Multilevel Monte Carlo simulation for Lévy processes based on the Wiener–Hopf factorisation

A. Ferreiro-Castilla [a,*], A.E. Kyprianou [a], R. Scheichl [a], G. Suryanarayana [b]

[a] *Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, UK*
[b] *Numerical Analysis and Applied Mathematics Section, KU Leuven, Celestijnenlaan 200a, Box 2402, 3001 Heverlee, Leuven, Belgium*

Abstract

In Kuznetsov et al. (2011) a new Monte Carlo simulation technique was introduced for a large family of Lévy processes that is based on the Wiener–Hopf decomposition. We pursue this idea further by combining their technique with the recently introduced multilevel Monte Carlo methodology. Moreover, we provide here for the first time a theoretical analysis of the new Monte Carlo simulation technique in Kuznetsov et al. (2011) and of its multilevel variant for computing expectations of functions depending on the historical trajectory of a Lévy process. We derive rates of convergence for both methods and show that they are uniform with respect to the "jump activity" (e.g. characterised by the Blumenthal–Getoor index). We also present a modified version of the algorithm in Kuznetsov et al. (2011) which combined with the multilevel methodology obtains the optimal rate of convergence for general Lévy processes and Lipschitz functionals. This final result is only a theoretical one at present, since it requires independent sampling from a triple of distributions which is currently only possible for a limited number of processes.
© 2013 Elsevier B.V. All rights reserved.

* Correspondence to: School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. Tel.: +44 0 20 7882 5471.
 E-mail addresses: aferreiro.c@gmail.com, a.ferreiro-castilla@qmul.ac.uk (A. Ferreiro-Castilla).

1. Introduction

A classical problem in mathematical finance deals with the ability to compute $\mathbb{E}[F(X)]$, where $X := \{X_s : s \in [0, t]\}$ is a stochastic process which models the underlying risky asset and F is a payoff function which may depend on the historical path of X. This is the typical setting for pricing a variety of exotic options in finance such as look back and barrier options. Starting with the early work of Madan and Seneta [30], a class of processes playing the role of X that have found prominence in this respect is the class of Lévy processes. An extensive overview of their position in mathematical finance can be found in the books [6,11, 33,34]. More recently Lévy processes have also been extensively used in modern insurance risk theory; see for example Asmussen and Albrecher [2] and Klüppelberg et al. [25]. In insurance mathematics, it is the Lévy process itself which models the surplus wealth of an insurance company until its ruin. There are also extensive applications of Lévy processes in queuing theory, genetics and mathematical biology as well as in stochastic differential equations (see e.g. [8,14,16,23]).

In both financial and insurance settings, a key quantity of generic interest is the joint law of the current position and the running maximum of a Lévy process at a fixed time, if not the individual marginals associated with the latter bivariate law. Consider the following example. If we define $\overline{X}_t = \sup_{s \le t} X_s$ then the pricing of barrier options boils down to evaluating expectations of the form $\mathbb{E}[f(x + X_t)\mathbf{1}_{\{x+\overline{X}_t > b\}}]$ for some threshold $b > 0$. Indeed if $f(x) = (K - e^x)^+$ and $K > 0$, then the latter expectation is related to the value of an "up-and-in" put. In credit risk one is predominantly interested in the quantity $\widehat{\mathbb{P}}(\overline{X}_t < x)$ as a function in x and t, where $\widehat{\mathbb{P}}$ is the law of the dual process $-X$. Indeed it is a functional of the latter probabilities that gives the price of a credit default swap not to mention the recently introduced financial instruments known as convertible contingencies (CoCos). See for example the recent book of Schoutens and Cariboni [34] as well as Corcuera et al. [12]. One is similarly interested in $\widehat{\mathbb{P}}(\overline{X}_t \ge x)$ in ruin theory, since these probabilities are also equivalent to the finite-time ruin probabilities; cf. Asmussen and Albrecher [2].

A widely used approach to compute expectations of functions depending on the historical trajectory of a Lévy process over the time horizon, say $[0, t]$, is to approximate the path by a random walk with n steps, each step covering t/n units of time, and therewith to perform a Monte Carlo (MC) simulation. Giles [18,19] introduced an adaptation of the straightforward MC methodology, the multilevel Monte Carlo method (MLMC), which is especially suited to the scenario we are interested in, but in the case that X is a pure diffusion process. Very recently, there has been increasing attention to the MLMC method also in the setting of Lévy processes, see Giles and Xia [20] for the jump–diffusion setting, and Dereich [14] and Dereich and Heidenreich [15] for more general Lévy processes. Generally speaking all these methods share a common approach, which consists in constructing an embedded sequence of grids that are made up of a mixture of deterministic and random points. The random points in these grids deal with the large jumps of the Lévy process and the deterministic points deal with the "small movements", that is to say, the diffusive part and/or the small jumps.

In this paper we consider an alternative method based entirely on a random grid. In particular we shall introduce an adaptation of the MLMC method based on a very recently introduced technique for performing MC simulations that appeals to the so-called Wiener–Hopf factorisation for one-dimensional Lévy processes. This last technique is called the Wiener–Hopf Monte Carlo (WHMC) simulation method and it was introduced in Kuznetsov et al. [28]. We will denote the

coupling of these two techniques the multilevel Wiener–Hopf Monte Carlo method (MLWH). Our analysis will focus on the particular setting that X is a one-dimensional Lévy process and F is a Lipschitz function which depends on the value of X and its past supremum at a fixed time $t > 0$.

The main result in this paper shows that the root mean square error in the MLWH method converges in general with order $\mathcal{O}(\nu^{-\frac{1}{4}})$ for processes of unbounded variation and $\mathcal{O}(\nu^{-\frac{1}{3}})$ for processes of bounded variation, where ν stands for computational cost measured as floating point operations of the algorithm. Unlike Dereich [14], Dereich and Heidenreich [15], and Giles and Xia [20], our method is robust in the sense that the convergence rate does not exhibit dependence on the nature of the jump structure (in particular small jumps) of the underlying Lévy process. This means that we can ensure a convergence rate of at least $\mathcal{O}(\nu^{-\frac{1}{4}})$ independently of the structure of the Lévy measure, which is important in the implementation of numerical methods for calibration purposes, as slightly different parameters in a model can lead to completely different behaviour of the process with respect to the jump structure. Moreover, the implementation of the algorithm is extremely straightforward, requiring only to be able to independently sample from two distributions. This robustness comes at the price of our algorithm not having optimal convergence rates, in particular when the underlying Lévy process has paths of bounded variation. For Lévy processes of unbounded variation we are able to derive regimes where our methodology performs better than the algorithms in Dereich [14], Dereich and Heidenreich [15] and Jacod et al. [22] and, as far as we know, better than any other existing algorithm. We will also show the existence of another multi-level algorithm, which is a modification of the original algorithm in Kuznetsov et al. [28], that achieves a convergence order $\mathcal{O}(\nu^{-\frac{1}{2}} \log^2 \nu)$ independently of the jump structure. As shown in Creutzig et al. [13], up to the logarithmic term this is in fact optimal for general Lévy processes. Unfortunately however, this algorithm requires sampling from a triplet of dependent distributions for which there are currently very few known examples of Lévy processes where this is possible. We shall elaborate further on this point later in this paper.

As alluded to above, we shall benchmark our MLWH results against those of Dereich and Heidenreich [15] and Dereich [14]. Note that, even though these two papers consider the more general setting of expectations of path functionals of solutions to SDEs driven by general Lévy processes, the scenario we consider here is permitted within their framework and we are currently working on extending our methodology to the more general setting.

The paper is organised as follows. In the next section we will review the general setting for the Wiener–Hopf factorisation of Lévy processes and describe the WHMC method introduced in Kuznetsov et al. [28]. Thereafter, in Section 3, we describe how an adaptation of the WHMC method can be used in the context of multilevel simulation, that is to say, we formally introduce the MLWH method. Section 4 is devoted to the numerical analysis of the one level (WHMC) and the multilevel (MLWH) method, in particular, giving mathematical justification to our claims concerning the performance of the MLWH method above. Moreover, we discuss how the MLWH method performs against other approaches that have been mentioned. Section 5 will introduce a theoretical variation of the original MLWH, which we can show achieves optimal convergence rates but which, unfortunately, cannot yet be implemented in practice until more substantial examples of the Wiener–Hopf factorisation are discovered. Finally, Section 6 provides some implementation details and numerical computations for a representative example which support the previous theoretical derivations and confirm the feasibility of the approach.

2. Preliminaries

2.1. Lévy processes and the Wiener–Hopf factorisation

We begin by briefly reviewing the definition of a one-dimensional Lévy process and its associated Wiener–Hopf factorisation. For a more in-depth account we refer the reader to the monographs of Bertoin [4], Kyprianou [29] or Sato [32]. Note that the Wiener–Hopf factorisation only exists for one-dimensional Lévy processes.

Recall that a one-dimensional Lévy process with law $\mathbb{P}$, henceforth denoted by $X := \{X_t : t \geq 0\}$, is a stochastic process issued from the origin which enjoys the properties of having stationary and independent increments with paths that are almost surely right-continuous with left limits. It is a well understood fact that, as a consequence of this definition, the law of every Lévy process is characterised through a triplet (a, σ, Π), where $a \in \mathbb{R}$, $\sigma \geq 0$ and Π is a measure concentrated on $\mathbb{R} \setminus \{0\}$ such that $\int_{\mathbb{R}} (1 \wedge x^2) \Pi(\mathrm{d}x) < \infty$. More precisely, for all $t \geq 0$,

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\theta X_t}] = \mathrm{e}^{-t\,\Psi(\theta)}, \quad \text{for all } \theta \in \mathbb{R},$$

where

$$\Psi(\theta) = \mathrm{i}a\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}} (1 - \mathrm{e}^{\mathrm{i}\theta x} + \mathrm{i}\theta x \mathbf{1}_{(|x|<1)}) \Pi(\mathrm{d}x)$$

is the so-called characteristic exponent of the process.

A property that is common to all Lévy processes is the so-called Wiener–Hopf factorisation. Suppose that for any $q > 0$, $\mathbf{e}(q)$ is an exponentially distributed random variable with mean q^{-1} that is independent of X. Recall that $\overline{X}_t = \sup_{s \leq t} X_s$ and let $\underline{X}_t := \inf_{s \leq t} X_s$. The Wiener–Hopf factorisation states that the random variables $\overline{X}_{\mathbf{e}(q)}$ and $\overline{X}_{\mathbf{e}(q)} - X_{\mathbf{e}(q)}$ are independent. Thanks to the so-called principle of duality, that is to say the equality in law of the pair $\{X_{(t-s)-} - X_t : 0 \leq s \leq t\}$ and $\{-X_s : 0 \leq s \leq t\}$, it follows that $\overline{X}_{\mathbf{e}(q)} - X_{\mathbf{e}(q)}$ is equal in distribution to $-\underline{X}_{\mathbf{e}(q)}$. This leads to the following factorisation of characteristic functions

$$\mathbb{E}(\mathrm{e}^{\mathrm{i}\theta X_{\mathbf{e}(q)}}) = \mathbb{E}(\mathrm{e}^{\mathrm{i}\theta \overline{X}_{\mathbf{e}(q)}}) \times \mathbb{E}(\mathrm{e}^{\mathrm{i}\theta \underline{X}_{\mathbf{e}(q)}}), \quad \text{for all } \theta \in \mathbb{R},$$

known as the Wiener–Hopf factorisation. Equivalently,

$$X_{\mathbf{e}(q)} \overset{d}{=} S_q + I_q, \tag{1}$$

where S_q and I_q are independent and equal in distribution to $\overline{X}_{\mathbf{e}(q)}$ and $\underline{X}_{\mathbf{e}(q)}$, respectively. Here we use the notation $\overset{d}{=}$ to mean equality in distribution.

We will work under the following assumptions:

(A1) $\int_{|x| \geq 1} x^2 \Pi(\mathrm{d}x) < \infty$,

(A2) the payoff function $F : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ is a Lipschitz function with Lipschitz constant assumed to be 1 for simplicity,

(A3) the computational time to sample from $\underline{X}_{\mathbf{e}(q)}$ and $\overline{X}_{\mathbf{e}(q)}$ is independent of the value of q.

The first assumption asks for much less than what is commonly accepted in the financial literature, where typically exponential moments of the truncated Lévy measure are required. Note that this condition ensures that for each $t \geq 0$, X_t has a finite second (and hence also a first) moment.

The second and third assumptions are equivalent to the corresponding conditions imposed in Dereich and Heidenreich [15] and Dereich [14]. A justification for the third assumption in the

case that X belongs to the so-called β-class of Lévy processes is provided in Section 6. As we shall discuss in more detail below, the β-class is a large family of Lévy processes which is both widely suitable for use in financial modelling, as well as for implementation of our algorithm.

