ell\}$ that

$$\begin{aligned}|x+\Delta(x,z)|^{2k} &= \big(x^\top x + \Delta(x,z)^\top(2x+\Delta(x,z))\big)^k \\ &= \sum_{\ell=0}^k \mathscr{C}_\ell\,|x|^{2(k-\ell)}\big(\Delta(x,z)(2x+\Delta(x,z))\big)^\ell\end{aligned}$$

and $\mathscr{C}_0 = 1$. Therefore, setting $L(z,x) = \Delta(x,z)^\top(2x+\Delta(x,z))$,

$$\begin{aligned}\mathrm{E}_x|X_{T\wedge\tau_c}|^{2k} &= |x|^{2k} + \int_0^T \mathrm{E}_x\Big(1_{\{s\le\tau_c\}}\big(\mathrm{G}^\top\mu + \frac{1}{2}\mathrm{tr}(\mathrm{H}\sigma\sigma^\top)\big)(X_s)\Big)\,\mathrm{d}s \\ &\quad + \sum_{\ell=1}^k\int_0^T \mathrm{E}_x\big(1_{\{s\le\tau_c\}}\mathscr{C}_\ell|X_s|^{2(k-\ell)}\Lambda(X_s)\int_{\mathcal{M}} L^\ell(z,X_s)\nu(\mathrm{d}z)\big)\mathrm{d}s\,.\end{aligned}$$

Since $xx^\top$ and $\sigma\sigma^\top$ are positive semidefinite, for some (finite) constant $K'_k > 0$,

$$\begin{aligned}\mathrm{tr}(\mathrm{H}\sigma\sigma^\top)(x) &\le \frac{1}{2}\big(K'_k|x|^{2(k-2)}\mathrm{tr}(xx^\top)\mathrm{tr}(\sigma\sigma^\top) + K'_k|x|^{2(k-1)}\mathrm{tr}(\sigma\sigma^\top)\big) \\ &\le K'_k(1+|x|^{2k}) \qquad (50)\end{aligned}$$

$$\mathrm{G}(x)^\top\mu(x) \le 2|x|^{2(k-1)}x^\top\mu(x) \le K'_k(1+|x|^{2k}) \qquad (51)$$

by Assumption 4.1. By the latter, Λ also has linear growth and combining this fact with Assumption 4.4 guarantees there is $C_\ell \ge 0$ such that

$$\begin{aligned}\Lambda(x)\int_{\mathcal{M}}|L^\ell(z,x)|\,\nu(\mathrm{d}z) &\le 2^\ell\Lambda(x)\int_{\mathcal{M}}\big(|x|^\ell|\Delta(x,z)|^\ell + |\Delta(x,z)|^{2\ell}\big)\,\nu(\mathrm{d}z) \\ &\le C_\ell(1+|x|^{2\ell})\,. \qquad (52)\end{aligned}$$

Assembling the above bounds, we obtain that for some $K_k, \gamma_k \in \mathbb{R}_+$,

$$\mathrm{E}_x\big(|X_{T\wedge\tau_c}|^{2k}\big) \le K_k + |x|^{2k} + \gamma_k\int_0^T \mathrm{E}_x\big(|X_{s\wedge\tau_c}|^{2k}\big)\,\mathrm{d}s < \infty$$

where finiteness is by definition of η_c above and Assumption 4.4. The latter is required since $|X_{s\wedge\tau_c}| \le c^2 + |\Delta(X_{T_n-}, Z_n)|$ on the set $\{T_n = s\wedge\tau_c\}$ which is bounded by that assumption. The γ_k depends on only on k through the $\{C_\ell\}_{\ell=1}^k$ in (52). The K_k also depends only on k through the K'_k in (50) and (51). By Grönwall's inequality, $\mathrm{E}_x\big(|X_{T\wedge\eta_c}|^{2k}\big) \le (K_k + |x|^{2k})\,e^{\gamma_k T}$. Also,

$$|X_{T\wedge\tau_c}| = 1_{\{T\wedge\tau\le\eta_c\}}|X_{T\wedge\tau}| + 1_{\{\eta_c<T\wedge\tau\}}|X_{\eta_c}| \ge c\,1_{\{\eta_c<T\wedge\tau\}},$$

so $\mathrm{P}_x(\eta_c < T\wedge\tau) \le |x|^2 e^{\gamma_1 T}/c^2 \downarrow 0$ as $c \uparrow \infty$. Then $T\wedge\tau_c \uparrow \tau \le T$ as $c\uparrow\infty$ almost surely (Borel-Cantelli). The claim now follows by Fatou's lemma. ■

References

Andersen, Leif & Dan Buffum (2003), 'Calibration and implementation of convertible bond models', *Journal of Computational Finance* **7**, 1–34.

