A Weak Approximation of Stochastic Differential Equations with Jumps through Tempered Polynomial Optimization

KENJI KASHIMA* AND REIICHIRO KAWAI†

Abstract

We present an optimization approach to the weak approximation of a general class of stochastic differential equations with jumps, in particular, when value functions with compact support are considered. Our approach employs a mathematical programming technique yielding upper and lower bounds of the expectation, without Monte Carlo sample paths simulations, based upon the exponential tempering of bounding polynomial functions to avoid their explosion at infinity. The resulting tempered polynomial optimization problems can be transformed into a solvable polynomial programming after a minor approximation. The exponential tempering widens the class of stochastic differential equations for which our methodology is well defined. The analysis is supported by numerical results on the tail probability of a stable subordinator and the survival probability of Ornstein-Uhlenbeck processes driven by a stable subordinator, both of which can be formulated with value functions with compact support and are not applicable in our framework without exponential tempering.

Keywords: exponential tempering, Lévy process, stable subordinator, Ornstein-Uhlenbeck process, semidefinite programming, tail probability estimation, survival probability estimation.
2010 Mathematics Subject Classification: 60H10, 65C30, 60G51, 90C22.

1 Introduction

Stochastic differential equations have long been used to build realistic models in economics, finance, biology, the social sciences, chemistry, physics and other fields. For practical applications such as moment and tail probability estimations, or expected utilities, we need to estimate the expected value of the solution of stochastic differential equations. The so-called weak approximations via the time discretization of Euler-Maruyama type has been the most standard approach, that is, $|\mathbb{E}[V(X_T)] - \mathbb{E}[V(X_T^\Delta)]| \le C\Delta^\beta$, where $\{X_t : t \in [0, T]\}$ is a solution of a suitable stochastic differential equation, V is a smooth function and X_T^Δ indicates the time discretization approximation of X_T with time step $\Delta > 0$. The theoretical investigation of time discretization schemes in diffusion settings has been thoroughly presented in Kloeden and Platen [7], while stochastic differential equations with jumps have also been studied, for example, in Protter and Talay [15].

In contrast to Monte Carlo simulations of discretized sample paths, methodologies leading to the computation of bounds for the expectation have been proposed and investigated in various fields of application by several authors, for example, Bertsimas, Popescu and Sethuraman [2], Eriksson and Pistorius [3], Helmes, Röhl and Stockbridge [4], Lasserre and Prieto-Rumeau [8], Lasserre, Prieto-Rumeau and Zervos [9], Suzuki, Miyoshi and Kojima [18], to mention just a few. It is known that there exist two types of formulation of this framework, both of which arrive at a *semi-definite programming* in the end. One is the so-called generalized

*Email address: kashima@mei.titech.ac.jp. Postal address: Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Tokyo, 152-8552, Japan.
†Email address: reiichiro.kawai@gmail.com. Postal address: Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK.

moment problem that makes use of the semi-definiteness of (localizing) moment matrices[1]. The other is a polynomial optimization approach for which sum-of-squares relaxation efficiently works. In this paper, our discussion is based on the latter formulation, mainly because it provides a more intuitive way to discuss how the proposed method works. Note also that from an optimization point of view, those two formulations are dual to one another. (See, for example, Nishihara, Yagiura and Ibaraki [10] for a related discussion.)

Our methodology can be described roughly as follows. First, one employs a function with arguments both in time and in state, which bounds from above (or from below) at time T the value function uniformly over the support of X_T. One further restricts the infinitesimal generator of the bounding function to be non-positive (or non-negative) over the whole space. Under all those constraints, the well-known Dynkin formula guarantees that the bounding function concentrated at the deterministic initial point $(0, X_0)$ serves as an upper (or lower) bound of the expectation. The final step is to minimize the upper bound (or maximize the lower bound). For this approach to make practical sense, one has to restrict the class of bounding functions to an extent where the optimization is solvable. In this respect, the most general class of differentiable functions for the Ito formula is too abstract to be tractable. To address this issue, the existing literature focuses on diffusion process with polynomial coefficients. Also, a more general class of stochastic differential equations including jumps is considered by the authors [5], for which standard Monte Carlo simulations are no longer implementable in terms of computational time, or often impossible due to nonavailability of simulation methods. The extension from the diffusion setting is not trivial due to the difference operator of the jump component in the infinitesimal generator.

The approach that we develop in this paper is a remarkable improvement of the methodology of [5], in particular when considering value functions with compact support. In [5], bounding functions must be in polynomial form to arrive at a polynomial programming, while in principle, any polynomial function necessarily explodes at infinity whenever it is constrained to be either non-positive or non-negative. Due to this explosion at infinity, bounds are likely to be very far from the true value in particular when considering stochastic differential equations with very heavy tails and a value function with compact support. Such situations are often of practical interest, for example, the tail probability estimation of a stochastic differential equation with jumps. To address this issue, we introduce the exponential tempering of bounding polynomial functions so that the explosion never occurs at infinity. It is the theoretical basis that the optimization problem with the exponential tempering employed can still be transformed into a solvable polynomial programming after a minor approximation. Moreover, it turns out that exponential tempering widens the class of stochastic differential equations, for which every step of our method is well defined.

The rest of this paper is organized as follows. Section 2 provides some basic exposition of the optimization approach of the authors [5] and illustrates why the optimization approach fails to yield tight bounds for functions with a compact support, for example, in the tail probability estimation. Section 3 introduces our approach based upon exponential tempering of bounding functions and investigates how to transform the tempered polynomial optimization problem to a solvable polynomial programming after a minor approximation. Section 4 presents two numerical examples to illustrate that our approach yields tight bounds in estimation of tail probabilities and survival probabilities. Section 5 indicates the direction of future research, including an investigation of estimation quality with respect to exponential tempering, an empirical study of its range of applicability, and the challenge of extending this work to multivariate stochastic differential equations. In the appendix, we provide a brief sketch of the method-of-moments approach, that is dual to polynomial optimization.

---

[1]Another approach to the generalized moment problem is the linear programming relaxation based on the Hausdorff moment conditions. See Lasserre and Prieto-Rumeau [8] for a comparison of the performances of SDP and LP approaches.

2 Motivation

Let us begin this section with general notations which will be used throughout the text. We define $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$, $\mathbb{R}_+ := (0, +\infty)$, and denote by $\mathbb{N}$ the set of positive integers. We write $\mathscr{B}(A)$ for the Borel σ-field of a set $A \subseteq \mathbb{R}$. For $k \in \mathbb{N}$, ∂_k indicates the partial derivative with respect to k-th argument. We denote by C^{k_t,k_x} the class of continuous functions which are k_t-times continuously differentiable with respect to the first variable and k_x-times continuously differentiable with respect to the second variable. We denote by C_p the class of polynomial functions in the form of

$$f_p(t,x) = \sum_{B(0,0)} c_{k_t,k_x} t^{k_t} x^{k_x}, \tag{2.1}$$

where

$$B(l,m) := \left\{ (k_t, k_x) \in \mathbb{N}^2 : k_t \geq l, \ k_x \geq m, \ k_t + k_x \leq K \right\},$$

for a fixed even natural number K, while $\{c_{k_t,k_x}\}_{B(0,0)}$ is a sequence of constants. Throughout the paper, we fix the even natural number K. (Note that the class C_p depends on this K.) We fix $(\Omega, \mathscr{F}, \mathbb{P})$ as our underlying probability space.