2.2. The Wiener–Hopf Monte Carlo method

Due to the independent increments of Lévy processes, the most common approaches to the Monte Carlo simulation of expectations involving the joint law of $(X_t, \overline{X}_t)$ work with random walk approximations to the Lévy process. This requires one either to be able to simulate the increments of the Lévy process exactly for a fixed time step or to be able to suitably approximate the Lévy process, typically by a jump–diffusion process.

Recently introduced by Kuznetsov et al. [28], the so-called Wiener–Hopf Monte Carlo method is related to the simulation of increments. However, it does not work with a fixed deterministic grid, but instead requires the underlying grid to be random with independent and exponentially distributed spacings (i.e. the arrival times of a compound Poisson process). This allows us to simulate the paths of a very large class of Lévy processes and thus to sample from the law of $(X_\tau, \overline{X}_\tau)$ where τ is a random time whose distribution can be concentrated arbitrarily close around t depending on a parameter chosen in the algorithm that controls the resolution of the random grid and thus the amount of work.

To describe the WHMC method in more detail, let us suppose that $\mathbf{e}_1(1), \mathbf{e}_2(1), \ldots$ are a sequence of i.i.d. exponentially distributed random variables with unit mean. The basis of the WHMC algorithm is the following simple observation, which follows directly from the Strong Law of Large Numbers. For all $t > 0$,

$$\sum_{i=1}^{n} \frac{t}{n} \mathbf{e}_i(1) \to t \quad \text{as } n \uparrow \infty \tag{2}$$

almost surely. Note that the random variable on the left hand side of (2) can also be written as the sum of n independent random variables with an exponential distribution having mean t/n and therefore is equal in law to a Gamma random variable with parameters n and n/t, henceforth written as $\mathbf{g}(n, n/t)$. For sufficiently large n, Kuznetsov et al. [28] argue that a suitable approximation to $\mathbb{P}(X_t \in \mathrm{d}x, \overline{X}_t \in \mathrm{d}y)$ is $\mathbb{P}(X_{\mathbf{g}(n,n/t)} \in \mathrm{d}x, \overline{X}_{\mathbf{g}(n,n/t)} \in \mathrm{d}y)$. Indeed, it is a triviality to note that, thanks to (2) and the independence of $\mathbf{g}(n, n/t)$ from X, the pair $(\overline{X}_{\mathbf{g}(n,n/t)}, X_{\mathbf{g}(n,n/t)})$ converges almost surely to $(\overline{X}_t, X_t)$ as $n \uparrow \infty$. We will compute rates of convergence in Section 4.

The following theorem is straightforward to prove using (1) together with the stationary and independent increments of the underlying Lévy process.

Theorem 2.1 (*Kuznetsov et al. [28, Thm. 1]*)**.** *Let* $\{S_{n/t}^j : j \geq 1\}$ *and* $\{I_{n/t}^j : j \geq 1\}$ *be i.i.d. sequences of random variables with common distribution equal to that of* $\overline{X}_{\mathbf{e}(n/t)}$ *and* $\underline{X}_{\mathbf{e}(n/t)}$, *respectively. Then, for all* $n \in \mathbb{N}$,

$$(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)}) \stackrel{d}{=} (V(n, n/t), J(n, n/t)),$$

where, for any $k \in \mathbb{N}$, *and setting* $V(0, n/t) := 0$ *we define*

$$V(k, n/t) = V(k-1, n/t) + \left(S_{n/t}^k + I_{n/t}^k\right), \tag{3}$$

$$J(k, n/t) = \bigvee_{j=1}^{k} \left\{ V(j-1, n/t) \; + \; S_{n/t}^{j} \right\}. \tag{4}$$

It is clear from earlier remarks on the convergence of $\left(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)}\right)$ that the pair $(V(n, n/t), J(n, n/t))$ converges in distribution to $(X_t, \overline{X}_t)$. Theorem 2.1 suggests that as soon as we are able to simulate i.i.d. copies of the distributions of $S_{n/t}$ and $I_{n/t}$, then by the simple functional transformations given in (3) and (4), we may produce an exact draw from the distribution of $(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)})$. Moreover, for a suitably nice function F, using standard Monte Carlo methods based on the Strong Law of Large Numbers, one may estimate $\mathbb{E}(F(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)}))$ by

$$\widehat{F}_{\mathrm{MC}}^{n,M} := \frac{1}{M} \sum_{i=1}^{M} F^{n,(i)}, \tag{5}$$

where $F^{n,(i)}$ is the i-th sample of

$$F^n := F\left(V(n, n/t), J(n, n/t)\right).$$

Indeed, we have $\lim_{M \uparrow \infty} \widehat{F}_{\mathrm{MC}}^{n,M} = \mathbb{E}\left(F(X_{\mathbf{g}(n,n/t)}, \overline{X}_{\mathbf{g}(n,n/t)})\right)$ almost surely, which in turn converges to $\mathbb{E}(F(X_t, \overline{X}_t))$ as $n \uparrow \infty$.

As alluded to above, the WHMC method is numerically feasible only if samples from the distributions of $S_{n/t}$ and $I_{n/t}$ are available. Until recently, this would have proved to be a significant stumbling block on account of there being few examples for which the aforesaid distributions are known in explicit form. However, developments in Wiener–Hopf theory for Lévy processes in the last couple of years (see for example Kuznetsov [26] or Kuznetsov et al. [27,28]) have provided a rich enough variety of examples for which the necessary distributional sampling can be performed. This family of processes is named meromorphic Lévy processes in Kuznetsov et al. [27,28]. One large subfamily of such processes is the β-class of Lévy processes, which also conveniently offers all the desirable properties of better known Lévy processes that are used in mathematical finance, such as CGMY processes, VG processes or Meixner processes; see for example the discussions in Ferreiro-Castilla and Schoutens [17] and Schoutens and van Damme [35].

2.3. The WHMC method in context

Let us now spend a little time contextualising the WHMC method for simulating the path of a Lévy process against related possible alternative schemes, pointing out in particular where we should expect to see improved efficiencies in working with the Wiener–Hopf factorisation. A general point of reference for further reading is the book of Cont and Tankov [11].

Generally speaking, there are three alternative approaches to simulate the path of a Lévy process. The first one relies on the ability to simulate the increments of the Lévy process exactly for a fixed (deterministic) time step and therefore construct a random walk as discussed at the beginning of Section 2.2. This method requires knowledge of the distribution of the Lévy process at a fixed time, either through an exact analytical formula, or via numerical inversion of the characteristic function. There are also general results on the simulation of infinite divisible distributions, for which the distribution of a Lévy process at a fixed time is an example, see e.g., Bondesson [5]. As mentioned earlier, approximating a Lévy process by an

embedded random walk may introduce significant errors on path functionals of X such as the running maximum which is of prime interest in applications we have in mind (cf. Broadie and Glasserman [7]).

The second approach is to use a time-dependent infinite series expansion to approximate the value of the Lévy process at each fixed time. A general result in this direction is found in Rosiński [31] for the case that an explicit expression for the Lévy measure is known in closed form. The aforesaid series representation converges uniformly and almost surely in any compact set of time. This lends itself better to sampling path functionals of the process than, perhaps the random walk approximation but it might make the numerical analysis difficult.

Finally, the third and most common approach is to approximate X via a jump–diffusion process; that is to say a Lévy process which can be written as a linear Brownian motion plus an independent compound Poisson process. This is done by truncating the Lévy measure, removing all small jumps below a certain threshold in magnitude and compensating for their removal by making an appropriate adjustment to the linear and/or Gaussian component. The truncation of small jumps ensures that the remaining jumps conform to a compound Poisson structure and hence one is left with simulating the path of a linear Brownian motion, interlaced with jumps distributed according to the normalised truncated Lévy measure arriving at an appropriate Poissonian rate. The method of truncating small jumps, originally due to Asmussen and Rosiński [3], is an obvious approach (cf. Cont and Tankov [11, Sect. 6.3] and the references therein) and it is the approach taken by Dereich and Heidenreich [15] in their design of multilevel methods for Lévy-driven stochastic differential equations. There it is observed that when the jump component of X is of finite variation, one may reasonably replace the small jumps by a linear trend. Further, in the case of an unbounded variation jump component it is seen that a more appropriate approximation is to replace the small jumps by a Gaussian process. This truncation method is also the approach used in Dereich [14]. Whilst [15,14] successfully demonstrate the convergence of this truncation method, there are limitations; see, for example, the discussion in Asmussen and Rosiński [3].

The main differences between these three methods and the WHMC simulation scheme are three-fold. Firstly, whilst the WHMC scheme is restricted to Lévy processes that allow sampling from the distributions of the variables $\overline{X}_{\mathbf{e}(q)}$ and $\underline{X}_{\mathbf{e}(q)}$, it is otherwise indifferent to the jump structure of the underlying Lévy process. Indeed, the approximation of a Lévy process by a compound Poisson process, which is the most popular approach, relies on the Lévy–Itô decomposition. The latter decomposes the process with respect to the endogenous jump structure. In contrast, the Wiener–Hopf factorisation concerns the decomposition of the path of a Lévy process according to the distribution of its maximum. Secondly, the WHMC method relies entirely on a randomised time grid. Note for comparison that the approach in [15,14] requires a deterministic grid interlaced by a random grid capturing the large jumps. Thirdly, the WHMC works on the principle of sampling the Lévy process over the "wrong" time horizon (but in a way which can be made arbitrarily close to the desired time horizon) purely in order to be able to sample from the exact maximum. Note that a similar idea of randomising the time horizon in a computation involving the expectation of a functional of the path of a linear Brownian motion that appears in the context of pricing American options is described in Carr [9]. We shall also see in the forthcoming numerical analysis that this idea of sampling over the "wrong" time horizon turns out to make the numerical analysis of the WHMC method, as well as its multilevel version, relatively straightforward. Indeed the rate of convergence of the WHMC and of the MLWH method can be expressed directly in terms of the rate of convergence of the randomised time horizon to the fixed time t.

Since the Wiener–Hopf technique naturally leads to a random walk with exponential time-spacings that will on average shrink as $n \uparrow \infty$, the WHMC and MLWH methods are in principle extendable to more general problems than the one considered here, for instance to approximate solutions of Lévy driven SDEs (as done in [15,14]).

3. Multilevel Wiener–Hopf Monte Carlo simulation

As noted already in the context of pure diffusion processes by Giles [18,19] and others, optimal convergence rates in simulating from random processes are only achievable by using multiple levels of time grids. This has its roots in multigrid techniques for deterministic differential equations. It relies on the fact that large features or longtime trends of a process can be approximated on a coarse time grid to a sufficient accuracy that can be related clearly to the time step size. Smaller features are then added as corrections to this longtime trend on a finer time grid. This can be done without introducing additional bias, and the benefit is that the variance of this correction on the finer grid is substantially smaller than the variance of the original process on the same grid, leading to a *variance reduction* and thus a smaller number of Monte Carlo samples, necessary to achieve a prescribed absolute tolerance. This variance reduction stems from the fact that our time discretisation scheme is a convergent process and, provided the convergence rates are known, the variance reduction, and thus the cost of the method, can be rigorously quantified. Applying this idea recursively on a sequence of ever finer grids, allows the simulation of the entire process with all small features in optimal computational complexity for any prescribed tolerance of the bias and the standard deviation. In more general terms, we are simulating from a random process by combining large numbers of very coarse and cheaply available paths with small numbers of fine paths that are expensive to compute. We refer to Giles [18,19] and Cliffe et al. [10] and the references therein for a detailed overview of the method, as well as to Giles and Xia [20], Dereich and Heidenreich [15] and Dereich [14] for applications to simulating jump–diffusions and Lévy processes. Our interest here is to construct a multilevel version of the WHMC method.

3.1. Definition of the multilevel method

Since the multilevel method relies on variance reduction, it is worth recalling the formula for the mean square error of the WHMC method for subsequent reference. The mean square error between the estimator $\widehat{F}_{\mathrm{MC}}^{n,M}$ and $\mathbb{E}[F(X_t, \overline{X}_t)]$ is

$$e(\widehat{F}_{\mathrm{MC}}^{n,M})^2 := \mathbb{E}[(\widehat{F}_{\mathrm{MC}}^{n,M} - \mathbb{E}[F(X_t, \overline{X}_t)])^2].$$

Since $\mathbb{E}[\widehat{F}_{\mathrm{MC}}^{n,M}] = \mathbb{E}[F^n]$ and $\mathbb{V}[\widehat{F}_{\mathrm{MC}}^{n,M}] = M^{-1}\mathbb{V}[F^n]$, where $\mathbb{V}[\cdot]$ denotes variance, it can be decomposed as follows:

$$\begin{aligned}
e(\widehat{F}_{\mathrm{MC}}^{n,M})^2 &= \mathbb{E}[(\widehat{F}_{\mathrm{MC}}^{n,M} - \mathbb{E}[\widehat{F}_{\mathrm{MC}}^{n,M}])^2] + (\mathbb{E}[\widehat{F}_{\mathrm{MC}}^{n,M}] - \mathbb{E}[F(X_t, \overline{X}_t)])^2 \\
&= \mathbb{V}[\widehat{F}_{\mathrm{MC}}^{n,M}] + \left(\mathbb{E}[\widehat{F}_{\mathrm{MC}}^{n,M}] - \mathbb{E}[F(X_t, \overline{X}_t)]\right)^2 \\
&= M^{-1}\mathbb{V}[F^n] + \left(\mathbb{E}[F^n - F(X_t, \overline{X}_t)]\right)^2.
\end{aligned} \tag{6}$$

The first term in the above decomposition is the variance of the WHMC simulation and the second one is the bias of the approximation induced by the randomised time horizon. Recall that F^n converges in distribution to $F(X_t, \overline{X}_t)$ as $n \uparrow \infty$. We shall say that F^n becomes a *finer* approximation of $F(X_t, \overline{X}_t)$ the larger n becomes. Conversely, the approximation F^n is said to become *coarser* the smaller the value of n.