Andersson, Patrik & Arturo Kohatsu-Higa (2017), 'Unbiased simulation of stochastic differential equations using parametrix expansions', *Bernoulli* **23**(3), 2028–2057.

Bally, Vlad & Arturo Kohatsu-Higa (2015), 'A probabilistic interpretation of the parametrix method', *The Annals of Applied Probability* **25**(6), 3095–3138.

Bally, Vlad & Denis Talay (1996), 'The law of the Euler scheme for stochastic differential equations I. Convergence rate of the distribution function.', *Probability Theory and Related Fields* **104**, 43–60.

Beskos, Alexander & Gareth Roberts (2005), 'Exact simulation of diffusions', *Annals of Applied Probability* **15**(4), 2422–2444.

Blanchet, Jose & Fan Zhang (2020), 'Exact simulation for multivariate itô diffusions', *Advances in Applied Probability* **52**(4), 1003–1034.

Bogachev, Vladimir I (2007), *Measure theory*, Vol. 1, Springer Science & Business Media.

Böttcher, Björn (2014), 'Feller evolution systems: Generators and approximation', *Stochastics and Dynamics* **14**(03), 1350025.

Buchmann, Fabian M. & Wesley P. Petersen (2006), 'An exit probability approach to solving high dimensional Dirichlet problems', *SIAM Journal on Scientific Computing* **28**(3), 1153–1166.

Carr, Peter & Dilip Madan (2010), ‘Local volatility enhanced by a jump to default’, *SIAM Journal on Financial Mathematics* **1**, 2–15.

Carr, Peter & Vadim Linetsky (2006), ‘A jump to default extended CEV model: an application of Bessel processes’, *Finance and Stochastics* **10**, 303–330.

Casella, Bruno & Gareth Roberts (2011), ‘Exact simulation of jump-diffusion processes with monte carlo applications’, *Methodology and Computing in Applied Probability* **11**, 449–473.

Chan, Tony F, Gene H Golub & Randall J LeVeque (1983), ‘Algorithms for computing the sample variance: Analysis and recommendations’, *The American Statistician* **37**(3), 242–247.

Chen, Guanting, Alexander Shkolnik & Kay Giesecke (2019), Unbiased simulation estimators for jump-diffusions, *in* ‘2019 Winter Simulation Conference (WSC)’, IEEE, pp. 890–901.

Chen, Guanting, Alexander Shkolnik & Kay Giesecke (2020), Unbiased simulation estimators for path integrals of diffusions, *in* ‘2020 Winter Simulation Conference (WSC), *forthcoming*’.

Chen, Nan & Zhengyu Huang (2013), ‘Localization and exact simulation of Brownian motion driven stochastic differential equations’, *Mathematics of Operations Research* **38**(3), 591–616.

Cheng, Peng & Olivier Scaillet (2007), ‘Linear-quadratic jump-diffusion modeling’, *Mathematical Finance* **17**(4), 575–598.

Crisan, Dan & Francois Delarue (2012), ‘Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations’, *Journal of Functional Analysis* **263**(10), 3024–3101.

Daley, Daryl J & David Vere-Jones (2003), *An introduction to the theory of point processes: volume I: elementary theory and methods*, Springer-Verlag, New York.

Duffie, Darrell, Jun Pan & Kenneth Singleton (2000), ‘Transform analysis and asset pricing for affine jump-diffusions’, *Econometrica* **68**, 1343–1376.

Duffie, Darrell & Peter Glynn (1995), ‘Efficient Monte Carlo estimation of security prices’, *Annals of Applied Probability* **4**(5), 897–905.

Errais, Eymen, Kay Giesecke & Lisa Goldberg (2010), 'Affine point processes and portfolio credit risk', *SIAM Journal of Financial Mathematics* **15**, 642–665.

Ethier, Stewart N & Thomas G Kurtz (2009), *Markov processes: characterization and convergence*, Vol. 282, John Wiley & Sons.

Giesecke, Kay & Alexander Shkolnik (2020), Reducing bias in event time simulations via measure changes. *Mathematics of Operations Research*, forthcoming.

Giesecke, Kay & Dmitry Smelov (2013), 'Exact simulation of jump-diffusions', *Operations Research* **4**(61), 894–907.

Giesecke, Kay & Pascal Tomecek (2005), Dependent events and changes of time. Working Paper, Stanford University.

Giesecke, Kay & Shilin Zhu (2013), 'Transform analysis for point processes and applications in credit risk', *Mathematical Finance* **23**(4), 742–762.