Let $T > 0$. Consider a one-dimensional stochastic differential equation

$$dX_t = a_0(t,X_t)\,dt + a_1(t,X_t)\,dW_t + \int_{|z| \leq 1} b(t, X_{t-}, z)\,(\mu - \nu)(dz, dt)$$
$$+ \int_{|z| > 1} b(t, X_{t-}, z)\,\mu(dz, dt), \quad t \in [0,T], \tag{2.2}$$

where the initial state X_0 is fixed at a constant in $\mathbb{R}$, $\{W_t : t \geq 0\}$ is a standard Brownian motion and μ is a Poisson random measure on $\mathbb{R}_0$ whose compensator is given by the Lévy measure ν, that is, a σ-finite measure defined on $\mathbb{R}_0$ satisfying

$$\int_{\mathbb{R}_0} \left(|z|^2 \wedge 1 \right) \nu(dz) < +\infty. \tag{2.3}$$

Here, we assume that for each $t \in [0,T]$, the functions $a_0(t,x)$, $a_1(t,x)$ and $b(t,x,z)$ in (2.2) satisfy the usual conditions such as at most linear growth and Lipschitz so that the solution of (2.2) is well defined. (For example, see Theorem 1.19 of Øksendal and Sulem [11] and Section 6.2 of Applebaum [1].) We henceforth equip our underlying probability space with the natural filtration $(\mathscr{F}_t)_{t \in [0,T]}$ generated by $\{X_t : t \in [0,T]\}$. Moreover, throughout this study, we assume that $b(t,x,z) \neq 0$ and $\nu \neq 0$ to avoid triviality. Moreover, we use the notation

$$\mathscr{X} := \inf \left\{ B \subseteq \mathbb{R} : \mathbb{P}\left(X_t \in B, t \in [0,T] \right) = 1, B \text{ connected} \right\},$$

where $\{X_t : t \in [0,T]\}$ is defined in (2.2). Note that by imposing that the set $\mathscr{X}$ is connected, it may be significantly larger than the state space of the sample paths. (For example, if $\{X_t : t \in [0,T]\}$ is a standard Poisson process, then its state space is $\mathbb{N} \cup \{0\}$, while the above definition yields $\mathscr{X} = [0, +\infty)$.) This larger space will be required for optimization problems.

Our interest throughout this study is in approximating the expectation

$$\mathbb{E}\left[V(\tau, X_\tau) \right].$$

Here, V is a function mapping from $[0,T] \times \mathbb{R}$ to $\mathbb{R}$, piecewise polynomial in t and x and such that its support is a bounded subset of $[0,T] \times \mathscr{K}_V$ where $\mathscr{K}_V$ is a bounded set and $\mathbb{E}[|V(\tau,X_\tau)|] < +\infty$. Note that the function V may have discontinuities. Moreover, τ is an $(\mathscr{F}_t)_{t \in [0,T]}$-stopping time taking its values in $[0,T]$. For the computation of $\mathbb{E}[V(\tau, X_\tau)]$, standard techniques include the Monte Carlo simulation of sample paths through the time discretization of stochastic differential equations, or even some exact knowledge of sample paths such as series representation of the Poisson jump component. Let us illustrate difficulties arising in

the approximation of stochastic differential equations with jumps in a simple setting, fix $X_0 > 0$, $\tau = T$, $a_0(t,x) = a_1(t,x) \equiv 0$, and $b(t,x,z) = xz$, that is, (2.2) reduces to a Doléans-Dade stochastic exponential

$$dX_t = X_{t-} \int_{\mathbb{R}_0} z(\mu - \nu)(dz,dt), \quad X_0 > 0, \tag{2.4}$$

which is a martingale with respect to its natural filtration. Assume further that the Lévy measure ν is supported on $(-1, +\infty)$. It is elementary that the Ito formula yields

$$d\ln X_t = -\int_{(-1,+\infty)} z\nu(dz)\,dt + \int_{(-1,+\infty)} \ln(1+z)\,\mu(dz,dt),$$

or equivalently,

$$X_t = X_0 \exp\left[-t\int_{(-1,+\infty)} z\nu(dz) + \int_0^t \int_{(-1,+\infty)} \ln(1+z)\,\mu(dz,ds)\right]. \tag{2.5}$$

It follows from (2.5) that $X_t > 0$, *a.s.* For the computation of $\mathbb{E}[V(T,X_T)]$, a standard technique is the Monte Carlo simulation with the sample generation of the marginal X_T. Let us discuss some typical drawbacks in simulation of X_T defined by (2.4).

(i) In the explicit solution (2.5), we are required to keep track of the Poisson random measure μ for simulation of the term $\int_0^T \int_{(-1,+\infty)} \ln(1+z)\mu(dz,ds)$, which can be done by using its shot noise series representation. (See Rosiński [16] for details.) It is however usually difficult to find a shot noise representation in a convenient form. In addition, it is often extremely expensive to use shot noise series representation for computational purposes, since shot noise series for an infinite Lévy measure is infinite as well.

(ii) We may instead rely on the time discretization of sample paths through (2.4), while the resulting marginal law may not be reasonably accurate. (In this example, it may take negative values, while initially $X_T > 0$, a.s. For details about the discretization error, see, for instance, Kloeden and Platen [7].) Estimation results for $\mathbb{E}[V(T,X_T)]$ could then be completely misleading.

(iii) It is rare to know how to simulate increments $\int_{t_1}^{t_2} \int_{\mathbb{R}_0} z(\mu - \nu)(dz,dt)$, with some exceptions such as gamma processes and stable processes. Moreover, Lévy processes have no scaling property, except for stable processes (including Brownian motion). Hence, it is often not as simple as in the diffusion setting to examine different stepsizes of the Euler-Maruyama scheme.

In the diffusion setting, the above issues often do not arise. Namely, increments of Brownian motion can easily be generated and have scaling property; stochastic differential equations are often explicitly solvable (see Chapter 4.4 of Kloeden and Platen [7] for various such examples). Those comparisons illustrate the difficulty of sample paths simulation for stochastic differential equations with jumps.

Meanwhile, completely different approaches have recently been investigated in Kashima and Kawai [5] for a general setting with jumps, based upon semi-definite mathematical programming providing us with upper and lower bounds of the expectation without generating random numbers. Let us first review their mathematical programming based approaches in brief. For $f \in C^{1,2}([0,T] \times \mathscr{X}; \mathbb{R})$, the Ito formula yields

$$df(t,X_t) = \mathscr{A}f(t,X_t)dt + \partial_2 f(t,X_t)a_1(t,X_t)dW_t + \int_{\mathbb{R}_0} B_z f(t,X_{t-})(\mu - \nu)(dz,dt), \quad a.s.,$$

where

$$\mathscr{A}f(t,x) := \partial_1 f(t,x) + \partial_2 f(t,x)a_0(t,x) + \frac{1}{2}\partial_2^2 f(t,x)a_1(t,x)^2$$
$$+ \int_{\mathbb{R}_0} \left(B_z f(t,x) - \partial_2 f(t,x)b(t,x,z)\mathbb{1}_{(0,1]}(|z|)\right)\nu(dz), \tag{2.6}$$

4

and for $z \in \mathbb{R}_0$,
$$B_z f(t,x) := f(t,x+b(t,x,z)) - f(t,x), \tag{2.7}$$
provided that for each $(t,x) \in [0,T] \times \mathscr{X}$,
$$\int_{|z|>1} |B_z f(t,x)| \, \nu(dz) < +\infty.$$

One of the important building blocks of our approach is the Dynkin formula;
$$\mathbb{E}[f(\tau,X_\tau)] - f(0,X_0) = \mathbb{E}\left[\int_0^\tau \mathscr{A} f(s,X_s)ds\right], \tag{2.8}$$
where τ is an $(\mathscr{F}_t)_{t\in[0,T]}$-stopping time taking its values in $[0,T]$. We briefly summarize the conditions under which the formula makes sense. Throughout this paper, we write (with some abuse of notation)
$$E_0 := \inf\{B \subseteq [0,T] \times \mathbb{R} : \mathbb{P}((\tau,X_\tau) \in B) = 1, B \text{ connected}\},$$
$$E_1 := \inf\{B \subseteq [0,T] \times \mathbb{R} : \mathbb{P}((t,X_t) \in B, t \in [0,\tau)) = 1, B \text{ connected}\},$$
$$E_2 := E_0 \cup E_1.$$

Lemma 2.1. *Let $f \in C^{1,2}(E_2;\mathbb{R})$ and assume that for each $(t,x) \in E_1$, the function $\mathscr{A} f(t,x)$ in (2.6) is well defined. Then, the Dynkin formula (2.8) holds if at least one of the following conditions is satisfied;*

   (i) $\mathbb{E}[|f(\tau,X_\tau)|] < +\infty$, and for almost surely all $t \in [0,\tau)$, $\mathbb{E}[|\mathscr{A} f(t,X_t)|] < +\infty$,

  (ii) $\mathbb{E}[\int_0^\tau (\partial_2 f(s,X_s)a_1(s,X_s))^2 ds] < +\infty$ and $\mathbb{E}[\int_0^\tau \int_{\mathbb{R}_0} (B_z f(s,X_s))^2 \nu(dz)ds] < +\infty$.