The MLWH method starts from the simple observation that we can write the expectation of the finest approximation F^{n_L} as a telescopic sum starting from a coarser approximation F^{n_0}, as well as intermediate ones:

$$\mathbb{E}[F^{n_L}] = \mathbb{E}[F^{n_0}] + \sum_{\ell=1}^{L} \mathbb{E}[F^{n_\ell} - F^{n_{\ell-1}}], \quad 1 \le n_0 < n_1 < \cdots < n_L. \tag{7}$$

A typical choice for n_ℓ/n_0, $\ell = 0, \ldots, L$, are powers of 2, or more generally of some integer $s \in \mathbb{N}$, i.e. $n_\ell := n_0 s^\ell$, for some $n_0 \in \mathbb{N}$. For the remainder of this paper, it suffices to simply think of $n_\ell := n_0 2^\ell$, for $\ell = 0, \ldots, L$. The MLWH method now consists in independently computing each of the expectations of the telescopic sum by a standard Monte-Carlo method, i.e.

$$\widehat{F}_{\mathrm{ML}}^{\mathcal{M}(n_0, L)} := \frac{1}{M_0} \sum_{i=1}^{M_0} F^{n_0,(i)} + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left(F^{n_\ell,(i)} - F^{n_{\ell-1},(i)} \right), \tag{8}$$

where $\mathcal{M}(n_0, L) := \{M_\ell\}_{\ell=0}^{L}$. Analogously to (6), we can again expand the mean square error for the multilevel estimator to obtain

$$e(\widehat{F}_{\mathrm{ML}}^{\mathcal{M}(n_0, L)})^2 := \frac{1}{M_0} \mathbb{V}[F^{n_0}] + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \mathbb{V}[F^{n_\ell} - F^{n_{\ell-1}}] + \left(\mathbb{E}[F^{n_L} - F(X_t, \overline{X}_t)] \right)^2. \tag{9}$$

Note that the bias term remains the same, i.e. we have not introduced any additional bias. However, by a judicious choice of $\mathcal{M}(n_0, L)$ it is possible to reduce the variance of the estimator (i.e. the first two terms) for any pre-chosen computational effort, or conversely reduce the computational cost for any pre-selected tolerance of the variance of the estimator.

The variance reduction stems from the fact that F^n converges to $F(X_t, \overline{X}_t)$ in mean square. This is equivalent to the mean square convergence of the Cauchy sequence $F^{n_\ell} - F^{n_{\ell-1}}$, and thus $\mathbb{V}[F^{n_\ell} - F^{n_{\ell-1}}] \to 0$. This implies that the contribution to the total variance from the finer levels (where the cost per sample is high) is decreasing with $n_\ell \to \infty$, allowing us to use only very few of these expensive samples, while on the coarser levels the cost of samples is smaller anyway. At this point in time, we have not yet established rigorously that F^n converges to $F(X_t, \overline{X}_t)$, not to mention at what rate. This will be dealt with in Section 4. However, let us assume mean square convergence for now.

Before continuing, it is first necessary to elaborate on another important point. One of the defining features of the multilevel method is that both approximations of the payoff function in simulating from the random variable $(F^{n_\ell} - F^{n_{\ell-1}})$ should come from the same draw. This means that once the increments of the Lévy process have been sampled to obtain a draw for F^{n_ℓ}, a deterministic transformation of the sample points and of the increments should take place to obtain a sample for $F^{n_{\ell-1}}$ (or vice versa). While this procedure is clear in the original setting of Giles [19], where a diffusion process is approximated by a random walk on a deterministic grid, it is not entirely trivial how to proceed in the case of a general Lévy process and with respect to random grids.

3.2. Thinning a Poisson random grid

Our approach to construct a sample for a coarser approximation from the draw of a finer approximation relies on the technique of Poisson thinning. A similar approach was considered in

Giles and Xia [20] in the case of jump–diffusion processes and can be traced back further (also for general Lévy processes) to, e.g., Glasserman and Merener [21].

The first step to sample from $(F^{n_\ell} - F^{n_{\ell-1}})$ on level ℓ is to construct a Poisson random grid with rate n_ℓ/t, i.e. the inter-arrival times are independently distributed as an exponential random variable with mean t/n_ℓ. Fix $\ell \in \mathbb{N}$ and $t > 0$ and let $N^\ell := \{N_s^\ell : s \geq 0\}$ denote a Poisson process with arrival rate n_ℓ/t; the arrival times in N^ℓ are denoted by $\{T_k^{N,\ell} : k \geq 0\}$, with $T_0^{N,\ell} = 0$. Recall that, without loss of generality, we restricted to the case where $n_\ell/n_{\ell-1} = 2$ and notice that under this notation we have $\mathbf{g}(n_\ell, n_\ell/t) = T_{n_\ell}^{N,\ell}$. It is a well established fact (cf. Kingman [24]) that if we censor arrivals in the process N^ℓ by tossing independent fair coins at each arrival and ignoring the arrival if the coin lands, say, on heads, then the resulting point process in time has the same law as $N^{\ell-1}$. In particular the inter-arrival times in this coarser grid are independent and exponentially distributed with mean $t/n_{\ell-1}$. The general case $n_\ell/n_{\ell-1} = s$, for some $s > 1$, could clearly be treated in an analogous manner by tossing a biased coin with probability of acceptance $1/s$.

To make the construction more precise, let us return to the case $s = 2$ and consider the sequences $\{S_{n_\ell/t}^j : j \geq 1\}$ and $\{I_{n_\ell/t}^j : j \geq 1\}$ of i.i.d. random variables with common distributions $\overline{X}_{\mathbf{e}(n_\ell/t)}$ and $\underline{X}_{\mathbf{e}(n_\ell/t)}$ respectively, where the exponential periods are taken from the Poisson process N^ℓ. Suppose we set $\kappa_0 = 0$, and for $i \geq 1$, we write κ_i for the index of the i-th accepted arrival in the process N^ℓ after censoring. Note that for $i \geq 1$, $\kappa_i - \kappa_{i-1}$ are a sequence of i.i.d. geometrically distributed random variables with parameter one half. Note also that, for $i \geq 1$, the time that elapses in the process N^ℓ between the $(i-1)$-th (after censoring) arrival and the i-th (after censoring) arrival is equal in distribution to precisely

$$\sum_{j=\kappa_{i-1}+1}^{\kappa_i} \mathbf{e}_j(n_\ell/t). \tag{10}$$

It is a straightforward computation to show that the above random variable is exponentially distributed with mean $t/n_{\ell-1}$, e.g. by computing its moment generating function.

Thanks to the Poisson thinning, we can now construct the independent sequences $\{S_{n_{\ell-1}/t}^i : i \geq 1\}$ and $\{I_{n_{\ell-1}/t}^i : i \geq 1\}$ of random variables, which are needed for sampling from $F^{n_{\ell-1}}$, via a deterministic transformation of the sequences $\{S_{n_\ell/t}^j : j \geq 1\}$ and $\{I_{n_\ell/t}^j : j \geq 1\}$ as follows:

$$S_{n_{\ell-1}/t}^i = \bigvee_{k=1}^{\kappa_i - \kappa_{i-1}} \left\{ \sum_{j=1}^{k-1} \left(S_{n_\ell/t}^{\kappa_{i-1}+j} + I_{n_\ell/t}^{\kappa_{i-1}+j} \right) + S_{n_\ell/t}^{\kappa_{i-1}+k} \right\} \tag{11}$$

$$I_{n_{\ell-1}/t}^i = \sum_{j=\kappa_{i-1}+1}^{\kappa_i} \left(S_{n_\ell/t}^j + I_{n_\ell/t}^j \right) - S_{n_{\ell-1}/t}^i, \tag{12}$$

where $S_{n_{\ell-1}/t}^i$ and $I_{n_{\ell-1}/t}^i$ are independent due to the Wiener–Hopf factorisation. Although it is now clear how to construct the infinite sequence of pairs $\{(S_{n_{\ell-1}/t}^i, I_{n_{\ell-1}/t}^i) : i \geq 1\}$ from $\{(S_{n_\ell/t}^j, I_{n_\ell/t}^j) : j \geq 1\}$, some explanation is still needed when dealing with finite sequences, such as those required by the MLWH method. When the Poisson random grid is stopped at the time horizon $\mathbf{g}(n_\ell, n_\ell/t)$, it will in general be necessary to first extend the number of exponential periods from the process N^ℓ beyond $\mathbf{g}(n_\ell, n_\ell/t)$ before thinning to get the corresponding $\ell - 1$

level grid. This is because we cannot ensure that after thinning, the total non-censored points in N^ℓ up to $\mathbf{g}(n_\ell, n_\ell/t)$ sum up to $n_{\ell-1}$. Alternatively, one can produce the remaining terms of the coarser level by sampling directly from $(S_{n_{\ell-1}/t}, I_{n_{\ell-1}/t})$ due to the thinning theorem.

Let us point out here another issue directly related to the fact that we are constructing our algorithm based on a completely random grid. Denote by $\check{F}^{n_{\ell-1}}$ a sample produced from F^{n_ℓ} by the thinning methodology described above. Since the telescopic sum in (7) has to cancel out, i.e. we do not want to introduce extra bias, it is important that the expectation of $F^{n_{\ell-1}}$ is the same as the expectation of $\check{F}^{n_{\ell-1}}$. This is called the consistency of the multilevel algorithm. Thanks to the thinning theorem this is straightforward to check in the above construction, however it will play a role in Section 5. Indeed, the above description asserts that thinning a Poisson process N^ℓ will lead in distribution to a Poisson process of rate $n_{\ell-1}/t$ and is thus equal in distribution to $N^{\ell-1}$. Therefore the random time $\mathbf{g}(n_{\ell-1}, n_{\ell-1}/t)$ has the same distribution as the resulting thinned version from $\mathbf{g}(n_\ell, n_\ell/t)$, say $\check{\mathbf{g}}(n_{\ell-1}, n_{\ell-1}/t)$. Finally, due to the fact that $\mathbf{g}(n_{\ell-1}, n_{\ell-1}/t)$ and $\check{\mathbf{g}}(n_{\ell-1}, n_{\ell-1}/t)$ are independent of the underlying Lévy process the consistency follows as $\check{F}^{n_{\ell-1}}$ and $F^{n_{\ell-1}}$ have the same law.

4. Analysis of Wiener–Hopf Monte Carlo simulation

Henceforth we shall use the following notation. We will write $a \lesssim b$ for two positive quantities a and b, if a/b is uniformly bounded independent of any parameters, such as t, L, $\{n_\ell\}_{\ell=0}^L$, or $\{M_\ell\}_{\ell=0}^L$, as well as of any parameters describing the underlying Lévy process. We will write $a \approx b$, if $a \lesssim b$ and $b \lesssim a$.

4.1. Abstract convergence theorems

There are two ways to quantify and compare the complexity of algorithms. We can either specify a target accuracy ε and then estimate the cost to achieve an error below this target accuracy with a particular method, or conversely, we can analyse the convergence of an algorithm as a function of the cost. The results are easily interchangeable. The first approach is the one chosen in Giles [19] and Cliffe et al. [10]. However, Dereich and Heidenreich [15] and Dereich [14] chose the second approach and since we will mainly compare our results with theirs, we will also choose this approach here. To fix notation, computational cost is measured in floating point operations here (but it could equally be CPU-time). Thus, for the following let us assume that the expected value of the computational cost $\mathcal{C}(\widehat{F})$ to compute a particular estimator $\widehat{F}$ of $\mathbb{E}[F(X_t, \overline{X}_t)]$ is bounded by v, i.e.

$$\mathbb{E}\big[\mathcal{C}(\widehat{F})\big] \lesssim v.$$

Then we want to bound the root mean square error $e(\widehat{F})$ in terms of v.