Glasserman, Paul (2003), *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, Berlin.

Glasserman, Paul & Nicolas Merener (2003), 'Numerical solution of jump-diffusion LIBOR market models', *Finance and Stochastics* **1**, 1–27.

Glasserman, Paul & Nicolas Merener (2004), 'Convergence of a discretization scheme for jump-diffusion processes with state-dependent intensities', *Proceedings of the Royal Society of London: Series A* **460**, 1–17.

Gonçalves, F. & G. O. Roberts (2014), 'Exact simulation problems for jump-diffusions', *Methodology and Computing in Applied Probability* **16**, 907–930.

Henry-Labordere, Pierre, Xiaolu Tan & Nizar Touzi (2015), Exact simulation of multi-dimensional stochastic differential equations. Working Paper.

Henry-Labordère, Pierre, Xiaolu Tan & Nizar Touzi (2017), 'Unbiased simulation of stochastic differential equations', *The Annals of Applied Probability* **27**(6), 3305–3341.

Ikeda, Nobuyuki & Shinzo Watanabe (1989), *Stochastic differential equations and diffusion processes*, North Holland, New York.

Johannes, Michael (2004), 'The statistical and economic role of jumps in continuous-time interest rate models', *Journal of Finance* **59**, 227–260.

Kloeden, Peter & Eckhard Platen (1999), *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin.

Kou, Steve (2002), 'A jump-diffusion model for option pricing', *Management Science* **48**, 1086–1101.

Kubilius, Kestutis & Eckhard Platen (2002), 'Rate of weak convergence of the euler approximation for diffusion processes with jumps', *Monte Carlo Methods and Applications* **8**, 83–96.

Kusuoka, Shigeo (2003), 'Malliavin calculus revisited', *J. Math. Sci. Univ. Tokyo* **10**, 261–277.

Kusuoka, Shigeo & Daniel Stroock (1985), 'Applications of Malliavin calculus ii', *J. Faculty Sci. Uni. Tokyo* **32**, 1–76.

Liu, X. Q. & C. W. Li (2000), 'Weak approximations and extrapolations of stochastic differential equations with jumps', *SIAM Journal on Numerical Analysis* **37**(6), 1747–1767.

Maghsoodi, Yoosef (1998), 'Exact solutions and doubly efficient approximations of jump-diffusion Itô equations', *Stochastic Analysis and Applications* **16**, 1049–1072.

Merton, Robert C. (1976), 'Option pricing when the underlying stock returns are discontinuous', *Journal of Financial Economics* **3**, 123–144.

Meyer, Paul-André (1971), Démonstration simplifée d'un théorème de Knight, *in* 'Séminaire de Probabilités V, Lecture Notes in Mathematics 191', Springer-Verlag Berlin, pp. 191–195.

Mikulevicius, Remigijus & Eckhard Platen (1988), 'Time discrete Taylor approximations for Ito processes with jump component', *Mathematische Nachrichten* **138**, 93–104.

Milstein, Grigori N. (1973), 'A method of second-order accuracy integration of stochastic differential equations', *Theory of Probability and Applications* **19**(3), 557–562.

Mordecki, E., A. Szepessy, R. Tempone & G. E. Zouraris (2008), 'Adaptive weak approximation of diffusions with jumps', *SIAM Journal on Numerical Analysis* **46**(4), 1732–1768.

Norros, Ilkka (1986), 'A compensator representation of multivariate life length distributions, with applications', *Scandinavian Journal of Statistics* pp. 99–112.

Platen, Eckhard & Nicola Bruti-Liberati (2010), *Numerical Solution of Stochastic Differential Equations With Jumps in Finance*, Springer-Verlag, Berlin.

Protter, Philip & Denis Talay (1997), 'The Euler scheme for Lévy driven stochastic differential equations.', *Annals of Probability* **25**(1), 393–423.

Rhee, Chang-han & Peter Glynn (2015), 'Unbiased estimation with square root convergence for SDE models', *Operations Research* **63**(5), 1026–1043.

Shaked, Moshe & J George Shanthikumar (1987), 'The multivariate hazard construction', *Stochastic Processes and Their Applications* **24**(2), 241–258.

Talay, Denis & Luciano Tubaro (1990), 'Expansion of the global error for numerical schemes solving stochastic differential equations', *Stochastic Analysis and Applications* **8**(4), 483–509.

Wagner, Wolfgang (1989), 'Unbiased monte carlo estimators for functionals of weak solutions of stochastic differential equations', *Stochastics An International Journal of Probability and Stochastic Processes* **28**(1), 1–20.