Proof. It is trivial that the formula holds when (i) is satisfied. Next, if (ii) is satisfied, then the stochastic process $\{f(t,X_t) - f(0,X_0) - \int_0^t \mathscr{A} f(s,X_s)ds : t \in [0,\tau]\}$ is a square-integrable local martingale with respect to the filtration $(\mathscr{F}_t)_{t\in[0,T]}$. $\qquad\square$

Remark 2.2. In case of no jump component, there exist other trivial conditions for the Dynkin formula to hold, such as the function f has compact support, the stopping time τ is the first exit time for a bounded subset of $\mathscr{X}$, and so on. Those conditions are no longer readily valid when the Poisson jump is involved, due to the local difference operator B_z of (2.7). $\qquad\square$

To proceed with our discussion, assume that both $a_0(t,x)$ and $a_1(t,x)$ are polynomial in t and x and the coefficient b is decomposed as $b(t,x,z) = b_1(t,x)b_2(z)$, where $b_1(t,x)$ is polynomial both in t and x, and where $b_2 : \mathbb{R}_0 \mapsto \mathbb{R}$ such that $\int_{\mathbb{R}_0} |b_2(z)|^k \nu(dz) < +\infty$, for $k = 2,\ldots,K$. Consider the following optimization formulation
$$\left|\begin{array}{ll} \min_{\{c_{k_t,k_x}\}_{B(0,0)}} & f_p(0,X_0) \\ \text{s.t.} & f_p(t,x) \geq V(t,x) \text{ on } E_0, \\ & \mathscr{A} f_p(t,x) \leq 0 \text{ on } E_1, \\ & f_p \in C_p(E_2;\mathbb{R}). \end{array}\right. \tag{2.9}$$

Let us emphasize that decision variables of this optimization problem (and all in what follows) are the coefficients $\{c_{k_t,k_x}\}_{B(0,0)}$ in the definition of the polynomial (2.1). By further assuming $\mathbb{E}[|X_t|^K] < +\infty$ for $t \in [0,T]$, we have
$$\mathscr{A} f_p(t,x) = \sum_{B(1,0)} c_{k_t,k_x} k_t t^{k_t-1} x^{k_x} + \sum_{B(0,1)} c_{k_t,k_x} t^{k_t} k_x x^{k_x-1} a_0(t,x)$$
$$+ \frac{1}{2} \sum_{B(0,2)} c_{k_t,k_x} t^{k_t} k_x(k_x-1) x^{k_x-2} a_1(t,x)^2$$
$$+ \sum_{B(0,2)} c_{k_t,k_x} t^{k_t} \sum_{k=0}^{k_x-2} {}_{k_x}C_k x^k b_1(t,x)^{k_x-k} \int_{\mathbb{R}_0} b_2(z)^{k_x-k} \nu(dz).$$

5

This implies that the optimization (2.9) is a polynomial programming problem. If the problem (2.9) is feasible, it provides us with an upper bound $f_p(0, X_0)$ of the expectation $\mathbb{E}[V(\tau, X_\tau)]$, in view of the inequalities

$$\mathbb{E}[V(\tau, X_\tau)] \leq \mathbb{E}[f_p(\tau, X_\tau)] \leq f_p(0, X_0),$$

due to (2.8).

In general, polynomial optimization problems are NP hard. However, if the degrees of f_p, that is, K is fixed, sums-of-squares relaxation enables us to solve the problem efficiently. (For details, we refer to Parrilo [12].) To obtain lower bounds of $\mathbb{E}[V(\tau, X_\tau)]$, we are only to find $g_p \in C_p(E_2; \mathbb{R})$ through the polynomial programming

$$
\begin{aligned}
\max \quad & g_p(0, X_0) \\
\text{s.t.} \quad & g_p(t, x) \leq V(t, x) \ \text{on } E_0, \\
& \mathscr{A} g_p(t, x) \geq 0 \ \text{on } E_1, \\
& g_p \in C_p(E_2; \mathbb{R}).
\end{aligned}
\tag{2.10}
$$

As previously mentioned, this optimization approach does not require the sample paths simulation at all towards the approximation of the expectation.

Let us close this section by pointing out some possible drawbacks of the above methods, which leads to the motivation of our study.

(i) In principle, the weak approximation problem is supposed to be well-posed, whenever $\mathbb{E}[|V(\tau, X_\tau)|] < +\infty$. On the contrary, most of the existing approaches listed above require that the marginals τ and X_τ have finite moments of higher order. However, this requirement may rule out some interesting problem settings, such as stochastic differential equations driven by a stable Lévy process, that we will deal with in Section 4. In addition, it is sometimes difficult to check whether the requirement is actually satisfied, in particular when stochastic differential equations are involved.

(ii) Suppose that we are interested in a probability estimation of a non-negative marginal X_T, that is, the stopping time τ is frozen at T and $\mathbb{E}[V(T, X_T)] = \mathbb{E}[\mathbb{1}(X_T \in [0, \theta])]$ for some $\theta > 0$. We will observe that $\lim_{x \uparrow +\infty} f_p(T, x) = +\infty$, since $f_p(T, x)$ is polynomial in x and must be no less than $V(T, x)$ over $\mathscr{X}$. When the marginal X_T has very heavy tail, bounds are likely to be very far from the true value, that is,

$$g_p(0, X_0) \leq \mathbb{E}[g_p(T, X_T)] \ll \mathbb{E}[V(T, X_T)] \ll \mathbb{E}[f_p(T, X_T)] \leq f_p(0, X_0).$$

By a similar reasoning, if $\mathscr{X} = \mathbb{R}$, for example, then $\lim_{|x| \uparrow +\infty} \mathscr{A} f_p(t, x) = -\infty$ due to the constraint $\mathscr{A} f_p(t, x) \leq 0$ over $[0, T] \times \mathscr{X}$. If this is the case, then we will observe

$$g_p(0, X_0) \ll \mathbb{E}[g_p(T, X_T)] \ll \mathbb{E}[V(T, X_T)] \ll \mathbb{E}[f_p(T, X_T)] \ll f_p(0, X_0).$$

Let us emphasize again that the explosion at infinity necessarily occurs, whenever the bounding functions are of a polynomial form. As an exception, the use of smooth piecewise polynomial may avoid the explosion at infinity. However, this method easily increases the computing burden and is not applicable to stochastic differential equations with jumps as discussed in Kashima and Kawai [6].

To address those issues, we introduce the exponential tempering of bounding functions in the next section.

3 Exponential Tempering of Bounding Functions

As before, we assume that both a_0 and a_1 are in $C_p(E_1; \mathbb{R})$ and the coefficient b can be decomposed as $b(t, x, z) = b_1(t, x) b_2(z)$. In principle, our approach is based on the replacement of the polynomial $f_p(t, x)$

with its tempered $f(t,x) = e^{-\beta x} f_p(t,x)$, and the optimization problem

$$
\begin{vmatrix}
\min & f(0,X_0) \\
\text{s.t.} & f(t,x) \geq V(t,x) \text{ on } E_0, \\
& \mathscr{A} f(t,x) \leq 0 \text{ on } E_1, \\
& f(t,x) = e^{-\beta x} f_p(t,x) \text{ on } E_2, \\
& f_p \in C_p(E_2; \mathbb{R}),
\end{vmatrix}
\tag{3.1}
$$

where the decision variable is again the set of coefficients $\{c_{k_t,k_x}\}_{B(0,0)}$ given in (2.1). With this optimization problem, in couple with a counterpart for lower bounds, we aim at finding bounds for $\mathbb{E}[V(\tau,X_\tau)]$ based upon a set of inequalities

$$
g(0,X_0) \leq \mathbb{E}[g(\tau,X_\tau)] \leq \mathbb{E}[V(\tau,X_\tau)] \leq \mathbb{E}[f(\tau,X_\tau)] \leq f(0,X_0),
\tag{3.2}
$$

where $g(t,x) = e^{-\beta x} g_p(t,x)$ for some $g_p \in C_p(E_2; \mathbb{R})$.

Prior to the development of our method, let us discuss conditions for the function $\mathscr{A} f(t,x)$ to be well defined. Rather than trying to cover a general class, we focus our attention on the class of Lévy-driven stochastic differential equations, which is sufficiently large for practical use in most situations. Throughout this paper, we will write

$$
\begin{aligned}
\mathscr{A}_\beta f_p(t,x) := {} & \partial_1 f_p(t,x) + (-\beta f_p(t,x) + \partial_2 f_p(t,x)) a_0(t,x) \\
& + \frac{1}{2} \left(\beta^2 f_p(t,x) - 2\beta \partial_2 f_p(t,x) + \partial_2^2 f_p(t,x)\right) a_1(t,x)^2 \\
& + \int_{\mathbb{R}_0} \left(e^{-\beta b_1(t,x) b_2(z)} f_p(t, x + b_1(t,x) b_2(z)) - f_p(t,x)\right. \\
& \qquad \left. - (-\beta f_p(t,x) + \partial_2 f_p(t,x)) b_1(t,x) b_2(z) \mathbb{1}_{(0,1]}(|z|)\right) \nu(dz).
\end{aligned}
\tag{3.3}
$$

Proposition 3.1. *Consider the optimization problem (3.1). Fix $(t,x) \in E_1$. Let $b_2(z) = z$ and let ν be supported on $\mathbb{R}_+$. If*

(i) $\beta b_1(t,x) > 0$, or
(ii) $\int_{z>1} e^{-\beta b_1(t,x) z} z^{(K \vee 1)} \nu(dz) < +\infty$,

then the function $\mathscr{A} f(t,x)$ is well defined and its sign is identical to that of $\mathscr{A}_\beta f_p(t,x)$.