Let us start by analysing the single-level WHMC method. The following result can be easily deduced from (6) with the modification that the cost per sample is itself a random variable. It will become clear in the proof of Theorems 4.6 and 5.3 why we need this generalisation.

Theorem 4.1. *Let $t > 0$ and let us assume that F^n converges in mean square to $F(X_t, \overline{X}_t)$. Suppose further that there exist positive constants $\alpha, \gamma > 0$ such that*

(i) $|\mathbb{E}[F^n - F(X_t, \overline{X}_t)]| \lesssim n^{-\alpha}$,
(ii) $\mathbb{E}[\mathcal{C}_n] \lesssim n^\gamma$,

where $\mathcal{C}_n$ represents the cost of computing a single sample of F^n on a Poisson grid with rate n/t. Then, for every $\nu \in \mathbb{N}$, there exist $n, M \in \mathbb{N}$ such that

$$\mathbb{E}\left[\mathcal{C}\left(\widehat{F}_{\mathrm{MC}}^{n,M}\right)\right] \lesssim \nu \quad and \quad e\left(\widehat{F}_{\mathrm{MC}}^{n,M}\right) \lesssim \nu^{-\frac{1}{2+\gamma/\alpha}}.$$

Proof. Assuming for the moment that $\mathbb{V}[F^n]$ is bounded independently of n, then balancing the two terms on the right hand side of (6) and using assumption (i), we see that we should choose $M \asymp n^{2\alpha}$. Now, since

$$\mathcal{C}\left(\widehat{F}_{\mathrm{MC}}^{n,M}\right) = M\,\mathcal{C}_n,$$

we can deduce from assumption (ii) that the expected value of the total cost of the estimator will be bounded by ν, if we choose $n \asymp \nu^{\frac{1}{2\alpha+\gamma}}$ and $M \asymp \nu^{\frac{2\alpha}{2\alpha+\gamma}}$, which leads to the required bound for $e\big(\widehat{F}_{\mathrm{MC}}^{n,M}\big)$.

It remains to bound $\mathbb{V}[F^n]$. First note that trivially $\mathbb{V}[F^n] \leq \mathbb{E}[(F^n)^2]$ and so

$$\frac{1}{2}\,\mathbb{V}[F^n] \leq \mathbb{E}[(F^n - F(X_t, \overline{X}_t))^2] + \mathbb{E}[F(X_t, \overline{X}_t)^2]. \tag{13}$$

The first term can be bounded independently of n using the Lipschitz continuity of F in Assumption (A2) and our assumption that F^n converges in mean square to $F(X_t, \overline{X}_t)$. The second term is bounded due to the Lipschitz continuity of F and Assumption (A1). $\square$

Thanks to the general considerations in Giles [19] and Cliffe et al. [10, Theorem 1], the analysis of the multilevel version follows in the same way. The proof of the following result is similar to that in [10, Appendix A], with the same modifications as above to deal with random computational cost.

Theorem 4.2. *Let $t > 0$ and $n_\ell = n_0 s^\ell$, for some $\ell, n_0 \in \mathbb{N}$ and $s > 1$, and suppose that there are positive constants $\alpha, \beta, \gamma > 0$ with $\alpha \geq \frac{1}{2}(\beta \wedge \gamma)$ such that*

(i) $|\mathbb{E}[F^{n_\ell} - F(X_t, \overline{X}_t)]| \lesssim n_\ell^{-\alpha}$
(ii) $\mathbb{V}(F^{n_\ell} - F^{n_{\ell-1}}) \lesssim n_\ell^{-\beta}$
(iii) $\mathbb{E}[\mathcal{C}_{n_\ell}] \lesssim n_\ell^{\gamma}$,

where $\mathcal{C}_n$ represents the cost of computing a single sample of F^n on a Poisson grid with rate n/t. Then, for every $\nu \in \mathbb{N}$, there exist a value L and a sequence $\mathcal{M}(n_0, L) = \{M_\ell\}_{\ell=0}^L$ such that

$$\mathbb{E}\left[\mathcal{C}\left(\widehat{F}_{\mathrm{ML}}^{\mathcal{M}(n_0,L)}\right)\right] \lesssim \nu \quad and \quad e\left(\widehat{F}_{\mathrm{ML}}^{\mathcal{M}(n_0,L)}\right) \lesssim \begin{cases} \nu^{-\frac{1}{2}}, & \text{if } \beta > \gamma, \\ \nu^{-\frac{1}{2}} \log \nu, & \text{if } \beta = \gamma, \\ \nu^{-\frac{1}{2+(\gamma-\beta)/\alpha}}, & \text{if } \beta < \gamma. \end{cases}$$

In order to apply Theorems 4.1 and 4.2 now, we need to find suitable α, β and γ such that the necessary assumptions are satisfied. We will verify in the next subsection that the values of α and β in the hypothesis of Theorems 4.1 and 4.2 are directly related to the convergence of the time horizon $\mathbf{g}(n, n/t)$ to t. As for the value of γ, Assumption (A3) in Section 2 implies that for both the single and the multilevel algorithm we always have $\gamma = 1$. This is clear for the single level case, since the cost is not random, but a proof is needed for the multilevel version as we might have to enlarge the finer grid in the thinning procedure as described in Section 3.2.

Lemma 4.3. *Under the notation of Theorem 4.2, assume that $\mathcal{C}(F^{n_\ell}) \lesssim n_\ell$. Then the expected cost of producing a sample of $F^{n_\ell} - F^{n_{\ell-1}}$ via the thinning methodology of Section 3.2 is*

$$\mathbb{E}[\mathcal{C}(F^{n_\ell} - F^{n_{\ell-1}})] \lesssim n_\ell.$$

Proof. Let us first focus on the case that $n_\ell/n_{\ell-1} = 2$ and the thinning methodology corresponds to tossing a fair coin. In this case, the probability of choosing i arrival times from the finer level when thinning is $\binom{n_\ell}{i} 2^{-n_\ell}$ and hence one easily concludes from the assumptions that

$$\mathbb{E}[\mathcal{C}(F^{n_\ell} - F^{n_{\ell-1}})] \lesssim n_\ell + \sum_{i=0}^{n_\ell/2} \binom{n_\ell}{i} \frac{1}{2^{n_\ell}} \mathcal{C}(F^{\frac{n_\ell}{2} - i})$$

$$\lesssim n_\ell + \sum_{i=0}^{n_\ell/2} \binom{n_\ell}{i} \frac{1}{2^{n_\ell}} \left(\frac{n_\ell}{2} - i\right)$$

$$\leq n_\ell + \frac{n_\ell}{2} \frac{1}{2^{n_\ell}} \sum_{i=0}^{n_\ell} \binom{n_\ell}{i} = \frac{3n_\ell}{2},$$

where $\mathcal{C}(F^{\frac{n_\ell}{2} - i})$ corresponds to the cost of producing the remaining terms on the coarser level after choosing i terms on the finer level.

In the general case where $n_\ell/n_{\ell-1} = s$ the probability $\binom{n_\ell}{i} 2^{-n_\ell}$ should be replaced by $\binom{n_\ell}{i} s^{-i} (1 - s^{-1})^{n_\ell - i}$ and a similar result holds using the binomial theorem. $\quad\square$

4.2. Explicit convergence rates

Let us write for convenience $\mathbf{X}_t = (X_t, \overline{X}_t)$, for $t \geq 0$, and let $\boldsymbol{\tau}$ and $\boldsymbol{\tau}'$ be any non-negative random variables that are independent of X (but potentially correlated). Then, thanks to the Lipschitz Assumption (A2) on F, it is straightforward to deduce that

$$|\mathbb{E}[F(\mathbf{X}_\tau) - F(\mathbf{X}_t)]| \leq \mathbb{E}[|\mathbf{X}_\tau - \mathbf{X}_t|] \leq \mathbb{E}[|X_\tau - X_t|] + \mathbb{E}[|\overline{X}_\tau - \overline{X}_t|] \tag{14}$$

and

$$\mathbb{V}(F(\mathbf{X}_\tau) - F(\mathbf{X}_{\tau'})) \leq \mathbb{E}[|\mathbf{X}_\tau - \mathbf{X}_{\tau'}|^2] \leq \mathbb{E}[(X_\tau - X_{\tau'})^2] + \mathbb{E}[(\overline{X}_\tau - \overline{X}_{\tau'})^2]. \tag{15}$$

To estimate the quantities on the right hand sides of (14) and (15) we will make use of the following lemma.

Lemma 4.4. *Suppose that X is a Lévy process which satisfies Assumption (A1). Let us denote by $\boldsymbol{\tau}$ any non-negative random variable, independent of X. Then, for any fixed $t > 0$,*

(i) $\mathbb{E}[(X_\tau - X_t)^2] = \mathbb{V}[X_1] \mathbb{E}[|\tau - t|] + (\mathbb{E}[X_1])^2 \mathbb{E}[(\tau - t)^2]$,

(ii) $\mathbb{E}[(\overline{X}_\tau - \overline{X}_t)^2] \leq 8\mathbb{V}[X_1] \mathbb{E}[|\tau - t|] + 2(\mathbb{E}[X_1] \vee 0)^2 \mathbb{E}[(\tau - t)^2]$.

Proof. (i) Due to the fact that increments of a Lévy process are i.i.d. we can write

$$X_\tau - X_t \overset{d}{=} \begin{cases} X_{\tau - t}, & \text{if } \tau \geq t, \\ -X_{t - \tau}, & \text{if } \tau < t. \end{cases}$$

Since X and $\boldsymbol{\tau}$ are assumed to be independent and since $\mathbb{E}[X_s^2] = \mathbb{V}[X_1] s + (\mathbb{E}[X_1])^2 s^2$, for all $s \geq 0$, this implies

$$\mathbb{E}[(X_\tau - X_t)^2] = \mathbb{V}[X_1]\mathbb{E}[|\tau - t|] + (\mathbb{E}[X_1])^2 \mathbb{E}[(\tau - t)^2].$$

(ii) To obtain the result for the mean square error of the supremum process we first note the following equality in distribution

$$\overline{X}_t \stackrel{d}{=} \overline{X}_s \vee (X_s + \overline{X}'_{t-s}),$$

where $s < t$ and $\overline{X}'_{t-s}$ is independent of $\{X_{s'} : s' \leq s\}$ and identically distributed to $\overline{X}_{t-s}$. Taking account of the duality property for Lévy processes, which tells us that $\underline{X}_s$ is equal in distribution to $X_s - \overline{X}_s$, we have that

$$\overline{X}_t - \overline{X}_s \stackrel{d}{=} 0 \vee (\underline{X}''_s + \overline{X}'_{t-s}) \leq \overline{X}'_{t-s},$$

where $\underline{X}''_s$ is independent of $\overline{X}'_{t-s}$ and equal in distribution to $\underline{X}_s$. It is now easy to check that

$$\mathbb{E}[(\overline{X}_\tau - \overline{X}_t)^2] \leq \mathbb{E}[\overline{X}^2_{|\tau-t|}].$$

Let us decompose X into its martingale part $X^* := \{X^*_t : t \geq 0\}$ and a drift, i.e. $X_t = X^*_t + t\mathbb{E}[X_1]$, for $t \geq 0$. Due to Assumption (A1) this is always possible. Then we have

$$\mathbb{E}[\overline{X}^2_t] = \mathbb{E}\left[\left(\sup_{s \leq t}(X^*_s + \mathbb{E}[X_1]s)\right)^2\right]$$

$$\leq \mathbb{E}[(\overline{X}^*_t + t(\mathbb{E}[X_1] \vee 0))^2]$$

$$\leq 2\left(\mathbb{E}[(\overline{X}^*_t)^2] + (\mathbb{E}[X_1] \vee 0)^2 t^2\right), \tag{16}$$

where $\overline{X}^*_t = \sup_{s \leq t} X^*_t$. Appealing to Doob's martingale inequality (see e.g. Sato [32, p. 167]) we end up with the inequality

$$\mathbb{E}[\overline{X}^2_t] \leq 2\left(4\mathbb{E}[(X^*_t)^2] + (\mathbb{E}[X_1] \vee 0)^2 t^2\right), \tag{17}$$

from which we may now write (again remembering that τ is independent of X)

$$\mathbb{E}[\overline{X}^2_{|\tau-t|}] \leq 8\mathbb{E}[(X^*_{|\tau-t|})^2] + 2(\mathbb{E}[X_1] \vee 0)^2 \mathbb{E}[(\tau - t)^2].$$

Since $\mathbb{E}[(X^*_{|\tau-t|})^2] = \mathbb{E}[(X^*_\tau - X^*_t)^2]$, we may use the conclusion of part (i) and finally write

$$\mathbb{E}[\overline{X}^2_{|\tau-t|}] \leq 8\mathbb{E}[(X^*_\tau - X^*_t)^2] + 2(\mathbb{E}[X_1] \vee 0)^2 \mathbb{E}[(\tau - t)^2]$$

$$= 8\mathbb{V}[X^*_1]\mathbb{E}[|\tau - t|] + 2(\mathbb{E}[X_1] \vee 0)^2 \mathbb{E}[(\tau - t)^2],$$

which concludes the proof since $\mathbb{V}[X^*_1] = \mathbb{V}[X_1]$. □

Using this lemma we can now verify the remaining hypotheses in Theorems 4.1 and 4.2 and in particular establish the values for the parameters α and β in the convergence bounds for the WHMC and MLWH methods.

Proposition 4.5. *Let $t > 0$ and X satisfies Assumption* (A1). *Then, for any $n \in \mathbb{N}$, we have*

(i) $\mathbb{E}[(X_{\mathbf{g}(n,n/t)} - X_t)^2] \approx n^{-1/2}$ *and* $\mathbb{E}[|X_{\mathbf{g}(n,n/t)} - X_t|] \lesssim n^{-1/4}$,
(ii) $\mathbb{E}[(\overline{X}_{\mathbf{g}(n,n/t)} - \overline{X}_t)^2] \lesssim n^{-1/2}$ *and* $\mathbb{E}[|\overline{X}_{\mathbf{g}(n,n/t)} - \overline{X}_t|] \lesssim n^{-1/4}$.

If, in addition, X has paths of bounded variation then we have the sharper bound

$$\mathbb{E}[|X_{\mathbf{g}(n,n/t)} - X_t|] \lesssim n^{-1/2} \quad \text{and} \quad \mathbb{E}[|\overline{X}_{\mathbf{g}(n,n/t)} - \overline{X}_t|] \lesssim n^{-1/2}.$$

Proof. It suffices to prove the results for the second moments in (i) and (ii). The results for the first moments then follow from Jensen's inequality. With a view to applying Lemma 4.4, we will show that for all $n \in \mathbb{N}$

$$\mathbb{E}[(\mathbf{g}(n, n/t) - t)^2] = \frac{t^2}{n} \tag{18}$$

$$\mathbb{E}[|\mathbf{g}(n, n/t) - t|] = 2t \mathrm{e}^{-n} \frac{n^n}{n!} \approx n^{-\frac{1}{2}}, \tag{19}$$

where the final equivalence in (19) follows from Stirling's formula.

First note that, since $\mathbb{E}[\mathbf{g}(n, n/t)] = t$, Eq. (18) is just the formula for the variance of the gamma distribution $\mathbf{g}(n, n/t)$. For Eq. (19), it is clear that

$$\mathbb{E}[|\mathbf{g}(n, n/t) - t|] = \int_0^t (t - s) g(n, n/t, s) \mathrm{d}s + \int_t^\infty (s - t) g(n, n/t, s) \mathrm{d}s$$

$$= 2 \int_0^t (t - s) g(n, n/t, s) \mathrm{d}s, \tag{20}$$

where $g(k, \theta, s) = \Gamma(k)^{-1} \theta^k s^{k-1} \mathrm{e}^{-\theta s}$. Recall the definition of the incomplete gamma function, $\gamma(k, u) := \int_0^u x^{k-1} \mathrm{e}^{-x} \mathrm{d}x$, so that

$$\int_0^t g(k, \theta, s) \mathrm{d}s = \frac{\gamma(k, t\theta)}{\Gamma(k)} \quad \text{and} \quad \int_0^t s g(k, \theta, s) \mathrm{d}s = \frac{\gamma(k+1, t\theta)}{\Gamma(k)\theta}.$$

We can now rewrite the expression (20) as

$$\int_0^t (t - s) g(n, n/t, s) \mathrm{d}s = t \int_0^t g(n, n/t, s) \mathrm{d}s - \int_0^t s g(n, n/t, s) \mathrm{d}s$$

$$= t \left(\frac{\gamma(n, n)}{\Gamma(n)} - \frac{\gamma(n+1, n)}{\Gamma(n+1)} \right). \tag{21}$$

From the representation of $\gamma(k, u)$ in, for example, Abramowitz and Stegun [1, Section 6.5.13], we can write for $k = 1, 2, \ldots,$

$$\frac{\gamma(k, u)}{\Gamma(k)} = \frac{1}{\Gamma(k)} \int_0^u x^{k-1} \mathrm{e}^{-x} \mathrm{d}x = 1 - \left(1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots + \frac{u^{k-1}}{(k-1)!} \right) \mathrm{e}^{-u},$$

which, when substituted into (21), together with (20) gives the equality (19).

If X has paths of bounded variation then we can return to the proof of Lemma 4.4, in particular the estimates in (16) and (17), and use similar computations together with Doob's martingale inequality for absolute moments of centred Lévy martingales (cf. Sato [32, p. 167]) to deduce that

$$\mathbb{E}[|\overline{X}_{\mathbf{g}(n,n/t)} - \overline{X}_t|] \leq \mathbb{E}[\overline{X}^*_{|\mathbf{g}(n,n/t)-t|}] + (\mathbb{E}[X_1] \vee 0) \mathbb{E}[|\mathbf{g}(n, n/t) - t|]$$

$$\leq 8 \mathbb{E}[|X^*_{|\mathbf{g}(n,n/t)-t|}|] + (\mathbb{E}[X_1] \vee 0) \mathbb{E}[|\mathbf{g}(n, n/t) - t|]. \tag{22}$$

Note also that $\mathbb{E}[|X_{\mathbf{g}(n,n/t)} - X_t|] = \mathbb{E}[|X_{|\mathbf{g}(n,n/t)-t|}|]$ is essentially bounded by a linear combination of $\mathbb{E}[|X^*_{|\mathbf{g}(n,n/t)-t|}|]$ and $\mathbb{E}[|\mathbf{g}(n, n/t) - t|]$. Taking account of (19), we therefore see that the proof is complete as soon as we show that $\mathbb{E}[|X^*_{|\mathbf{g}(n,n/t)-t|}|] \lesssim n^{-1/2}$.

To this end, note that every Lévy process of bounded variation can be written as the difference of two independent subordinators. Accordingly we shall write, for any $t \geq 0$, $X_t^* = X_t' - X_t''$. Assumption (A1) ensures that both X' and X'' have finite first and second moments at all fixed times. Therefore, using a similar derivation as above, as well as (19), we have

$$\mathbb{E}[|X_{|\mathbf{g}(n,n/t)-t|}^*|] \leq \mathbb{E}[X_{|\mathbf{g}(n,n/t)-t|}'] + \mathbb{E}[X_{|\mathbf{g}(n,n/t)-t|}'']$$
$$= \mathbb{E}[|\mathbf{g}(n, n/t) - t|]\mathbb{E}[|X_1'| + |X_1''|] \approx n^{-1/2}. \quad \square$$

We are now ready to state the main result of the paper which gives the convergence rates of the WHMC and the MLWH method.

Theorem 4.6. *Suppose that Assumptions* (A1)–(A3) *are satisfied. Then, the hypotheses in Theorems* 4.1 *and* 4.2 *hold with* $\alpha = 1/4$, $\beta = 1/2$ *and* $\gamma = 1$. *In the case that X has paths of bounded variation we have* $\alpha = 1/2$. *Thus, for any* $v \in \mathbb{N}$ *and under the constraint that the expected value of the total computational cost is* $\mathcal{O}(v)$ *operations, the root mean square errors for the single and multilevel methods are*

$$e(\widehat{F}_{\mathrm{MC}}^{n,M}) \lesssim \begin{cases} v^{-\frac{1}{6}}, & \text{if } X \text{ is of unbounded variation,} \\ v^{-\frac{1}{4}}, & \text{if } X \text{ is of bounded variation,} \end{cases} \tag{23}$$

and

$$e(\widehat{F}_{\mathrm{ML}}^{\mathcal{M}(n_0,L)}) \lesssim \begin{cases} v^{-\frac{1}{4}}, & \text{if } X \text{ is of unbounded variation,} \\ v^{-\frac{1}{3}}, & \text{if } X \text{ is of bounded variation.} \end{cases} \tag{24}$$

Proof. This follows immediately from Theorems 4.1 and 4.2, Proposition 4.5, Eqs. (14) and (15) and the fact that $\gamma = 1$ (cf. Assumption (A3) and Lemma 4.3). $\quad \square$

Remark 4.7. Note that Assumption (A2) implicitly transforms the weak error estimate in Theorems 4.1 and 4.2 into a strong error estimate through inequality (14). This means that by using Proposition 4.5 we only get a lower bound estimate of the asymptotic parameter α. By noting that the time horizon $\mathbf{g}(n, n/t)$ is an unbiased estimator of t one easily sees that $|\mathbb{E}[X_{\mathbf{g}(n,n/t)} - X_t]| = 0$. Although this does not hold for the supremum process in general, for particular cases and under regularity constraints on the function F, estimates for the weak convergence[1] might be computable, see for instance the approach taken in Jacod et al. [22] in the case of functions depending on the terminal value of the solution of a SDE. We have chosen to follow the approach taken in Dereich [14] and Dereich and Heidenreich [15] which allows for more general functions F with the penalisation that we will have only strong estimates. We will compute numerically the weak error in the numerical implementation of the algorithm in Section 6.1 and show that there is a substantial improvement of the estimates of parameter α.

We finish this section by comparing the convergence results of Theorem 4.6 with those of Dereich and Heidenreich [15], Dereich [14] and Jacod et al. [22]. Since all these alternative methods show a significant dependence on the jump structure of the underlying Lévy process, let

---

[1] Weak convergence refers here to the behaviour of $|\mathbb{E}[F(\mathbf{X}_{\mathbf{g}(n,n/t)})] - \mathbb{E}[F(\mathbf{X}_t)]|$ as $n \uparrow \infty$.
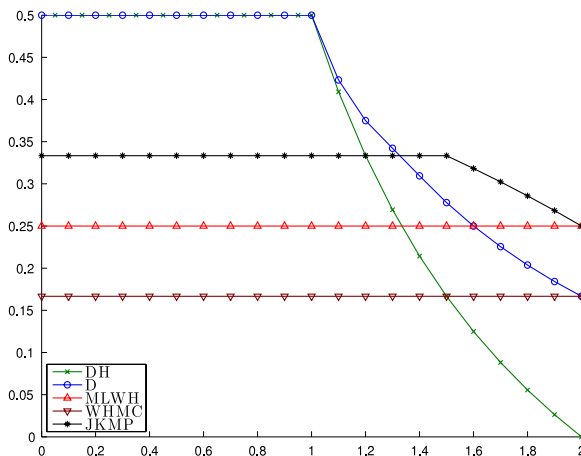
Fig. 1. Order of convergence with respect to the Blumenthal–Getoor index.

us first recall the definition of the Blumenthal–Getoor index,

$$\rho := \inf \left\{ \beta > 0 : \int_{(-1,1)} |x|^\beta \, \Pi(\mathrm{d}x) < \infty \right\} \in [0, 2].$$

The higher this index, the more important small jumps are. Lévy processes with finite activity will always have Blumenthal–Getoor index $\rho = 0$.

The approach in Dereich and Heidenreich [15] is to approximate the underlying Lévy process via a jump–diffusion process, to remove all the small jumps and to replace them by an additional linear component. While this method works optimally for processes with jump component of bounded variation, i.e. Blumenthal–Getoor index $\rho \leq 1$, the convergence rate degenerates rapidly when ρ gets larger and it becomes arbitrarily bad as $\rho \to 2$. See Fig. 1 for a plot of the rate of convergence of this method (denoted by **DH** in the figure). As shown in Dereich [14], a better performance is possible when a Brownian correction is added in addition to the linear trend to better take account of the discarded small jumps. This is the curve denoted by **D** in Fig. 1. It converges even for $\rho = 2$, but the convergence rate in this case decreases to $\nu^{-\frac{1}{6}}$.