Proof. Observe that

$$
\mathscr{A} f(t,x) = e^{-\beta x} \mathscr{A}_\beta f_p(t,x).
$$

Hence, we investigate $\mathscr{A}_\beta f_p(t,x)$. The drift and the diffusion components are clearly well defined. The jump component can be rewritten as

$$
\begin{aligned}
& \int_{\mathbb{R}_0} \left(e^{-\beta b_1(t,x) b_2(z)} - 1 + \beta b_1(t,x) b_2(z) \mathbb{1}_{(0,1]}(|z|)\right) \nu(dz) \sum_{B(0,0)} c_{k_t,k_x} t^{k_t} x^{k_x} \\
& + b_1(t,x) \int_{\mathbb{R}_0} b_2(z) \left(e^{-\beta b_1(t,x) b_2(z)} - \mathbb{1}_{(0,1]}(|z|)\right) \nu(dz) \sum_{B(0,1)} k_x c_{k_t,k_x} t^{k_t} x^{k_x-1} \\
& + \sum_{B(0,2)} c_{k_t,k_x} t^{k_t} \sum_{k=0}^{k_x-2} {}_{k_x} C_k x^k b_1(t,x)^{k_x-k} \int_{\mathbb{R}} e^{-\beta b_1(t,x) b_2(z)} b_2(z)^{k_x-k} \nu(dz).
\end{aligned}
\tag{3.4}
$$

Hence, $\mathscr{A}_\beta f_p(t,x)$ is well defined, if all the integrals with respect to ν are well defined. Note that the third integral appears only when $K \geq 2$. With this in mind, we suppose so during this proof.

First, as $z \downarrow 0$, the integrands of (3.4) behave like z^2, z^2 and z^k, $k = 2, \ldots, K$, respectively. Hence, by (2.3), they integrate near the origin. (Note that the conditions (i) and (ii) are irrelevant to the integrability near the origin.)

Next, as $z \uparrow +\infty$, the integrands of (3.4) are $O(1)$, $o(1)$ and $o(1)$, respectively, if $\beta b_1(t, x) > 0$. Hence, again due to (2.3), the claim holds when (i) is satisfied. If (i) is not satisfied, then the integrands behave respectively like $e^{-\beta b_1(t,x)z}$, $z e^{-\beta b_1(t,x)z}$ and $z^k e^{-\beta b_1(t,x)z}$, $k = 2, \ldots, K$, at infinity.

The last claim holds by $\mathscr{A}f(t,x) = e^{-\beta x}\mathscr{A}_\beta f_p(t,x)$ and $e^{-\beta x} > 0$. $\qquad\square$

Let us return to the optimization problem (3.1). The minimization of $f(0, X_0)$ with respect to $\{c_{k_t, k_x}\}_{B(0,0)}$ is obviously an operation identical to the minimization of $f_p(0, X_0)$, multiplied by $e^{-\beta X_0}$ afterward, since $e^{-\beta X_0}$ is independent of $\{c_{k_t, k_x}\}_{B(0,0)}$. Next, we deal with the constraint $\mathscr{A}f(t,x) \leq 0$ over E_1. The following result provides a verifiable condition under which we can safely replace the constraint with $\mathscr{A}_\beta f_p(t,x) \leq 0$.

Proposition 3.2. *If $\beta > 0$ and if $b_1(t,x)$ is constant over E_1, then $\mathscr{A}_\beta f_p(t,x)$ is polynomial in t and x.*

Proof. All the components, but the jump component, of $\mathscr{A}_\beta f_p(t,x)$ are clearly polynomial in t and x. Concerning the jump component, the integrals in (3.4) are independent of (t,x) when the conditions are imposed. $\qquad\square$

Remark 3.3. It seems almost necessary to have that $b_1(t,x)$ is constant over E_1. As an illustrative example, we take the Lévy measure $\nu(dz) = e^{-z}/z\,dz$, $z \in \mathbb{R}_+$, of the gamma process. We then have

$$\int_0^{+\infty} \left(e^{-\beta b_1(t,x)z} - 1 + \beta b_1(t,x)z \right) \nu(dz) = -\ln\left(1 + \beta b_1(t,x)\right) - \beta b_1(t,x),$$

that cannot be polynomial in x no matter what polynomial b_1 is chosen. $\qquad\square$

We have so far fixed most ingredients of the problem setting. Let us finalize the validity of the optimization problem (3.1).

Proposition 3.4. *Let $\beta > 0$, let $b_1(t,x)$ be constant over E_1, let $\mathscr{X} \subseteq \mathbb{R}_+ \cup \{0\}$, and let $f(t,x) = e^{-\beta x} f_p(t,x)$ where $f_p \in C_p(E_2; \mathbb{R})$. Assume that for each $(t,x) \in E_1$, the function $\mathscr{A}f(t,x)$ is well defined. Then, the Dynkin formula (2.8) holds.*

Proof. Thanks to the polynomial form of $\mathscr{A}_\beta f_p(t,x)$, it suffices to check if $\mathbb{E}[e^{-\beta X_t} X_t^k] < +\infty$ for a suitable $k \in \mathbb{N}$. For any well-defined non-negative random variable X and for each $k \in \mathbb{N}$, $\mathbb{E}[e^{-\beta X} X^k] < +\infty$, since $e^{-\beta x} x^k \leq e^{-k}(k/\beta)^k < +\infty$ for $x \in \mathbb{R}_+ \cup \{0\}$. From this fact and $\mathscr{X} \subseteq \mathbb{R}_+ \cup \{0\}$, it follows that $\mathbb{E}[|f(\tau, X_\tau)|] < +\infty$, and for each $t \in [0, \tau)$, $\mathbb{E}[|\mathscr{A}f(t, X_t)|] < +\infty$. Hence, by Lemma 2.1 (i), the claim holds. $\qquad\square$

This claims that when a non-negative process is considered, our methodology is well defined as soon as $\beta > 0$, with no additional condition on the law of the solution of stochastic differential equations. Therefore, the exponential tempering widens the class of stochastic differential equations for which our methodology is well defined. For example, the integrals in (3.4) with respect to the stable Lévy measure are well defined when $\beta > 0$, while not as soon as $\beta \leq 0$. Also, when $\beta < 0$, one would need to check the conditions presented in Lemma 2.1, that are often not verifiable in particular when a stochastic differential equation is involved. We will henceforth assume that the Dynkin formula (2.8) holds and the optimization problem (3.1) is well defined.

Now, coming back to the optimization problem (3.1), we have shown in Proposition 3.1 and 3.2 that the constraint $\mathscr{A}_\beta f_p(t,x) \leq 0$ is equivalent to $\mathscr{A}f(t,x) \leq 0$. The optimization problem (3.1) is now transformed into an equivalent form

$$e^{-\beta X_0} \times \left|\begin{array}{ll} \min & f_p(0, X_0) \\ \text{s.t.} & f_p(t,x) \geq e^{\beta x} V(t,x) \ \text{ on } E_0, \\ & \mathscr{A}_\beta f_p(t,x) \leq 0 \ \text{ on } E_1, \\ & f_p \in C_p(E_2; \mathbb{R}). \end{array}\right.$$

8

Finally, the constraint $f_p(t,x) \geq e^{\beta x} V(t,x)$ over E_0 remains non-polynomial. There seem to be no exact methods to transform $e^{\beta x} V(t,x)$ into a (piecewise) polynomial form. (Recall that $V(t,x)$ may have discontinuities.) We instead try to replace this with a (piecewise) polynomial constraint, which is a little more conservative than the original one. It suffices to approximate the exponential $e^{\beta x}$ by polynomial on a bounded set $\mathcal{K}_V$ thanks to the support of the function V. It is not very difficult to find a (piecewise) polynomial $u(t,x)$ such that $u(t,x) \geq e^{\beta x} V(t,x)$ over E_0. For example, for a small tolerance $\varepsilon > 0$, one first finds a (piecewise) polynomial $p(t,x)$ such that $\sup_{(t,x) \in E_0} |e^{\beta x} V(t,x) - p(t,x)| \leq \varepsilon$, and then set $u(t,x) = p(t,x) + \varepsilon$ and $l(t,x) = p(t,x) - \varepsilon$.

Suppose that we have found polynomials $u(t,x)$ and $l(t,x)$ such that $l(t,x) \leq e^{\beta x} V(t,x) \leq u(t,x)$ over E_0. Then, for each $\beta > 0$, we arrive at a polynomial programming problem

$$H_U(\beta) := e^{-\beta X_0} \times \left| \begin{array}{ll} \min & f_p(0, X_0) \\ \text{s.t.} & f_p(t,x) \geq u(t,x) \text{ on } E_0, \\ & \mathscr{A}_\beta f_p(t,x) \leq 0 \text{ on } E_1, \\ & f_p \in C_p(E_2; \mathbb{R}), \end{array} \right. \tag{3.5}$$

that is a fairly close approximation of the original problem (3.1). For lower bounds, we compute

$$H_L(\beta) := e^{-\beta X_0} \times \left| \begin{array}{ll} \min & g_p(0, X_0) \\ \text{s.t.} & g_p(t,x) \leq l(t,x) \text{ on } E_0, \\ & \mathscr{A}_\beta g_p(t,x) \geq 0 \text{ on } E_1, \\ & g_p \in C_p(E_2; \mathbb{R}). \end{array} \right. \tag{3.6}$$

The parameter β above can be chosen arbitrarily. We will look at its choice later in numerical examples.

Remark 3.5. Notice that the exponential term $e^{-\beta x}$ serves as an *exponential tilting* when applied to the whole real line. It is the *exponential tempering* only when the state space of the stochastic differential equation is bounded on at least one side. It seems tempting to apply the exponential tempering $e^{-\beta x^2}$ of second order (or of a higher even order) since it tempers any polynomial function on both sides and its derivative is still as simple as a product of a polynomial and the tempering term itself. We have however observed that the application of a higher order tempering is not of practical use for the reason that $e^{\beta x^2}$ (in the constraint $f_p(t,x) \leq e^{\beta x^2} V(t,x)$) grows so fast that the polynomial approximation is very difficult even over a bounded set. $\qquad \square$

Under mild conditions such as the moment determinate property, it is straightforward to theoretically guarantee the convergence of the gap to zero, when considering the dual problem summarized in the Appendix. (See, for example, Theorem 7 of [9].) In practice, however, it is truly impossible to take the degrees of polynomial arbitrarily large due to the well known curse of dimensionality of the semi-definite programming.

Before proceeding to numerical examples, let us present a result on a transform of the value function V, of practical interest.

Proposition 3.6. *Let v be in $C^1([0,T]; \mathbb{R})$. For f in $C^{1,2}(E_2; \mathbb{R})$ such that*

$$f(t,x) \geq V(t,x) - v(t) \text{ on } E_0,$$

$$\mathscr{A} f(t,x) \leq -\frac{d}{dt} v(t) \text{ on } E_1,$$

it holds that

$$\mathbb{E}\left[V(\tau, X_\tau) \right] \leq f(0, X_0) + v(0).$$

Proof. The result holds by

$$\mathbb{E}\left[V(\tau,X_\tau)-v(\tau)\right]\leq\mathbb{E}\left[f(\tau,X_\tau)\right]=f(0,X_0)+\mathbb{E}\left[\int_0^\tau\mathscr{A}f(t,X_t)dt\right]$$

$$\leq f(0,X_0)-\mathbb{E}\left[\int_0^\tau\frac{d}{dt}v(t)dt\right],$$

where we have used the assumptions imposed on f. $\qquad\square$

This result is primarily meant for the case V does not have bounded support, while $V(t,x)-v(t)$ does. In order for this to be actually valid in our framework, we need to look closely at the above assumptions on f. Since f is of the form $e^{-\beta x}f_p(t,x)$ and $\mathscr{A}f(t,x)=e^{-\beta x}\mathscr{A}_\beta f_p(t,x)$, we get

$$f_p(t,x)\geq e^{\beta x}\left(V(t,x)-v(t)\right)\quad\text{on }E_0,$$

$$\mathscr{A}_\beta f_p(t,x)\leq -e^{\beta x}\frac{d}{dt}v(t)\quad\text{on }E_1.$$

This implies that the result is valid if the following two conditions are satisfied;

    (i) $V(\tau,x)-v(\tau)$ has compact support in terms of x almost surely, or the set E_0 is bounded,
   (ii)   the set $\{(t,x)\in E_1:(d/dt)v(t)\neq 0\}$ is bounded.

Examples in the following section are related to Proposition 3.6.

4   Numerical Illustrations

In this section, we test numerically our method on a probability estimation of a stable subordinator and on a survival probability of an Ornstein-Uhlenbeck process driven by a stable process, both without diffusion component, that is, $a_1(t,x)\equiv 0$. Throughout this section, we set $c_{k_t,k_x}\equiv 0$ for $k_t+k_x>K$ where K is an even natural number, and approximate $e^{\beta x}$ by the standard Taylor expansion $\sum_{k=0}^K(\beta x)^k/k!$ around the origin of order K. The resulting approximation gap $\varepsilon:=\max_{(t,x)\in E_0}|e^{\beta x}V(t,x)-u(t,x)|$ is satisfactorily small in each example, where $u(t,x)=V(t,x)\sum_{k=0}^K(\beta x)^k/k!$ is then (piecewise) polynomial.

4.1   Tail Probability of Stable Subordinator

For illustrative purpose, we first test our method on a toy example; the tail probability estimation for a stable subordinator at time T. Set $X_0=0$, $a_1(t,x)=0$, $b_1(t,x)=1$, $b_2(z)=z$, and

$$\nu(dz)=\frac{1}{z^{1+\alpha}}dz,\quad z\in\mathbb{R}_+,\tag{4.1}$$

for some $\alpha\in(0,1)$, and $a_0(t,x)=\int_{z\in(0,1]}z\nu(dz)$. Then, the stochastic differential equation reduces to a stable subordinator

$$X_t=\int_0^t\int_{\mathbb{R}_+}z\mu(dz,ds).\tag{4.2}$$

In this setting, we have

$$\mathscr{A}_\beta f_p(t,x)=\partial_1 f_p(t,x)+f_p(t,x)\int_{\mathbb{R}_+}\left(e^{-\beta z}-1\right)\nu(dz)$$

$$+\sum_{B(0,1)}c_{k_t,k_x}t^{k_t}\sum_{k=0}^{k_x-1}{}_{k_x}C_k x^k\int_{\mathbb{R}_+}e^{-\beta z}z^{k_x-k}\nu(dz),\tag{4.3}$$

10

where the integrals are explicit as

$$\int_0^{+\infty} \left(e^{-\beta z} - 1 \right) \nu(dz) = -\frac{\beta^\alpha}{\alpha} \Gamma(1-\alpha),$$

and

$$\int_0^{+\infty} e^{-\beta z} z^{k_x-k} \nu(dz) = \frac{\Gamma(k_x-k-\alpha)}{\beta^{k_x-k-\alpha}},$$

that are well defined if and only if $\beta > 0$. Hence, fix $\beta > 0$. By $X_0 = 0$, it holds that $\mathscr{X} = \mathbb{R}_+$. Moreover, $E_0 = \{T\} \times \mathscr{X}$ and $E_1 = [0,T) \times \mathscr{X}$. By Proposition 3.4, the Dynkin formula holds and our methodology is well defined.

Now, consider the asymptotic tail probability, as $\lambda \uparrow +\infty$,

$$\mathbb{P}(X_1 > \lambda) \sim \frac{1}{\alpha \lambda^\alpha}.$$

To this end, the value function should be set in the form of $V(t,x) = \mathbb{1}(x > \lambda)$, while this value function does not have compact support. To apply Proposition 3.6, we set $v(t) \equiv 1$ so that $\mathbb{1}(x > \lambda) = v(t) - V(t,x)$. (To comply with the formulation of Proposition 3.6, the transform should instead be $-\mathbb{1}(x > \lambda) = V(t,x) - v(t)$, while the sign is certainly not fundamental.) Obviously, this transform is valid in our exponential tempering approach since $V(t,x) - v(t)$ has compact support in terms of x and $(d/dt)v(t) \equiv 0$.