Our Wiener–Hopf based methods, **WHMC** (single-level) and **MLWH** (multilevel), are indifferent to the type of jump structure and thus entirely independent of the Blumenthal–Getoor index ρ. The price to pay for this robustness is that, in the bounded variation regime (i.e. for Lévy processes without Gaussian component and with $\rho < 1$), the multilevel version is not quite optimal; recall that for a very general class of Lévy processes X and payoff functions F the optimal rate of convergence is $\nu^{-\frac{1}{2}}$ (cf. Creutzig et al. [13] and Dereich [14] for comments to that effect). Indeed, for $\rho < 1$ and $\sigma = 0$, the multilevel algorithm performance does not match the rate achieved by the methods of Dereich, but it coincides with the rate reported in Jacod et al. [22] (depicted as **JKMP** in Fig. 1). The performance of the single level algorithm in the bounded variation regime improves to the rate of **MLWH** for the general case. We do not depict these lines in Fig. 1. Let us instead focus on the regime where the Blumenthal–Getoor index is $\rho > 1$, i.e. the Lévy process has unbounded variation paths. In this case our single-level algorithm **WHMC** performs better than the method in Dereich and Heidenreich [15] for $\rho > 4/3$, but it is not able to beat the improved method in Dereich [14]. The multilevel version

MLWH, on the other hand, does outperform Dereich [14] for $\rho > 8/5$ and seems to be at least as good as the best currently available method for $\rho = 2$. Let us reiterate though that the class of functions F we have so far analysed is particular in that they depend only on the terminal value X_t and the maximum $\overline{X}_t$ of X, whereas in Dereich [14] and Dereich and Heidenreich [15] more general functionals of the entire path of X are allowed. Nevertheless, in the region where the Blumenthal–Getoor index is close to 2 our multilevel version **MLWH** is better or equal to any other existing algorithm. This becomes important when we do not have any *a priori* information of the Lévy process to be used in a numerical application, and hence a robust method with respect to the jump structure is crucial. For example, one often faces the problem of fitting a parametric family of Lévy processes to given data with a range of parameters allowing unbounded variation jump component. In such cases the method proposed by Dereich and Heidenreich [15] may not converge in any reasonable time since the numerical performance of the method is restricted by the worst convergence rate in Fig. 1.

As mentioned above, we believe that our approach can also be adapted to the more general scenario considered in those publications, involving SDEs and path dependent functionals F, and still retain the same rates of convergence. We leave this issue to a further piece of work. In the context of more special functionals, it is also worth recalling the results in Jacod et al. [22]. Our multilevel algorithm only matches the performance of their method for $\rho = 2$. However, the assumptions in [22] are more restrictive than our assumptions and those in Dereich et al. [15], since only functionals of the terminal value X_t are allowed. Moreover, their functions are required to be four times continuously differentiable, highlighting some of the drawbacks and advantages when considering weak instead of strong estimates (cf. Remark 4.7).

The next section is devoted to a description of another Wiener–Hopf-based algorithm, which achieves optimal convergence rates up to a logarithmic term for all values of the Blumenthal–Getoor index. Despite the fact that we are able to derive theoretical rates of convergence, the algorithm requires the ability to sample from a triplet of dependent distributions coming out of the Wiener–Hopf factorisation, for which currently no known significant examples are known.

5. The random time $\mathbf{T}(n, t)$

A careful look at Section 2.2 reveals that there is nothing special about the random time $\mathbf{g}(n, n/t)$ in the description of the WHMC method of Theorem 2.1 other than it being a sum of independent and exponentially distributed random variables converging almost surely to t. Therefore one may investigate more efficient ways to approximate t by a sum of exponentials. We now present an alternative random time that exhibits faster convergence rates.

Let $N := \{N_s : s \geq 0\}$ be a Poisson process independent of X with intensity t/n, where $t > 0$ and $n \in \mathbb{N}$, associated with the Wiener–Hopf random walk (3). Let T_j^N be the time of the j-th arrival in N, that is to say $T_j^N = \inf\{s > 0 : N_s = j\}$ and define $\mathbf{T}(n, t) = T_{N_t+1}^N$. Then, it is well known and intuitively clear from the lack of memory property of the exponentially distributed inter-arrival times of N, that $\mathbf{T}(n, t) - t$ follows an exponential distribution with mean t/n. Moreover, $\lim_{n \uparrow \infty} \mathbf{T}(n, t) = t$ almost surely, and hence $(X_{\mathbf{T}(n,t)}, \overline{X}_{\mathbf{T}(n,t)})$ also converges to $(X_t, \overline{X}_t)$ almost surely. Therefore the same arguments as those found in Theorem 2.1 suggest that one may try to approximate $(X_{\mathbf{T}(n,t)}, \overline{X}_{\mathbf{T}(n,t)})$ by $(V(N_t+1, n/t), J(N_t+1, n/t))$. Unfortunately, current knowledge of the Wiener–Hopf factorisation only gives us information about the pair $(\overline{X}_{\mathbf{e}(q)}, X_{\mathbf{e}(q)} - \overline{X}_{\mathbf{e}(q)})$, rather than the triplet $(\overline{X}_{\mathbf{e}(q)}, X_{\mathbf{e}(q)} - \overline{X}_{\mathbf{e}(q)}, \mathbf{e}(q))$. Therefore we are not able to perform a simultaneous exact sample from $(V(N_t + 1, n/t), J(N_t + 1, n/t), N_t + 1)$,

hence the random time $\mathbf{T}(n, t)$ cannot be used in a practical implementation with the theory developed in Kuznetsov et al. [28].

The time horizon $\mathbf{T}(n, t)$ leads to a slightly different construction or modification of the WHMC estimator described in Theorem 2.1 and eventually a different construction of the MLWH estimator and its associated algorithm, but for the sake of simplicity the Monte-Carlo estimate of $\mathbb{E}(F(\overline{X}_t, X_t))$ is still denoted by $\widehat{F}_{\mathrm{MC}}^{n,M}$ and written as in (5). However, for any $i \in \mathbb{N}$, we now understand $F^{n,(i)}$ as the i-th draw of the random variable

$$F^n = F(V(N_t + 1, n/t), J(N_t + 1, n/t)). \tag{25}$$

Remark 5.1. A third time horizon which constitutes only a minor variation of $\mathbf{T}(n, t)$, is to consider the random time $\widetilde{\mathbf{T}}(n, t) := T_{N_t}^N$, in which case $t - \widetilde{\mathbf{T}}(n, t)$ has the same distribution as $\mathbf{e}(n/t) \wedge t$. It is an easy exercise to see that all results applying to $\mathbf{T}(n, t)$ will hold also for $\widetilde{\mathbf{T}}(n, t)$ and therefore we avoid any further comments on $\widetilde{\mathbf{T}}(n, t)$.

5.1. Thinning and convergence theorems

For the time horizon $\mathbf{T}(n, t)$ the thinning procedure follows analogously to the description in Section 3.2 and Eqs. (11) and (12) remain true. The issue of consistency is not entirely straightforward however. Following the notation in Section 3.2, let ℓ be one of the levels in the multilevel algorithm and denote by $\check{\mathbf{T}}(n_{\ell-1}, t)$ the time horizon produced by thinning the Poisson process N^ℓ. We now show that $\check{\mathbf{T}}(n_{\ell-1}, t)$ has the same distribution as $\mathbf{T}(n_{\ell-1}, t)$. It is clear from the algorithm that

$$\mathbf{T}(n_\ell, t) \le \check{\mathbf{T}}(n_{\ell-1}, t). \tag{26}$$

Indeed, due to the coin tossing either the two time horizons in (26) are equal or differ in a geometric sum of exponentials. Since $\mathbf{T}(n_\ell, t) - t \overset{d}{=} \mathbf{e}(n_\ell/t)$ we note that

$$\check{\mathbf{T}}(n_{\ell-1}, t) - t = \mathbf{T}(n_\ell, t) - t + \sum_{i=1}^{\kappa} \mathbf{e}_i(n_\ell/t) \overset{d}{=} \sum_{i=0}^{\kappa} \mathbf{e}_i(n_\ell/t)$$

$$\overset{d}{=} \mathbf{e}(n_{\ell-1}/t) \overset{d}{=} \mathbf{T}(n_{\ell-1}, t) - t , \tag{27}$$

where κ is an independent geometric distribution on $\{0, 1, 2, \ldots\}$ with parameter $1/2$. We then conclude that, as in Section 3.2, $\check{F}^{n_{\ell-1}}$ and $F^{n_{\ell-1}}$ have the same law.

Eqs. (26) and (27) suggest an interesting simplification. Given the pair $(X_{\mathbf{T}(n_\ell,t)}, \overline{X}_{\mathbf{T}(n_\ell,t)})$, rather than extending the random grid on level ℓ and using the thinning algorithm together with (11) and (12) to produce a coarser sample, we only have to flip a coin once. If it is a head, then $(X_{\mathbf{T}(n_{\ell-1},t)}^{(i)}, \overline{X}_{\mathbf{T}(n_{\ell-1},t)}^{(i)}) = (X_{\mathbf{T}(n_\ell,t)}^{(i)}, \overline{X}_{\mathbf{T}(n_\ell,t)}^{(i)})$ and thus the entire sample $F^{n_\ell,(i)} - F^{n_{\ell-1},(i)} = 0$. On the other hand, if it is a tail, we just need to simulate one independent draw $(S_{n_{\ell-1}/t}^{(i)}, I_{n_{\ell-1}/t}^{(i)})$ of $(\overline{X}_{\mathbf{e}(n_{\ell-1}/t)}, \underline{X}_{\mathbf{e}(n_{\ell-1}/t)})$ and set

$$\overline{X}_{\mathbf{T}(n_{\ell-1},t)}^{(i)} = \overline{X}_{\mathbf{T}(n_\ell,t)}^{(i)} \vee (X_{\mathbf{T}(n_\ell,t)}^{(i)} + S_{n_{\ell-1}/t}^{(i)}) , \tag{28}$$

$$X_{\mathbf{T}(n_{\ell-1},t)}^{(i)} = X_{\mathbf{T}(n_\ell,t)}^{(i)} + S_{n_{\ell-1}/t}^{(i)} + I_{n_{\ell-1}/t}^{(i)}. \tag{29}$$

Observe that Eqs. (28) and (29) significantly simplify the multilevel algorithm for the time horizon $\mathbf{T}(n, t)$.

Proposition 5.2. *Let $t > 0$ and suppose X satisfies Assumption* (A1). *Then, for any $n \in \mathbb{N}$,*

(i) $\mathbb{E}[(X_{\mathbf{T}(n,t)} - X_t)^2] \eqsim n^{-1}$ *and* $\mathbb{E}[|X_{\mathbf{T}(n,t)} - X_t|] \lesssim n^{-1/2}$,
(ii) $\mathbb{E}[(\overline{X}_{\mathbf{T}(n,t)} - \overline{X}_t)^2] \lesssim n^{-1}$ *and* $\mathbb{E}[|\overline{X}_{\mathbf{T}(n,t)} - \overline{X}_t|] \lesssim n^{-1/2}$.

Proof. As in the proof of Proposition 4.5, it is enough to prove the second moment results in (i) and (ii), since the results for the first moments then follow from Jensen's inequality. Thanks to the lack of memory property, it is well known that $\mathbf{T}(n,t) - t$ follows an exponential distribution with mean t/n. It follows that, for all $n \in \mathbb{N}$,

$$\mathbb{E}[(\mathbf{T}(n,t) - t)^2] = \frac{2t^2}{n^2} \quad \text{and} \quad \mathbb{E}[|\mathbf{T}(n,t) - t|] = \mathbb{E}[\mathbf{T}(n,t) - t] = \frac{t}{n}. \tag{30}$$

The desired result now follows by appealing to Lemma 4.4. □

As for Theorem 4.6, the next theorem follows immediately from Theorems 4.1 and 4.2, Proposition 5.2, and the fact that $\gamma = 1$ (cf. Assumption (A3)). Note that in this case the cost of producing a single path F^{n_ℓ} is random, as the Wiener–Hopf walk will have a random number of iterations. It is nevertheless clear that $\gamma = 1$ in Theorem 4.1, but also in Theorem 4.2, since we will have to produce at most one additional exponential time step with rate $n_{\ell-1}$ in order to construct the coarser sample (as shown in Eq. (27)).

Theorem 5.3. *Suppose that Assumptions* (A1)–(A3) *are satisfied and that the random time horizon is* $\mathbf{T}(n,t)$. *Then, for any* $\nu \in \mathbb{N}$ *and under the constraint that the expected value of the total computational cost is* $\mathcal{O}(\nu)$ *operations, the root mean square error for the WHMC and the MLWH method can be bounded respectively by*

$$e(\widehat{F}_{\mathrm{MC}}^{n,M}) \lesssim \nu^{-\frac{1}{4}} \quad \text{and} \quad e(\widehat{F}_{\mathrm{ML}}^{\mathcal{M}(n_0,L)}) \lesssim \nu^{-\frac{1}{2}} \log \nu.$$

Note that the multilevel algorithm with time horizon $\mathbf{T}(n,t)$ achieves optimal convergence rate up to a logarithm factor. Recalling the discussion in Section 4.2 and Fig. 1, the above methodology would offer optimal convergence throughout the entire range of the Blumenthal–Getoor index, thereby beating any of the methods considered there.

6. Implementation and numerical results

The aim of this section is to provide an example of the application of the WHMC and the new MLWH method to demonstrate the feasibility of both algorithms and to confirm the theoretical results of the preceding sections.