Using the selfsimilarity of stable processes $\{h^{-1/\alpha} X_{ht} : t \geq 0\} \overset{\mathscr{L}}{=} \{X_t : t \geq 0\}$, we set $V(t,x) = \mathbb{1}(x > t^{1/\alpha}\lambda)$ with $\{X_t : t \in [0,T]\}$ for some small $T > 0$, rather than $V(t,x) = \mathbb{1}(x > \lambda)$ with $\{X_t : t \in [0,1]\}$. We then compute the probability through the equation

$$\mathbb{P}(X_1 \leq \lambda) = 1 - \mathbb{P}(X_1 > \lambda) = \mathbb{E}[v(T) - V(T,X_T)],$$

with $\lambda = T^{-1/\alpha}$. We will rather report numerical results for the probability $\mathbb{P}(X_1 \leq \lambda)$ since this is what our optimization procedure actually deals with.

Numerical results with fixed polynomial degrees are presented in Table 1 and 2. For example, concerning the case $\alpha = 0.3$ and $T = 0.001$ in Table 2, we can choose the smallest upper bound and the largest lower bound, that is,

$$0.996189 \leq \mathbb{P}\left(X_1 \leq T^{-1/\alpha} \right) \leq 0.997045,$$

both of which happen to be from $\beta = 3.0$. Since $T^{-1/\alpha}$ is sufficiently large, we have

$$\mathbb{P}\left(X_1 \leq T^{-1/\alpha} \right) \approx 1 - \frac{T}{\alpha} = 0.996667 \in [0.996189, \ 0.997045].$$

This also implies that the estimation of this tail probability requires one to deal with a highly rare event simulation. To compare this result with the Monte Carlo framework, observe that

$$\mathrm{Var}\left(\mathbb{1}\left(X_1 \leq T^{-1/\alpha} \right) \right) = \mathbb{E}\left[\mathbb{1}\left(X_1 \leq T^{-1/\alpha} \right)^2 \right] - \mathbb{E}\left[\mathbb{1}\left(X_1 \leq T^{-1/\alpha} \right) \right]^2$$

$$\approx 1 - \frac{T}{\alpha} - \left(1 - \frac{T}{\alpha} \right)^2 = 0.05763^2.$$

Denoting by $\widehat{F}_n$ a Monte Carlo estimator for the random variable $\mathbb{1}(X_1 \leq T^{-1/\alpha})$ based on n iid replications, its 99.9999%-confidence interval is approximately given by

$$\left[\widehat{F}_n - 4.89\frac{0.05763}{\sqrt{n}}, \ \widehat{F}_n + 4.89\frac{0.05763}{\sqrt{n}} \right].$$

To narrow the length of this approximate confidence interval to that of $[0.996189, 0.997045]$, we need at least $n = 433539$ of iid samples. This number does not support the use of Monte Carlo methods in full.

Let us discuss further on a superiority of our method over Monte Carlo methods. On one hand, although the strong law of large number guarantees the almost sure convergence of a well-defined Monte Carlo estimator to the true value *in theory*, it is rather usual that a limiting value is very far from the true value and is extremely sensitive to the seed selected for a random number generator *on computer*. On the other hand, our approach is based on a purely deterministic optimization method, free of random elements. Unlike confidence intervals in the Monte Carlo framework, it is guaranteed that the true value sits somewhere within the obtained optimality gap. Therefore, the optimality gap $[0.996189, 0.997045]$ above can be considered as a 100%-confidence interval, which certainly predominates the Monte Carlo simulation with any large sample size. In fact, the 100%-confidence interval is simply impossible in the Monte Carlo framework unless the estimator is simply degenerate.

		$\alpha = 0.3$			$\alpha = 0.6$		
			K				
		6	8	10	6	8	10
T	0.01	0.97247 (0.96667) 0.95981	0.97154 (0.96667) 0.96147	0.97132 (0.96667) 0.96172	0.98829 (0.98333) 0.97568	0.98765 (0.98333) 0.97693	0.98761 (0.98333) 0.97699
	0.001	0.997220 (0.996667) 0.995924	0.997124 (0.996667) 0.996097	0.997104 (0.996667) 0.996119	0.998831 (0.998333) 0.997594	0.998772 (0.998333) 0.997706	0.998771 (0.998333) 0.997707

Table 1: Numerical Results for the estimation of $\mathbb{P}(X_1 \leq T^{-1/\alpha})$ with $\beta = 2$ for different polynomial degrees K. Each consists of an upper bound, the theoretical asymptotic value (in the parentheses), and a lower bound.

		$\alpha = 0.3$			$\alpha = 0.6$		
			β				
		1.0	2.0	3.0	1.0	2.0	3.0
T	0.01	0.97110 (0.96667) 0.96045	0.97132 (0.96667) 0.96172	0.97076 (0.96667) 0.96265	0.98846 (0.98333) 0.97501	0.98761 (0.98333) 0.97699	0.98696 (0.98333) 0.97823
	0.001	0.997221 (0.996667) 0.995922	0.997104 (0.996667) 0.996119	0.997045 (0.996667) 0.996189	0.998874 (0.998333) 0.997463	0.998771 (0.998333) 0.997707	0.998701 (0.998333) 0.997840

Table 2: Numerical Results for the estimation of $\mathbb{P}(X_1 \leq T^{-1/\alpha})$ with $K = 10$ for different exponential tempering parameters β. Each consists of an upper bound, the theoretical asymptotic value (in the parentheses), and a lower bound.

Remark 4.1. We have observed through the numerical results in Table 2 that the choice of β may not largely affect the quality of the estimation. Note, however, that different β's lead to different polynomial optimization problems. In view of intrinsic nature of the numerical optimization, the semi-definite relaxation of polynomial optimization and so on, it is difficult to raise theoretical reasons why this family of optimization problems gave similar bounds for different β's. Our method may however tighten the optimality gap through the choice of β, without increasing the size of mathematical programming with larger polynomial degree K. It seems that numerical results tend to be more sensitive to the choice of β as the marginal X_T has lighter tails. This would be an interesting topic as a future work. $\square$

4.2 Survival Probability of Ornstein-Uhlenbeck Processes

Here, we test our method on the survival probability estimation of an Ornstein-Uhlenbeck process driven by a stable subordinator. (See Suzuki, Miyoshi and Kojima [18] for a similar problem in the diffusion setting.) Compared with the example of Section 4.1, this example entails a stopping time τ, rather than the frozen terminal time T.

Set $\nu(dz) = z^{-1-\alpha}dz$, $z \in \mathbb{R}_+$ as given by (4.1), and set $a_0(t,x) = -\lambda x + \int_{z \in (0,1]} z\nu(dz)$ for some $\lambda > 0$, $a_1(t,x) = 0$, $b_1(t,x) = 1$, $b_2(z) = z$, and $X_0 > 0$. Then, the stochastic differential equation reduces to

$$dX_t = -\lambda X_t dt + \int_{\mathbb{R}_+} z\mu(dz, dt), \tag{4.4}$$

which is called an Ornstein-Uhlenbeck process. Its solution is given by

$$X_t = e^{-\lambda t}X_0 + \int_0^t \int_{\mathbb{R}_+} e^{-\lambda(t-s)}z\mu(dz, ds).$$

Its sample paths can be simulated in the exact sense over a finite horizon $[0,T]$, $T > 0$, by using the series representation of stable subordinators as

$$\{X_t : t \in [0,T]\} \overset{\mathscr{L}}{=} \left\{ e^{-\lambda t}X_0 + \sum_{k=1}^{+\infty} e^{-\lambda(t-T_k)}\left(\frac{\alpha\Gamma_k}{T}\right)^{-1/\alpha} \mathbb{1}(T_k \in [0,t]) : t \in [0,T] \right\}, \tag{4.5}$$

where $\{\Gamma_k\}_{k \in \mathbb{N}}$ is a sequence of arrival times of a standard Poisson process and $\{T_k\}_{k \in \mathbb{N}}$ is a sequence of iid uniform random variables on $[0,T]$. However, this simulation method is extremely expensive and not of practical use due to the infinite sum for each sample path. In addition, the Euler-Maruyama discretization of the stochastic differential equation is an expensive yet only approximative method, which in general produces an estimation error due to the discrete monitoring.

In this setting, we have

$$\mathscr{A}_\beta f_p(t,x) = \partial_1 f_p(t,x) - \lambda x(-\beta f_p(t,x) + \partial_2 f_p(t,x)) + f_p(t,x)\int_{\mathbb{R}_+}\left(e^{-\beta z} - 1\right)\nu(dz)$$

$$+ \sum_{B(0,1)} c_{k_t,k_x} t^{k_t} \sum_{k=0}^{k_x-1} {}_{k_x}C_k x^k \int_{\mathbb{R}_+} e^{-\beta z}z^{k_x-k}\nu(dz),$$

where the integrals are well defined if $\beta > 0$ and are explicit as presented in Section 4.1. Henceforth, we fix $\beta > 0$. By $X_0 > 0$, it holds that $\mathscr{X} = \mathbb{R}_+$. By Proposition 3.4, the Dynkin formula (2.8) holds and thus our optimization approach is well defined. Note that the Dynkin formula fails to hold as soon as $\beta \leq 0$, since for each $t > 0$,

$$\mathbb{E}[X_t] \geq e^{-\lambda t}X_0 + e^{-\lambda t}\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_+} z\mu(dz, ds)\right] = +\infty.$$

We consider the survival probability of $\{X_t : t \in [0,T]\}$ with finite time horizon out of the bounded set

$$E_1 = [0,T) \times [0,U],$$

where $U > X_0$. Then, τ is the $(\mathscr{F}_t)_{t \in [0,T]}$-stopping time defined by

$$\tau := \inf\{t \geq 0 : X_t \notin E_1\} \wedge T,$$

that is, the first exit time out of E_1. The random vector (τ, X_τ) indicates the exit location, where E_0 can split into two disjoint sets

$$E_u := [0,T] \times (U, +\infty),$$
$$E_r := \{T\} \times [0,U].$$

It is enough to have $E_u \cup E_r$ for the exit location, since $\{X_t : t \in [0,T]\}$ is almost surely non-negative. The unboundedness of the set E_u corresponds to the unbounded jump size of the stable subordinator.

To estimate the survival probability $\mathbb{P}((\tau, X_\tau) \in E_r)$, we set $V(t,x) := \mathbb{1}((t,x) \in E_r)$. Since $V(t,x)$ has compact support, we do not have to apply Proposition 3.6. Tempered polynomial optimization problems are formulated as

$$
\left|
\begin{aligned}
\min \quad & e^{-\beta X_0} f_p(0, X_0) \\
\text{s.t.} \quad & f_p(T,x) \geq u(x) \text{ on } [0,U], \\
& f_p(t,x) \geq 0 \text{ on } E_u, \\
& \mathscr{A}_\beta f_p(t,x) \leq 0 \text{ on } E_1, \\
& f_p \in C_p(E_2; \mathbb{R}),
\end{aligned}
\right.
\quad \text{and} \quad
\left|
\begin{aligned}
\max \quad & e^{-\beta X_0} g_p(0, X_0) \\
\text{s.t.} \quad & g_p(T,x) \leq l(x) \text{ on } [0,U], \\
& g_p(t,x) \leq 0 \text{ on } E_u, \\
& \mathscr{A}_\beta g_p(t,x) \geq 0 \text{ on } E_1, \\
& g_p \in C_p(E_2; \mathbb{R}),
\end{aligned}
\right.
\tag{4.6}
$$

where $u(x)$ and $l(x)$ are polynomial functions bounding $e^{\beta x}$ respectively from above and below uniformly over E_r.

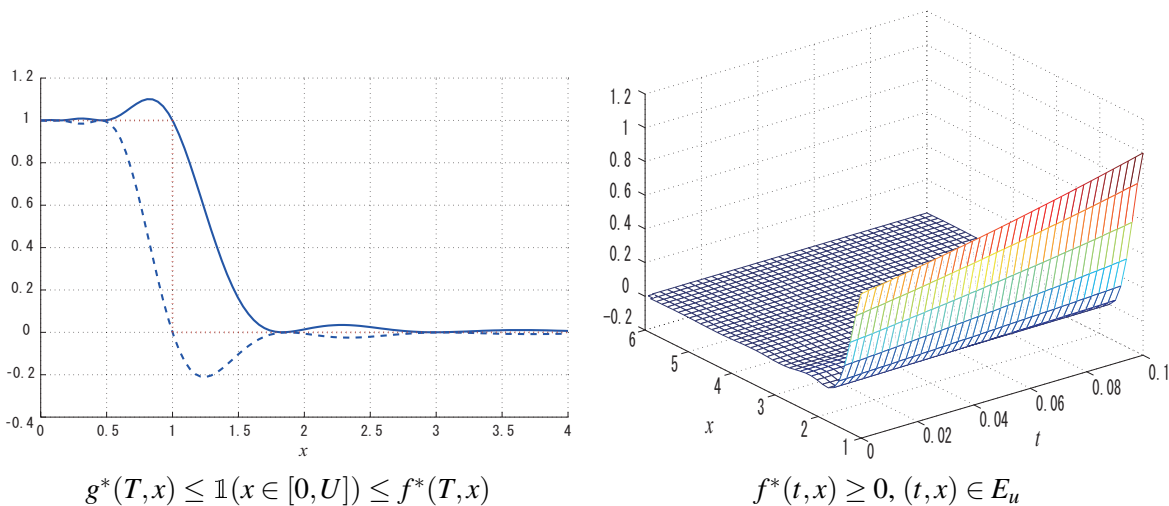$$g^*(T,x) \leq \mathbb{1}(x \in [0,U]) \leq f^*(T,x) \qquad\qquad f^*(t,x) \geq 0, \ (t,x) \in E_u$$

Figure 1: Optimal Tempered Polynomial Functions of the Optimization Problems (4.6) with $\lambda = 1$, $T = 0.1$, $U = 1$, and $\beta = 3$.

In Figure 1, we draw optimal bounding functions f^* and g^* to illustrate the advantage of our exponential tempering approach. First, the left figure indicates that at terminal time T, the bounding functions bound the step function $\mathbb{1}(x \in [0,U])$ from above and below. In contrast to the polynomial case with explosion at infinity, it also indicates that they satisfy the ideal property that for each $t \in [0,T]$,

$$\lim_{x \uparrow +\infty} f^*(t,x) = \lim_{x \uparrow +\infty} g^*(t,x) = 0.$$

This avoids bounding functions to be far from one another and helps to achieve a tight optimality gap. Next, the right figure draws f^* on the set E_u. Here, we are required to deal with two somewhat conflicting requirements on the positive function f^*; it is desired to be as close to zero as possible uniformly over E_u, while the inequality $f^*(T,U) \geq 1$ must hold at the boundary point (T,U). As can be observed, however, our approach works effectively under such a complex circumstance in such a way that the tempered bounding function f^* tends rapidly to vanish at infinity. Let us emphasize again that this never comes true with bounding functions of polynomial form, that explode at infinity.

Remark 4.2. In this problem setting, other quantities can also be estimated by our approach. One example is the moment of the exit time, investigated in [4]. First, consider the value function $V(t,x) = t^k$ for

$\mathbb{E}[V(\tau, X_\tau)] = \mathbb{E}[\tau^k]$. The tempered optimization problem for this setting,

$$
\begin{vmatrix}
\begin{aligned}
\min \quad & e^{-\beta X_0} f_p(0, X_0) \\
\text{s.t.} \quad & f_p(t,x) \geq t^k e^{\beta x} \ \text{ on } E_0, \\
& \mathscr{A}_\beta f_p(t,x) \leq 0 \ \text{ on } E_1,
\end{aligned}
\quad \text{and} \quad
\begin{aligned}
\max \quad & e^{-\beta X_0} g_p(0, X_0) \\
\text{s.t.} \quad & g_p(t,x) \leq t^k e^{\beta x} \ \text{ on } E_0, \\
& \mathscr{A}_\beta g_p(t,x) \geq 0 \ \text{ on } E_1,
\end{aligned}
\end{vmatrix}
$$

is not solvable due to the exponential tempering because the set E_0 is unbounded and the value function $V(t,x)$ does not have bounded support. Instead, by applying Proposition 3.6 with $v(t) = t^k$, we arrive at the tempered optimization problems are now formulated as

$$
\begin{vmatrix}
\begin{aligned}
\min \quad & e^{-\beta X_0} f_p(0, X_0) \\
\text{s.t.} \quad & f_p(t,x) \geq 0 \ \text{ on } E_0, \\
& \mathscr{A}_\beta f_p(t,x) \leq -e^{\beta x} k t^{k-1} \ \text{ on } E_1,
\end{aligned}
\quad \text{and} \quad
\begin{aligned}
\max \quad & e^{-\beta X_0} g_p(0, X_0) \\
\text{s.t.} \quad & g_p(t,x) \leq 0 \ \text{ on } E_0, \\
& \mathscr{A}_\beta g_p(t,x) \geq -e^{\beta x} k t^{k-1} \ \text{ on } E_1.
\end{aligned}
\end{vmatrix}
$$