We will study X_t and $\overline{X}_t$, as well as the price of a barrier option assumed to belong to a market driven by a member of the β-class. We use this class because there is an explicit form of the Wiener–Hopf factorisation and because, through appropriate parameter choices, it contains examples of processes with paths of bounded and unbounded variation. Moreover Kuznetsov et al. [28, Section 4] argue that a large class of Lévy processes can be approximated by a member of the β-class plus a compound Poisson process. It is beyond the scope of this paper to discuss the β-class in a financial framework, but we refer the reader to Ferreiro-Castilla and Schoutens [17] and Schoutens and van Damme [35] where the β-class is studied in relation with other popular models in finance and where a comprehensive numerical analysis of the computation of the

extrema for this family is done. We will simply introduce the β-class as a ten parameter Lévy process with triplet (a, σ, Π), such that the Lévy measure is absolutely continuous with density

$$\Pi(\mathrm{d}x) = c_1 \frac{\mathrm{e}^{-a_1 b_1 x}}{(1 - \mathrm{e}^{-b_1 x})^{\lambda_1}} \mathbf{1}_{\{x>0\}} \mathrm{d}x + c_2 \frac{\mathrm{e}^{a_2 b_2 x}}{(1 - \mathrm{e}^{b_2 x})^{\lambda_2}} \mathbf{1}_{\{x<0\}} \mathrm{d}x , \tag{31}$$

and $a_i > 0$, $b_i > 0$, $c_i \geq 0$ and $\lambda_i \in (0, 3)$, for $i = 1, 2$. The expressions for the distribution of the extrema can be found in Kuznetsov [26, Section 4] and are given by the following infinite product representation

$$\mathbb{E}[e^{iz \overline{X}_{\mathbf{e}(q)}}] = \frac{1}{1 + \frac{iz}{\zeta_0^-(q)}} \prod_{n \leq -1} \frac{1 + \frac{iz}{b_1(n+1-a_1)}}{1 + \frac{iz}{\zeta_n(q)}} \quad \text{and}$$

$$\mathbb{E}[e^{iz \underline{X}_{\mathbf{e}(q)}}] = \frac{1}{1 + \frac{iz}{\zeta_0^+(q)}} \prod_{n \geq 1} \frac{1 + \frac{iz}{b_2(n-1+a_2)}}{1 + \frac{iz}{\zeta_n(q)}} ,$$

where $\zeta_0^-(q)$ and $\{\zeta_n(q)\}_{n \leq -1}$ correspond to the negative zeros and $\zeta_0^+(q)$ and $\{\zeta_n(q)\}_{n \geq 1}$ correspond to the positive zeros of $\Psi(-iz) + q = 0$. The numerical simulation of $\overline{X}_{\mathbf{e}_q}$ and $\underline{X}_{\mathbf{e}_q}$ requires truncating the infinite products above, but it can be shown using [26, Theorem 10] and [17, Section 3.1] that the cost to obtain a sample via this process (to an arbitrary degree of accuracy) is independent of the value of q, thus verifying Assumption (A3). In fact this claim holds even for the more general family of processes introduced in Kuznetsov et al. [28] known as meromorphic Lévy processes. From (31), one sees that processes belonging to the β-class have Lévy measures which decay exponentially and therefore have finite moments of all orders, in particular they verify Assumption (A1) for the entire range of parameters.

For the numerical results below, we choose $a = c_i = a_i = b_i = 1$ and set $\lambda_2 = \lambda_1 = \lambda$. We then study the behaviour and the robustness of our methods for various choices of σ and λ. For instance, if we set $\lambda = 0.5$ then we have a process of finite activity and thus with Blumenthal–Getoor index $\rho = 0$. By letting $\lambda = 1.5$, we have an infinite activity process with jump component of bounded variation (i.e. $\rho \in (0, 1)$), and finally by setting $\lambda = 2.5$, we have an infinite activity process with index $\rho > 1$. We are free to chose any value of the Gaussian coefficient σ. When $\sigma = 0$ only the case with $\lambda = 2.5$ will produce a process with paths of unbounded variation. When $\sigma^2 > 0$ all processes have paths of unbounded variation.

There are not many interesting pay-off functions for which Assumption (A2) is satisfied. Therefore, we present two case studies to show one example to check the theory presented in this paper and another example which is more relevant to financial mathematics but do not satisfy Assumption (A2). In the first case study our pay-off function is

$$\mathbb{E}[((x_0 + \overline{X}_t) - K(x_0 + X_t))^+] \tag{32}$$

for some $K > 0$ and $x_0 \in \mathbb{R}$ which corresponds to the price of a floating lookback option under the assumption that the underlying follows a Lévy process. The pay-off function in our second case study relates to the pricing of barrier options, i.e. we are interested in computing

$$\mathbb{E}[(\exp(x_0 + X_t) - K)^+ \mathbf{1}_{\{x_0 + \overline{X}_t < b\}}] \tag{33}$$

for some $K, b > 0$ and $x_0 \in \mathbb{R}$. Barrier options play an important role in the computation of exotic options and are the building blocks for more complex derivatives such as Credit Default Swaps, see for instance Schoutens and Cariboni [34]. Even though the pay-off function in (33) is
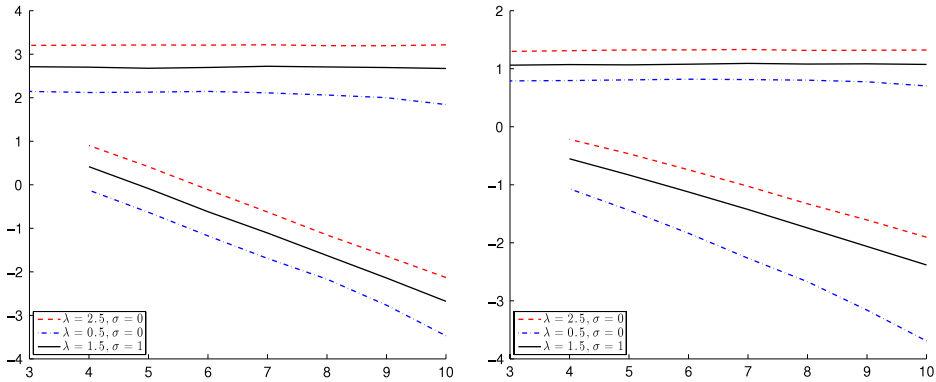
Fig. 2. $\mathbb{V}(X_{\mathbf{g}(n_\ell,n_\ell/t)} - X_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)})$ and $\mathbb{V}(X_{\mathbf{g}(n_\ell,n_\ell/t)})$ (left figure), as well as $\mathbb{E}[|X_{\mathbf{g}(n_\ell,n_\ell/t)} - X_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)}|]$ and $\mathbb{E}[|X_{\mathbf{g}(n_\ell,n_\ell/t)}|]$ (right figure) plotted against n_ℓ (using $\log_2$-scales).

not Lipschitz continuous, and thus does not satisfy Assumption (A2), we will see below that the theoretically predicted performance is nevertheless achieved.

The programs were implemented in C++, in long double precision and executed on a 64-bit Intel(R) Xeon(R) CPU at 2.66 GHz running 32-bit Linux 2.6.32-41-generic-pae using the GNU gcc compiler (version 4.8). Although currently running in a single threaded environment, (multilevel) Monte Carlo algorithms are of course ideally posed to be parallelised, as individual samples and simulations at different levels are independent of each other.

6.1. Numerical results

We start in Figs. 2 and 3 by plotting, for various pairs of λ and σ and for $n_\ell = 2n_{\ell-1}$, the mean and the variance of $(X_{\mathbf{g}(n_\ell,n_\ell/t)} - X_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)})$ and $(\overline{X}_{\mathbf{g}(n_\ell,n_\ell/t)} - \overline{X}_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)})$. For reference we also include mean and variance of $X_{\mathbf{g}(n_\ell,n_\ell/t)}$ and $\overline{X}_{\mathbf{g}(n_\ell,n_\ell/t)}$. We see from the left plots in those figures that the convergence of $\mathbb{V}(X_{\mathbf{g}(n_\ell,n_\ell/t)} - X_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)})$ with respect to $n_\ell^{-1/2}$ is linear as predicted in Proposition 4.5 and slightly better than predicted for $\mathbb{V}(\overline{X}_{\mathbf{g}(n_\ell,n_\ell/t)} - \overline{X}_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)})$.

Since the rate of the Cauchy sequence $\mathbb{E}[|X_{\mathbf{g}(n_\ell,n_\ell/t)} - X_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)}|]$ dictates the rate for $\mathbb{E}[|X_{\mathbf{g}(n_\ell,n_\ell/t)} - X_t|]$, we can also verify that the convergence of $\mathbb{E}[|X_{\mathbf{g}(n_\ell,n_\ell/t)} - X_t|]$ and $\mathbb{E}[|\overline{X}_{\mathbf{g}(n_\ell,n_\ell/t)} - \overline{X}_t|]$ is no worse than $\mathcal{O}(n_\ell^{-1/4})$. In fact, in the absence of a Gaussian component and in the case bounded variation jump activity, the convergence of $\mathbb{E}[|X_{\mathbf{g}(n_\ell,n_\ell/t)} - X_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)}|]$ seems to be even linear in $n_\ell^{-1/2}$, which is not surprising given the piecewise linear nature of this model. More surprisingly though, the bias for the running maximum seems to converge faster than predicted for any choice of λ and σ.

The behaviour for our output functional in (32) is in complete accordance with the conclusions in Theorem 4.6. As the plots in Fig. 4 show, the rate of decay in $\mathbb{V}(F^{n_\ell} - F^{n_{\ell-1}})$ is approximately 0.52 for $\lambda = 1.5$ and $\sigma = 1$ and the decay is 0.62 for $\lambda = 0.5$ and $\sigma = 0$, while Theorem 4.6 claims $\beta = 1/2$. Similarly, $\mathbb{E}[|F^{n_\ell} - F^{n_{\ell-1}}|]$ behaves better than $\mathcal{O}(n_\ell^{-1/4})$, confirming that $\alpha = 1/4$ at least. We also see that $\mathbb{V}(F^{n_\ell})$ remains constant as shown in (13). In the right plot of Fig. 4 we also include the weak error, i.e. $|\mathbb{E}[F^{n_\ell}] - \mathbb{E}[F^{n_{\ell-1}}]|$, which is indeed much smaller than the strong error and decays much faster (cf. Remark 4.7).
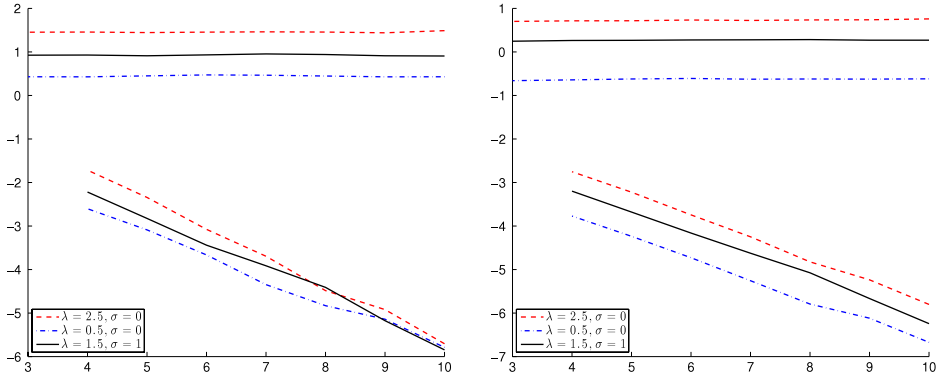
Fig. 3. $\mathbb{V}(\overline{X}_{\mathbf{g}(n_\ell,n_\ell/t)} - \overline{X}_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)})$ and $\mathbb{V}(\overline{X}_{\mathbf{g}(n_\ell,n_\ell/t)})$ (left figure), as well as $\mathbb{E}[|\overline{X}_{\mathbf{g}(n_\ell,n_\ell/t)} - \overline{X}_{\mathbf{g}(n_{\ell-1},n_{\ell-1}/t)}|]$ and $\mathbb{E}[|\overline{X}_{\mathbf{g}(n_\ell,n_\ell/t)}|]$ (right figure) plotted against n_ℓ (using $\log_2$-scales).
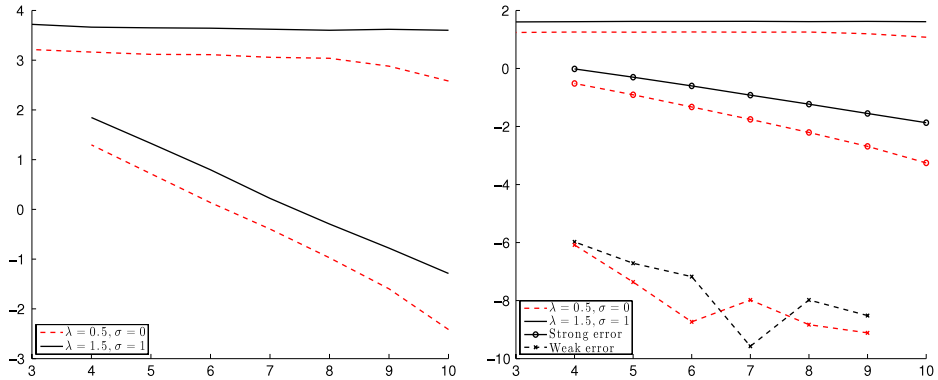

Fig. 4. $\mathbb{V}(F^{n_\ell} - F^{n_{\ell-1}})$ and $\mathbb{V}(F^{n_\ell})$ (left figure), as well as strong error $\mathbb{E}[|F^{n_\ell} - F^{n_{\ell-1}}|]$, weak error $|\mathbb{E}[F^{n_\ell}] - \mathbb{E}[F^{n_{\ell-1}}]|$ and $\mathbb{E}(|F^{n_\ell}|)$ (right figure) plotted against n_ℓ (using $\log_2$-scales). Plots with respect to the pay-off function (32).