Those are valid formulations since then $V(t,x) - v(t) \equiv 0$ and the set E_1 is bounded. $\qquad\square$

Remark 4.3. We have so far considered exponential tempering with $\beta > 0$, while a negative β may also be chosen as soon as all required conditions are satisfied. For instance, consider the Ornstein-Uhlenbeck process (4.4) with μ being a Poisson random measure whose compensator is given by the truncated stable Lévy measure $\nu(dz) = z^{-1-\alpha} dz$ supported on $(0, \eta]$, for some $\eta > 0$. Due to its bounded size of jumps, the overshoot out of the set E_1 in is certainly bounded, that is, the set E_u is bounded; $E_u = [0, T] \times (U, U + \eta)$. Contrary to our initial motivation, the optimization problems (4.6) can then be formulated with every $\beta \in \mathbb{R}$. In connection with Remark 4.1, however, this provides us with some possibility of obtaining a tighter optimality gap. $\qquad\square$

5 Concluding Remarks

In this paper, we have proposed an improvement of the optimization methodology of [5] through exponential tempering of bounding functions when value functions have compact support. Our approach yields lower and upper bounds of the expectation for stochastic differential equations with jumps without Monte Carlo sample paths simulation, which often requires extremely expensive computing effort. We have shown that the tempered polynomial optimization can be transformed into a polynomial optimization problem after the polynomial approximation of the exponential function on a compact set. Moreover, exponential tempering widens the class of stochastic differential equations to which our methodology is actually applicable.

As discussed in Remark 4.1 and 4.3, the estimation quality may be improved by choosing the parameter β wisely, without increasing the size of mathematical programming. We did not present an exhaustive study of its range of applicability relative to different value functions or underlying stochastic differential equations, which is significantly large. In fact, encouraged by the quality of numerical results and by the wide applicability, our methodology is expected to be a standard tool for the weak approximation of a general class of stochastic differential equations with jumps. We should also remark that this framework may not be as robust as Monte Carlo methods when underlying stochastic differential equations are multivariate. These issues will be addressed in subsequent papers.

Appendix: Generalized Problem of Moments

In principle, this approach is aimed at maximizing and minimizing the expectation, rather than its bounds as in our approach, where the underlying probability measure implicitly serves as the decision variable, under the so-called moment conditions that reflect necessary conditions for a set of scalars to be identified with moments of the probability measure. Due to the nature of methods, they are often called a method-of-moments approach altogether in the literature.

To be more mathematically precise, we illustrate this approach under the setting of Section 4.2. Define the distribution of the exit location by

$$\nu_0(B) := \mathbb{P}((\tau, X_\tau) \in B), \quad B \in \mathscr{B}(E_0),$$

and the expected occupation measure on the set E_1 until the exit by

$$\nu_1(B) := \mathbb{E}\left[\int_0^\tau \mathbb{1}((t, X_t) \in B)dt\right], \quad B \in \mathscr{B}(E_1).$$

With the measures ν_0 and ν_1, the Dynkin formula (2.8) reads

$$\int_{E_r} f(t,x)\nu_0(dt,dx) = f(0,X_0) - \int_{E_u} f(t,x)\nu_0(dt,dx) + \int_{E_1} \mathscr{A}f(t,x)\nu_1(dt,dx),$$

which is the so-called basic adjoint equation, named by Helmes, Röhl and Stockbridge [4]. We have shown that for each non-negative integer k and l,

$$b_{k,l}^{(\text{top})} := \int_{E_u} e^{-\beta x} t^k x^l \nu_0(dt,dx),$$

$$b_l^{(\text{rig})} := \int_{E_r} e^{-\beta x} x^l \nu_0(dt,dx),$$

$$m_{k,l} := \int_{E_0} e^{-\beta x} t^k x^l \nu_1(dt,dx).$$

are well defined. In the method-of-moments, the quantities above serve as decision variables in the resulting semi-definite programming. First, by plugging a tempered monomial of the form

$$f(t,x) = e^{-\beta x} t^k x^l,$$

into the basic adjoint equation, we obtain a family of equality constraints that moment sequence should meet. Second, we constrain that the adequately defined moment and localizing moment matrices are positive semi-definite. It is a necessary condition for decision variables to be moment sequence with respect to *some* measure on E_u, E_r, and E_0. Under these two constraints, the maximal and minimal values of $\sum_k c_k b_k^{(\text{rig})}$ are respectively upper and lower bounds for the desired survival probability, where the sequence $\{c_k\}_k$ of real numbers must satisfy, respectively, $\sum_k c_k x^k \geq e^{\beta x}$ and $\sum_k c_k x^k \leq e^{\beta x}$ uniformly over $[0, U]$.

References

[1] Applebaum, D. (2004) *Lévy Processes and Stochastic Calculus*, Cambridge University Press.

[2] Bertsimas, D., Popescu, I., Sethuraman, J. (2000) Moment problems and semidefinite programming, In: *Handbook on semidefinite programming: Theory, Algorithms, and Applications*, H. Wolkovitz, ed., 469-509.

[3] Eriksson, B., Pistorius, M. (2008) A method of moments approach to pricing double barrier contracts driven by a general class of jump diffusions, available at arXiv:0812.4548v1.

[4] Helmes, K., Röhl, S., Stockbridge, R.H. (2001) Computing moments of the exit time distribution for Markov processes by linear programming, *Operations Research*, **49**(4) 516-530.

[5] Kashima, K., Kawai, R. (2011) An optimization approach to weak approximation of stochastic differential equations with jumps, *Applied Numerical Mathematics*, **61**(5) 641-650.

[6] Kashima, K., Kawai, R. (2010) An optimization approach to weak approximation of Levy-driven stochastic differential equations, *Perspectives in Mathematical System Theory, Control, and Signal Processing: A Festschrift in Honor of Yutaka Yamamoto on the Occasion of his 60th Birthday* (Lecture Notes in Control and Information Sciences 398), J. C. Willems, S. Hara, Y. Ohta and H. Fujioka, Eds. Spinger, 263-272.

[7] Kloeden, P.E., Platen, E. (1999) *Numerical Solution of Stochastic Differential Equations*, Third Printing, Springer, Berlin.

[8] Lasserre, J.B., Prieto-Rumeau, T. (2004) SDP vs. LP relaxations for the moment approach in some performance evaluation problems, *Stochastic Models*, **20**(4) 439-456.

[9] Lasserre, J.B., Prieto-Rumeau, T., Zervos, M. (2006) Pricing a class of exotic via moments and SDP relaxations, *Mathematical Finance*, **16**(3) 469-494.

[10] Nishihara, M., Yagiura, M., Ibaraki, T. (2007) Duality in option pricing based on prices of other derivatives, *Operations Research Letters*, **35**(2) 165-171.

[11] Øksendal, B., Sulem, A. (2007) *Applied Stochastic Control of Jump Diffusions* (2nd Edition) Springer-Verlag Berlin Heidelberg.

[12] Parrilo, P.A. (2003) Semidefinite programming relaxations for semialgebraic problems, *Mathematical Programming Series B*, **96**(2) 293-320.

[13] Prajna, S., Papachristodoulou, A., Seiler, P., Parrilo, P.A. (2004) *SOS-TOOLS: Sum of squares optimization toolbox for MATLAB*.

[14] Primbs, J.A. (2008) Optimization based option pricing bounds via piecewise polynomial super- and sub-martingales, In: *2008 American Control Conference*.

[15] Protter, P. Talay, D. (1997) The Euler scheme for Lévy driven stochastic differential equations, *Annals of Probability*, **25**, 393-423.

[16] Rosiński, J. (2001) Series representations of Lévy processes from the perspective of point processes, In: *Lévy Processes - Theory and Applications,* Eds. Barndorff-Nielsen, O.-E., Mikosch, T., Resnick, S.I., Birkhäuser, 401-415.

[17] Sturm, J. (2006) SeDuMi version 1.1., available at http://sedumi.ie.lehigh.edu/

[18] Suzuki, K., Miyoshi, N., Kojima, M. (2008) A numerical method for survival probability of diffusion processes using semidefinite programming (in Japanese), *Transactions of the Operations Research Society of Japan*, **51**(1) 25-43.