The behaviour for our output functional in (33), which violates Assumption (A2), is also very close to the conclusions in Theorem 4.6. As the plots in Fig. 5 show, the rate of decay in $\mathbb{V}(F^{n_\ell} - F^{n_{\ell-1}})$ is approximately 0.36. Similarly, $\mathbb{E}[|F^{n_\ell} - F^{n_{\ell-1}}|]$ behaves roughly like $\mathcal{O}(n_\ell^{-1/4})$. We also check that $\mathbb{V}(F^{n_\ell})$ remains constant and include the weak error for comparison.

Finally, in Fig. 6 we look at actual CPU times in running our WHMC and MLWH algorithms. The x-axis describes the estimated root mean square error (RMSE) e. The y-axis describes the cost ν. In each data point of the plot, the finest level in the MLMC algorithm matches the approximation level in the WHMC. Next to each vertical pair of points we have labelled that value. This ensures that the bias error is the same in both algorithms. The sampling error in the MLMC error is split between levels accordingly to [19, Equation (12)]. Since the weak error is much smaller than the variance reduction for the pay-off function (33), as seen in Fig. 5, the multilevel algorithm outperforms the single level version only for relatively small accuracies.
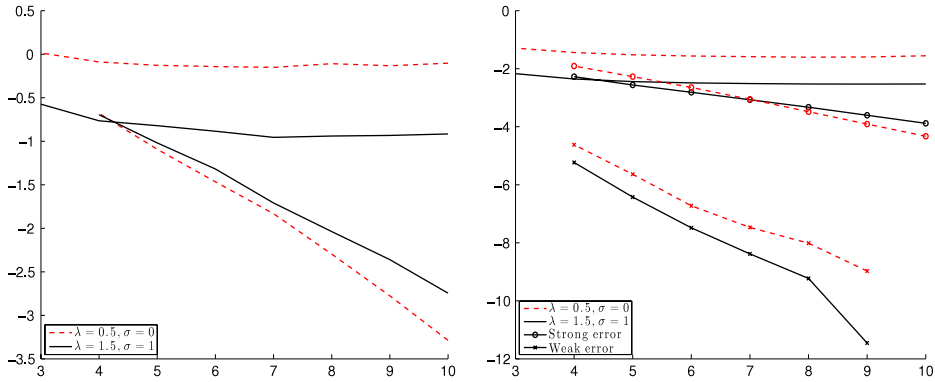
Fig. 5. $\mathbb{V}(F^{n_\ell} - F^{n_{\ell-1}})$ and $\mathbb{V}(F^{n_\ell})$ (left figure), as well as strong error $\mathbb{E}[|F^{n_\ell} - F^{n_{\ell-1}}|]$, weak error $|\mathbb{E}[F^{n_\ell}] - \mathbb{E}[F^{n_{\ell-1}}]|$ and $\mathbb{E}(|F^{n_\ell}|)$ (right figure) plotted against n_ℓ (using $\log_2$-scales). Plots with respect to the pay-off function (33).
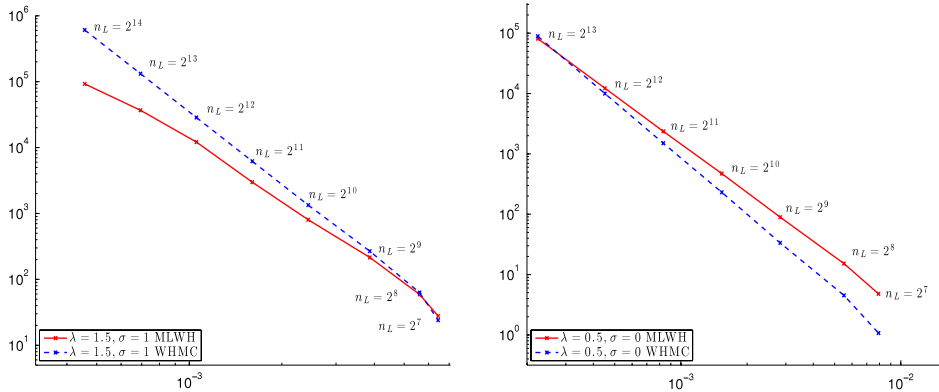

Fig. 6. Total CPU-time (in seconds) for the WHMC and MLWH methods plotted against the estimated root mean square error e ($\log_{10}$-scales). Left: pay-off function (32). Right: pay-off function (33).

However, from the right plot in Fig. 6 one also sees that the slopes of the two lines clearly play into the hands of the MLMC. The variance curves in Fig. 3 suggest that for functionals that depend more strongly on the maximum $\overline{X}_t$, the MLWH method would be faster than the single level method already for $n_L = 2^8$ as seen in the left plot of Fig. 6 with the pay-off function (32).

For the sake of completeness, Fig. 7 plots the cost of computing one single sample $F^{n,(i)}$ of F^n using the Wiener–Hopf approach, to confirm also numerically Assumption (A3). The cost clearly grows linearly with n, independently of any parameters in the Lévy process. As mentioned in the previous section, this can in fact be proved rigorously.

Finally, we also have not yet experimented with the factor s at which we coarsen the rates from one level to the next, i.e. $n_\ell = sn_{\ell-1}$. We have only studied the case $s = 2$, while in Giles [19] it was shown that the optimal value for standard multilevel Monte Carlo methods for SDEs with Gaussian noise is $s = 7$. In any case, the results in Fig. 6 (right) show that Wiener–Hopf Monte Carlo techniques allow the pricing of options in a market driven by a Lévy process to an accuracy of $\mathcal{O}(10^{-2})$ in less than a second.
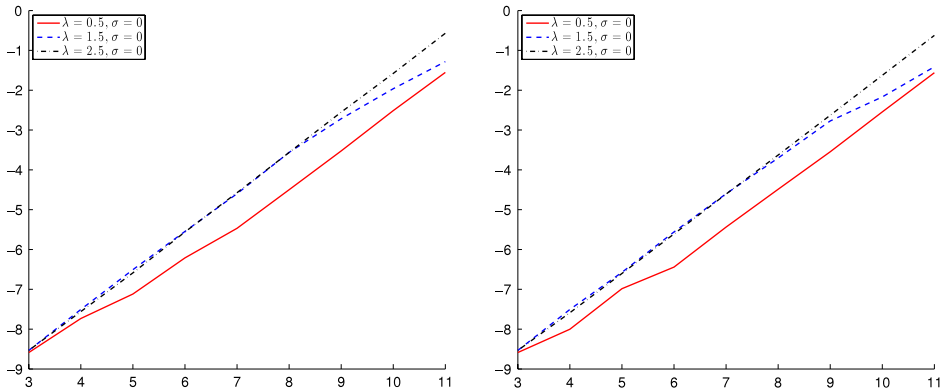
Fig. 7. Average CPU-time (in milliseconds) to compute one sample $F^{n,(i)}$ plotted against $n = n_L$ ($\log_2$-scales). Left: pay-off function (32). Right: pay-off function (33).

Acknowledgements

References

[1] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover Publications, Inc., New York, 1970.

[2] S. Asmussen, H.-J. Albrecher, Ruin Probabilities, second ed., World Scientific Press, River Edge, NJ, 2010.

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

[4] J. Bertoin, Lévy Processes, in: Cambridge Tracts in Mathematics, vol. 121, Cambridge Univer. Press, Cambridge, 1996.

[5] L. Bondesson, On simulation from infinitely divisible distributions, Adv. Appl. Probab. 14 (4) (1982) 855–869.

[6] S.I. Boyarchenko, S.Z. Levendorskii, Non-Gaussian Merton–Black–Scholes Theory, in: Advanced Series on Statistical Science & Applied Probability, vol. 9, World Scientific Press, 2002.

[7] M. Broadie, P. Glasserman, A continuity correction for discrete barrier options, Math. Finance 7 (4) (1997) 325–348.

[8] M.-E. Caballero, A. Lambert, G. Uribe, Proof(s) of the Lamperti representation of continuous-state branching processes, Probab. Surv. 6 (2009) 62–89.

[9] P. Carr, Randomization and the American put, Rev. Fin. Stud. 11 (3) (1998) 597–626.

[10] K.A. Cliffe, M.B. Giles, R. Scheichl, A.L. Teckentrup, Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients, Comput. Vis. Sci. 14 (1) (2011) 3–15.

[11] R. Cont, P. Tankov, Financial Modeling with Jump Processes, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[12] J. Corcuera, J. de Spiegeleer, A. Ferreiro-Castilla, A.E. Kyprianou, D. Madan, W. Schoutens, Efficient pricing of contingent convertibles under smile conform models, 2011 Preprint.

[13] J. Creutzig, S. Dereich, T. Müller-Gronbach, K. Ritter, Infinite-dimensional quadrature and approximation of distributions, Found. Comput. Math. 9 (2009) 391–429.

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

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

[16] L. Doering, M. Barczy, A jump type SDE approach to positive self-similar Markov processes, 2011. arXiv:1111.3235v1 [math.PR].

[17] A. Ferreiro-Castilla, W. Schoutens, The $\boldsymbol{\beta}$-Meixner model, J. Comput. Appl. Math. 236 (9) (2012) 2466–2476.

[18] M.B. Giles, Improved multilevel Monte Carlo convergence using the Milstein scheme, in: A. Keller, S. Heinrich, H. Niederreiter (Eds.), Monte Carlo and Quasi–Monte Carlo methods 2006, Springer, Berlin, 2007, pp. 343–358.

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

[20] M.B. Giles, Y. Xia, Multilevel path simulation for jump-diffusion SDEs, in: L. Plaskota, H. Woźniakowski (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2010, Springer, Berlin, 2012.

[21] P. Glasserman, N. Merener, Numerical solution of jump-diffusion LIBOR market models, Finance Stoch. 7 (1) (2003) 1–27.

[22] J. Jacod, T.G. Kurtz, S. Méléard, P. Protter, The approximate Euler method for Lévy driven stochastic differential equations, Ann. Inst. Henri Poincaré Probab. Stat. 41 (2005) 523–558.

[23] O. Kella, W. Whitt, Useful martingales for stochastic storage processes with Lévy input, J. Appl. Probab. 29 (1992) 396–403.

[24] J.F.C. Kingman, Poisson Processes, Oxford University Press, Oxford, 1993.

[25] C. Klüppelberg, A.E. Kyprianou, R. Maller, Ruin probabilities and overshoots for general Lévy insurance risk process, Ann. Appl. Probab. 14 (2004) 1766–1801.

[26] Kuznetsov, Wiener–Hopf factorization and distribution of extrema for a family of Lévy processes, Ann. Appl. Probab. 20 (2010) 1801–1830.

[27] A. Kuznetsov, A.E. Kyprianou, J.C. Pardo, Meromorphic Lévy processes and their fluctuation identities, Ann. Appl. Probab. 22 (3) (2012) 1101–1135.

[28] A. Kuznetsov, A.E. Kyprianou, J.C. Pardo, K. van Schaik, A Wiener–Hopf Monte Carlo simulation technique for Lévy process, Ann. App. Probab. 21 (6) (2011) 2171–2190.

[29] A.E. Kyprianou, Introductory Lectures on Fluctuations of Lévy processes with Applications, Springer, Berlin, 2006.

[30] D. Madan, E. Seneta, The variance gamma (V.G.) model for share market returns, J. Bus. 63 (4) (1990) 511–524.

[31] J. Rosiński, in: O. Barndorff-Nielsen, T. Mikosch, S. Resnick (Eds.), Levy Processes Theory and Applications, in: Series Representation of Lévy Processes from the Perspective of Point Process, Birkhauser, Boston, 2001.

[32] K.-I. Sato, Lévy processes and infinitely divisible distributions, Cambridge Univer. Press, Cambridge, 1999.

[33] W. Schoutens, Lévy Processes in Finance: Pricing Financial Derivatives, Wiley, Chichester, 2003.

[34] W. Schoutens, J. Cariboni, Lévy Processes in Credit Risk, Wiley, Chichester, 2009.

[35] W. Schoutens, G. van Damme, The $\boldsymbol{\beta}$-variance gamma model, Rev. Deriv. Res. 14 (3) (2010) 263–282.